Systematic Correlation of Structural, Thermodynamic and Residual Solvation Properties of Hydrophobic Substituents in Hydrophobic Pockets Using Thermolysin as a Case Study

Dissertation

**zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)**

**dem
Fachbereich Pharmazie der
Philipps-Universität Marburg
vorgelegt**

**von
Diplom-Pharmazeut**

Stefan Günter Krimmer

**aus
Villingen-Schwenningen**

Marburg/Lahn 2017

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von August 2011 bis März 2017 durchgeführt.

Erklärung zum Eigenanteil

Titel der Publikation und Journal (O: Originalarbeit, Ü: Übersichtsartikel)	Autoren	geschätzter Eigenanteil
O: Methyl, Ethyl, Propyl, Butyl: Futile But Not for Water, as the Correlation of Structure and Thermodynamic Signature Shows in a Congeneric Series of Thermolysin Inhibitors *ChemMedChem* **2014**, 9: 833–846	Krimmer SG, Betz M, Heine A, Klebe G	70%
Ü: Thermodynamics of protein–ligand interactions as a reference for computational analysis: How to assess accuracy, reliability and relevance of experimental data *J. Comput. Aided Mol. Des.* **2015**, 29: 867–883	Krimmer SG, Klebe G	70%
O: Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein Bound Ligands *J. Med. Chem.* **2016**, 59: 10530–10548	Krimmer SG,* Cramer J,* Betz M, Fridh V, Karlsson R, Heine A, Klebe G	35%
O: Elucidating the Origin of Long Residence Time Binding for Inhibitors of the Metalloprotease Thermolysin *ACS Chem. Biol.* **2017**, 12: 225–233	Cramer J,* Krimmer SG,* Fridh V,* Wulsdorf T, Karlsson R, Heine A, Klebe G	25%
O: High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits *Structure* **2016**, 24: 1398–1409	Schiebel J, Krimmer SG, Röwer K, Knörlein A, Wang X, Park AY, Stieler M, Ehrmann FR, Fu K, Radeva N, Krug M, Huschmann FU, Glöckner S, Weiss MS, Muller U, Klebe G, Heine A	10%
O: Experimental Active-Site Mapping by Fragments - Hot Spots Remote from the Catalytic Center of Endothiapepsin *J. Med. Chem.* **2016**, 59: 7561–7575	Radeva N, Krimmer SG, Stieler M, Fu K, Wang X, Ehrmann FR, Metz A, Huschmann FU, Weiss MS, Mueller U, Schiebel J, Heine A, Klebe G	10%
O: Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study *ACS Chem. Biol.* **2016**, 11: 1693–1701	Schiebel J, Radeva N, Krimmer SG, Wang X, Stieler M, Ehrmann FR, Fu K, Metz A, Huschmann FU, Weiss MS, Mueller U, Heine A, Klebe G	5%
O: Impact of Surface Water Layers on Protein-Ligand Binding: How Well Are Experimental Data Reproduced by Molecular Dynamics Simulations in a Thermolysin Test Case *J. Chem. Inf. Model.* **2016**, 56: 223–233	Betz M, Wulsdorf T, Krimmer SG, Klebe G	5%
O: Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads to Address a Catalytic Dyad *J. Med. Chem.* **2016**, 59: 9743–9759	Radeva N, Schiebel J, Wang X, Krimmer SG, Fu K, Stieler M, Ehrmann FR, Metz A, Rickmeyer T, Betz M, Winquist J, Park AY, Huschmann FU, Weiss MS, Mueller U, Heine A, Klebe G	5%

* Diese Autoren haben einen gleichwertigen Beitrag geleistet

....
Stefan G. Krimmer Prof. Dr. Gerhard Klebe

Abstract

Water molecules participate besides protein and ligand as an additional binding partner in every *in vivo* protein–ligand binding process. The displacement of water molecules from apolar surfaces of solutes is considered the driving force of the hydrophobic effect. It is generally assumed that the mobility of the water molecules increases through the displacement, and, as a consequence, entropy increases. This explanation, which is based on experiments with simple model systems, is, however, insufficient to describe the hydrophobic effect as part of the highly complex protein–ligand complex formation process. For instance, the displacement of water molecules from apolar surfaces that already exhibit an increased mobility before their displacement can result in an enthalpic advantage. Furthermore, it has to be considered that by the formation of the protein–ligand complex a new solvent-exposed surface is created, around which water molecules have to rearrange. The present thesis focuses on the impact of the latter effect on the thermodynamic and kinetic binding properties of a given ligand.

A congeneric ligand series comprised of nine ligands binding to the model protein thermolysin (TLN) was analyzed to determine the impact of the rearrangement of water molecules around the surface of a newly formed protein–ligand complex on the thermodynamic binding properties of a ligand. The protein–ligand complexes were characterized structurally by X-ray crystallography and thermodynamically by isothermal titration calorimetry (ITC). The only structural difference between the ligands was their strictly apolar P_2' substituent, which changed in size from a methyl to a phenylethyl group. The P_2' group interacts with the flat, apolar, and well-solvated S_2' pocket of TLN. Depending on the bound ligand, the solvent-exposed surface of the protein–ligand complex changes. The ITC measurements revealed strong thermodynamic differences between the different ligands. The structural analysis showed ligand-coating water networks pronounced to varying degrees. A pronounced water network clearly correlated with a favorable enthalpic and less favorable entropic term, and overall resulted in an affinity gain.

Based on these results, new P_2' substituents were rationally designed with the aim to achieve stronger stabilization of the adjacent water networks and thereby further increase ligand affinity. First, the quality of the putative water networks was validated using molecular dynamics (MD) simulations. Subsequently, the proposed ligands were synthesized,

crystallized in complex with TLN, and analyzed thermodynamically. Additionally, a kinetic characterization using surface plasmon resonance (SPR) was performed. The crystallographically determined water networks adjacent to the P_2' substituents were in line with their predictions conducted by MD simulations. The ligands showed increasingly pronounced water networks as well as a slight enthalpy-drive affinity increase compared to the ligands from the initial study. The ligand with the highest affinity showed an almost perfect water network as well as a significantly reduced dissociation constant.

To analyze the influence of the ligand-coating water networks on the kinetic binding properties of a ligand, seventeen congeneric TLN ligands exhibiting different P_2' groups were kinetically (by SPR) and crystallographically characterized. The different degree of the water network stabilization showed only a minor influence on the binding kinetic properties. By contrast, the strength of the interaction between the ligand and Asn112 proved crucial for the magnitude of the dissociation rate constant. A strong interaction resulted in a considerably prolonged residence time of the ligand by hindering TLN to undergo a conformational transition that is necessary for ligand release.

In the last study, the reason for the exceptionally high affinity gain for addressing the deep, apolar S_1' pocket of TLN with apolar ligand portions was investigated. Therefore, a congeneric TLN ligand series substituted with differently large apolar P_1' substituents (ranging from a single hydrogen atom to an *iso*-butyl group) was analyzed. The exchange of the hydrogen atom at the P_1' position with a single methyl group already results in a 100-fold affinity increase of the ligand. To elucidate the molecular mechanism behind this considerable affinity gain, the solvation state of the S_1' pocket was carefully analyzed. The results strongly indicate that the S_1' pocket is completely free of the presence of any water molecules. Thus, the huge affinity gain was attributed to the absence of an energetically costly desolvation step.

The data presented in this thesis show that to describe the thermodynamic signature of the hydrophobic effect it is necessary to explicitly consider the change of the thermodynamic properties of every involved water molecule. Solely considering the buried apolar surface area and assigning an entropic term to it is not sufficient. The increasing stabilization of the water network adjacent to the protein-bound ligand represents a promising approach — quite independent of specific properties of the target protein — to optimize the thermodynamic profile of a given ligand. This approach also allows fine-tuning of the kinetic binding parameters.

Zusammenfassung

Wassermoleküle nehmen neben Protein und Ligand als zusätzlicher Bindungspartner an jedem *in vivo* Protein–Ligand Bindungsprozess teil. Die Verdrängung von Wassermolekülen von apolaren Oberflächen gelöster Moleküle in die umgebende flüssige Phase wird als treibende Kraft des hydrophoben Effekts angesehen. Die gängige Annahme ist, dass die Beweglichkeit der Wassermoleküle durch ihre Verdrängung erhöht wird, was einen Anstieg der Entropie zur Folge hat. Diese Erklärung, die auf Experimenten mit einfachen Modellsystemen beruht, reicht jedoch nicht aus, um den hydrophoben Effekt als Teil eines hochgradig komplexen Protein–Ligand Komplexbildungsprozesses zu beschreiben. So kann zum Beispiel die Verdrängung von Wassermolekülen von apolaren Oberflächen, die bereits vor der Verdrängung eine erhöhte Beweglichkeit aufweisen, zu einem enthalpischen Vorteil führen. Desweiteren muss berücksichtigt werden, dass durch die Bildung des Protein–Ligand Komplexes eine neue, zur Wasserphase hin exponierte Oberfläche ausgebildet wird, um die sich Wassermoleküle neu anordnen müssen. Diese Arbeit behandelt den Einfluss dieses Effektes auf die thermodynamische und kinetische Bindungseigenschaft eines Liganden.

Um den Einfluss der Neuanordnung von Wassermolekülen um die Oberfläche eines neu gebildeten Protein–Ligand Komplexes auf das thermodynamische Bindungsprofil eines Liganden zu ermitteln, wurde eine homologe Serie aus neun Liganden, die an das Modellprotein Thermolysin (TLN) binden, analysiert. Die Protein–Ligand Komplexe wurden strukturell mittels Röntgenkristallographie und thermodynamisch mittels isothermaler Titrationskalorimetrie (ITC) charakterisiert. Der einzige strukturelle Unterschied zwischen den Liganden stellte deren strikt apolarer P_2' Substituent dar, dessen Größe von einer Methylgruppe bis zu einer Phenylethylgruppe variiert wurde. Die P_2' Gruppen adressieren die flache, apolare und gut solvatisierte S_2' Tasche von TLN. Je nach gebundenem Ligand ändert sich die Form der wasserexponierten Oberfläche des Protein–Ligand Komplexes. Die ITC Messungen ergaben starke thermodynamische Unterschiede zwischen den einzelnen Liganden. Die strukturelle Analyse zeigte eine unterschiedlich starke Ausprägung der die gebundenen Liganden überziehenden Wassernetzwerke. Ein ausgeprägtes Wassernetzwerk korrelierte deutlich mit einem günstigen enthalpischen und einem ungünstigen entropischen Beitrag, und insgesamt mit einem Affinitätsanstieg.

Auf diesen Ergebnissen aufbauend wurden zusätzliche P_2' Substituenten rational entworfen mit dem Ziel eine stärkere Stabilisierung der angrenzenden Wassernetzwerke zu erreichen

und dadurch die Ligandaffinität weiter zu erhöhen. Zunächst wurde die Qualität der putativen Wassernetzwerke mittels molekulardynamischer (MD) Simulationen validiert. Anschließend wurden die vorgeschlagenen Liganden synthetisiert, im Komplex mit TLN kristallisiert und thermodynamisch analysiert. Eine kinetische Charakterisierung mittels Oberflächenplasmonresonanzspektroskopie (SPR) wurde ebenfalls durchgeführt. Die an die P_2' Substituenten angrenzenden, kristallographisch bestimmten Wassernetzwerke entsprachen den Vorhersagen der MD Simulationen. Die Liganden zeigten ausgeprägtere Wassernetzwerke sowie eine leichte, enthalpiegetriebene Affinitätserhöhung im Vergleich zu den Liganden der ersten Studie. Der Ligand mit der höchsten Affinität zeigte ein nahezu perfektes Wassernetzwerk sowie eine signifikant reduzierte Dissoziationskonstante.

Um den Einfluss der ligandüberziehenden Wassernetzwerke auf die kinetischen Bindungseigenschaften eines Liganden zu untersuchen, wurden 17 homologe TLN Liganden mit unterschiedlichen P_2' Gruppen kinetisch mittels SPR und kristallographisch charakterisiert. Die Wassernetzwerkstabilisierung zeigte nur einen geringen Einfluss auf die Bindungskinetik. Im Gegensatz dazu erwies sich die Stärke der Interaktion zwischen Ligand und Asn112 als bestimmend für die Größe der Dissoziationskonstante. Eine starke Wechselwirkung führte zu einer deutlichen Verlängerung der Aufenthaltszeit des Liganden durch die Erschwerung einer konformativen Änderung von TLN, die jedoch notwendig für die Freisetzung des Liganden ist.

In der letzte Studie wurde die Ursache für den außergewöhnlich großen Affinitätsanstieg bei einer Adressierung der tiefen, apolaren S_1' Tasche von TLN mit apolaren Ligandgruppen untersucht. Dafür wurde eine homologe TLN Ligandserie mit unterschiedlich großen, apolaren P_1' Gruppen (von einem Wasserstoffatom bis zu einem Isobutylrest) analysiert. Bereits der Austausch des Wasserstoffatoms an der P_1' Position durch eine Methylgruppe führte zu einem 100-fachen Affinitätsanstieg des Liganden. Um den molekularen Mechanismus hinter dieser Affinitätssteigerung aufzuklären, wurde die S_1' Tasche genau auf ihren Solvatationszustand hin untersucht. Die Ergebnisse sprechen deutlich dafür, dass die S_1' Tasche komplett frei von Wassermolekülen ist. Die Affinitätssteigerung wurde deshalb auf das Fehlen eines energetisch kostspieligen Desolvatationsschrittes zurückgeführt.

Die im Rahmen dieser Arbeit diskutierten Daten legen dar, dass für die thermodynamische Beschreibung des hydrophoben Effekts die Änderung des thermodynamischen Zustandes jedes einzelnen involvierten Wassermoleküls explizit berücksichtigt werden muss. Die alleinige Berücksichtigung der vergrabenen Oberfläche und die Zuweisung eines entropischen Beitrags ist nicht ausreichend. Die Erhöhung der Stabilisierung der Wasserstruktur, die angrenzend an einen proteingebundenen Liganden ausgebildet wird, stellt ein generelles und planbares — und vom Zielprotein relativ unabhängiges — Konzept dar, um das thermodynamische Profil eines gegebenen Liganden zu optimieren. Zu gewissem Maße kann dadurch auch Einfluss auf die Bindungskinetik genommen werden.

Contents

Abbreviations

$\Delta C_\mathrm{p}°$	change in standard heat capacity at constant pressure
$\Delta G°$	change in standard Gibbs free energy
$\Delta H°$	change in standard enthalpy
$\Delta S°$	change in standard entropy
τ	residence time
Å	Ångström (1 Å = 10^{-10} m)
ACES	N-(2-acetamido)-2-aminoethanesulfonic acid
ADP	atomic displacement parameter
clogP	calculated logarithm of the n-octanol/water partition coefficient
CW	capping water
DA	dummy atom
DCM	dichloromethane
DEAD	diethyl azodicarboxylate
DIPEA	N,N-diisopropylethylamine
DMSO	dimethyl sulfoxide
DPPA	diphenyl-phosphoryl azide
e^-	electron
EC number	Enzyme Commission number
EDC	1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide
EEC	enthalpy/entropy compensation
EI-MS	electron ionization mass spectrometry
EP	endothiapepsin
ESI-MS	electron spray ionization mass spectroscopy
FBLD	fragment-based lead discovery
FOM	figure of merit
GOL	glycerol
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOBt	hydroxybenzotriazole
HPLC	high-performance liquid chromatography
HTS	high throughput screening

ITC	isothermal titration calorimetry
K_a	association constant at equilibrium
k_a	association rate constant
K_d	dissociation constant at equilibrium
k_d	dissociation rate constant
kDa	kilodalton
kJ	kilojoule
k_{off}	dissociation rate constant
k_{on}	association rate constant
LE	ligand efficiency
MAD	multi-wavelength anomalous diffraction
MD	molecular dynamics
MES	2-(N-morpholino)ethanesulfonic acid
MOPSO	3-morpholino-2-hydroxypropanesulfonic acid
MPD	methyl-2,4-pentanediol
MX	macromolecular X-ray crystallography
n-BuLi	n-butyllithium
NMR	nuclear magnetic resonance
PDB	protein data bank
PEG	polyethylene glycol
PIPES	piperazine-N,N-bis(2-ethanesulfonic acid)
R	universal gas constant (8.314 J mol^{-1} K^{-1})
RCF	relative centrifugal force
RMSD	root-mean-square deviation
RU	resonance units
SASA	solvent accessible surface area
SD	standard deviation
SE	standard error
SPR	surface plasmon resonance
TLN	thermolysin
tris	tris(hydroxymethyl)aminomethane
vdW	van der Waals
W	water molecule

Chapter 1

**General Introduction
and Thesis Outline**

1.1 Molecular recognition in rational drug design

Despite many years of research and considerable efforts in the field of rational computer-aided drug design, it is still challenging to predict from scratch how molecules have to be composed to bind to proteins of a known three-dimensional structure [1, 2]. Likewise, the prediction of the affinity of a known protein–ligand complex structure using computational methods is unreliable and not precise [3–5]. A virtual screening campaign is usually already defined as successful if 10% of the proposed compounds show weak activities in a bioassay [5]. The reason is that molecular recognition (or more specifically in the context of drug design, protein–ligand recognition) is a very complex topic that involves a multitude of steps and interactions ultimately determining the binding affinity and binding mode. Many of these effects compensate or amplify each other, rendering predictions of net energies — and particularly the binding thermodynamics — difficult. It is now generally accepted that water needs to be considered as a third "active player" in the protein–ligand recognition process [6–9]. Life on earth is impossible without water and it is present in every organism [10, 11]. Water molecules participate in every *in vivo* protein–ligand binding event: in the simplest case, they need at least to be displaced from the binding interface of protein and ligand to make space for the association process [12]. Unfortunately, the involvement of water molecules tremendously increases the complexity of the protein–ligand recognition process. However, in rational computer-aided drug design, simply ignoring water or applying over-simplified water models will result in false predictions [13]. To achieve progress in the field of rational computer-aided drug design and improve currently applied computational approaches, more high-quality, reliable and interpretable experimental studies are necessary as a basis to develop a profound knowledge and deeper understanding of protein–ligand recognition processes [14].

1.2 Protein–ligand recognition and the strength of complex formation

Protein–ligand recognition describes the process of complex formation between a ligand (e.g. a drug molecule) and its target protein via specific, non-covalent interactions [15, 16]. This process holds fundamental importance for many biological processes (e.g. for enzyme regulation) and thus life, as well as the mode-of-action of most drug molecules. The magnitude of the binding affinity between protein and ligand is defined highly specific at the molecular level [17]. It can be described thermodynamically under equilibrium conditions as the dissociation constant K_d, which is given as the concentration ratio between the uncomplexed protein and ligand to the formed protein–ligand complex. The dissociation constant is directly related to the change in the Gibbs free energy (ΔG) via Eq. 2.1, Chapter 2.

A prerequisite for a protein–ligand complex actually being formed is a negative Gibbs free energy change ΔG, which means that the free energy must decrease upon complex formation. The change in the Gibbs free energy partitions into enthalpic (ΔH) and entropic ($-T\Delta S$) energy changes (as described by Eq. 2.2, Chapter 2). As least one of them must be favorable upon complex formation, rendering the overall process (ΔG) favorable. The enthalpy of binding describes the change of the internal energy and is determined by the difference of non-bonded interactions (mainly hydrogen bonds, hydrophobic interactions, ionic interactions and metal complexation) between the uncomplexed and the complexed state of protein and ligand [18, 19]. The binding entropy — which describes how evenly the energy is distributed over the system — is influenced by changes in the conformational, translational and rotational degrees of freedom of protein and ligand upon complexation, and particularly by solvent effects [18, 19].

1.3   Why water is so important

Water is the solvent in which all biological reactions occur. However, the role of water extends much further than simply being an excellent solvent. Water is indispensable for protein folding, maintaining protein stability and dynamics as well as protein function [6, 20, 21]. The first layer of water molecules adjacent to a protein surface is 10–20% denser than the bulk phase and these waters show a longer residence time than bulk water molecules [6]. In a high resolution crystal structure, a hydrogen bond connected network of water molecules is observed spanning over the surface of the protein [22]. This network of water molecules clearly also plays a major role in protein–ligand recognition. Water molecules make such a significant contribution because they have a large dipole moment and thus they will be involved in electrostatic interactions and, most importantly, hydrogen bonds [8, 21, 23]. Water molecules can act as either hydrogen bond donors or acceptors, establishing up to four hydrogen bonds simultaneously. As a result of the constrained hydrogen-bonding geometry and enabled by their small size, they arrange in the energetically most favorable three-dimensional structures connected by a network of hydrogen bonds. In the bulk water phase, this is a highly fluxional, approximately tetrahedral arrangement [23] in which the number of hydrogen bonds is maximized. Experimental data suggests that on average 3.5 hydrogen bonds are established in bulk water. However, this number remains subject to dispute [23]. Furthermore, the existence of entropy-maximized, hydrogen bond distorted regions in the bulk water phase at ambient conditions has been reported [24]. This intuitively makes sense since either enthalpy or entropy can favorably contribute to an overall favorable Gibbs free energy state (Eq. 2.2, Chapter 2) and further illustrates the complexity of the subject of "water" in this process.

1.4 Thermodynamic dissection of the protein–ligand complex formation process

In the following, the protein–ligand recognition process is briefly depicted qualitatively with respect to the energetically favorable and unfavorable steps involved, along with their partitioning in enthalpy and entropy. The aim is to illustrate the inherent complexity of protein–ligand binding.

Upon binding, the conformation that the ligand has to adopt can be different from the energetically minimum that the ligand adopts in the solvated, uncomplexed state. Often an enthalpically less favorable conformation has to be adopted, roughly fitting to the shape of the protein binding pocket. Furthermore, conformational degrees of freedom of ligand and protein are sacrificed upon complex formation, resulting in an unfavorable entropic term [25], aside from a loss given that two independent particles merge into one. This can be attenuated by an appropriate ligand pre-organization [26, 27]. Pre-organized ligands are more rigid and already adapt in solution a conformation resembling the conformation required in the bound state. Thus, less conformational degrees of freedom are lost upon binding, which indirectly increases affinity. However, data suggest that the explanation is not that simple: ligand pre-organization can also be entropically counterproductive and enthalpically beneficial due to the formation of stronger polar interactions [28, 29].

In due course of ligand binding to a protein, both binding partners have to shed part of their solvation shell before covering the binding interface. Subsequently, after the formation of the protein–ligand complex, water molecules re-arrange around the newly-formed, solvent-exposed surface. Desolvation and solvent effects hold utmost importance for ligand affinity (solvent effects with respect to the hydrophobic effect are discussed in Chapter 1.4.1). For instance, to determine the net free energy change of the formation of a hydrogen bond between ligand and protein, it is important to consider that polar groups of ligand and protein already establish hydrogen bonds to water molecules in their unbound states. These hydrogen bonds must be broken before binding. The thermodynamic consequence for the displacement of these bound water molecules is highly context dependent. For instance, displacing tightly bound water molecules (establishing 3–4 hydrogen bonds) will most likely result in a loss in enthalpy (partly compensated by a gain in entropy [30]) and can lead overall to an unfavorable binding free energy. By contrast, the displacement of loosely bound water molecules is less costly or even energetically favorable [31–33]. In conclusion, the hydrogen bond established between ligand and protein must (over-)compensate for any unfavorable desolvation effects, which often requires optimal geometry and distance of the hydrogen bond between protein and ligand. Suboptimal geometry or distance can result in no net increase or even a decrease in binding affinity [34]. However, even if the formation of a hydrogen bond between protein and ligand over-compensates unfavorable desolvation

effects, it does not necessarily translate into a significant gain in affinity. The reason is that the establishment of enthalpically favorable interactions can be paralleled by an entropically unfavorable term, resulting in a hardly changed overall binding free energy. An intuitive explanation for this phenomenon was provided by Dunitz, indicating that a stronger fixation (enthalpic beneficial) automatically results in a lowered mobility and thus a decrease of entropy [35]. This phenomenon is well known as enthalpy–entropy compensation [36, 37] and must be overcome to achieve high-affinity ligands [34, 38].

Another aspect making it so difficult to describe molecular interactions is that they are frequently non-additive [39–41]. Accordingly, this means that most interactions established between ligand and protein influence each other in their magnitude. In this context, the involvement of water molecules also plays an important role [42]. Due to non-additivity, it is difficult to assign contributions to individual interactions. Moreover, it also makes the contribution of functional groups to the binding free energy highly context dependent. Depending on how a ligand is already substituted, the introduction of new functional groups will have a different strong impact on the overall binding affinity.

1.4.1 The hydrophobic effect(s)

The hydrophobic effect describes the phenomenon whereby non-polar solutes aggregate with each other when exposed to an aqueous solution. As the driving force of the association process, it discusses the displacement of adjacent water molecules from apolar solutes' surfaces into the bulk water phase [43, 44]. In the course of a protein–ligand complex formation, hydrophobic ligand portions are frequently buried in nonpolar protein pockets. The contribution of this burial to the change in Gibbs free energy of binding is estimated to be on average about 75% [45]. Thus, it is *the* driving force of most protein–ligand association processes. The prerequisite that water is displaced from an apolar surface is steric pressure; for instance, by a ligand that has a "shape complementary" to the binding pocket of a protein. The better the steric complementarity, the more water molecules can be released from the surfaces. Shape complementarity is described by several models [19], the oldest one being the "lock and key" model proposed by Emil Fischer [46]. More modern ones follow the idea of an "induced fit" model [47] and a "conformational selection" model [48]. The newer models describe that the shape of the protein binding pocket can adapt upon ligand binding to some extent with respect to its *apo* structure, ranging from simple loop movements to whole hinge bending motions [19, 49]. The "induced fit" model describes that this conformational change is actually induced by the binding of the ligand, whereas in the "conformational selection" model the ligand binds selectively to one "active" protein conformation out of many conformations that the uncomplexed protein is already adopting in solution, thereby shifting the equilibrium in the direction to adopt more of the binding-competent conformation.

In the "classical" model of the hydrophobic effect, the displacement of water molecules from apolar surfaces at room temperature is responded by a favorable increase in entropy, which translates into a decrease of free energy [18, 43, 44, 50, 51]. An explanation for this observation is provided by the "iceberg model", which describes the water structure adjacent to apolar solutes as ice-like and more ordered compared to the arrangement in the bulk solvent. A cage of water molecules (clathrate) is formed around the hydrophobic solute. These water molecules are stabilized in a conformation where they can establish strong hydrogen bonds to other water molecules. As a result, they are comparable to bulk phase water molecules with respect to enthalpy, although they are entropically unfavorable due to reduced mobility. Upon release into the bulk — for instance, due to the reduction of the solvent-exposed surface owing to the aggregation of molecules — the mobility of these water molecules and thus entropy increases, overall decreasing the free binding energy. However, this explanation is increasingly questioned, e.g. by neutron diffraction experiments [52–54], which do not confirm that water molecules adopt a more ordered state adjacent to apolar surfaces [45].

Besides the "classical" hydrophobic effect, an enthalpically-driven hydrophobic effect — termed "non-classical" hydrophobic effect — has been observed in protein–ligand binding [31, 39, 55, 56]. This is frequently associated with the presence of mobile ("disordered"), enthalpically unfavorable water molecules that lack favorable hydrogen-bonding options within a hydrophobic binding pocket [57–59]. Upon displacement into the bulk, the mobility of such water molecules remains the same (and thus does entropy), although the displaced water molecule can now establish favorable hydrogen bonds to other bulk water molecules. Thus, enthalpy increases upon displacement. Another explanation is suboptimally hydrated binding pockets, where solvent effects are relatively small or even neglectable. Upon complex formation, the thermodynamic profile is thus mainly determined by (enthalpic) van der Waals interactions [56, 60].

It is increasingly realized that the hydrophobic effect as part of highly complex protein–ligand interactions cannot be simply categorized into "classical" or "non-classical", and that model studies investigating small and simple systems, e.g. exhibiting solely a convex, concave, or flat surface shape [61, 62], cannot be transferred to comprehensively describe the hydrophobic effect as observed in protein–ligand binding [45, 63]. By contrast, data suggest that the hydrophobic effect is determined by subtle structural differences of the water networks adjacent to protein and ligand before and after complex formation [45, 64–66]. The structure of the water network is determined by the surface topography of protein and ligand and the formed protein–ligand complex, thus being context dependent. Upon complexation, some water molecules will become displaced into the bulk solvent, some will re-arrange around the newly-formed surface of the protein–ligand complex and in some cases new water molecules will be recruited from the bulk solvent to form a more elaborate network of

water molecules. Consequently, the enthalpy/entropy partitioning of the hydrophobic effect is determined by the net difference of the thermodynamic profiles of all involved water molecules before and after complex formation [66].

1.5   Techniques to experimentally analyze protein–ligand interactions

Nowadays, many sophisticated biophysical methods are available for the experimental characterization of protein–ligand interactions [67]. All of them provide the experimenter with different information about the protein–ligand recognition process and they have their individual strengths and weaknesses. This must be considered in experiment design, data analysis and interpretation. Furthermore, it must be considered whether experimental data are used to develop and validate computational methods for the simulation of protein–ligand interactions. In the following, selected techniques out of the many available ones and their key features with respect to studying protein–ligand interactions are briefly summarized.

The most widely-employed method for determining protein structural information (the "workhorse" of structural biology) is **macromolecular X-ray crystallography** [68, 69]. This technique has allowed invaluable insights into the structure of proteins as well as protein–ligand complexes [70, 71]. Even though the data quality of crystal structures has overall remained relatively constant during recent years [72], the improvement of synchrotrons — for instance the development of brighter light sources, faster detectors, robots for automatic crystal mounting, improved software for diffraction data collection and processing — and the emergence of automatic refinement pipelines has dramatically reduced the time required for data acquisition and crystal structure determination [73–76]. This renders the application of X-ray crystallography feasible as a primary screening technique, for instance, in a fragment-based drug discovery project [76–78]. Nevertheless, prerequisites for a high-resolution dataset include well-diffracting crystals, which can be very demanding or even impossible to obtain, for instance, in the case of membrane proteins like G protein-coupled receptors (GPCRs) [79]. The primary information resulting from a diffraction experiment is the electron density [69], which reports the average electron distribution at defined locations in the unit cells. Only sufficiently ordered and periodically arranged atoms result in clearly defined electron density signals. By contrast, electron densities from mobile, "disordered" atoms — for example, water molecules in solvent channels in crystals or density from the superposition of different molecules — are difficult or often impossible to interpret. It must be considered that the finally assigned static crystal structure model deposited in the PDB [80] is an interpretation of the obtained electron density by the crystallographer and thus it can potentially contain incorrectly modeled or biased features [81, 82]. For instance, water

molecules can be placed into densities that are actually ions, or they can even be modeled into "noise". Just recently, a case was discussed in which a ligand was apparently modeled into noise [83, 84]. Obviously, the use of such erroneous models can result in incorrect conclusions and a waste of efforts and resources. Inspection of the electron density and the validation report, both available from the PDBe server [85], can help to assess the quality of putatively incorrectly or ambiguously modeled atoms. The higher the resolution of the dataset, the more likely it is that the assigned crystal structure is unbiased and correctly modeled [81, 86]. The resolution of the dataset also determines the amount of details that can be observed (alternative conformations, the amount of visible water molecules, partially occupied molecules) and determines the accuracy of the assigned atom coordinates [81]. Hydrogen atoms are usually not observed in the electron density and thus not modeled, or their positions in the model are based solely on theoretical considerations. The reason is that hydrogen atoms usually contribute to the diffraction signal very weakly (because they have only one electron), and cannot be distinguished from background noise [87]. Exceptions are ultra-highly resolved datasets [88]. Consequently, the ionization state of ligands and amino acids cannot be determined experimentally and the positions of hydrogens in a crystal structure model have to be assumed [81]. If a crystal structure should be correlated with data obtained from other techniques (e.g. ITC-derived thermodynamic data or SPR-derived binding kinetics), it must be considered that the buffer composition (especially the pH) and the crystallization protocol (co-crystallization versus soaking) can potentially influence the binding mode of the ligand [89]. Furthermore, crystal structures are usually collected at cryogenic temperature at 100 K [90], which can potentially result in artifacts not observed at room temperature [91]. Standard macromolecular X-ray crystallography has the disadvantage that it provides the experimenter with a mainly static picture. Limited information about protein dynamics is accessible through refined B factors, visible alternative conformations or the "missing" of electron density in the crystal structure, indicating highly disordered regions.

Solution nuclear magnetic resonance (NMR) spectroscopy allows studying protein–ligand interactions under conditions close to physiological conditions via protein-observed NMR [67, 92]. Protein-observed NMR records properties like chemical shifts, the transfer of magnetization, coupling constants and relaxation phenomena of protein nuclei and their perturbations resulting from ligand binding. If signals are assigned to individual NMR active protein atoms, it is possible to determine the three-dimensional structure of the protein and the binding site of the ligand. It is also feasible to identify the positions of water molecules [93]. Furthermore, since the protein is dissolved in solution, this method is well suited for studying the dynamics of the protein [92]. In the majority of cases, protein-observed NMR is performed by heteronuclear 2D NMR (^{1}H-^{13}C or ^{1}H-^{15}N). To increase the occurrence of the magnetically active nuclei ^{13}C or ^{15}N, isotope labeling of the protein is necessary.

Furthermore, the size of the protein is usually limited to about 40 kDa [92], since this method is dependent on fast molecular tumbling. Another disadvantage of protein-observed NMR is that it is very labor-intensive, especially the individual protein resonance assignments.

Single crystal neutron diffraction can be applied to determine the position of hydrogen and deuterium atoms in protein structures very accurately [94]. This is possible because neutrons are diffracted by the nucleus of an atom and the scattering power of small atoms like hydrogen or deuterium is quite large compared to the diffraction power of X-rays, which correlates with the number of electrons around an atom. For instance, this allows studying the protonation state of particular molecules, the orientation of water molecules along with their rotational properties, as well as the geometry of hydrogen bond interactions. The exchange of hydrogen by their isotope deuterium (deuteration with D_2O) can increase the diffraction power of the crystal and improve the signal-to-noise ratio due to the reduced inelastic scattering of deuterium [95]. In a neutron diffraction experiment, hydrogen (negative peaks) and deuterium atoms (positive peaks) can be clearly distinguished from each other. This enables studying the hydrogen/deuterium exchange rate of molecules, which, for instance, informs about the solvent-accessibility of protein regions of particular interest. A significant disadvantage of neutron diffraction is the much higher experimental demand compared to X-ray diffraction [94]. To date, neutron sources are not as readily accessible to researchers as X-ray sources. Moreover, due to a low neutron flux of most reactors, 10–100 times larger crystals are necessary. The low flux also results in a long and costly data collection time of up to several weeks or longer.

Isothermal titration calorimetry (ITC) is the first and hitherto only technique that allows the accurate determination of the thermodynamic binding profile (Gibbs free energy, enthalpy and indirectly entropy) of a protein–ligand association process at a given temperature in one experiment [1]. Earlier employed methods — for instance, binding enthalpy determination by van't Hoff analysis (discussed in detail in Chapter 2) — frequently resulted in questionable accuracy [39, 96]. Since the thermodynamic profile determined by ITC reflects the net inventory of all interactions of a protein–ligand complex formation process including any solvent effects, it is almost unavoidable to use **congeneric ligand series** for the partitioning and subsequent determination of thermodynamic contributions of individual functional groups and interactions [39]. In such a series, the parent scaffold of all ligands remains unchanged and only one group is varied; for instance, the size of a hydrophobic portion is systematically increased by the addition of methyl groups. Resulting from the identical parent scaffold, all ligands adopt a similar binding mode at the protein binding site. Thus, the observed differences between the thermodynamic profiles of the ligands can be attributed to the modification of the one particular group. It is important to wisely select a **model system**, which must exhibit the structural properties under consideration (for instance, a solvent-exposed, apolar binding pocket if the hydrophobic

effect should be studied). Preferentially, the model system is as simple (1:1 binding stoichiometry) and rigid as possible. High rigidity allows the exclusion of conformational adaptations of the protein during complex formation; otherwise, any conformational adaptation of the protein would have to be considered for the interpretation of the thermodynamic profile. Furthermore, since ITC requires a considerable amount of protein material, the latter should be readily accessible. Moreover, to determine most accurate thermodynamic parameters, it should be as pure as possible and highly active. A newly-developed application of ITC is "kinITC" [97, 98], which allows extracting binding kinetic information from conventional ITC isotherms. A recent comparative study of kinITC and SPR-derived binding kinetics resulted in a fair to excellent agreement between the two methods, depending on the studied system [98]. Thus, this new method represents an alternative to the binding kinetic parameter determination by SPR, with the advantage that the measurements do not require any immobilization of the protein and are recorded under the same conditions as the thermodynamic analysis.

Surface plasmon resonance (SPR) spectroscopy allows the real-time observation of a protein–ligand binding event and the binding kinetic parameters (K_d, k_{on}, k_{off}, and the residence time $\tau = 1/k_{off}$), as well as the stoichiometry of the binding reaction [99]. Performing experiments at multiple temperatures allows determining van't Hoff-based binding thermodynamic parameters [100]. In the pharmaceutical industry, SPR is one of the most widely-employed methods for (high-throughput) ligand affinity screenings and lead optimization [67, 101]. It is necessary to immobilize the protein on a chip surface, although no labeling of the inhibitor is required. For the immobilization, many different techniques and assay kits are available [102]. Nevertheless, depending on the protein, its stable immobilization on the chip surface can be difficult. However, once immobilized, the amount of protein required for a measurement is significantly smaller than for ITC and after a measurement the immobilized protein can be regenerated and used for further measurements. In contrast to ITC, SPR works under continuous flow and steady-state conditions, which can result in an inherent offset between the measurement results of the two methods. If a binding event is very rapid, the rate-limiting step can be confined by mass transportation, rendering the measured k_{on} slower than is actually the case [103]. Kinetic binding parameters can further be artificially altered due to protein immobilization (e.g. sterically blockage of the active site) and unspecific binding of the inhibitor to the chip surface [103].

1.6 Thesis outline

This thesis can be divided into three parts. The first part (Chapters 1 and 2) serves as an introduction to protein–ligand recognition and to the biophysical methods commonly employed for the experimental analysis of protein–ligand interactions. A focus is placed on ITC (Chapter 2). In the second, main part of this thesis (Chapters 3–5), congeneric phosphonamidate ligands binding to the model protein thermolysin are analyzed with the aim to obtaining further insights into the hydrophobic effect and its impact on the thermodynamic and binding kinetic properties of the protein–ligand complex formation process (see Chapter 8.1.1 for a discussion of the model protein thermolysin). In this context, especially the influence of the first solvation layer of water molecules arranging around the newly formed surface of a protein–ligand complex on the binding properties of the ligand is carefully elucidated. For this, an interdisciplinary approach was chosen, including ligand synthesis, analysis of the thermodynamic (ITC) and binding kinetic (SPR) parameters, molecular dynamics (MD) simulations and crystal structure determination. As common in multidisciplinary research projects nowadays, this was undertaken in collaboration with other researchers. The obtained data were correlated and interpreted, resulting in a convincing overall picture of the studied effects and the formulation of a working hypothesis, which was subsequently experimentally confirmed. In the third part of this thesis, additionally performed studies are described. In Chapter 6, the molecular origin behind the extraordinary strong affinity gain upon apolar group insertion into the apolar S_1' pocket of thermolysin is analyzed. Chapter 7 describes my contributions to a crystallographic fragment screening of a 361-entry fragment library toward the target protein endothiapepsin.

I obtained the experimental skills and knowledge necessary to write **Chapter 2** (publication [96]) during the work for Chapter 3. I start Chapter 2 by reviewing the basic principles of ITC, discuss which measurement protocol is appropriate under which conditions and point out common sources of error. Provided with this information, one should be able to establish an ITC measurement protocol which results in the best possible parameter accuracy under the given conditions. Furthermore, provided with this information, non-experimenters should also be able to assess the quality of ITC data. A very important part of this chapter is the discussion of the significance of thermodynamic data in different contexts. Whereas the analysis of thermodynamic data on a relative scale obtained for a congeneric series measured under similar conditions makes perfect sense, global (and thus absolute-scale) comparisons of thermodynamic data publicly available in the literature or from online databases has to be considered highly problematic.

Chapter 3 (publication [104]) describes our attempts to obtain further insight into the hydrophobic effect in the context of a protein–ligand complex formation process. For this, we analyzed a congeneric series of nine thermolysin ligands exhibiting apolar P_2' portions of

different size addressing the apolar, solvent-exposed S_2' pocket of thermolysin by X-ray crystallography and ITC. As a model system, the zinc metalloprotease thermolysin was selected. The collected data suggest that the way in which water molecules arrange around a newly formed protein–ligand complex surface has a significant impact on the affinity of the ligand and even more affects the partitioning into enthalpy and entropy. Whereas the stabilization of the water structure results in a beneficial enthalpic contribution to binding, entropy increases if the water network is disrupted. It is very likely that similar effects also apply to other systems beyond the one studied here. However, in most cases, data are simply missing or the data quality is not sufficient for studying such details.

Chapter 4 (publication [66]) is essentially the proof-of-concept study to Chapter 3. According to our hypothesis, it should be possible to optimize the ligand binding profile by increasing the ability of solvent-exposed groups to stabilize the adjacent water network of the first solvation layer. Consequently, based on the data gathered in Chapter 3, we designed new thermolysin ligands exhibiting P_2' groups potentially better suited for water network stabilization. MD simulations validated the expected water networks. Because the MD simulation predictions appeared very promising, the ligands were subsequently synthesized and structurally, thermodynamically and kinetically characterized. Indeed, the new ligands showed more pronounced water networks adjacent to their P_2' groups, reflected in increased affinities. The ligand with the highest affinity showed a virtually perfect water network, composed of three interconnected water clusters. Notably, this ligand also showed a significantly prolonged residence time as measured by SPR, possibly resulting from the additional stabilization and fixation of the ligand in its complex with the protein by the confining water network. We conclude that the ligand optimization strategy to increasingly stabilize the adjacent water network is a feasible approach to optimize thermodynamic and probably also kinetic binding parameters. Even in cases where data quality is poor, the applied MD simulation protocol could help to identify the most suited substituents.

Chapter 5 (publication [105]) describes the results of the kinetic binding parameter analysis by SPR of seventeen congeneric thermolysin ligands. Our collaboration partners at GE Heathcare, Uppsala, Sweden helped in the kinetic analysis. I had the opportunity to visit their site and participate in the measurements, benefiting from my experience with thermolysin. The gathered kinetic data allowed us to formulate a mechanism for the protein–ligand complex dissociation process. Especially the role of Asn112 was elucidated. Its steric restriction in a conformational step required for ligand release results in a dramatic prolongation of the residence time. Furthermore, similar to the first observed hints described in Chapter 4, data suggest that a well-established water network adjacent to the bound ligand can prolong its residence time.

Chapter 6 elucidates the role of a dry, non-solvated binding pocket on the protein–ligand recognition process. The S_1' pocket of thermolysin is a deep, largely apolar pocket. Upon insertion of apolar ligand portions into this pocket, an extraordinary high, enthalpically driven affinity increase is observed. To understand the molecular mechanism behind this affinity boost, exact knowledge of the pocket solvation state is essential. Applying routine crystallography, the pocket appears empty. However, the strong enthalpic signal could potentially be the result of the displacement of a highly mobile water molecule. Such a water molecule would elude conventional crystallographic detection. To unambiguously clarify the molecular mechanism behind the remarkable affinity boost, we performed an in-depth analysis of a MAD phased electron density, which allows tracing down even mobile water molecules. Unexpectedly, we found no evidence for the existence of a highly mobile water molecule. The gathered data strongly suggest that the S_1' pocket is actually empty. We conclude that due to the missing of a desolvation step particularly favorable van der Waals interactions are established, resulting in the considerable affinity boost.

The results presented in **Chapter 7** have been incorporated into four different publications [76, 106–108]. In this chapter, I describe how the aspartic protease endothiapepsin was used as a model protein to crystallize the enzyme with a 361-entry fragment library. Due to the heavy workload, several researchers had to work as a team on this project. Out of these 361 entries, initially 64 fragment binding modes were discovered. Nine of these fragments were discovered and refined by me. During the refinement, I noticed that an elaborate refinement of a crystal structure results in a strong increase in the quality of the electron density. This makes it much easier and in some cases makes it possible at all, to identify weakly bound, potentially partly occupied fragments. Stimulated by this observation, a second, this time automated round of elaborate crystal structure refinement was performed, followed by investigation for further putatively bound fragments. Thereby, it was possible to discover additional 22 fragment binding poses.

In **Chapter 8,** I relate the key findings of the previous chapters to each other and create a coherent picture of these studies and their results.

Chapter 2

Thermodynamics of protein–ligand interactions as a reference for computational analysis: How to assess accuracy, reliability and relevance of experimental data

Stefan G. Krimmer,[1] and Gerhard Klebe[1]

[1]Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

2.1 Abstract

For a conscientious interpretation of thermodynamic parameters (Gibbs free energy, enthalpy and entropy) obtained by isothermal titration calorimetry (ITC), it is necessary to first evaluate the experimental setup and conditions at which the data were measured. The data quality must be assessed and the precision and accuracy of the measured parameters must be estimated. This information provides the basis at which level discussion of the data is appropriate, and allows insight into the significance of comparisons with other data. The aim of this article is to provide the reader with basic understanding of the ITC technique and the experimental practices commonly applied, in order to foster an appreciation for how much measured thermodynamic parameters can deviate from ideal, error-free values. Particular attention is paid to the shape of the recorded isotherm (c-value), the influence of the applied buffer used for the reaction (protonation reactions, pH), the chosen experimental settings (temperature), impurities of protein and ligand, sources of systematic errors (solution concentration, solution activity, and device calibration) and to the applied analysis software. Furthermore, we comment on enthalpy–entropy compensation, heat capacities and van't Hoff enthalpies.

2.2 Introduction: basic thermodynamic relationships

Many computational approaches make use of thermodynamic properties. Most important among all is the binding affinity, usually the target property used for scoring and ranking solutions generated in computational docking simulations and key to all virtual screening applications [109]. However, what kind of a property is "affinity" and how do we obtain experimental information to characterize this property? How good is the quality of the experimental data usually consulted to describe the affinity of a compound, how is its precision and accuracy, particularly if such data are intended for further usage in the development of computational models [110–113]? The aim of this article is to provide foundations necessary to understand which experimental protocols are commonly applied to perform an isothermal titration calorimetry (ITC) measurement and how critically different setups can influence the recorded binding parameters.

The affinity of a ligand binding to its target protein is described by the change in the Gibbs free energy of the system before and after the binding event. Only changes in the Gibbs free energy are detectable, whereas absolute values for individual states cannot be measured. Once equilibrium is attained for the reaction between protein 'P' and ligand 'L' forming the protein–ligand complex 'PL', $P+L \rightleftharpoons PL$, the association constant K_a (L mol^{-1} or M^{-1}) describes the ratio between the concentration of the protein–ligand complex [PL] and the product of the free protein [P] and free ligand [L] concentrations: $K_a=[PL]/[P][L]$. In contrary, the dissociation equilibrium constant K_d (mol L^{-1} or M) is the inverse of the association constant K_a, i.e. $K_d=[P][L]/[PL]$.

$$\Delta G° = -RT \ln K_a \qquad (2.1)$$

$$\Delta G° = \Delta H° - T\Delta S° \qquad (2.2)$$

As described in Eq. 2.1, at equilibrium the Gibbs free energy of binding $\Delta G°$ (kJ mol^{-1}) is logarithmically related to the association constant K_a, weighted by the ideal gas constant R (8.314 J mol^{-1} K^{-1}) and the absolute temperature T (K). It consists of two components (Eq. 2.2): a change in enthalpy $\Delta H°$ (kJ mol^{-1}) and a change in entropy $\Delta S°$ (kJ mol^{-1}), the latter weighted by the absolute temperature. The change in enthalpy describes the amount of heat released (exotherm, negative $\Delta H°$) or absorbed (endotherm, positive $\Delta H°$) as bonds and intermolecular contacts (e.g. hydrogen bonds, electrostatic interactions, van der Waals contacts) are established and broken between protein, ligand, water and other buffer components resulting in the formation of a protein–ligand complex. The difference in entropy describes the change in ordering parameters and the distribution of the system over multiple accessible states. A positive $\Delta S°$ describes an increase in entropy and thus an increase in disorder and in the number of accessible states. The change in entropy is not only

related to conformational changes of the ligand and the protein, but for instance also to the water molecules which play a major role in the binding process. A classic example is the displacement of water molecules from apolar surfaces and the related increase in entropy, which is considered to be the driving force of association in the so-called hydrophobic effect [44]. It should be mentioned that $\Delta G°$, $\Delta H°$ and $\Delta S°$ are all state functions – their values depend only on the two thermodynamic equilibrium states referred to, and not on the route by which these states are accessed.

The superscript '°' (pronounced "naught") is attached to indicate that the binding free energy value refers to its standard state. However, this sign is frequently omitted. The necessity for referring to a standard state is to achieve comparability between measurements on the same scale. At standard state, the binding free energies are described for the conversion of 1 mole protein and 1 mole ligand to 1 mole of protein–ligand complex, in a hypothetical ideal solution (infinitely diluted), with a unit activity coefficient at a constant pressure of $p° = 10^5$ Pa. The temperature is not part of the standard state and therefore has to be specified. While K_a and $\Delta H°$ are determined experimentally in an ITC experiment (see below), $\Delta G°$ is calculated according to Eq. 2.1. This requires the use of the natural logarithm of K_a, which makes it necessary to convert K_a to a unitless value. To achieve this, the standard concentration $c°$ is used, which is by convention 1 M. Depending on the reference concentration scale (e.g. M, mM, µM), the magnitude of the calculated ΔG changes. For example, for a K_a of 10^6 M^{-1} and a reference concentration of 1 M, the result for ΔG is $-13.8 \times RT$, calculated from $-RT \times \ln(10^6$ M$^{-1} \times 1$ M$)$. On the other hand, for the same K_a, applying a reference concentration of 1 mM, ΔG results in $-6.9 \times RT$, calculated from $-RT \times \ln(10^3$ mM$^{-1} \times 1$ mM$)$. Consequently, it is necessary to specify the reference concentration applied, which in the case of the standard state is $c° = 1$ M. With this information, the standard entropy change $\Delta S°$ is calculated according to Eq. 2.2.

At this point, the first approximation must already be regarded. In principle, we try to describe the number of particles actively involved in the considered equilibrium by the "concentrations". However, this is only correct if we are dealing with so-called ideal solutions, which correspond to infinite dilutions. Real solutions deviate in their actual concentrations and instead we would have to consider "activities" which usually correspond to a smaller number of particles compared to the theoretically achievable concentrations [114]. In biological systems only a small number of validation studies have been performed to estimate how "ideal" the investigated solutions really are. It has been suggested to perform ITC investigations at different concentrations to estimate the extent to which the measured properties are affected. In one reported study the binding of 2'-cytidine monophosphate to ribonuclease was investigated [115]. Deviations in the binding constants as large as 40% were reported by increasing the used protein concentration from diluted 0.0145 mM to more concentrated 0.65 mM, which corresponds to a 44-times higher protein concentration.

Further down in this article, we will describe another measurement from our own research giving an idea by how much the thermodynamic signature can vary on an absolute scale with concentration. In practical application, an appropriate amount of protein is dissolved in a buffer by the experimentalist and the resulting thermodynamic properties are referenced to this "concentration". As long as data are compared relative to each other across a series using unchanged protein "concentrations" (better: "activities"), data interpretation will unlikely be strongly affected. However, if data are taken from different proteins and measured at largely deviating concentrations, analysis on an absolute scale can easily become quite problematic.

2.3 Which energetic contributions of the protein–ligand binding reaction are measured by ITC?

Isothermal titration calorimetry (ITC) allows the determination of K_a, $\Delta G°$, $\Delta H°$, $\Delta S°$ and n of a binding event in a single experiment at one given measurement temperature. This is done without any need for labeling by simply measuring heat changes related to a reaction [115–119]. By performing the measurement at varying temperatures, the heat capacity change ΔC_p can also be obtained. This review will focus on the thermodynamics measured by ITC as a source of experimental information about protein–ligand interactions, assuming a single-site 1:1 binding event without major conformational changes of the target protein. It is important to note that ITC records the entire binding event, starting with the separately solvated binding partners (ligand and protein), and detects any alteration giving rise to a heat signal until the formation of the final complex. The process is affected by all changes involving the surrounding buffer, conformational transitions and, importantly enough, modulations of the solvation structure. The picture produced becomes quite complex as many steps on the molecular level can compensate in their thermodynamic signature and thus make it extremely difficult to factorize the ITC results into the discrete contributions of each separate interaction formed between a ligand and its target protein. Correlation of thermodynamic parameters with structural features [29], for example those obtained by X-ray crystallography, must therefore be performed very carefully. This often requires the interpretation within a narrow ligand series involving small variations, for instance the exchange of a moiety, a functional group or even only the addition or removal of a single methyl group [39]. The small variations between two ligands can then be attributed to the observed changes in their thermodynamic profiles. Conversely, unchanged thermodynamic signatures of two closely related ligands do not necessary mean that the binding modes of these ligands are identical, as seen in a series of thrombin inhibitors [120]. The mutual compensation of thermodynamic effects can result in identical thermodynamic signatures with simultaneous changes in the ligands' binding modes. Therefore, a structural inspection

is essential. Considering the classification of binding as enthalpy or entropy-driven binding, the selection of enthalpically favored lead structures for subsequent affinity optimization has been suggested as desirable [34, 121, 122]. However, an unambiguous classification with respect to such profiles is rather problematic in light of the large impact the rearrangement of the residual water solvation pattern has on the thermodynamic signature, e.g. for the binding of low affinity fragments [39, 123, 124]. Introduction of only small structural modifications can lead to major changes in the fragment's thermodynamic binding signature. In a recent review [125], it is even concluded that thermodynamically guided compound optimization is not feasible in most cases due to the complexity of the parameters enthalpy and entropy and the difficulties with their assignment to specific interactions.

2.4   How does an ITC measurement work and how does ITC raw data look like?

The general principle of an ITC measurement is that two reaction partners, for instance a protein and its ligand, are mixed with each other in a step-wise fashion, and the heat signal associated with the binding event is recorded. Figure 2.1a displays a schematic representation of an ITC device. The instrument consists of a sample cell and a reference cell, both in a jacket which is kept below measurement temperature. Both cells are maintained at the constant measurement temperature by applying a thermal heating device using very sensitive and highly regulated electric heating control units. The reference cell contains a solvent of a similar heat capacity to the one used in the sample cell (usually water or buffer). For the measurement, the protein solution is released into the sample cell and the ligand solution is gradually added via a rotating syringe, which also functions as a stirring rod. Typically, about 10−30 injections of the ligand solution are added into the sample cell until all active sites of the protein are saturated. The change in the heat signal in the sample cell resulting from the complex formation is quantified by analyzing the difference in thermal power ($\mu J\ s^{-1}$) necessary to keep the sample cell at the same temperature as the reference cell. For an exothermic reaction in the sample cell, the required thermal power is reduced compared to the reference cell, whereas for endothermic reactions the required thermal power increases. These differences in power over time are recorded and evaluated to quantify the event in the sample cell. Differences in heat as low as 0.1 μJ are detectable with the most sensitive ITC devices. In the ITC thermogram (Figure 2.1b), an exothermic binding reaction between protein and ligand is indicated by a series of "downward" peaks, whereas an endothermic reaction produces "upward" peaks. For the first injections of the measurement, the protein in the sample cell has a sufficient amount of unoccupied binding sites so that all injected ligand molecules can find a vacant binding pocket. This results in equally large heat signals. With increasing amount of injected ligand, the concentration of uncomplexed protein molecules
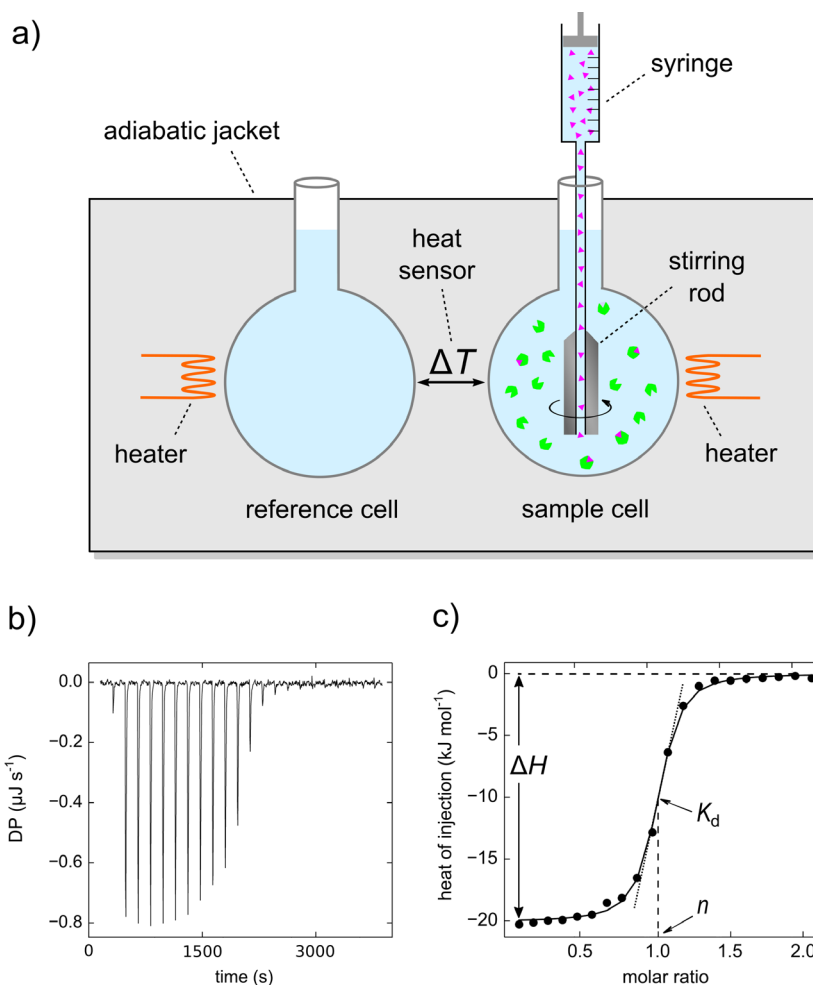
Figure 2.1. a) Schematic depiction of an ITC device. A solution containing dissolved ligand molecules (magenta) is step-wise injected into the sample cell containing a solution with dissolved protein (green). The heat released from the binding reaction in the sample cell between protein and ligand is recorded with respect to a reference cell. **b)** Raw thermogram of an ITC measurement, the differential power (DP) in $\mu J\,s^{-1}$ of the electric device keeping both cells at constant temperature is plotted against time. **c)** Integrated raw data and isotherm. The molar change in enthalpy observed for the injections is plotted against the molar ratio of the binding reaction. A 1:1 binding model is fitted to the data, from which $\Delta H°$, K_a and the stoichiometry n of the reaction are extracted.

becomes progressively smaller, allowing fewer ligand molecules to bind, which results in a gradual decrease of the heat signal. Due to chemical equilibrium conditions, further added free ligand molecules start to displace already bound ligand molecules from the protein. After several further ligand injections, all protein molecules are saturated by ligand molecules and under the regime of equilibrium, an increasing concentration of uncomplexed ligand molecules builds up. At the end of the titration, well beyond the 1:1 binding stoichiometry, only very small peaks of equal size remain which represent the heat of mixing of the solutions in the cell and in the syringe. Integration over these peaks can be used to define the zero

baseline and to correct for the heat of dilution. For data analysis, all measured peaks until those purely resulting from dilution must be integrated. Integration in this manner gives the total amount of heat originating from each injection, which is then related to the amount of injected ligand. To achieve this, the measured heats are plotted against the molar ratio between ligand and protein concentrations in the sample cell (Figure 2.1c). An appropriate binding model is fitted to the data points, in the simplest case a single-site 1:1 binding model. More complicated cases such as a two-site or triple-site binding or a competitive binding require different models [126, 127]. The selection of the binding model must be performed carefully and ideally under the control of independent experimental results obtained by other techniques, e.g. knowledge of the binding mode from a crystal structure. After curve fitting, the thermodynamic parameters are then extracted from the model curve.

2.5 How to get which data from the ITC isotherm?

In Figure 2.1c, a typical ITC isotherm is displayed. An appropriate model, in this case clearly a one-site binding model, was fitted to the data points of the integrated heat peaks via a nonlinear least squares fitting process. From the curve fitted to the data points, we obtain the change in enthalpy $\Delta H°$, the equilibrium constant K_a, and, by use of the latter value and application of Eq. 2.1, the Gibbs free energy $\Delta G°$ of the studied reaction [119, 128]. The change in enthalpy is related to the observed heat signal, while the K_a value is obtained from the slope at the inflection point. The location of the inflection point on the molar ratio axis describes the binding stoichiometry n, which is also referred to as the "site parameter". Importantly, the entropic term $-T\Delta S°$ of binding is not available from an independent experiment but must be calculated as the numerical difference between $\Delta G°$ and $\Delta H°$, using Eq. 2.2. Accordingly, any error affecting the experimental determination of $\Delta G°$ or $\Delta H°$ will directly influence the calculated magnitude of the entropy.

2.6 Which requirements must a curve fulfill to enable the extraction of reliable thermodynamic parameters?

Optimally, a binding isotherm should show a sigmoidal curvature with plateaus at the beginning and end of the titration. Experimental uncertainties can be further reduced during the integration step by ensuring an adequate signal-to-noise level, as well as by observing significant differences between peaks resulting from the binding reactions compared to peaks from buffer mismatch reactions. A buffer mismatch reaction between syringe and sample cell buffer can result in huge mixing heat signals in addition to those originating from ligand binding. In order to avoid buffer mismatch, dialysis of protein and ligand solutions against

the same buffer can be performed. A buffer mismatch is often the result of an inappropriate adjustment of the dimethyl sulfoxide (DMSO) concentration in cell and syringe, or due to a mismatch of the buffers' pH value. DMSO is a dipolar, low reactive solvent frequently added to increase ligand solubility. Furthermore, it is used for the preparation of ligand stock solutions (pure solutions of DMSO containing a high amount of ligand, typically between 10 mM−100 mM). Such solutions are used for storage and efficient use of sometimes precious compound material. Prior to measurement, the stock solution is diluted with buffer to obtain the desired concentration, usually resulting in high concentration accuracy. However, it is recommended to keep the concentration of DMSO during the ITC measurement as low as possible. A maximum concentration of 5% (v/v) should not be exceeded, which already corresponds to the high molar concentration of 0.7 mol L^{-1}. It must be kept in mind that even concentrations of 0.5−1% (v/v) DMSO were reported to significantly influence protein−ligand binding parameters [129]. Furthermore, in one of our thermolysin crystal structures (PDB code 4D91), DMSO was found in complex with the active site of the protein. Therefore, at least for the metalloprotease thermolysin, DMSO actively competes for the binding site with any other ligand present and hence influences the measured binding parameters of the ligand. A further source of dilution peaks is the dissociation of ligand aggregates within the sample cell, which can occur upon injection into the larger volume of the sample cell.

Owing to these experimental deficiencies, an extraction of the heat signals originating solely from the protein−ligand complex formation is necessary. For the correction of the buffer mismatch peaks, it is considered as best practice to subtract the average of all constant dilution peaks recorded, which appear after the system has reached sufficient saturation, from all measured peaks [130]. Another possibility for correction is to perform control titrations: for the first control, the ligand solution is titrated into the sample cell containing pure buffer. For the second control, pure buffer is titrated into the sample cell containing the protein. Both control titrations are then subtracted from the actual titration curve of ligand into protein. Either way, the corrected integrals of the peaks are then fitted to an appropriate model curve. For the extraction of reliable thermodynamic parameters from the binding isotherm, the shape of the curve resulting from the fit is critical and can be described by evaluating the so-called c-value [115]:

$$c = n\, K_a\, M_{\text{tot}} \tag{2.3}$$

The parameter n describes the stoichiometry of the reaction (the molar ratio between syringe reactant and cell reactant inside the sample cell at which the inflection point of the titration curve occurs). K_a refers to the association constant of the ligand and M_{tot} (mol L^{-1}) to the total concentration of the macromolecule in the sample cell. In Figure 2.2a, ITC isotherms with c-value between 0.01 and 1000 are shown.
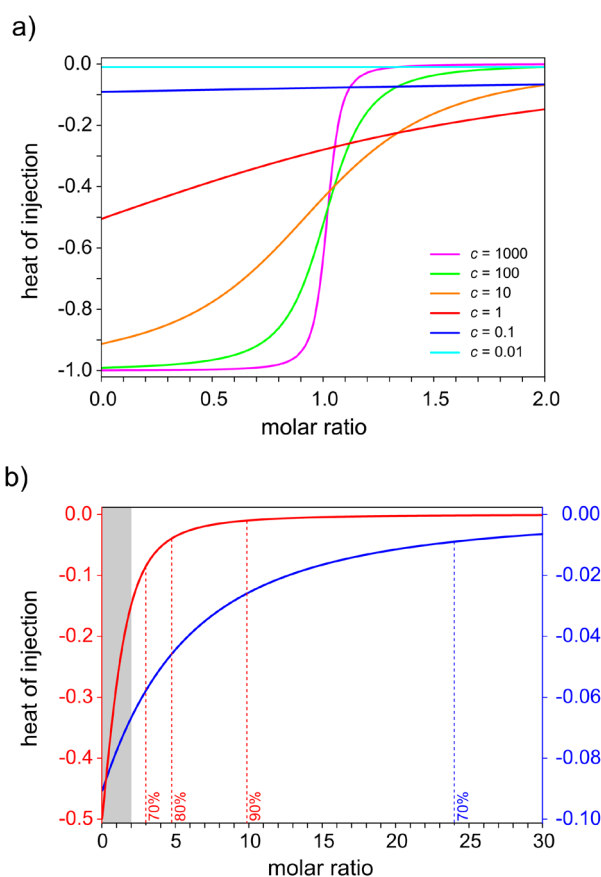
Figure 2.2. a) ITC isotherms of exothermic 1:1 binding reactions showing curvatures with *c*-values between 0.01 and 1000. The titration curves are shown up to a molar excess of two of the ligand over the protein. The arbitrarily chosen heat of injection of −1 corresponds to the exothermic heat signal for complete binding of the injected ligand. The isotherms were simulated with a modified version of a tool for modeling ITC curves of a perfusion calorimeter [131], not considering volume change or overflow of the sample cell. **b)** Curves with *c*-values of 1 (red) and 0.1 (blue), displayed with different scales for the heat of injection, are titrated up to a molar ratio of 30 between ligand and protein. The part of the curves in the grayed area describe the curves resulting from a titration up to a molar ratio of two between ligand and protein. The dashed lines indicate the degree of protein saturation for a given molar ratio. Protein saturation was calculated with the fractional occupancy calculator [123].

Obviously, curves with high *c*-values show a clear sigmoidal curvature, whereas curves described by low *c*-values appear flat. In practical experience, ligands with an affinity of about 10^4 M^{-1} to 10^8 M^{-1}, corresponding to K_d values between 100 μM and 10 nM, yield curves with *c*-values between 10–500, which is frequently considered as optimal by experimenters [115, 117, 132–135]. For such compounds, the experimental setup is often designed according to the so-called prevailing "standard-protocol" [136], consisting of about 25 injections and a molar ratio between ligand and protein of two at the end of the titration (Figure 2.1b). For compounds in the "optimal" affinity window, this protocol usually results in sufficient heat signals for each injection and in sufficient protein saturation at the end of the titration,

leading to well-analyzable titration curves (Figure 2.1c). However, rather than the usually applied 25 injections, it was found that titration curves with the highest precision are achievable by designing the measurement with only 10 injections of equal volume, also resulting in a reduced runtime of the measurement [137]. A first, small injection, as visible in Figure 2.1b, is usually performed and later discarded for data evaluation due to inaccuracy in the heat signal frequently observed for the first injection. However, it was shown that the inaccuracy in the heat signal is the result of an injection volume error originating from the syringe plunger's drive mechanism. The inaccuracy is observed directly after the drive direction of the plunger changed, as is the case between the filling (up) and the ejection (down) movement of the plunger [138]. Thus, even if the first injection is deleted from the data evaluation, it is inaccurate to assume that the whole volume was actually injected into the sample cell. A simple solution to this problem is to perform a short 'down' movement of the plunger after the syringe filling but prior to the actual measurement. Thereby, volume errors can be significantly reduced [138].

For ligands with affinities lower than 10^4 M^{-1} (100 μM), titration curves exhibiting c-values below 10 are usually observed. As a matter of fact, such curves do not show a clear sigmoidal shape but rather a more simple one (Figure 2.2a) without a clearly defined inflection point or a baseline at the beginning of the titration. In theory the c-value can be adjusted for every K_a by simply adjusting the concentration of the macromolecule participating in the reaction, according to Eq. 2.3. However, for the analysis of a ligand with the low affinity of 1 mM, this would require a protein concentration in the sample cell of about 10 mM in order to achieve a c-value of 10 [133]. In practice, this strategy is usually hampered owing to too low protein solubility and limited availability of protein material. In addition, at such high concentrations, deviation from an ideal solution will likely occur, resulting in reduced protein activity [114]. Therefore, a modified experimental setup must be applied – the so-called "low c-value titration" [134, 139]. In such a scenario, the low c-value curves are actually used for parameter analysis. The critical step of such a titration is to achieve sufficient reaction between protein and ligand [134]. A protein saturation of at least 70% at the end of the titration has been suggested to be the lower limit [123, 134]. In order to achieve sufficient saturation, decreasing ligand affinity must be compensated with increasing ligand excess to favor the formation of the protein–ligand complex. For a low affinity ligand giving rise to a curve with a c-value of 0.01, this corresponds to a 24-fold molar ligand excess over the macromolecule in the sample cell (Figure 2.2b). Consequently, ligand solubility is the main issue at this point for achieving the required ligand concentration in the syringe solution [123, 133, 134]. As mentioned before, curves exhibiting low c-values below 10 do not show a clear sigmoidal shape but a more simple one (Figure 2.2a and b). Because a clear inflection point is missing, it is impossible to determine the value of n experimentally [133]. Nevertheless, in order to still get access to the thermodynamic parameters, the stoichiometry

n of the binding reaction must be fixed according to an independently determined value. For an accurate determination of *n*, the concentration of protein and ligand as well as the binding ratio between protein and ligand must be exactly known. It has been shown that the error in $\Delta H°$ is strongly dependent on the error in *n* [137]. On the other hand, the determination of the affinity constant K_a turns out to be almost independent from the stoichiometry *n* [139]. Therefore, in a low *c*-value titration, even if the determination of accurate values for *n* and $\Delta H°$ fails, the affinity constant K_a can still be measured. Furthermore, because for low affinity ligands only a fraction of the injected ligand actually binds to the protein and thus produces a measurable heat signal, the observed signals are usually very low. Accordingly, low *c*-value titrations should be performed with a small amount of injections (e.g. only 4–5), but with a large injection volume [140]. Additionally, it is of advantage to vary the injection volume during the titration. As more protein becomes saturated and as a smaller fraction of the injected ligand binds, the gradual decrease in the heat signal can be compensated by increasing the injection volume.

Conversely to low *c*-value curves, curves with *c*-values above 1000 also create some problems in the analysis. For curves with *c*-values >500, the uncertainty for the K_a determination increases [135]. Such curves no longer show sigmoidal curvature (Figure 2.2a), but instead a more rectangular shape, which makes the determination of the slope at the inflection point unreliable. According to Eq. 2.3, in order to obtain an optimal *c*-value for high affinity ligands (K_a >10^8 M^{-1}), measurement with very low protein concentration is required. This, however, can lead to injection peaks below the sensitivity range of the ITC instrument. In contrast to the assignment of K_a, the enthalpy of binding $\Delta H°$ is easily determinable for curves with high *c*-values. For $\Delta H°$, the molar heat signal of a complete binding reaction of all injected ligand molecules has to be determined. This can be reliably extracted from the step-like titration curves which show clearly defined plateaus at the beginning and end of the titration.

The displacement titration is an alternative strategy that has been developed in order to yield reliable microcalorimetry data from ligands across a wider range of affinities. This strategy is available for both low [123, 141] and high affinity binders [142]. For weak binding ligands, the protein is first saturated with the low-affinity ligand of interest, which is subsequently displaced by a previously characterized high-affinity reference ligand. Therefore, the reference ligand must bind competitively to the same protein site as the low-affinity ligand. As a result, a thermogram in the optimal *c*-value range is obtained. The amount by which this new competitive binding signal differs from the signal of the reference ligand alone depends on the amount of heat required to displace the low affinity ligand, which in turn relates to the latter ligand's binding signature. As a disadvantage, any uncertainties and experimental errors in the determination of the thermodynamic parameters of the reference ligand will also affect the parameters of the low-affinity ligand. The displacement strategy for high-

affinity ligands follows the same concept as the displacement titration of weak binders, however with the important difference that a weak to medium potent ligand is used to preincubate the protein. This ligand must be previously characterized thermodynamically and serves as a reference ligand [142]. This strategy also allows the titration curve to shift into a c-value range that results in proper sigmoidal isotherms. From this, the stoichiometry and the K_a value are extracted. Unlike the characterization of weak-binding ligands, the $\Delta H°$ value is taken from a separate titration curve of the strong binding ligand showing a rectangular shape. As mentioned earlier, the rectangular shape is no obstacle for the $\Delta H°$ determination, and using this curve avoids error propagation that can occur due to uncertainties in the characterization of the reference ligand. Therefore, displacement titrations applied to ligands with a high affinity yield much more accurate data than displacement titrations of weak binders.

2.7   The interdependence of enthalpy and entropy

As mentioned, the ITC measurement determines only $\Delta H°$ and K_a experimentally, while changes in entropy are calculated from the numerical difference between $\Delta G°$ and $\Delta H°$ according to Eq. 2.2 [109]. Thus, considering that the values determined for K_a and, in consequence, $\Delta G°$ are less error-prone than those for $\Delta H°$ [133], the error of $-T\Delta S°$ will always depend on the error of $\Delta H°$. Calculating the $-T\Delta S°$ value from the numerical difference will propagate any error affecting the $\Delta H°$ measurement. In consequence, inaccurate measurement of $\Delta H°$ can lead to "artificial" enthalpy−entropy compensation (EEC) [36, 143]. In contrast to artificial EEC, "intrinsic" EEC describes the phenomenon by which enthalpy and entropy are really compensating each other and ultimately hardly influence the overall Gibbs free energy of binding, for example in a drug optimization scenario [35, 144]. It is intuitive to understand that during ligand optimization, EEC occurs at least to some extent: stronger fixation, which leads to a higher enthalpic contribution, leads to less flexibility and therefore lowers entropy and vice versa. However, due to inaccuracies in $\Delta H°$ and thus in $-T\Delta S°$ determination, the extent of EEC can be overestimated. An EEC purely imposed owing to experimental inaccuracies is particularly dangerous if a global analysis [145] of available thermodynamic data (for instance derived from the *BindingDB* [146], the *SCORPIO* [147] or from the *PDBcal* [148] online databases) is conducted, and thermodynamic data across different proteins and ligand series are compared on an absolute scale [109]. It must be considered that such thermodynamic data result from measurements conducted under deviating experimental conditions. The experiments are possibly performed at different temperatures, and buffers of deviating ionization enthalpies are used without applying the required correction. Protein concentrations are selected in very different ranges making direct comparison on absolute scale problematic. Furthermore, the experiments are

obviously performed by different persons in different laboratories using different devices, leading to systematic deviations and uncertainties in the data of an unknown magnitude. A remarkable test case on ITC data accuracy has been studied across several laboratories. In the ABRF-MIRG'02 study [149], identical samples were thermodynamically characterized by 14 independent laboratories. Surprisingly enough, a plot of the determined $-T\Delta S°$ versus $\Delta H°$ values of the identical reaction performed 14 times suggests a nice EEC [36]. This study demonstrates how careful one must be when making a comparison of global data, and how easily such comparisons can be misleading. Before discussing the corrections necessary to reveal accurate comparative information, we want to argue that data evaluated and compared relative to each other across a congeneric series of ligands can yield reliable information. In the case when congeneric ligand series are measured under the same conditions with concentrations falling in a narrow window and corrected for putative differences in the heats of ionization (see below), $\Delta H°$ can be determined with very high precision but also accuracy and the influence of an error-prone EEC can be minimized. In one of our studies of a congeneric ligand series binding to the well-established model system thermolysin [104], measurements were performed by the same operator over a short period of time using the same ITC device. Experimental conditions such as pH and temperature were kept constant. Moreover, it was also possible to keep the concentration of the protein and ligand solutions constant. The applied ligands were checked for high purity and the protein solutions were all prepared from the same batch. DMSO-free protein and ligand solutions were freshly prepared for each measurement. As a result, the extended ligand series showed ITC isotherms with c-values in the narrow range between 11 to 158 and an average stoichiometry n of 0.753±0.04. The important characteristic of the observed stoichiometry is that it remains constant throughout all measurements. Deviations from the theoretical value of 1.000 are due to partial protein inactivity, which, however, has no effect on the accuracy of the measurement parameters and could easily be corrected by adjusting the protein concentration to the measured activity level of the respective batch. Again, this has no advantage for the measurement itself, but would mask the otherwise obvious partial protein activity. In our opinion, the ITC isotherms of such studies must be documented in the supporting information of a publication as they can be a proof of the accuracy of the measured data [104]. In our study, we examined how the rearrangement of surface water molecules during the protein–ligand binding reaction affects the thermodynamic signature of the complex formation by analyzing a series of closely related ligands. Because we were dealing with relatively small changes in their thermodynamic profiles, high precision — particularly with respect to a relative comparability — was very important.

2.8 Heat effects from proton transfer reactions between protein, ligand and buffer

The observed heat signal resulting from an ITC measurement (ΔH°_{obs}) is the sum of the heat signals produced by the actual binding event (the intrinsic change in binding enthalpy ΔH°_{bind}) plus any additional effects contributed by the entire system [150]. Most important are heat changes resulting from a proton transfer (protonation or deprotonation) between the formed protein–ligand complex and the surrounding buffer ($n_{H^+} \Delta H^{\circ}_{ion}$):

$$\Delta H^{\circ}_{obs} = \Delta H^{\circ}_{bind} + n_{H^+}\Delta H^{\circ}_{ion} \qquad (2.4)$$

The explanation for such an occurrence, also known as 'proton linkage', can be found in a shift of the pK_a value of ionizable functional groups of the protein residues and/or the ligand during complex formation, as these groups are brought into a novel environment with different dielectric properties [151]. Depending on the buffer compounds (different buffers show different ionization enthalpies ΔH°_{ion} upon proton exchange) and the involved functional groups of protein and/or ligand, the heat of ionization ($n_{H^+}\Delta H^{\circ}_{ion}$) can have a significant impact on the observed heat of binding (ΔH°_{obs}). By running the binding reaction in buffers exhibiting deviating heats of ionization, the resulting enthalpies will be different between buffers, but the association constant K_a and thus affinity data (ΔG°) are usually not significantly affected [152, 153]. The thermodynamic binding profiles of a ligand measured in different buffers showing proton linkage thus show similar affinities, but their enthalpic and entropic terms vary depending on the applied buffer, and resemble enthalpy–entropy compensation. This exemplifies how arbitrary an absolute scale comparison of such data would be — the uncorrected enthalpies are rather meaningless on such a scale due to the superimposed buffer effects. At best, it is still possible to uncover trends, but it is difficult to detect more subtle correlations. Therefore, if proton linkage occurs, enthalpies must be corrected for the heat of ionization before they are ready for comparative analysis on an absolute scale, or even a relative scale in some cases (see below). By measuring the heat signal of ligand binding in buffers of different ionization enthalpies at the same pH, e.g. in Tris, ACES, HEPES and PIPES buffers as performed in the experiment described in Figure 2.3a–e, the enthalpic contribution from each buffer's heat of ionization can be determined. The number of protons exchanged during the reaction (n_{H^+}) can also be identified (Figure 2.3d), and the observed enthalpies can be corrected for their buffer contributions (Figure 2.3e). To achieve this correction, the experimentally obtained enthalpies are plotted against the ionization enthalpies of the buffers considering the values referenced in literature [154]. A linear regression is performed, and its intercept with the y-axis reveals the buffer-corrected enthalpy.
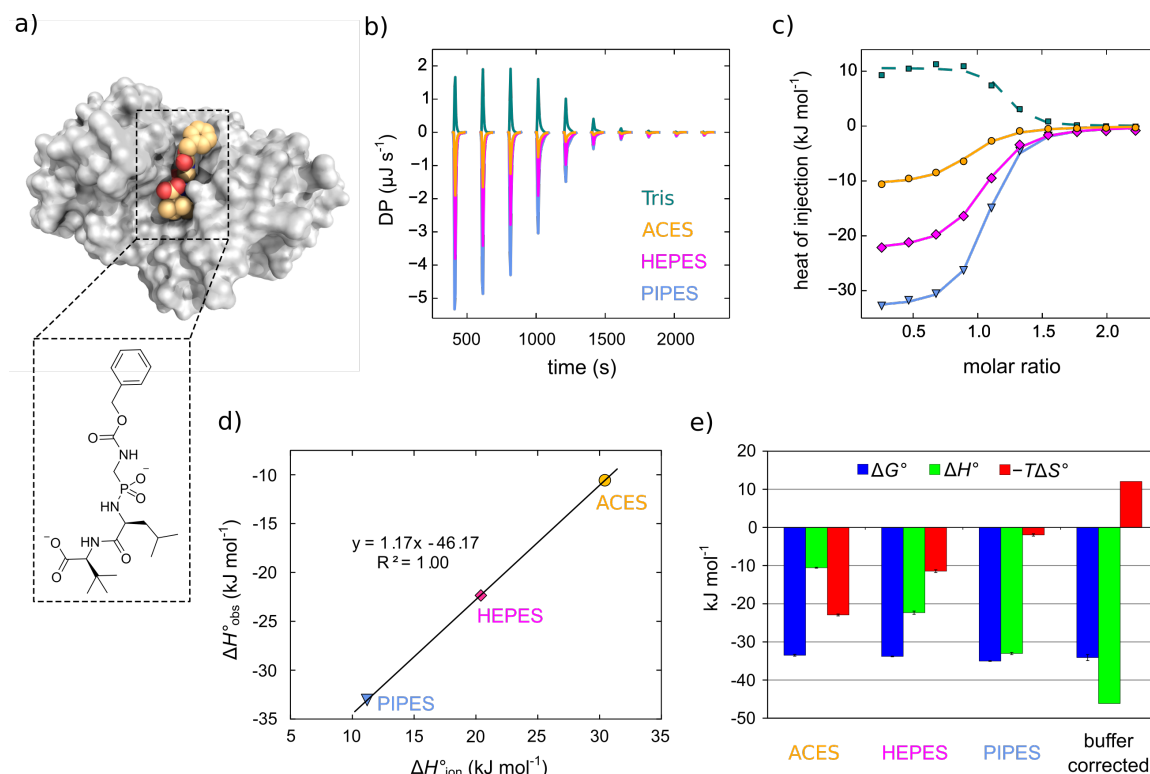
Figure 2.3. Determination of the heat of ionization for the binding reaction between thermolysin and a phosphonamidate ligand. **a)** The crystal structure of thermolysin (Connolly surface in white) in complex with the analyzed ligand clearly reveals a 1:1 binding mode (PDB code 5DPE). **b)** Overlay of the ITC raw thermograms of the binding reaction measured in Tris, ACES, HEPES and PIPES buffer. Only extracted heat peaks (without baselines) are displayed, as performed by the peak shape analysis algorithm of *NITPIC* [155]. Except for the buffer substance, an identical experimental setup was applied for all titrations in order to guarantee comparability of the resulting heat signals. Whereas the binding reaction in Tris buffer results in an overall endothermic reaction (upward peaks), the complex formation in the other buffers is exothermic (downward peaks), its signal increasing from ACES to HEPES to PIPES. **c)** Integrated data of the heat signals observed for the measurements in the four different buffers. The legend for the data is shown in panel b. The 1:1 binding model curve does not fit perfectly to the integrated data points of the titration in Tris buffer (dotted lines), suggesting a more complex scenario, likely due to an active displacement of Tris from the active site of the protein during ligand binding. Consequently, the titration in Tris buffer was not applied for the calculation of the heat of ionization. In contrast, the 1:1 binding model perfectly fits to the data points of ACES, HEPES and PIPES, confirming the chosen model in these cases. **d)** Calculation of the heat of ionization. The experimentally observed enthalpies $\Delta H°_{obs}$ are plotted against the heat of ionization $\Delta H°_{ion}$ of the respective buffers. The slope of the straight line describes the proton uptake during the formation of the protein–ligand complex (on average 1.17 mole), whereas its interception with the y-axis describes the buffer corrected enthalpy of the binding reaction ($\Delta H°_{corrected} = -46.2$ kJ mol^{-1}). **e)** Thermodynamic profiles of the complex formation in ACES, HEPES and PIPES buffers as well as the buffer corrected thermodynamic profile. For the buffer corrected profile, the change in the Gibbs free energy $\Delta G°$ is calculated as the average of $\Delta G°$ observed in the three buffers, $\Delta H°$ is derived as described in panel d, and the entropic term is calculated from the numerical difference between $\Delta G°$ and $\Delta H°$. More experimental details are given in the supplementary material.

Interestingly, ligands with more entropy-driven binding profiles are better measureable if they have an ionization reaction superimposed onto the actual binding event. Without the ionization reaction, the enthalpic signal of the binding reaction can be below the detection limit of the ITC device. This was the case in a ligand binding reaction to thrombin, which showed a buffer corrected enthalpy of -1.4 kJ mol^{-1} — a value impossible to detect, if not the nicely measurable buffer uncorrected heat signals of -29.0 kJ mol^{-1} (Tris buffer), -17.4 kJ mol^{-1} (TRICINE buffer) and -14.3 kJ mol^{-1} (HEPES buffer) [156] would have occurred.

It should be noted that the buffer correction is performed under the often unfounded assumption that interactions between protein and ligand do not change with the various buffers and additives [114], even though salts can significantly influence the activity of the protein [114, 157], according to the Hofmeister series [158]. Furthermore, it must also be considered that, for instance in the case of the aspartic proteases [159], the pH used for the measurement can have a significant influence on the actual protonation state of residues and functional groups (e.g. on the catalytic dyad). The protonation state can influence the molar quantity of protons transferred, which in turn affects the heat of ionization. Thus, the enthalpy from the ligand binding process can vary, and the Gibbs free energy can also be altered. Interestingly, a method has been described where the pH dependence of binding affinity is exploited to provide access to affinity data for binding that is too tight to be measured directly at the pH of interest. In this method, affinities are measured at pH values showing less tight binding, and are subsequently extrapolated to obtain the affinity at the pH of interest [151, 160].

In special cases, buffer-uncorrected enthalpies can be used for a relative comparison, particularly across a narrow compound series and if all studied ligands induce the same change in their protonation states. This may occur if the site where the ligands are structurally modulated is remote from the site where the protonation transfer occurs. All binding events will be influenced by the superimposed protonation change in similar fashion, but in a relative comparison across the series this contribution cancels out. For example, in a study of a congeneric series of phosphonamidate thermolysin ligands like the one shown in Figure 2.3a [64, 104], we observed a buffer dependency of the enthalpic term (Figure 2.3a–e). It was possible to identify Glu143 as the site which entraps one proton upon ligand binding. However, because the parent scaffold of the congeneric ligand series remains unchanged next to Glu143 and varies only at a site remote to it, all ligands are equally affected by the heat of ionization. In this example, only relative changes of the thermodynamic signatures of the ligands were of interest, and not their absolute values and thus the heat of ionization contributions fall out of the correlation. However, it must be underscored that such data cannot be used in a global correlation of thermodynamic properties on an absolute scale. In the described thermolysin example, it was sufficient to perform the ITC measurements in only one buffer. However, it would be rather meaningless and arbitrary to compare these

results with data measured in other buffers or with ligands showing a deviating basic scaffold next to Glu143.

The presence of ionization effects upon complex formation is not always so obvious. In a study on thrombin inhibitors [156], mutual compensation of protonation effects between ligand and protein occurred upon ligand binding. The imidazole moiety of thrombin's His57 released 0.6 mole of protons, whereas a primary amino function of the inhibitor picked up an equal amount of protons, resulting in a negligible detectable net proton exchange. However, a ligand where the ionizable amino function was replaced by a non-ionizable amide function revealed the proton exchange upon complex formation — a release of 0.6 mole of protons only attributable to His57. It was gathered that the same proton release occurs during binding of the ligand with the amino function, but in this case it is masked by the superimposed proton uptake of the latter group, and therefore the expected buffer dependence is not apparent. In such a case, buffer ionization corrections will be difficult to make and accordingly cannot be successfully performed without further studies. One strategy to at least reduce the contribution of a superimposed proton linkage is to perform the measurement in a buffer with low heat of ionization (e.g. acetate buffer, $\Delta H^{\circ}_{ion} = 0.41$ kJ mol^{-1}). Thus, the buffer contribution will be negligible. However, the contribution added by the group of the protein or ligand which displays the partner in the proton exchange reaction will still show a heat effect.

Are any further effects expected to modulate the heat contribution? Ions are often involved in ligand binding, and in some cases can be detected in the formed crystal structure [161]. The entrapment or the release of such ions most likely has a heat contribution, representing a possible artifact superimposed to the binding process which must be corrected. Further influences can originate from the salt as a component of the buffer. In recent studies on a host–guest system comprising a hydrophobic binding site [162, 163], the thermodynamics of binding is strongly influenced by different salts. The measurements in buffers containing NaF or NaCl ('kosmotropic' salts) result only in a slight increase in affinity with minor changes in enthalpy and entropy. However, ITC measurements in buffers containing NaClO$_4$, NaSCN, NaClO$_3$ or NaI ('chaotropic' salts) result in significantly decreasing K_a values, involving a major decrease in enthalpy and increase in entropy. It was shown that the chaotropic anions competitively bind to the hydrophobic pocket of the host and thereby modulate the thermodynamics of binding [163]. In our own studies, we analyzed a thermolysin–ligand binding reaction in buffers containing 200 and 1000 mM NaSCN (Figure 2.4a and b). As a result, between 200 to 1000 mM NaSCN, the enthalpic term increases, whereas the entropic term decreases. Nonetheless, the Gibbs free energy is not significantly affected (Figure 2.4a).

Therefore, especially if chaotropic salts are used as a buffer additive (e.g. for increasing the solubility of an otherwise not sufficiently soluble protein, so-called 'salting-in' effect of
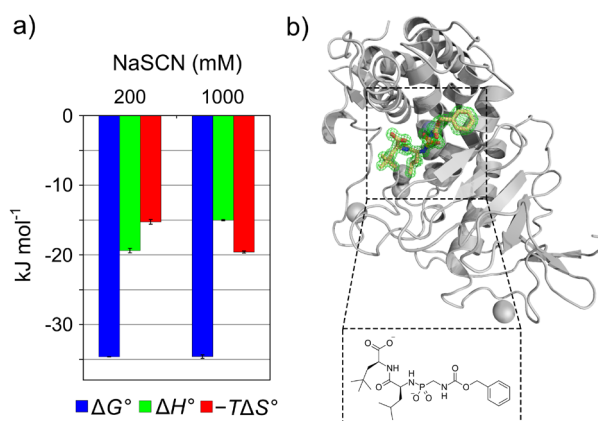
Figure 2.4. a) Effect of salt concentration on the thermodynamics of binding of the same thermolysin–ligand binding reaction performed in buffers containing the chaotropic salt NaSCN at concentrations of 200 and 1000 mM. Standard deviations are given for the measurements performed in duplicate. Experimental details are given in the supplementary material. **b)** Crystal structure of the protein in complex with the analyzed ligand (PDB code 5DPF). Thermolysin is displayed as cartoon, the ligand binding to the active site is displayed as stick representation (orange) and the F_o–F_c omit electron density of the ligand is shown at a contour level of 3σ as green mesh. The crystal structure clearly reveals a 1:1 binding mode.

chaotropes) and the active site of a protein contains a hydrophobic concave surface as a binding site for the chaotropic anions [164], the binding profile can be significantly influenced by the added salt. Again, in the case of congeneric series of ligands where all studied ligands show the same effects, the contribution will cancel out in a relative comparison.

2.9 Temperature-dependency of $\Delta H°$, change in heat capacity ΔC_p and van't Hoff analysis of $\Delta H°$

It is well recognized that chemical processes are dependent on temperature. In consequence, chemical equilibria and the corresponding association or dissociation constants are temperature-dependent. As the Gibbs free energy is related to the latter constants (Eq. 2.1), also this property will in general be dependent on temperature. $\Delta G°$ factorizes into enthalpy and entropy, whereby entropy is weighted with the absolute temperature (Eq. 2.2). Likewise, $\Delta H°$ and $\Delta S°$ change with temperature. The partial derivative of the enthalpy with respect to temperature while holding the pressure constant reveals the above-mentioned change in heat capacity ΔC_p (kJ mol^{-1} K^{-1}) of a reaction:

$$\Delta C_\mathrm{p}=\left(\frac{\partial \Delta H°}{\partial T}\right)_\mathrm{p} \qquad (2.5)$$

The change in heat capacity ΔC_p describes the amount of heat which is necessary for a temperature change of the system of 1 K. In other words, it describes how well the system can absorb or release heat, attributable to the available degrees of freedom.[39] Empirically, a correlation of increasing ΔC_p with an increasing burial of apolar and polar surfaces between macromolecules has been found, which is associated to the displacement of water molecules upon complex formation [165]. According to Eq. 2.5, for the analysis of the change in heat capacity ΔC_p of a protein–ligand complex formation, the change in $\Delta H°$ at different temperatures needs to be determined. Interestingly, in biological systems, ΔC_p of a protein–ligand complex formation almost exclusively exhibits negative values, and usually adopts values differing from zero. Accordingly, the complex exhibits a lower heat capacity compared to the sum of heat capacities of protein and ligand in their uncomplexed state. With respect to enthalpy and entropy, this general behavior results in the finding that protein–ligand complex formation becomes more exothermic (enthalpic) with increasing temperature and simultaneously entropically less favorable [150]. This observation can be exploited in ITC measurements. The property measured in an ITC experiment is a heat signal resulting from the enthalpic component of binding. Thus, a predominantly entropically driven process hardly produces any measurable effect. If such a situation is experienced, the titration should be repeated at a temperature 5 or 10 K higher or lower. Then, usually a detectable signal can be recorded. On the other hand, this observation clearly demonstrates that the thermodynamic properties are not temperature independent, even in the small windows accessible with biological systems. It also implies that values of $\Delta G°$, $\Delta H°$ and $-T\Delta S°$ measured at different temperatures can hardly be compared directly. Furthermore, it indicates that some care is needed to define a process as '*enthalpy or entropy-driven*', as it matters at which temperature the process has been recorded [166]. This means, in a discussion of thermodynamic properties, we should only compare series of complexes measured at the same temperature relative to each other and regard them in the comparative analysis as '*enthalpically or entropically more favored*' in their formation.

Popular evaluations of thermodynamic properties make use of the so-called van't Hoff evaluation [39, 167, 168]. For this, usually the biological system under consideration is studied at for example three different temperatures by evaluating well-recordable signals such as a change in a spectroscopic property or shifts in resonance signals. The recorded signals are then used to quantify the concentrations (or better: activities) of the unbound and bound species involved in the equilibrium. However, therefore, the binding event has to follow a two-state transition between the free and bound state and the change in the recorded spectrometric signals, subsequently used to assign the binding constant, has to consider all of the free and bound molecules involved in the complex formation reaction [39, 109, 150]. Considering the recent results found by simulations to describe binding kinetics, at least questions this assumption quite strongly, as usually multistep mechanisms have to be

discussed [169]. At this point the burden is on the experimentalist to correctly assign the concentrations at equilibrium, however, usually it is by no means trivial to ensure this assumption. The measurements of the binding constants are in the following performed at different temperatures and for the evaluation the integrated form of the van't Hoff equation is used [39]:

$$\ln\left(\frac{K_2}{K_1}\right) = \frac{1}{R}\int_{T_1}^{T_2}\frac{\Delta H°(T)\,\mathrm{d}T}{T^2} \tag{2.6}$$

The binding constants K_1 and K_2 for a reaction describe the measurement at the two different temperatures T_1 and T_2. Frequently, for the evaluation a "simpler" form, the so-called linear form of the van't Hoff equation is used (Eq. 2.7), which, however, only arises if $\Delta H°$ is assumed to be temperature independent, as only then it can be taken out of the integral:

$$\ln\left(\frac{K_2}{K_1}\right) = \frac{\Delta H°}{R}\int_{T_1}^{T_2}\frac{\mathrm{d}T}{T^2} = -\frac{\Delta H°}{R(T_2 - T_1)} \tag{2.7}$$

Applying this latter form, the binding constants are plotted against the reciprocal of the temperature and evaluated by a linear fit, where the slope of the straight line describes the van't Hoff enthalpy. However, as described above, this is usually a non-valid assumption, as experience shows that ΔC_p deviates from zero and thus $\Delta H°$ is actually temperature dependent. To circumvent this, integration of the differential form of the equation requires some kind of approximation to describe the temperature dependency of $\Delta H°(T)$, for example as a Taylor expansion, to achieve a non-linear fit [39].

The advantage of ITC experiments is that they are performed at one temperature and reflect the entire binding process. They including all heat signals produced, even if binding passes through multiple states. From this, $\Delta G°$ and $\Delta H°$ become available. At first glance, heat capacity changes appear as an ideal property to relate structural properties and molecular degrees of freedom with thermodynamic entities. However, measurements of ΔC_p require ITC experiments to be performed across a temperature range. As a matter of fact, the complexity of multicomponent systems like protein–ligand complexes, including the surrounding aqueous buffer environment, is so large that the changes of heat capacity are very difficult to interpret on molecular level [39]. It should not be forgotten that the ubiquitously present water in biological systems is a substance with one of the largest heat capacities known, and most likely the changes with temperature while studying biological processes involve major changes in the surrounding water environment superimposed or inherently correlated with the changes of the biological system.

2.10 The importance of high ligand purity and accurately known ligand and protein concentrations

The importance of ligand purity and the determination of the exact ligand concentration is well appreciated [39, 134, 149, 157]. Inaccurate ligand concentration can be the result of solution preparation directly based on a ligand sample's weight if the sample contains unexpected impurities. Water is a common impurity for hygroscopic powders in particular; impurities may also originate from the synthesis. Even without impurities, accurate weighting in can be a serious problem, especially for electrostatically charged ligand powders. This problem can be addressed by an antistatic device, which, however, is not available in many laboratories. Another concern is that ligands in solution (for instance in a ligand stock solution) can suffer from chemical instabilities like partial hydrolysis over time during storage. Inaccurate or inadequate methods to determine the ligand solution concentration, for example via HPLC, might also impose a problem. Incorrect ligand and protein concentrations both have only minor consequences for the accurate determination of $\Delta G°$ [133, 134, 157]. However, for titrations in the c-value range of 10−500, incorrect ligand concentrations have a huge impact on $\Delta H°$ in particular [133, 134, 157], because the measured heat signal is attributed to a false amount of injected ligand. Errors resulting from ill-defined ligand concentrations must be classified as systematic, mostly unrecognized errors [149, 157, 170]. On the other hand, for low c-value titrations, it is the inaccuracy in the actually active protein concentration that lead to inaccurate $\Delta H°$ determinations [133, 134]. In addition to their effects on concentration, ligand impurities also lead to unpredictable heat reactions. The first indication about ligand purity is the stoichiometry n of the binding reaction available from proper sigmoidal titration curves (described by the 'incomplete fraction' parameter in the program *SEDPHAT*), especially in studies of ligand series binding to the same protein. Assuming that the protein shows unchanged activity in each measurement (which can be achieved by using protein material from the same batch), the stoichiometry should remain unchanged throughout the measurement of the whole series. If this is not the case, ligand impurity may provide an explanation. It must be noted that the experimentally determined stoichiometry will hardly match exactly 1.00, even in a simple one-site binding reaction, due to partial degradation or denaturation of the protein. If the protein activity is controlled by an independent experiment and n is found to be significantly lower than the expected value, the most likely reason is a higher than expected ligand concentration. If the stoichiometry cannot be determined experimentally, as is the case for low c-value titrations with fixed stoichiometry, a thorough purity validation must be conducted. This can be done using mass spectrometry (MS), quantitative nuclear magnetic resonance spectroscopy (qNMR), high−performance liquid chromatography (HPLC) or elemental analysis.

2.11 How accurate are ITC results, what is the true error and which systematic errors exist?

As mentioned, a comparative study across 14 independent laboratories has investigated the simple one-to-one binding reaction between carbonic anhydrase II and 4-carboxybenzensulfonamide. This study gave a $\Delta H°$ of -43.5 ± 10.5 kJ mol^{-1} and a K_a of $1.00 \pm 0.22 \cdot 10^6$ M^{-1} [149]. The reported values suggest rather worrying uncertainties of more than 20% in $\Delta H°$ as well as in K_a! For the determination of K_a, c-values of the isotherms below 20 were found to be the main source of such pronounced uncertainty. However, due to the logarithmic relationship between K_a and $\Delta G°$ (Eq. 2.1), uncertainties in the values of K_a are of minor influence for the calculation of accurate $\Delta G°$ values [133]. For the determination of $\Delta H°$, an accurate ligand concentration was found to be particularly critical. In a reanalysis of this study [170], the observed ligand concentration uncertainties were found to amount to about 10%, while it was stated that based on all precision limiting steps, uncertainties of below 1% in the ligand concentration could have been achieved for this reaction. If quantifiable, uncertainties in the ligand concentration should be stated together with other errors as a total uncertainty value [170]. However, the errors reported for ITC experiments are often simply taken from the nonlinear least square fit of a model curve to the data points. Alternatively, standard deviations are given for multiply performed measurements, which state the repeatability of the measurement by one person, but not its reproducibility by independent persons and over independent laboratories [111]. Hence, the observed deviations of more than 20% in the study must be considered as systematic errors which would otherwise have never been reported, and the errors detected by all independent laboratories would have been greatly underestimated. One way to discover systematic errors in $\Delta H°$ originating from deficiencies in the execution of the measurement is the use of enthalpy standards, which can also uncover uncertainties originating from the ITC instrumental setup itself, for instance from devices such as VP-ITC, ITC$_{200}$, and the Nano ITC-III. One proposed enthalpy standard reaction is the titration of 5 mM NaOH into the cell containing 0.5 mM HNO$_3$ at 25 °C [171]. The performance of chemical calibration has been suggested in addition to the routine electric calibration in order to avoid the occurrence of undetectable, systematic calibration errors of the ITC instrument and thus lowered accuracy [172]. Electrical calibration errors of 5% were reported as not uncommon [173]. This is of particular importance if the determination of thermodynamic data on an absolute scale is intended. However, for ligands with affinities in the optimal range, especially for relative comparison, congeneric ligand series of well-characterized systems are expected to show deviations in $\Delta H°$ smaller than 1 kJ mol^{-1} [39]. This estimation is in reasonable agreement with reported achievable deviation of 1% for K_a and $\Delta H°$ without the inclusion of systematic errors, and of 3% with systematic errors consideration [136].

Recently, certain weaknesses in the commonly applied protocol for analyzing K_a (and thus $\Delta G°$ and $-T\Delta S°$) from the experimentally obtained enthalpies have been pointed out [114]. For the correct calculation of the equilibrium constant, all components involved in the equilibrium must be considered. In addition to protein−ligand interactions, this involves interactions between buffer components and protein molecules, as well as protein−protein interactions. Clearly, such interactions are almost never considered in the parameter determination. Furthermore, as mentioned in the beginning, activities of the solvated components must be taken rather than concentrations because the studied solutions are likely not ideal. Typically, the differences between concentrations and their real activities are considered negligible. However, the concentrations of especially weak binding ligands and protein solutions can differ significantly from their activities. For instance, the activity coefficient of protein molecules can be influenced by the applied buffer (buffer salt, pH, additives). The ligand solution activity can be influenced by partial insolubility or ligand aggregation, especially of hydrophobic compounds. One option for considering the possible influence of activities instead of concentrations is the implementation of ITC measurements over a protein concentration range and in different buffers [114]. This should show whether the recorded equilibrium constants are equal for every measurement. Strong concentration dependencies would suggest a necessity to determine the real protein and ligand activities, for instance via equilibrium dialysis or potentiometric titration [174]. We performed ITC titrations of the same protein−ligand binding reaction with different thermolysin concentrations between 50−300 µM (Figure 2.5). As a result, the magnitude of $\Delta G°$, $\Delta H°$ and $-T\Delta S°$ significantly decreases with increasing protein concentration, whereas the relative difference between $\Delta H°$ and $-T\Delta S°$ remains constant. Accordingly, over the studied concentration range, the measured protein solutions are no ideal mixtures and their
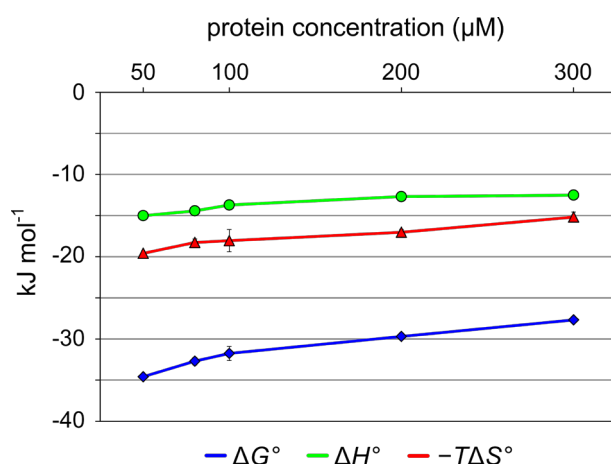
Figure 2.5. Thermodynamic binding profiles of the same protein−ligand binding reaction measured at different thermolysin concentrations of 50, 80, 100, 200 and 300 µM. The chemical structure of the analyzed ligand is shown in Figure 2.4b. Experimental details are given in the supplementary material.

concentrations are not equal to their real activities. Thus, comparison of data on an absolute scale and from measurements based on different protein concentrations cannot be performed accurately without knowledge of the activity coefficients.

Further systematic errors can originate from numerous sources, including solvent evaporation during the measurement, adsorption of reactive components at the cell wall, mechanical effects (e.g. from the stirring of the syringe paddle), metal corrosion of the device [173], smaller volume of the sample cell than usually assumed [175], as well as the temperature dependency of the buffer pH [176]. These factors will not be discussed here in detail.

2.12 Comparison of available analysis software

For the analysis of raw data, several analysis programs are available. For instance, *Origin* 7 SR4 (OriginLab Corporation, Northampton, MA, USA) is useful for peak integration and model fitting. Alternatively, *NITPIC* [155] can be used for peak integration in combination with its companion program *SEDPHAT* [177] for model curve fitting. Another option is *AFFINImeter* (*Software for Science Developments,* Santiago de Compostela, A Coruña, Spain), a web-based tool for model fitting of integrated data. In our own experience, *Origin* gives comparable results to *NITPIC/SEDPHAT* for titrations showing strong heat peaks. However, for smaller peaks with poor signal-to-noise ratio or a less well-defined base line, analysis can be tricky using *Origin*. Manual adjustment of baseline and integration limits is frequently required, and can easily induce undesired bias, especially in the hands of unexperienced users. We have found that the shape analysis and integration of heat peaks by *NITPIC* in combination with model fitting by *SEDPHAT* delivers the most unbiased, well-defined thermograms. The achievability of quality improved isotherms by *NITPIC* compared to *Origin* has also been described in literature [178]. For further improvement in data precision, *SEDPHAT* offers the combined analysis of several ITC isotherms ('global ITC'), and even offers the analysis in combination with data originating from other biophysical techniques ('global multi-method analysis') such as surface plasmon resonance (SPR) [179].

2.13 Conclusion

For the estimation of the quality of data obtained by isothermal titration calorimetry (ITC), it is necessary to develop a basic understanding of the method itself. Then, under the assumption that sufficient experimental details can be extracted from the measurement protocol, judgement about accuracy and uncertainty of thermodynamic data can be drawn. Analysis of the shape and curvature of the ITC isotherm and of the stoichiometry n of the

reaction provides information about data accuracy — the shape of the fitted model curve relative to the data points informs whether the chosen binding model is appropriate and how well the fit can actually be achieved, whereas the curvature (c-value) informs whether a reliable extraction of the thermodynamic binding parameters from the curve is possible. Curvatures that are too flat or rectangular can lead to inaccuracies in the parameter extraction and require to apply special ITC techniques, such as low c-value titrations or displacement titrations. The stoichiometry, only experimentally available for sigmoidal titration curves of a binding reaction, especially when comparable across a series of ligands, can be an indicator for the purity of both ligand and protein. Inaccuracies in the latter will likely affect the accuracy of the recorded thermodynamic parameters. A very important point is the dependence of the thermodynamic parameters on the applied measurement conditions, especially if comparison of data on an absolute scale is intended, which to our opinion is hardly possible to achieve. Nevertheless, this has been frequently done in literature, particularly to derive general rules about thermodynamic properties and optimization strategies in medicinal chemistry. A lot of care is needed in the interpretation to establish such correlations. Protonation reactions superimposed onto the actual binding event can strongly affect the measured enthalpic contribution to binding. If this is the case, the buffer effect must be corrected prior to data usage. A comparison of thermodynamic data including different, uncorrected heats of protonation will induce vast systematic errors, and artificial enthalpy–entropy compensation will arise from this lack of proper data correction. Trends can disappear in such arbitrarily correlated data. Furthermore, thermodynamic measurements have to be performed at the same temperature if mutual comparison is intended.

The best data quality can be achieved by using an experimental setup that is optimized with respect to the number of injections and the injection volume (resulting in strong heat signals and a sigmoidal curvature of the isotherm), the ratio between ligand and protein at the end of the titration (sufficient protein saturation) and the buffer conditions (small heat of dilution, experimentally determined heat of ionization). Usage of the same protein batch with unchanged concentrations across the entire experimental series and highly pure ligand, measurement at a constant temperature, and performing all steps with the same operator and ITC device are also important. If necessary, heats of ionization must be corrected. Considering the complexity of ITC experiments and the large variety of possibly superimposed systematic effects, it is highly recommended to use ITC data only for a relative comparison within narrow congeneric compound series. In our eyes, only such evaluations make sense and can lead to relevant and reliable conclusions. We also believe that classifications of ligands as "enthalpic" or "entropic" binders should only be done as relative comparisons of closely matching pairs. In any case such relative classifications have to be limited to "*more* enthalpic" or "*more* entropic" in light of the fact that with increasing

Chapter 2

temperature protein–ligand binding becomes in general more enthalpy-driven and ITC experiments are usually performed at 25 °C and not at body temperature.

For the assessment of the data quality, we rely on detailed experimental protocols provided by the experimenter. They describe the measurement parameters, raw thermograms, report ITC isotherms, assessment of possibly superimposed ionization reactions, and prove ligand purity. Unfortunately, this is often not given, even though it should be self-evident to include such data in the publication or in the supplementary material. Accordingly, putative reviewers of paper submissions are prompted to request such information from the authors. Only this will enable others to validate whether the data are suitable and reliable enough for their purposes, for instance for a computational study.

2.14 Acknowledgements

2.15 Supporting Information

2.15.1 Experimental part: Correction of the heat of ionization (Figure 2.3)

ITC measurements were performed with a Microcal ITC_{200} from GE Healthcare (Piscataway, New Jersey). Freeze-dried thermolysin powder was commercially obtained from Calbiochem (EMD Biosciences) with an activity of 10000 U mg^{-1} and a purity of 60%, additionally containing $Na(CH_3COO)$ and $Ca(CH_3COO)_2$. The protein powder was used without further processing, as no significant difference was observed in the heats of dilution without and with salt removal (via dialysis). The synthesis of the measured ligand (Figure 2.3a of the main text) was performed and the purity was confirmed as described previously [64]. The binding reaction between the ligand and thermolysin was analyzed in triplicate in four buffers differing in their heats of ionization. The four applied buffers were composed of 20 mM buffer substance (Tris, ACES, HEPES or PIPES), 2 mM $CaCl_2$ and 200 mM NaSCN. In order to increase thermolysin solubility, NaSCN was used as salt [180]. The buffers were adjusted to pH 7.5, filtered at 0.22 µm and finally degassed under vacuum for 10 minutes. Between 0.3−0.4 mg compound material and 1.5−2 mg protein powder was freshly weighted in for each measurement so that the weighting error would be included in the standard deviation of the result of the measurement. For weighting in the samples, a MX5 balance from Mettler Toledo (Switzerland) with a readability of d=1 µg and a repeatability of 0.8 µg was used. Ligand and thermolysin powder were dissolved prior to measurement in each of the four buffer solutions, resulting in each case in a clear solution. The protein solution in the sample cell was adjusted to 50 µM, and the ligand concentration in the syringe to 1 mM. Filling of the syringe with ligand solution was followed by two purge and refill steps, and then a 0.03 inch down movement of the syringe plunger prior to measurement. The same measurement scheme was applied for every measurement in order to guarantee comparability of the resulting heat signals (Figure 2.3b). Measurements were performed at a cell temperature of 298.15 K, a reference power of 5 µcal s^{-1}, and an initial delay of 200 s with 200 s spacing between each main injection. The stirring speed of the syringe was set to 1000 rpm. A first injection of 0.5 µL over a period of 1 s was followed by 10 injections of 2.1 µL with a duration of 4.2 s. This injection strategy results in a molar ration of ligand to protein of 2.2 at the end of the titration, and curves described by a *c* value of 50. Peak integration was performed with *NITPIC* [155]. One-binding site models were fitted with *SEDPHAT* (Figure 2.3c) [177], ignoring the first injection peak. Finally, ITC raw thermograms and isotherms were plotted with the graphical output tool *GUSSI*. $\Delta G°_{corrected}$ was calculated as the average of the respective $\Delta G°_{obs}$ values obtained from the measurements performed in ACES, HEPES and PIPES buffers. For the calculation of the buffer corrected thermodynamic binding enthalpy $\Delta H°_{corrected}$, the experimentally determined $\Delta H°_{obs}$ were plotted against the heats of ionization from literature [154] of ACES ($\Delta H°_{ion}$ = 30.4 kJ mol^{-1}), HEPES ($\Delta H°_{ion}$ = 20.4 kJ mol^{-1}) and

PIPES (ΔH°_{ion} = 11.2 kJ mol^{-1}) (Figure 2.3d). A straight line was fitted to the data points by linear regression. The slope of the straight line describes the proton uptake during complex formation, and the interception with the y-axis describes the buffer corrected enthalpy. Subsequently, the corrected standard entropy change was calculated according $-T\Delta S^\circ_{corrected}=\Delta G^\circ_{corrected}-\Delta H^\circ_{corrected}$. Measurement results are summarized in Table S2.1 and Figure 2.3e. The thermograms as a result of the measurements in Tris buffer were not considered for the correction of the heat of ionization, due to an insufficient fit of the 1:1 binding model. The fitting issue suggests a more complex scenario, likely due to an active displacing of bound Tris by the ligand from the active site of thermolysin. Raw ITC thermograms and isotherms of all measurements are displayed in Figure S2.1.

Table S2.1. Thermodynamic results of the thermolysin–ligand reaction measured in different buffers, and the resulting thermodynamic binding profile corrected for the heat of ionization.

Measurement	n	K_d (µM)	ΔG° (kJ mol^{-1})	ΔH° (kJ mol^{-1})	$-T\Delta S^\circ$ (kJ mol^{-1})
ACES (1)	0.853	1.339	−33.5	−10.7	−22.8
ACES (2)	0.906	1.230	−33.7	−10.5	−23.2
ACES (3)	0.873	1.489	−33.3	−10.5	−22.8
	0.877±0.027	**1.353±0.130**	**−33.5±0.2**	**−10.6±0.1**	**−22.9±0.3**
HEPES (1)	0.980	1.150	−33.9	−22.5	−11.4
HEPES (2)	0.952	1.235	−33.7	−21.9	−11.9
HEPES (3)	0.962	1.183	−33.8	−22.7	−11.2
	0.965±0.014	**1.190±0.043**	**−33.8±0.1**	**−22.4±0.4**	**−11.5±0.3**
PIPES (1)	0.983	0.705	−35.1	−33.2	−2.0
PIPES (2)	0.982	0.773	−34.9	−31.6	−3.3
PIPES (3)	1.020	0.733	−35.0	−32.7	−2.3
	0.992±0.025	**0.731±0.026**	**−35.0±0.1**	**−33.1±0.3**	**−2.0±0.3**
Buffer corrected			**−34.1±0.7**	**−46.0**	**+12.0**
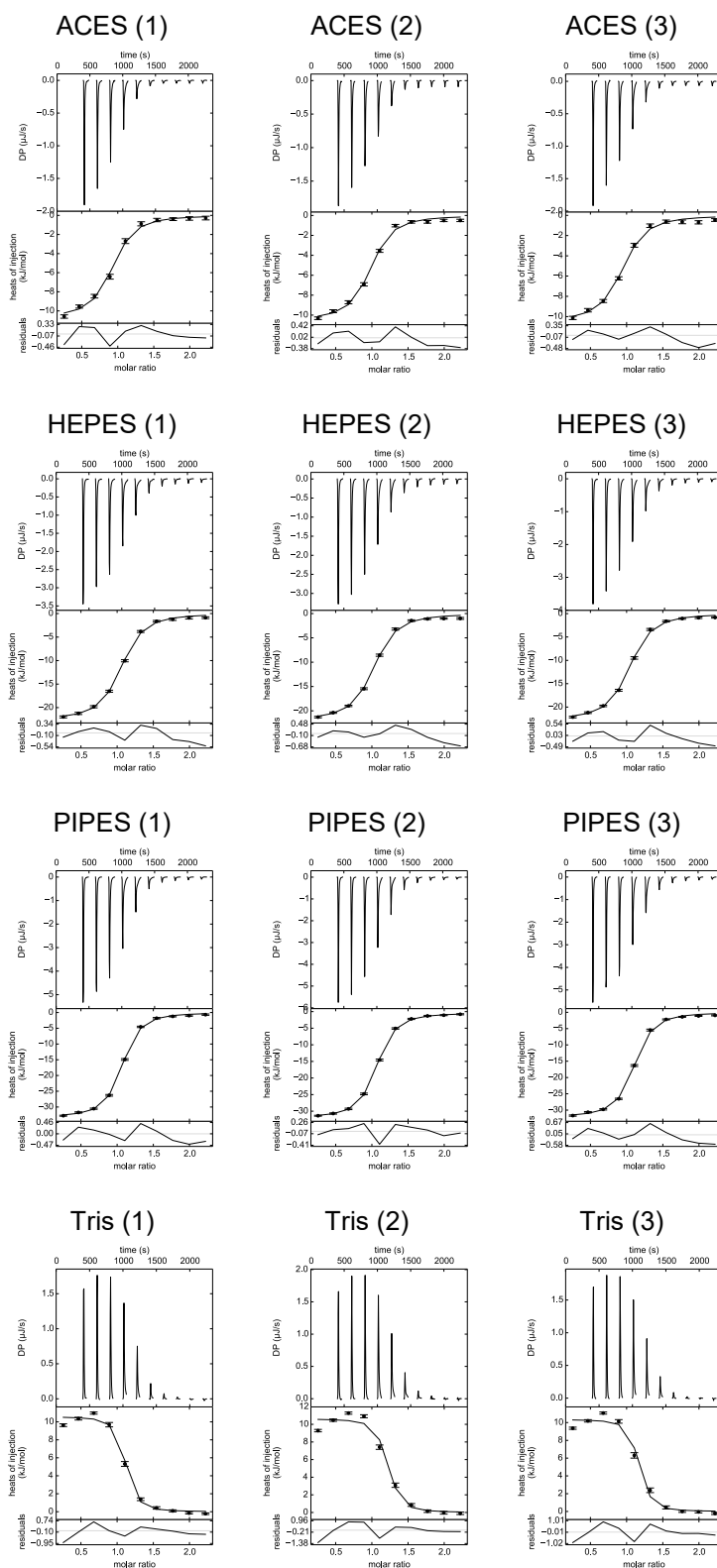
Figure S2.1. Raw ITC thermograms and binding isotherms measured in different buffers for the determination of the heat of ionization.

2.15.2 ITC measurements at different concentrations of NaSCN (Figure 2.4) and at different concentrations of thermolysin (Figure 2.5)

The ligand applied for the measurement is shown in Figure 2.4b, synthesized as reported previously [64]. For all measurements, a filtered (0.22 μm) and degassed buffer composed of 20 mM HEPES, and 2 mM CaCl$_2$, adjusted to a pH of 7.5 containing NaSCN in a concentration of either 200 or 1000 mM (Table S2.2). For the measurements of the different protein concentrations, the unusually high salt concentration of 1 M NaSCN was necessary in order to ensure protein solubility at the highest measured concentration of thermolysin of 300 μM. The measurements were performed with a ligand solution adjusted to a concentration of 1000 μM or 3000 μM (Table S2.2). Thermolysin powder was freshly weighted in for each measurement (between 2–12 mg) and diluted to the desired concentration. The same protein batch was used throughout all measurements. The syringe fill procedure was performed as described for the measurements of the correction of the heat of ionization. All measurements were performed in duplicate at a temperature of 298.15 K with the stirring speed of the syringe set to 1000 rpm. Experimental parameters were adjusted according to Table S2.2. For the measurement '200 μM (1)', the last main injection was discarded due to an overlaying measurement artefact. Heat peak integration was performed with *NITPIC*, 1:1 model fitting and parameter extraction with *SEDPHAT* and data plotting with *GUSSI*. ITC measurement results are listed in Table S2.3 and displayed in Figure S2.2.

Table S2.2. ITC measurement parameters applied for the titration of the different salt and thermolysin concentrations.

Measurement (protein conc.)	Ligand mM)	NaSCN (mM)	First injection (μL)[a]	Main injections (μL)[a]	Injection spacing (s)	Isotherm c-value[b,c]	Final molar ratio[b]
50 μM (1)	1	200	0.5	2	160	58	2.1
50 μM (2)	1	200	0.5	2	160	58	2.1
50 μM (3)	3	1000	0.3	0.7	160	53	2.2
50 μM (4)	3	1000	0.3	0.7	140	61	2.2
80 μM (1)	3	1000	0.5	1.0	150	38	2.0
80 μM (2)	3	1000	0.5	1.0	150	49	2.0
100 μM (1)	3	1000	0.5	1.3	160	37	2.1
100 μM (2)	3	1000	0.5	1.3	160	60	2.1
200 μM (1)	3	1000	0.5	3.6	160	31	2.6[d]
200 μM (2)	3	1000	0.5	3.0	160	33	2.4
300 μM (1)	3	1000	0.5	3.6	180	19	1.9
300 μM (2)	3	1000	0.5	3.6	160	23	1.9

[a]1 μL ligand solution is injected within 2 s. [b]Calculated with the calculator published earlier [131]. [c]For the calculation, the K_d value of the respective measurement was applied according to Table S2.3. [d]Molar ratio in the sample cell reached after the first small and 9 main injections.

Table S2.3. ITC measurement results obtained for different NaSCN and thermolysin concentrations according to Table S2.2.

Measurement	n	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
50 μM (1)	1.012	0.864	−34.6	−19.6	−15.0
50 μM (2)	0.933	0.855	−34.6	−19.1	−15.5
	0.973±0.056	**0.859±0.006**	**−34.6±0.0**	**−19.4±0.3**	**−15.3±0.3**
50 μM (3)	1.000	0.935	−34.4	−15.0	−19.5
50 μM (4)	0.881	0.810	−34.8	−15.1	−19.7
	0.941±0.084	**0.873±0.089**	**−34.6±0.3**	**−15.0±0.1**	**−19.6±0.2**
80 μM (1)	0.866	2.133	−32.4	−14.4	−18.0
80 μM (2)	0.896	1.640	−33.0	−14.5	−18.6
	0.881±0.021	**1.886±0.348**	**−32.7±0.5**	**−14.4±0.0**	**−18.3±0.4**
100 μM (1)	0.854	2.728	−31.8	−13.7	−18.1
100 μM (2)	0.988	1.673	−33.0	−13.0	−20.0
	0.854±0.095	**2.200±0.746**	**−31.8±0.9**	**−13.7±0.5**	**−18.0±1.4**
200 μM (1)	0.930	6.472	−29.6	−12.8	−16.8
200 μM (2)	0.898	6.053	−29.8	−12.5	−17.3
	0.914±0.023	**6.263±0.296**	**−29.7±0.1**	**−12.7±0.2**	**−17.0±0.3**
300 μM (1)	0.898	15.652	−27.4	−12.7	−14.7
300 μM (2)	0.896	12.711	−27.9	−12.3	−15.6
	0.897±0.001	**14.182±2.079**	**−27.7±0.4**	**−12.5±0.2**	**−15.2±0.6**
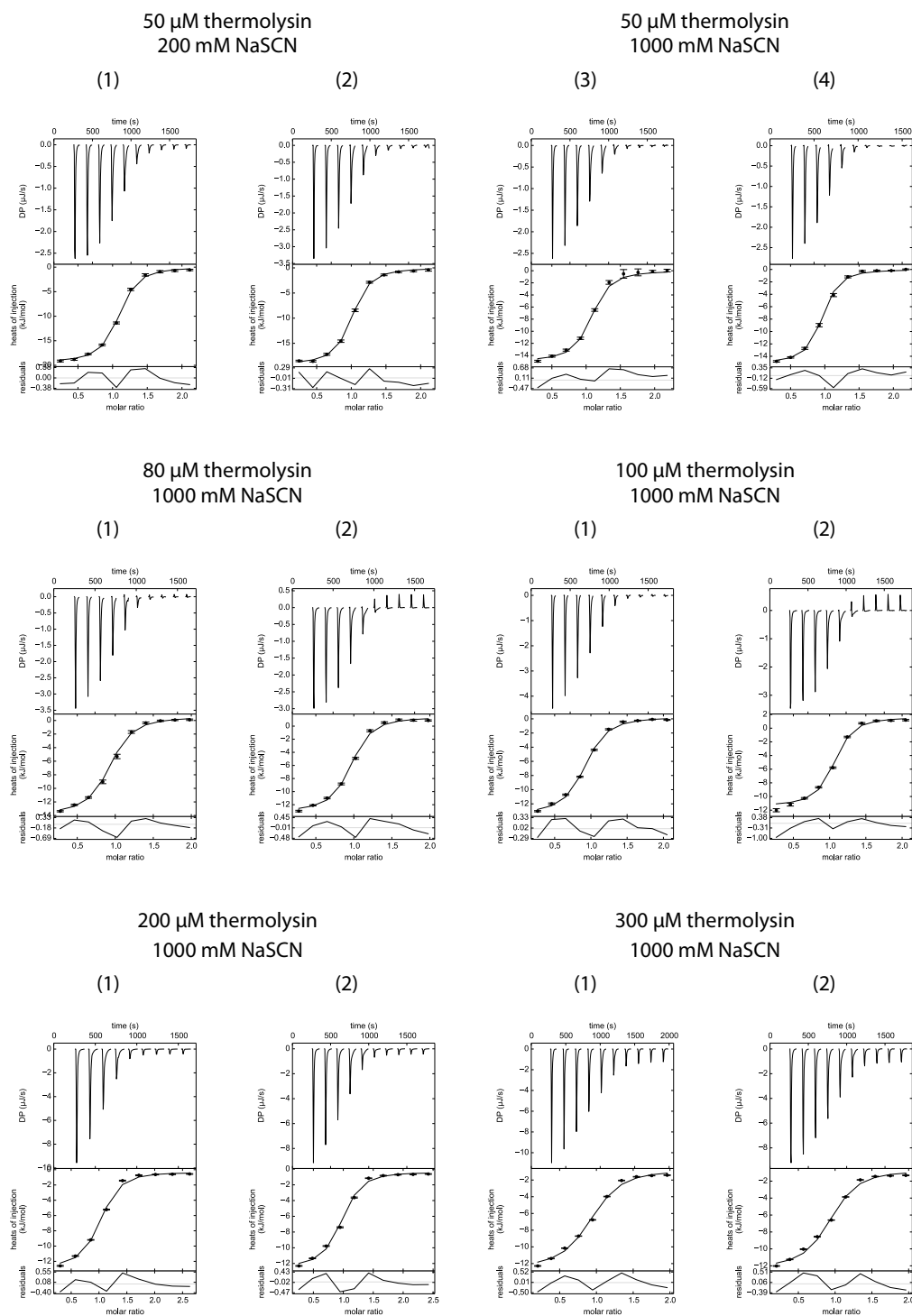
Figure S2.2. Raw ITC data and binding isotherms measured at different salt and protein concentrations.

Methyl, Ethyl, Propyl, Butyl: Futile But Not for Water, as the Correlation of Structure and Thermodynamic Signature Shows in a Congeneric Series of Thermolysin Inhibitors

Stefan G. Krimmer,[1] Michael Betz,[1] Andreas Heine,[1] Gerhard Klebe[1]

[1]Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

3.1 Abstract

Water is ubiquitously present in any biological system and has therefore to be regarded as an additional binding partner in the protein–ligand binding process. Upon complex formation, a new solvent-exposed surface is generated and water molecules from the first solvation layer will arrange around this newly formed surface. So far, the influence of such water arrangements on the ligand binding properties is unknown. In this study, the binding modes of nine congeneric phosphonamidate-type inhibitors with systematically varied, size-increasing hydrophobic P_2' substituents (from methyl to phenylethyl) addressing the hydrophobic, solvent-exposed S_2' pocket of thermolysin were analyzed by high-resolution crystal structures and correlated with their thermodynamic binding profiles as measured by isothermal titration calorimetry. Overall, $\Delta\Delta G$ spreads over 7.0 kJ mol^{-1}, $\Delta\Delta H$ varies by 15.8 kJ mol^{-1}, and $-T\Delta\Delta S$ by 12.1 kJ mol^{-1}. Throughout the series, these changes correlate remarkably well with the geometric differences of water molecules arranged adjacent to the P_2' substituents. Ligands with medium-sized P_2' substituents exhibit highest affinities, presumably because of their optimal solvation patterns around these complexes. The addition, removal, or rearrangement of even a single methyl group can result in a strong modulation of the adjacent water network pattern shifting from enthalpy to entropy-driven binding. In conclusion, the quality of a water network assembled around a protein–ligand complex influences the enthalpy/entropy signature and can even modulate affinity to a surprising extent.

Keywords: crystal structure analysis, isothermal titration calorimetry, protein–ligand interactions, structure–activity relationships, water network formation

3.2   Introduction

At the later stages of a drug development program a particular lead structure is optimized for binding to its target protein by systematically varying substituents of a promising parent scaffold. Congeneric series of candidate molecules are synthesized and structural variations are correlated with trends in binding affinity. Structure–activity relationships established this way seek for rational explanations of the observed tendencies. Usually medicinal chemists explore binding properties in a systematic fashion by producing congeneric series attaching methyl, ethyl, propyl, butyl, and other suitable aliphatic and aromatic substituents at a given position of the scaffold [181]. Frequently, these variations are regarded as rather futile. Nonetheless, even though the molecules grow gradually in their hydrophobic properties, the resulting structure–activity relationships often do not reflect these gradual changes at all. Instead, some remarkable variations are observed rendering a particular member of the series, often surprisingly, as optimal. Usually, the discovered trends are then explained by a better or inferior fit of the produced molecules to the binding pocket of the target protein.

Consistently, it appears remarkable that nature itself uses the principle of varying aliphatic side chains extensively. The arsenal of amino acids with aliphatic side chains is quite comprehensive (Gly, Ala, Val, Leu, Ile), whereas examples of basic or acidic residues appear limited, even though we believe that polar residues are much more important for molecular recognition and, thus, the drug binding process.

In all these considerations it has been disregarded that protein–ligand binding and complex formation involves a third binding partner — the ubiquitously present water. Biology has developed in water, thus it appears self-evident that water and the structural arrangement of water has taken a determinant influence on the structural properties of biomolecules. Water and aliphatic compounds appear to be species of opposite character, however, opposites often attract. Such effects have frequently been observed. For example, water produces a relatively open framework to host methane in clathrate arrangements of impressive architecture [182]. The systematic and rather complex arrangement of water networks in terms of polygonal constructions has been described at the surface of proteins [183, 184]. In consequence, we have to ask the question of how much impact the water arrangement, wrapping around the surface of proteins, has on the ligand binding.

It is well appreciated that water molecules are involved in ligand binding [31, 56]. They frequently show up in structural studies at interstitial positions, filling unoccupied gaps, triggering selectivity [185] and promiscuity [186, 187], or mediating interactions between ligands and proteins. But what about water molecules arranging on the surface of proteins as a proximate solvent layer bridging from the binding site to the bulk water phase? Do they make any impact on ligand binding? Clearly, every newly formed protein–ligand complex creates a new and altered surface next to the binding site. Water molecules will have to

arrange adjacent to this altered surface and a deviating fit in terms of the establishment of better or inferior fitting water networks will result. By how much do these properties influence the binding characteristics of ligands and can they provide some explanations for the irregularities in the structure–activity relationships observed for congeneric series with alkyl substituents of varying topology?

In the present study, we investigated the binding profile of a congeneric series of thermolysin inhibitors systematically modified at the P_2' position by attaching different hydrophobic side chains. A detailed thermodynamic analysis has been performed and correlated with the structural properties determined by high-resolution crystal structure analysis. As a target the zinc metalloprotease thermolysin (EC number 3.4.24.27) has been selected [188–190]. This protein has frequently been used in the past as a model system to validate and understand novel concepts [191–193]. It exhibits three specificity pockets of hydrophobic character [194]. The S_1 pocket is rather nonspecific and accommodates quite generally hydrophobic ligand portions. The neighboring S_1' subsite forms a deep pocket and preferentially recognizes the side chains of amino acids such as Val, Leu, Ile, and Phe. The adjacent S_2' pocket opens as a shallow, more bowl-shaped depression of hydrophobic nature, however it prefers virtually the same residues as the S_1' pocket. The S_2' site is more easily accessible by water molecules and highly variable water patterns characterize the binding of molecular portions to this subsite. Besides, thermolysin appears to be an ideal target for the attempted studies, as it crystallizes readily, exhibits a rather rigid geometry, and crystal structures with high resolution in the range of 1.1–1.6 Å can be determined routinely.

In this study, we succeeded in determining a series of nine complex structures with P_2'-modified phosphonamidate-type inhibitors showing detailed differences in the network pattern of 10–15 water molecules next to the S_2' pocket. Apart from the P_2' substituents, the ligands exhibit highly conserved binding poses within the substrate binding cleft of the protease. Remarkably, the observed differences in the water patterns correlate with affinity changes and, even stronger, with enthalpic and entropic binding contributions. They explain irregular trends in the structure–activity relationship and demonstrate why neither the smallest nor the largest hydrophobic P_2' side chains represent the optimum but instead medium-sized substituents appear to be most favorable. The addition, removal, or rearrangement of even a single methyl group at the P_2' substituent can result in a strong modulation of the adjacent water network pattern and, thus, of the thermodynamic signature of ligand binding.

3.3 Results

3.3.1 Crystal structures of the protein–ligand complexes

Structural comparison of the basic binding modes of the congeneric ligands

The nine ligands presented in this study share the same basic carboxybenzyl-Gly-(PO_2)-L-Leu-NH_2-P_2' scaffold, but differ in their substituents P_2' = methyl (**1**), ethyl (**2**), *n*-propyl (**3**), *iso*-butyl (**4**), 2-methylbutyl (**5**), *iso*-pentyl (**6**), *neo*-hexyl (**7**), *neo*-pentyl (**8**), and phenylethyl (**9**) as shown in Figure 3.1. All nine crystal structures were determined at high resolution between 1.13 and 1.60 Å (Table 3.1) and show well defined F_o–F_c difference electron densities of the bound ligands. As evident from a superposition of the complex structures with ligands **1** to **9** (Figure 3.2a; the protein–ligand complexes are referred to as TLN-**1** to TLN-**9**), the conserved scaffold adopts virtually the same binding pose in thermolysin, which has been described previously [42, 64, 194]. In contrast, the P_2' group addressing the S_2' pocket induces interesting differences. Whereas the positioning of water molecules adjacent to the conserved
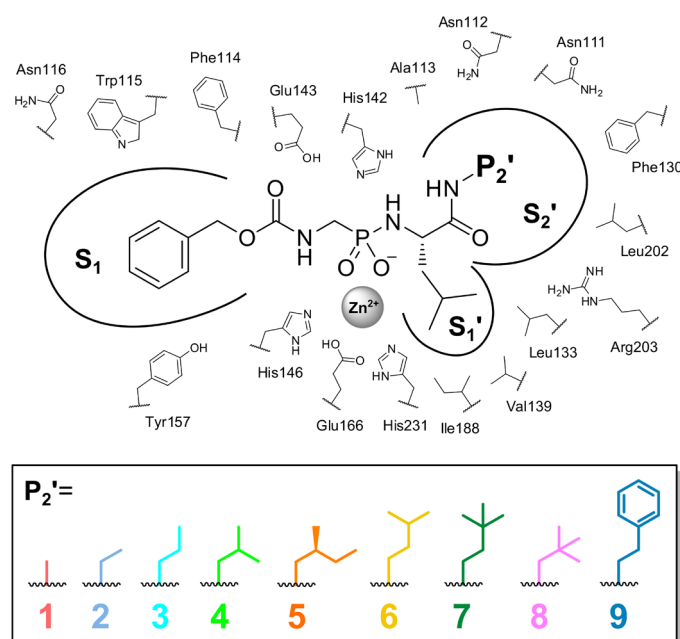
Figure 3.1. Schematic binding mode of ligands **1**–**9** within the substrate binding cleft of thermolysin. The S_2' pocket forms a hydrophobic, bowl-shaped cavity easily accessible to water molecules. The hydrophobic P_2' group of the ligands addressing the thermolysin S_2' pocket was varied systematically, whereas the basic ligand scaffold remained unchanged and shows a virtually identical binding mode to thermolysin throughout the nine complexes. The carboxybenzyl part of the basic ligand scaffold is addressing the S_1 pocket, which forms a nonspecific, hydrophobic cavity. The S_1' pocket forms a deep, hydrophobic cavity which is occupied by the isobutyl side chain of the ligand.

Table 3.1. Data collection and refinement statistics for crystal structures TLN-**3** to TLN-**9**.[a]

	Complex (PDB code)						
	TLN-**3** (4MXJ)	TLN-**4** (4MTW)	TLN-**5** (4MZN)	TLN-**6** (4N5P)	TLN-**7** (4N4E)	TLN-**8** (4N66)	TLN-**9** (4MWP)
A) Data collection and processing							
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: a, b, c (Å)	92.6, 92.6, 130.8	92.6, 92.6, 131.1	92.5, 92.5, 131.2	92.5, 92.5, 130.9	92.5, 92.5, 131.5	92.2, 92.2, 130.1	92.5, 92.5, 131.0
Matthews coefficient (Å³ Da⁻¹)[b]	2.4	2.4	2.4	2.4	2.4	2.3	2.4
Solvent content (%)[b]	48	48	48	48	48	47	48
B) Diffraction data							
Resolution range (Å)	50.00–1.35 (1.37–1.35)	50.00–1.32 (1.34–1.32)	50.00–1.17 (1.19–1.17)	50.00–1.25 (1.27–1.25)	50.00–1.13 (1.15–1.13)	50.00–1.44 (1.46–1.44)	50.00–1.23 (1.25–1.23)
Unique reflections	70475 (3410)	76659 (3237)	110592 (4798)	90096 (3866)	118776 (4789)	59295 (2786)	92439 (3704)
$R(I)_{sym}$ (%)	6.9 (47.5)	6.5 (42.1)	7.2 (30.9)	5.5 (27.3)	6.4 (28.2)	5.7 (44.5)	6.8 (20.7)
Completeness (%)	96.4 (95.1)	98.0 (84.1)	99.3 (87.9)	98.5 (86.4)	95.9 (78.7)	99.6 (96.1)	96.2 (78.2)
Redundancy	4.7 (4.4)	4.1 (2.2)	9.2 (3.1)	12.7 (5.7)	10.4 (6.9)	9.0 (8.4)	3.6 (1.6)
$<I/\sigma(I)>$	17.9 (3.8)	20.1 (2.3)	29.1 (3.8)	42.9 (6.2)	32.6 (6.3)	39.2 (5.0)	17.8 (3.9)
C) Refinement							
Resolution range (Å)	38.32–1.35	34.20–1.32	43.62–1.17	34.16–1.25	31.83–1.13	38.16–1.44	43.62–1.23
Reflections used (work/free)	70427 (66886/3541)	76600 (72736/3864)	110501 (104969/5532)	90046 (85517/4529)	118733 (112782/5951)	59267 (56279/2988)	92405 (87770/4635)
Final R value for all reflections (work/free) (%)	11.6/15.1	12.1/15.4	11.6/13.8	11.3/13.5	12.1/13.8	11.2/14.5	11.8/14.4
Protein residues	316	316	316	316	316	316	316
Calcium/zinc ions	4/1	4/1	4/1	4/1	4/1	4/1	4/1
Inhibitor atoms	27	28	29	29	30	29	32
Water molecules	464	449	476	473	462	403	433
RMSD from ideality:							
Bond lengths (Å)	0.007	0.007	0.006	0.007	0.007	0.007	0.007
Bond angles (°)	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Ramachandran plot:[c]							
Residues in most favored regions (%)	88.9	88.5	88.1	88.1	88.1	88.1	88.9
Residues in additionally allowed regions (%)	10.0	10.4	10.7	10.7	10.7	11.1	10
Residues in generously allowed regions (%)	0.7	0.7	0.7	0.7	0.7	0.4	0.7
Residues in disallowed regions (%)[d]	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Mean B factor (Å²):[e]							
Protein	10.7	11.4	10.7	11.0	10.1	12.0	10.5
Inhibitor	11.3	11.3	11.0	11.8	11.2	13.3	13.3
Water molecules	29.5	29.8	29.3	29.4	28.3	30.8	28.3

Values in parenthesis describe the highest-resolution shell. [a]Crystal structures of TLN-**1** (3T73) and TLN-**2** (3T8F) have already been published [42]. [b]Matthews coefficients and solvent contents were calculated with the *Matthews_coef* program from the *CCP4* suite version 6.3.0 [195]. [c]Ramachandran plots were calculated according to *PROCHECK* [196]. [d]The Ramachandran outlier Thr26 detectable in every complex structure of TLN is a well-known case in literature [197]. [e]Mean *B* factors were calculated with *MOLEMAN* [198].
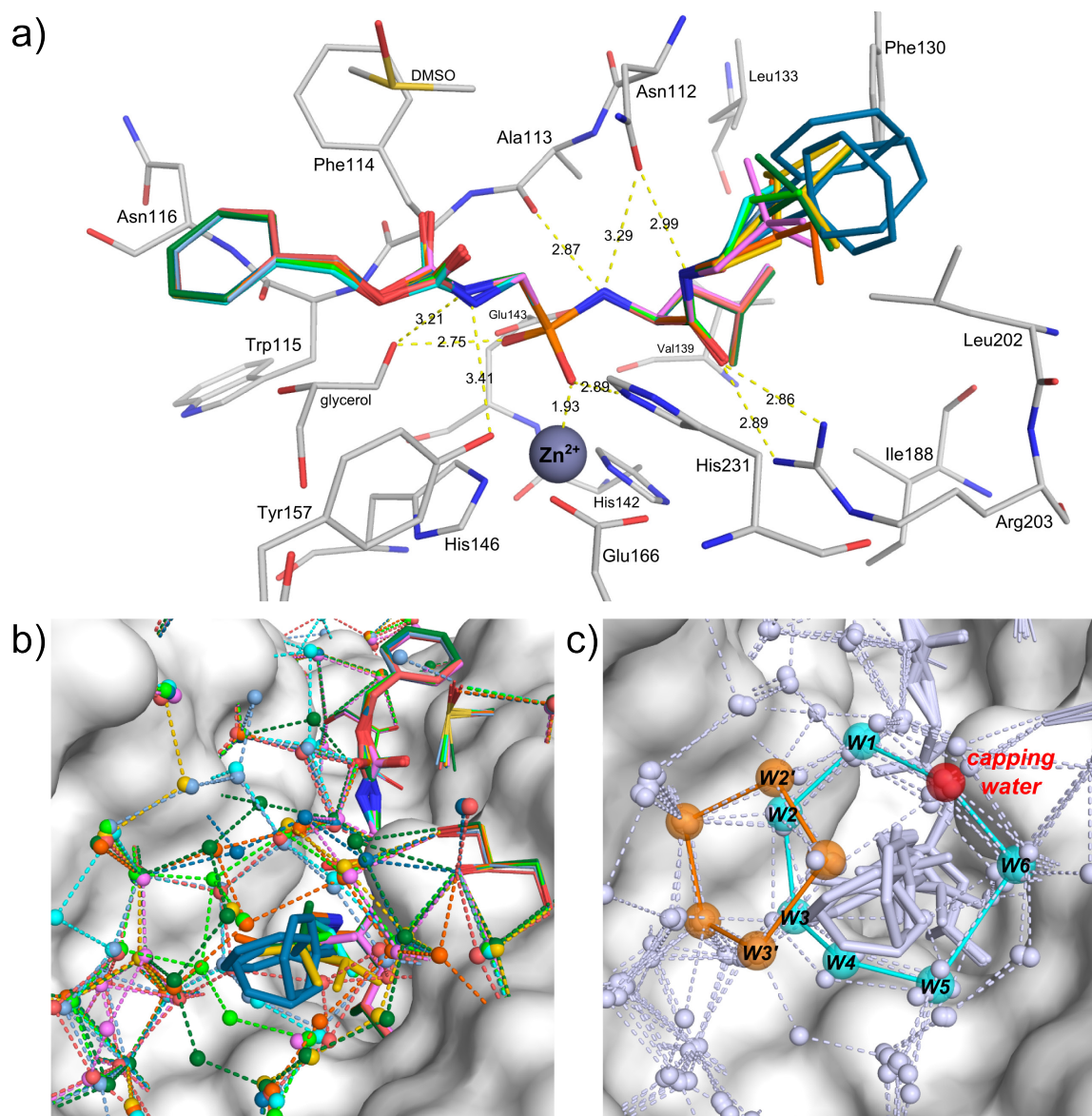
Figure 3.2. Superposition of complexes TLN-**1** to TLN-**9**. **a)** Protein residues and molecules within a distance of 4.2 Å to the superimposed ligands are displayed in gray with oxygen atoms in red and nitrogen atoms in blue. Yellow dotted lines labeled with their length in Å indicate polar interactions between protein residues and ligands. The ligands' binding modes are virtually identical, except for the varying P$_2$' substituents. All ligands show two conformations of their carbamate groups in the crystal structures. **b)** The Connolly surface of thermolysin is displayed in white, the ligands with their corresponding water molecules of the complexes in varying colors. Water molecules in the upper part of the complexes occupy virtually identical positions. In contrast, changes in the water structures are obvious in the lower left part of the complexes. **c)** Nomenclature of water molecules repeatedly observed throughout complexes TLN-**1** to TLN-**9** occupying comparable positions. The Connolly surface of the superimposed thermolysin structures is displayed in white, ligands and water molecules are displayed in pale blue. Water molecules accumulate with only slight shifts at seven distinct positions, which can be described by the red sphere (capping water) and the six cyan spheres (*W1–W6*). The straight lines in cyan visualize where H-bonds could be formed between positions of water molecules. As indicated in orange, a second pentagonal water network is repeatedly observed for some of the complexes. However, the cyan and the orange networks cannot exist simultaneously, as the orange water network is the result of a shift of the water molecules *W2* and *W3* (shifted positions described as *W2'* and *W3'*).

parts of the ligands is identical throughout the complexes (except for one water molecule in TLN-**7**), major shifts of water molecules surrounding the varying P_2' groups are obvious (Figure 3.2b). Interestingly, all nine structures contain a glycerol molecule originating from the cryo buffer between the protein surface and the carboxybenzyl group of the ligands as well as a DMSO molecule at the right rim of the complexes relative to the orientation shown in Figure 3.2b and Figure 3.3. Close to the S_2' pocket, an additional glycerol molecule is observed in all nine crystal structures at an identical position. This glycerol forms hydrogen bonds with up to two water molecules which are directly interacting with the ligands' P_2' groups via van der Waals (vdW) interactions. Physicochemically, the hydroxy groups of glycerol are considered as surrogates for water molecules [194]. We thus expect that the glycerol does not exert any artificial influence on the establishment of the water structures surrounding the S_2' pocket. In a glycerol-free crystal structure of thermolysin in complex with **7** (PDB code 4OI5, described in the Supporting Information), water molecules are observed at the same positions as the hydroxy groups of the two glycerol molecules next to the binding cleft in TLN-**7**. Additionally, no significant differences are detectable between water network patterns established in the thermolysin S_2' pocket of the glycerol-containing and glycerol-free structures.

Binding modes of the P_2' groups and their adjacent solvation patterns

The high resolution of the crystal structures TLN-**1** to TLN-**9** (Table 3.1) is critical for a reliable analysis of the water patterns, as the number of observed water molecules in a crystal structure strongly depends on its resolution [199]. Only the comprehensive detection of water networks around the protein–ligand complexes throughout the ligand series enables a reliable correlation with their thermodynamic profiles and therefore allows a conclusive interpretation.

To conveniently describe the observed water patterns and the ligand-induced changes in these water networks, we use a common nomenclature for the water molecules located adjacent to the P_2' substituents (Figure 3.2c). This assignment reflects positions at which water molecules were repeatedly observed with only minimal spatial variations throughout the nine complexes. In the following, we refer to these water molecules as capping water (*CW*), and *W1* to *W6*. As these water molecules are conserved among the different complexes, special stabilization seems to be given. Depending on the size of the P_2' substituent of a ligand, some of these seven positions may not be occupied by a water molecule in the respective complex. Water molecules which experienced only a slight shift from their conserved positions are marked with an apostrophe.
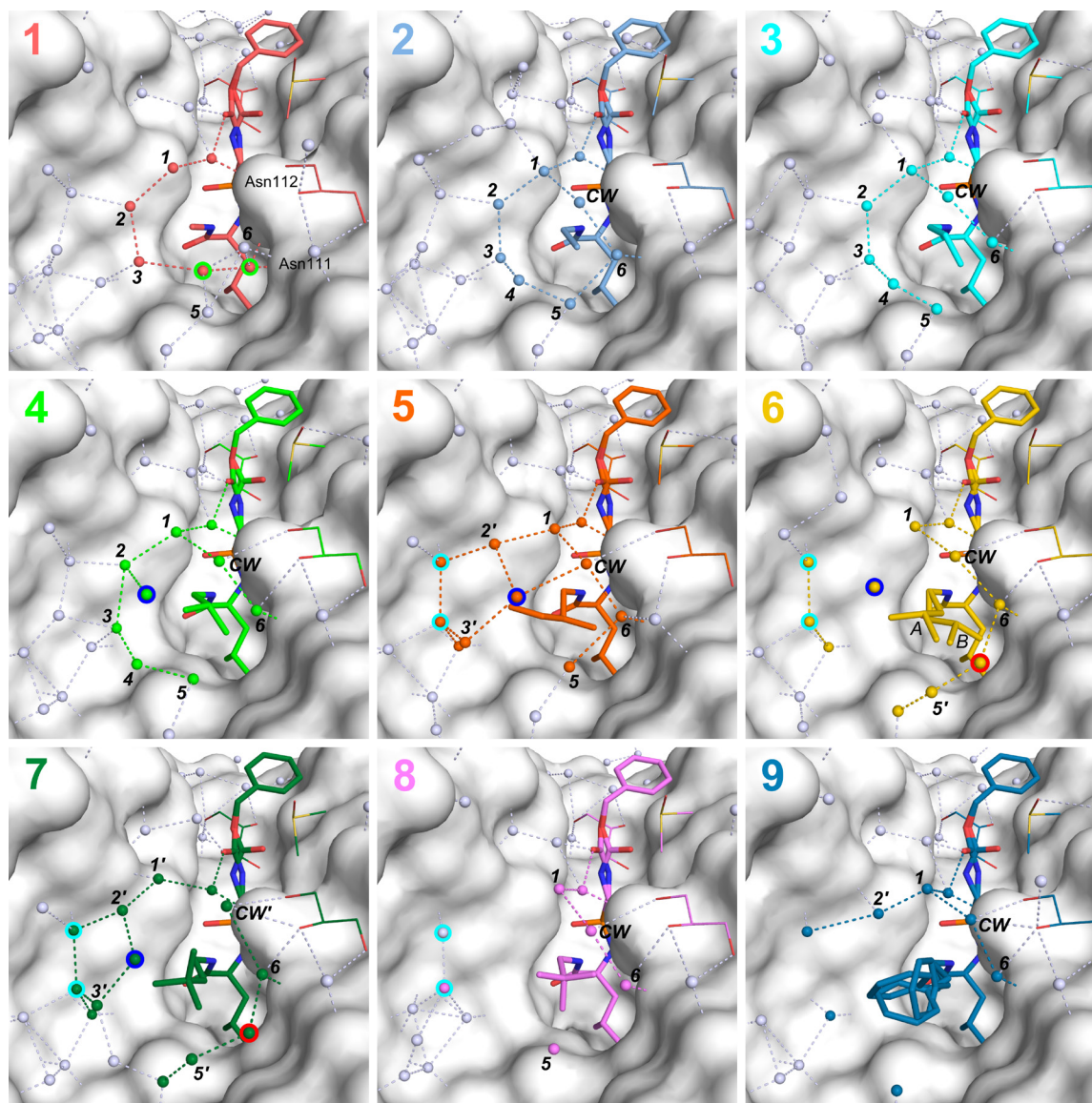
Figure 3.3. Crystal structures of complexes TLN-**1** to TLN-**9**. The Connolly surface (solvent excluded surface) of thermolysin is displayed in white. Ligands' carbon atoms are displayed with complex-assigned colors. The water molecules of each complex in vdW contact distance to the ligand (up to 4.6 Å) are illustrated by spheres in the color of their respective ligand. Water molecules beyond vdW interaction distance to the ligand are illustrated by pale blue spheres. Hydrogen bond interaction distances are displayed as dotted lines. Important water molecules, which are referred to in the text, are labeled with a number (*CW* = capping water) or encircled in varying colors.

The water network patterns around the P_2' substituents of **1** and **2** have already been described previously [42, 200]. In summary, the terminal methyl group of **1** is framed by six water molecules located within 4.0 Å (Figure 3.3), a distance where favorable vdW interactions can be established to the P_2' group. All six water molecules are arranged in a way that optimal H-bonds, at distances of 2.6–3.1 Å, can be formed, leading to a contiguous six-membered chain of water molecules. The water molecules at the two termini of this chain form H-bonds to the backbone carbonyl oxygen of Asn111 and to the side-chain nitrogen of Asn112, respectively. Thereby, a closed water network wrapping around the P_2' substituent is formed.

Ligand **2** places its terminal ethyl group into the S_2' pocket (Figure 3.3). This time, a total of eight water molecules with vdW interactions to the P_2' ethyl group (distances between 3.5–4.3 Å) form an H-bond network around this group. The involved water molecules are again located in optimal H-bond distances to each other (between 2.5–3.0 Å). Relative to **1**, the P_2' substituent of **2** replaces two water molecules (Figure 3.3 and Figure 3.4a, encircled in green). On the other hand, two new water molecules are picked up and stabilized at new positions (the capping water and water molecule *W4*), which are not found in TLN-**1**. The capping water bridges across the P_2' substituent, which leads to a cyclic polygonal and a more complex water network than the one formed by TLN-**1**. Interestingly, all further complexes exhibit this capping water (TLN-**2** to TLN-**9**, in TLN-**7** with a slight shift). Furthermore, water molecules *W5* and *W6*, which were already present in TLN-**1**, are now merged with the extended water network wrapping around the hydrophobic P_2' substituent of **2**.

The terminal propyl group of **3** fits perfectly into the right cleft of the S_2' pocket (Figure 3.5). Compared with **2**, the additional terminal methyl group forms further vdW interactions to the protein with distances between 3.7–4.1 Å (Asn112, Asn111, Phe130, and Leu202). Moreover, additional contacts between ligand and adjacent water molecules are established via this methyl group. The water networks wrapping around the P_2' substituents of **3** and **2** are virtually identical. Only slight shifts of the positions of the capping water (0.3 Å) and *W6* (0.7 Å) going away from the P_2' group are observed in TLN-**3** relative to their positions in TLN-**2**. Consequently, the distance between *W5* and *W6* increases from 3.0 Å to 3.8 Å, leading to a large expansion of the usual H-bond distance observed between these two water molecules in TLN-**2**.

The crystal structure of TLN-**4** shows nine water molecules with vdW interactions to the P_2' substituent (Figure 3.3). The binding modes of **3** and **4** are almost identical, except for the placement of the additional methyl group of **4**. However, in the lower left part of the binding pocket deviations between the locations of water molecules are clear. In comparison with TLN-**3**, the positions of water molecules *W3* and *W4* are shifted in TLN-**4** by 0.9 Å and 1.0 Å, respectively. The most remarkable change in the water network pattern adjacent to the P_2'
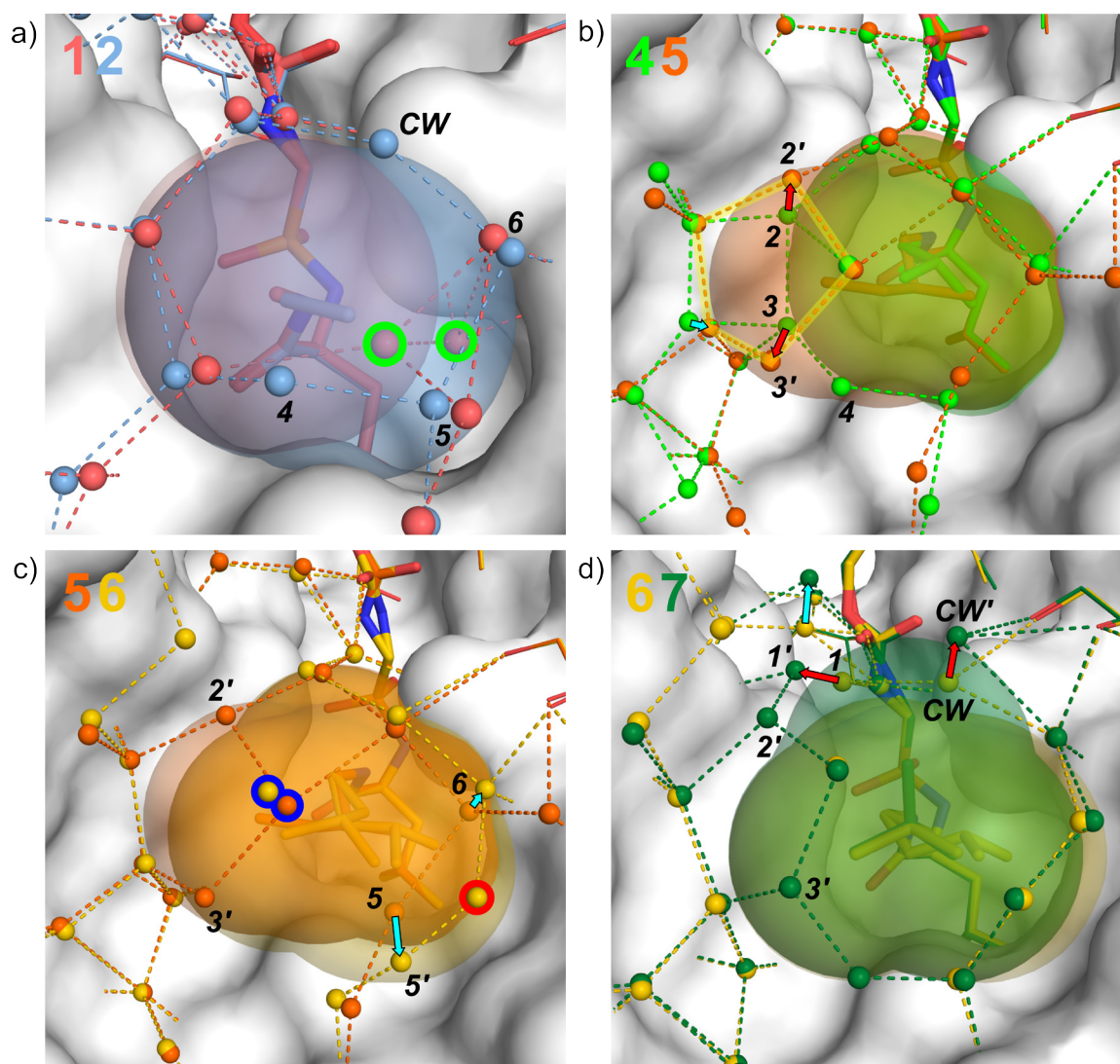
Figure 3.4. Superposition of thermolysin–ligand complexes with related water network patterns. The Connolly surface of thermolysin is displayed in white, ligands are shown with the semitransparent Lee-Richards surfaces of their P_2' groups, water molecules are illustrated as spheres and H-bond interactions are depicted as dotted lines. **a)** Superposition of TLN-**1** (light blue) and TLN-**2** (red). The two water molecules of TLN-**1** encircled in green are not detectable in TLN-**2**. Water molecules *W5* and *W6* are present in both complexes, but only in TLN-**2** they are incorporated in the network within vdW interaction distance to the P_2' group. Two additional water molecules are observed in TLN-**2**, the capping water (*CW*) and water molecule *W4*. **b)** Superposition of TLN-**4** (light green) and TLN-**5** (orange). Relative to TLN-**4**, water molecules *W2* and *W3* of TLN-**4** are shifted in TLN-**5** (*W2'* and *W3'*) as indicated by the red arrows. Additionally, one water molecule in TLN-**4** is dragged closer to the P_2' group of **5**, as indicated by the cyan arrow, whereby a pentagonal water network is established (highlighted in yellow). **c)** Superposition of TLN-**5** (orange) and TLN-**6** (yellow). In TLN-**6**, water molecules *W2'* and *W3'* of the pentacyclic water network motif observed in TLN-**5** are missing. At the lower right rim of the complex, the water network of TLN-**6** is expanded by one water molecule (encircled in red) and two water molecules are shifted (cyan arrows). The water molecule encircled in blue is observable in both complexes. **d)** Superposition of TLN-**6** (yellow) and TLN-**7** (dark green). Relative to TLN-**6**, in TLN-**7** the capping water and *W1* are shifted as displayed by the red arrows, as well as one water molecule next to the conserved part of the ligands (cyan arrow). The pentagonal water network motif on the left rim is fully detectable for TLN-**7**, whereas in TLN-**6** the water molecules *W2'* and *W3'* are missing.
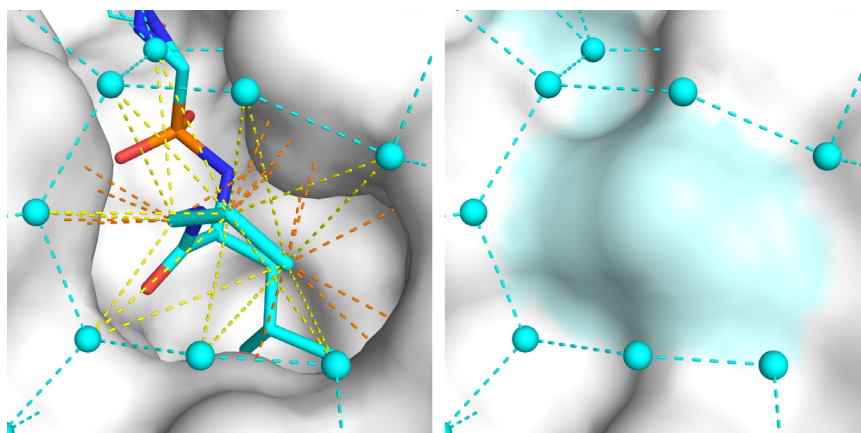
Figure 3.5. Thermolysin (Connolly surface in white) in complex with **3** (cyan). Water molecules are displayed as cyan spheres and H-bond interactions between them are depicted as dotted lines. Left panel: The orange dotted lines show vdW interactions between ligand and protein, whereas the yellow dotted lines illustrate vdW interactions between the aliphatic P_2' group and water molecules. Right panel: The surface created by **3** binding to the active site is shown in light cyan. The inhibitor perfectly fills the binding pocket and creates a new and extended surface which is capable of stabilizing water molecules wrapping around it.

substituent is the stabilization of a newly recruited water molecule (encircled in blue) with a distance of 3.4 Å to the additional methyl group of **4**. This water forms an H-bond to $W2$, as well as weak H-bonds to the capping water (4.0 Å) and to $W4$ (3.5 Å) as displayed in Figure 3.6. Relative to TLN-**3**, the distance between $W5$ and $W6$ in TLN-**4** increases by 0.3 Å to give a total of 4.0 Å due to a shift of $W5$, clearly beyond the agreed H-bonding limit of 3.5 Å [31, 166, 201]. The water molecules at the left rim of TLN-**3** ($W2$, $W3$, and $W4$) exhibit better defined electron densities and significantly lower B factors than in TLN-**4** (Figure 3.6).The P_2' substituent of **5** adopts a binding mode rotated by ~180° relative to **4** (Figure 3.3). Thus, the chiral carbon atom of the P_2' group of **5** penetrates deeper into the protein binding pocket, whereas in TLN-**4**, the substituent remains closer to the exterior. Relative to **4**, inhibitor **5** comprises an additional methyl group which is oriented toward the left side of the S_2' pocket. Interestingly, we observed major shifts in the water network adjacent to this additional methyl group (Figure 3.4b). The positions of water molecules $W2$ and $W3$ in TLN-**4** are shifted to new positions in TLN-**5** (referred to as $W2$' and $W3$') as illustrated by the red arrows in Figure 3.4b. Together with the two water molecules encircled in cyan and the water molecule encircled in blue (Figure 3.3), a pentagonal water network is established in vdW distance to the left terminal methyl group, which is characterized by H-bond distances between 2.8–3.1 Å and low to average B factors between 13–34 Å^2 (mean B factors of all water molecules in TLN-**5**: 29 Å^2). The B factor of the water molecule encircled in blue is significantly lower in TLN-**5** (B factor of 34 Å^2) than in TLN-**4** (B factor of 45 Å^2), as expected as the former is involved in more hydrogen bonds. The positions of the remaining
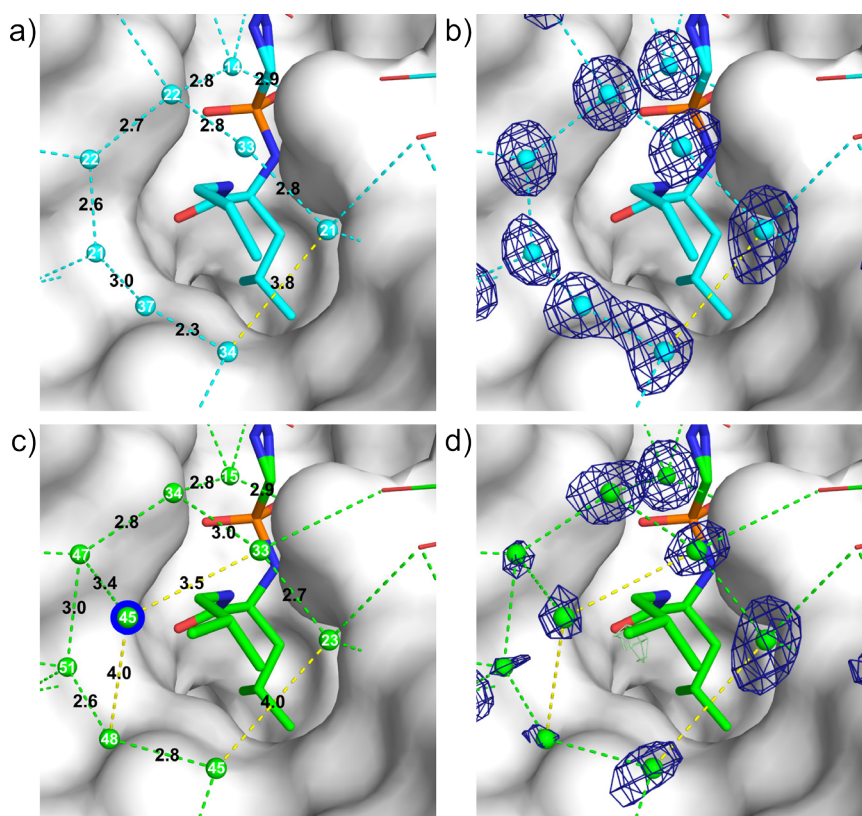
Figure 3.6. Thermolysin in complex with **3** (cyan) and **4** (green). The Connolly surface of thermolysin is displayed in white, water molecules are illustrated as spheres, H-bond interactions as dotted lines. The yellow dotted lines indicate distances beyond the maximum H-bond distance. In panels **a)** and **c)**, the B factors ($Å^2$) of the water molecules are displayed within the spheres and the H-bonds are labeled with their lengths (Å). In panels **b)** and **d)**, the $2F_o–F_c$ (blue) and the $F_o–F_c$ (green) electron density maps are shown at 0.5 e $Å^{-3}$ ($2F_o–F_c$ contour level of 0.99 σ for TLN-**3** and 1.01 σ for TLN-**4**; $F_o–F_c$ contour level of 4.51 σ for TLN-**3** and 4.45 σ for TLN-**4**). The $2F_o–F_c$ density maps of protein, ligand, and glycerol are omitted for clarity. The water molecule encircled in blue is additionally stabilized in TLN-**4** relative to TLN-**3**.

four water molecules (*W1*, *W5*, *W6*, and the capping water) not involved in the pentagonal motif adjacent to **5** are observed at virtually identical positions as for TLN-**4** (Figure 3.3). However, *W4* is no longer detectable in TLN-**5**, whereas in TLN-**4** difference electron density is detectable, sufficient to justify placement of a water molecule (B factor of 48 $Å^2$). Another difference of TLN-**5** relative to TLN-**4** is that *W5* is found at a distance of 3.2 Å to *W6* which allows the formation of a hydrogen bond in TLN-**5**.

In case of **6**, a second conformation of the P$_2$' group is visible in the $F_o–F_c$ difference electron density map at a contour level of 5.21 σ. Therefore, two conformations *A* and *B* were assigned (Figure 3.3), even though the resulting $2F_o–F_c$ electron density of conformation *B* is weak. Conformation *A* is oriented toward the left (64% occupancy), whereas conformation *B* is oriented toward the right and refines to 36% occupancy. Conformation *B* adopts an

orientation with rather close contacts to the protein surface allowing vdW interactions at distances of 3.3–4.1 Å. The water network adjacent to TLN-**6** is partly similar to that of TLN-**5** (Figure 3.4c). On the lower right hand side of the binding pocket of TLN-**6**, the water network is expanded by one water molecule (encircled in red) which sneaks in between *W5* and *W6*, simultaneously shifting these two water molecules toward the protein surface. This enables the network to wrap efficiently around the bulkier P_2' group. Consequently, attractive vdW interactions can be formed with both conformations of **6**. The pentagonal water motif on the left rim of TLN-**5** is partially preserved in TLN-**6**, although water molecules *W2*' and *W3*' (Figure 3.3 and Figure 3.4c) are no longer observable. Consequently, the *B* factors of the three remaining water molecules of the former pentagonal motif are enhanced relative to TLN-**5**. In TLN-**6**, the water molecule encircled in blue especially shows a diffuse electron density and a relatively high *B* factor of 42 Å^2 was obtained after refinement (*B* factor of 34 Å^2 in TLN-**5**). Similarly, the capping water in TLN-**6** exhibits a *B* factor of 43 Å^2 in comparison with 32 Å^2 in TLN-**5**.

The P_2' substituent of **7** shows the same binding mode as conformation *A* of **6** (Figure 3.4d). The additional methyl group of **7** relative to **6** sticks out of the S_2' pocket toward the solvent. The S_2' water network pattern is similar to those observed in TLN-**6** and TLN-**5** (Figure 3.3 and Figure 3.4c, d). On the right-hand side of the P_2' substituent, the water molecules show the same extended network also observed in TLN-**6**. On the left hand side, five water molecules are arranged in a similar pentagonal network as observed for TLN-**5** (water molecules *W2*' and *W3*' of the pentagonal motif are missing in TLN-**6**). However, in TLN-**7** the capping water and *W1* are strongly shifted relative to their positions in all other complexes (except TLN-**1**) as illustrated by the red arrows in Figure 3.4d. The new positions of these two water molecules expand their mutual distance to 4.7 Å, which lies beyond the H-bonding limit of 3.5 Å. Furthermore, a water molecule interacting with the carbonyl oxygen of the carboxybenzyl group of the ligand is also shifted as indicated by the cyan arrow (Figure 3.4d). The displacement of this water molecule interacting with the conserved part of the ligand is unique for TLN-**7** compared with the other members of the series. This shift likely occurs due to the movement of *W1*, even though there is no immediate steric conflict. At the lower rim of the TLN-**7** pocket, the water network is contiguously connected, very similar to that found in TLN-**6** (Figure 3.3).

Inhibitor **8** has a terminal *neo*-pentyl group instead of the terminal *iso*-butyl group of **4**. Due to the additional methyl group of **8** with respect to **4**, its P_2' substituent adopts a different conformation (Figure 3.3 and Figure 3.7). In consequence, only *W1* (slightly shifted), the capping water, *W5*, and *W6* are detected at their preferred positions. The remaining parts of the water network frequently observed in the other complexes is completely removed in TLN-**8**. On the left side of the S_2' pocket, several water molecules are detectable (Figure 3.3).
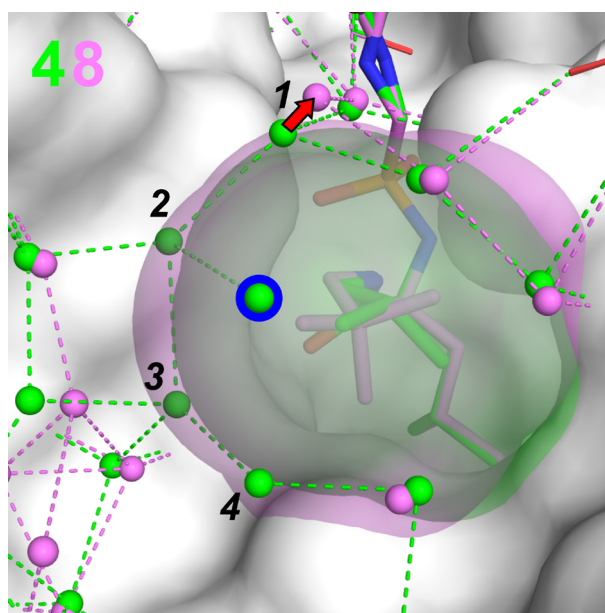
Figure 3.7. Superposition of TLN-**4** (pink) and TLN-**8** (green). Water molecules are illustrated as spheres in the color of the corresponding ligand, H-bonds between them as dotted lines. The Connolly surface of thermolysin is shown in white, whereas the semitransparent Lee-Richards surfaces of the P₂' groups of **4** and **8** are shown in the color of the respective ligand. Water molecules *W2*, *W3*, *W4*, and the water molecule encircled in blue are detectable in TLN-**4** but not in TLN-**8**. The red arrow indicates the shift of *W1* in TLN-**8** relative to the position in TLN-**4**.

All except for one of these water molecules are positioned beyond the vdW contact distance of the side chain and only interactions with the protein surface are experienced.

The P₂' phenylethyl side chain of **9** is scattered over at least three distinct conformations as indicated by the electron density map. At the lower left hand side of the S₂' pocket, the water network is almost completely disrupted (Figure 3.3). As in TLN-**8**, the capping water, *W1*, and *W6* are still found at their usual positions. Additionally, *W2* is slightly shifted and stabilized on top of the π-electron system of one of the conformations of the phenyl group.

3.3.2 Thermodynamic signatures upon complex formation

Figure 3.8 and Table 3.2 show the thermodynamic profiles of complexes TLN-**1** to TLN-**9** measured by isothermal titration calorimetry (ITC) [115, 119]. Overall, enthalpy/entropy compensation [1, 35, 144, 202] is predominantly observed across the series. The absolute changes in the enthalpic and entropic contributions are therefore much larger than the residual changes in the Gibbs free energy. TLN-**1**, TLN-**2**, and TLN-**3** gradually increase in ligand potency. This increase results from a growing exothermic binding, which is partly compensated by a descending entropic term. The improvement of $\Delta\Delta G$ from TLN-**1** to TLN-**2** is −1.9 kJ mol⁻¹ ($\Delta\Delta H_{1/2}$ =−5.1 kJ mol⁻¹, $-T\Delta\Delta S_{1/2}$ = +3.2 kJ mol⁻¹), whereas from TLN-**2** to

TLN-**3** $\Delta\Delta G$ increases twice as much, by −3.8 kJ mol⁻¹ ($\Delta\Delta H_{2/3}$ = −6.1 kJ mol⁻¹, $-T\Delta\Delta S_{2/3}$ = +2.3 kJ mol⁻¹). The largest enthalpy/entropy difference in the series is observed for TLN-**3**. From TLN-**3** to TLN-**4**, ΔG remains nearly unchanged ($\Delta\Delta G_{3/4}$ = +0.4 kJ mol⁻¹), however, the enthalpic contribution is decreased while the entropic contribution augments. TLN-**5** shows an almost identical thermodynamic profile to TLN-**4**. Overall, **3**, **4**, and **5** are the most potent inhibitors of the series and show only minor differences in ΔG. From TLN-**6** to TLN-**9**, the enthalpic term gradually decreases which is partly compensated by a beneficial entropic term. Overall, a net enhancement of the negative Gibbs free energy is found among these four ligands.
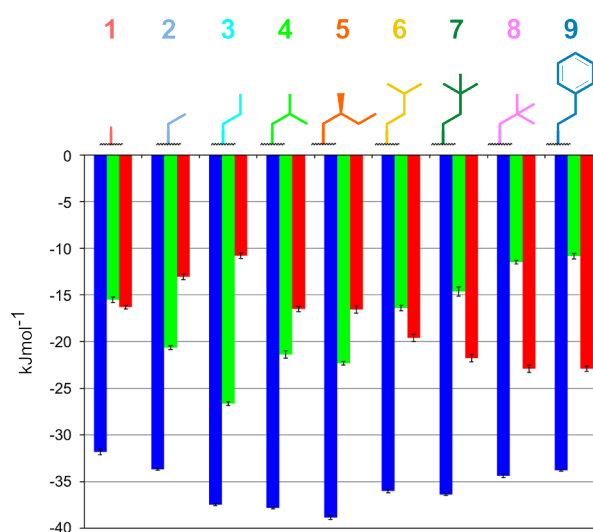
Figure 3.8. Thermodynamic profiles upon TLN-**1** to TLN-**9** complex formation obtained by ITC. ΔG (blue), ΔH (green), and $-T\Delta S$ (red) are displayed in kJ mol⁻¹ with their error bars showing standard deviations calculated from three measurements.

Table 3.2. Thermodynamic profiles of ligands **1** to **9** as determined by ITC.

Ligand	ΔG (kJ mol⁻¹)	ΔH (kJ mol⁻¹)	$-T\Delta S$ (kJ mol⁻¹)
1[a]	−31.8±0.3	−15.5±0.3	−16.3±0.2
2[b]	−33.7±0.1	−20.6±0.2	−13.1±0.3
3	−37.5±0.1	−26.7±0.2	−10.8±0.3
4	−37.9±0.1	−21.4±0.4	−16.5±0.3
5	−38.8±0.2	−22.3±0.2	−16.6±0.4
6	−36.0±0.2	−16.4±0.3	−19.6±0.4
7	−36.4±0.1	−14.6±0.5	−21.8±0.4
8	−34.4±0.2	−11.5±0.2	−22.9±0.4
9	−33.8±0.1	−10.9±0.3	−22.9±0.3

[a] Previously published data [42]. [b] Previously published data [42], which was re-measured in this study.

3.3.3 Buried and solvated molecular surface areas

The solvent exposed and buried Connolly surface areas of the P_2' groups of ligands **1**–**9** in complex with thermolysin are displayed in Figure 3.9. A continuous increase of the buried surface areas is observed for ligands **1** to **5**, followed by approximately constant buried surface areas (ligands **5** to **9**). In contrast, the solvent exposed surfaces are continuously increasing throughout the ligands, with an exception for ligand **8** whose solvated surface area falls in-between those of **4** and **5**.
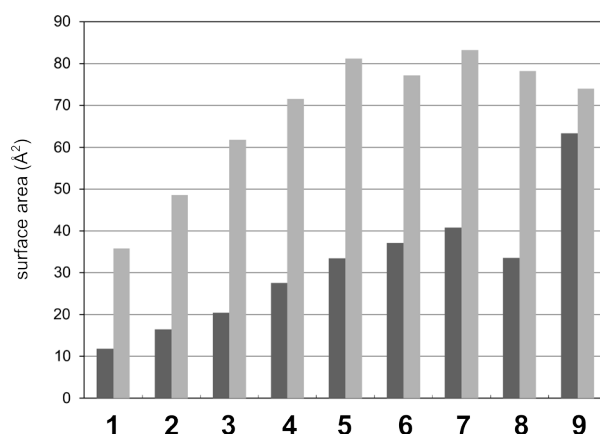
Figure 3.9. Solvent exposed (dark gray) and buried (light gray) Connolly surface areas of the P_2' groups of ligands **1**–**9** while in complex with thermolysin.

3.4 Discussion

Despite the high similarity of the inhibitors investigated in this study, we observe remarkable differences in their thermodynamic profiles. As all ligands share a common scaffold and binding mode, any changes in the ligands' thermodynamic profiles can be attributed to the varying P_2' substituents. Overall, the Gibbs free energy varies by $\Delta\Delta G = 7.0$ kJ mol^{-1} with enthalpic and entropic modulations of $\Delta\Delta H = 15.8$ kJ mol^{-1} and $-T\Delta\Delta S = 12.1$ kJ mol^{-1}.

It is well established that the behavior of water molecules critically influences the energetics of hydrophobic protein–ligand interactions [29, 56, 58, 65, 120, 203]. The rather drastic changes in ΔH and $T\Delta S$ observed in our ligand series may partially be explained by desolvation effects of either protein or ligand [56, 204]. Figure 3.9 shows the solvent exposed and buried molecular surface areas of the P_2' groups of ligands **1**–**9**. Clearly, the changes in the buried lipophilic surface areas do not correlate with the observed energetic changes during complex formation (Figure 3.8). According to the "classical" hydrophobic effect [44, 205], the entropy would be expected to increase during ligand binding due to the

displacement of water molecules from the hydrophobic binding pocket and ligand surface. However, increasing buried hydrophobic surface going from **1** to **5** is responded to by the increasing enthalpy and decreasing entropy for ligands **1** to **3** and almost identical thermodynamic profiles for ligands **4** and **5**. The virtually identical buried hydrophobic surface areas of ligands **6** to **9** also do not fit the observed thermodynamic trend of increasing entropy and decreasing enthalpy if only the conventional hydrophobic effect would be applied.

Besides the burial of parts of the hydrophobic P_2' surfaces during binding, the substituents are still partly exposed to the solvent (Figure 3.9) and therefore induce considerable changes in the water networks close to the protein surface. Clearly, these water rearrangements will affect the thermodynamic signature of ligand binding. As previously reported [42, 64, 200], the binding of a ligand creates a new local solvent-accessible surface of the protein–ligand complex. The altered exposure toward the surrounding solvent can lead to the stabilization as well as destabilization of adjacent water molecules. The stabilization of water molecules should lead to an improved enthalpic contribution. Simultaneously, the entropy will be decreased due to a significant loss of degrees of freedom as more water molecules become ordered on the surface of the protein–ligand complex. In contrast, enthalpic losses will be experienced if water molecules are displaced and electrostatic interactions between two interaction partners are ruptured (for example, H-bonds or vdW interactions), simultaneously leading to increased entropy due to decreased fixation of the network. Finally, enthalpy and entropy will compensate each other up to a certain extent [35, 206]. Based on these considerations, we attempt to correlate the changes in the observed thermodynamic signatures with changes in the water network patterns adjacent to the varying P_2' substituents. It must be emphasized that the Gibbs free energy of two ligands can match exactly, even though the partitioning of the thermodynamic signature and the adjacent water network around the P_2' residue may be significantly different (for example, as it is the case for ligands **3** and **4**). Therefore, the consideration of the Gibbs free energy alone will not enable a conclusive interpretation of the structures and can even be rather misleading.

The successive addition of methyl groups from **1** over **2** to **3** results in a remarkable increasing enthalpic gain and smaller entropic loss, leading overall to a beneficial Gibbs free energy enhancement. Interestingly, the comparison of TLN-**1** and TLN-**2** reveals an increasing complexity of the water networks surrounding the P_2' groups [42, 200]. TLN-**1** and TLN-**2** are superimposed in Figure 3.4 a showing the Lee-Richards surfaces (solvent accessible surface area (SASA), surface radius obtained by the vdW radius of the respective atom plus additional 1.4 Å for the radius of a water molecule) [207] of the ligands P_2' groups. Importantly, water molecules can only exist on or outside the Lee-Richards surface of a given molecule, but not within the surface due to repulsive interactions. Two water molecules of TLN-**1** (encircled in green) are thus displaced by **2** (semitransparent blue surface), as they

would occupy energetically unfavorable positions within the Lee-Richards surface of **2**. In contrast, water molecules *W5* and *W6*, also observed in TLN-**1**, coincide with the Lee-Richards surface of **2** and become incorporated into a water network directly wrapping around this ligand. Moreover, the capping water and *W4* are stabilized at new positions due to the extended hydrophobic surface of **2**, enabling the formation of additional methyl-water contacts. A distance of ~1.6 Å separates the Lee–Richards surface of **1** and the position where the capping water is observed in TLN-**2** (distance Lee-Richards surface of **2** to the capping water of TLN-**2**: 0.8 Å). As this water molecule is lacking in TLN-**1**, this complex is clearly unable to stabilize the capping water. Only once the additional methyl group is present as in **2**, is the water stabilized. In a previous study, we observed a remarkable decrease in the enthalpic contribution to the Gibbs free energy for complexes which do not stabilize the capping water compared to complexes which stabilize a water molecule at this position [64]. We concluded that the occupancy of this position by a water molecule is enthalpically very beneficial and may overcompensate entropy costs, which may lead to a gain in affinity. Overall, in TLN-**1**, six water molecules and nine hydrogen bonds are present to establish the water network adjacent to the P_2' group whereas TLN-**2** involves eight water molecules and eleven hydrogen bonds. The stronger fixation of water molecules by **2** compared to **1** explains the more exothermic ΔH term. The concurrently decreased water mobility is reflected by lower B factors of the water molecules in TLN-**2**. On the other hand, the decreased residual mobility of these water molecules results in a reduced entropic contribution as a consequence of a loss of degrees of freedom. However, this effect only partly compensates the observed gain in enthalpy and therefore an overall improvement in $\Delta\Delta G_{1/2}$ by –1.9 kJ mol^{-1} is remaining.

The water networks of TLN-**3** and TLN-**2** are almost identical. From TLN-**2** to TLN-**3**, the protein–ligand complex's surface only slightly expands, because the additional methyl group of TLN-**3** is pointing to the protein surface. Nevertheless, the thermodynamic profile of TLN-**3** is the most enthalpic one amongst the ligand series and the largest splitting of enthalpy/entropy is observed for this ligand (Figure 3.8 and Table 3.2). This inhibitor is also among the most potent of the series. Compared with **2**, the additional methyl group of **3** binds with optimal vdW interaction distances to the protein surface residues (Figure 3.5) and stabilizes the water network. This may lead to the observed increase of the enthalpic contribution upon TLN-**3** formation. Overall, the enthalpic term impressively overcompensates the entropic term and leads to a gain in Gibbs free energy of $\Delta\Delta G_{2/3}$ =–3.8 kJ mol^{-1}.

The size of the hydrophobic P_2' substituent grows gradually in the S_2' pocket from ligand **1** and **2** to **3**. As the observed trend in binding affinity is dominated by enthalpy, it can be attributed to a "nonclassical" hydrophobic effect [55, 208].

The thermodynamic profiles of TLN-**4** and TLN-**5** (Figure 3.8 and Table 3.2) are almost identical with a slight improvement in enthalpy in favor of TLN-**5** ($\Delta\Delta H_{4/5}$ = –0.9 kJ mol^{-1}) which also translates into a free energy gain. Relative to TLN-**3**, the enthalpic terms of TLN-**4** and TLN-**5** are decreased, but compensated by a growing entropic contribution. The free energy for all three inhibitors remains almost unchanged ($\Delta\Delta G_{3/4}$ = –0.4 kJ mol^{-1}, $\Delta\Delta G_{4/5}$ = –0.9 kJ mol^{-1}). As the thermodynamic profile of TLN-**3** differs strongly from those of TLN-**4** and TLN-**5**, their water networks have to be inspected carefully. Indeed, the water patterns of TLN-**4** and TLN-**5** reveal some impressive similarities as expected based on their comparable thermodynamic profiles, whereas major alterations relative to the water network of TLN-**3** are evident. TLN-**4** and TLN-**5** exhibit the same amount of water molecules and a similar number of hydrogen bonds. An additional water molecule (encircled in blue, Figure 3.3) compared to TLN-**3** is stabilized by both ligands. This water molecule is fixed by the increased molecular surfaces of **4** and **5**. Interestingly, the additional methyl group of **5** compared to **4** only reshuffles the adjacent water molecules, but no displacement can be observed (Figure 3.4b). This clearly results in an unchanged thermodynamic profile. The loss in enthalpy and gain in entropy with respect to TLN-**3** may be partly explained by the increased disruption of the water network next to the lower rim of the pocket of TLN-**4** and TLN-**5**. More importantly, however, might be that the water network observed in TLN-**3** results in a rather tight network nearly perfectly encapsulating the P$_2$' substituent. Figure **6** shows the electron densities of the water molecules of TLN-**3** and TLN-**4**. The four water molecules at the lower left rim of TLN-**3** (*W2*, *W3*, *W4*, and *W5*) exhibit well-defined electron densities with rather low *B* factors (21–37 Å^2). In comparison, the five water molecules at the lower left rim of TLN-**4** (*W2*, *W3*, *W4*, *W5*, and the water molecule encircled in blue) reveal relatively weak electron densities and enhanced *B* factors (45–51 Å^2). The network of TLN-**3** should therefore be enthalpically more beneficial than the increasingly extended, partly disrupted water arrangements seen in TLN-**4** and TLN-**5**, although the total amount of stabilized water molecules is larger for the latter two complexes.

With respect to free energy, inhibitors **3**, **4**, and **5** are conspicuously superior to the remaining complexes, even though they exhibit comparatively small hydrophobic P$_2$' substituents. It is remarkable that only in these complexes, one of the terminal methyl groups fills perfectly one portion of the S$_2$' pocket at the right rim next to the protein surface. This clearly leads to a considerable gain in enthalpy. The improved affinity of **5** over **4** is based on a more favorable enthalpy ($\Delta\Delta H_{4/5}$ = –0.9 kJ mol^{-1}) and can be attributed to the formation of a pentagonal water network (Figure 3.4b). Such arrays, in which water molecules form a contiguous hydrogen bonding network, are known to provide additional input to the stability of the complex [183].

The remaining inhibitors of our congeneric ligand series (ligands **6–9**) display a stepwise decrease of the enthalpic and an increase of the entropic signature. The gradual increase of the P$_2$' hydrophobicity leads to a prevalent entropic signal and a loss in binding affinity. This can therefore formally be assigned to a "classical" hydrophobic effect [44, 205].

In the crystal structure of TLN-**6**, the P$_2$' substituent is disordered and adopts two conformations, a phenomenon which should be entropically beneficial and will lower the enthalpic signal as a smaller amount of favorable directional interactions are experienced. Interestingly, the water network of TLN-**6** adjacent to the P$_2$' substituent mimics that of TLN-**5** (Figure 3.4c). However, water molecules *W2*' and *W3*' are no longer stabilized in TLN-**6**, leading to the disruption of the pentagonal motif, presumably due to the two conformations of the P$_2$' group. At the positions of missing water molecules *W2*' and *W3*' in TLN-**6**, positive F_o–F_c electron density blobs are detectable. However, these electron densities are not significantly stronger than noise, and therefore, no water molecules were assigned to them. We assume that these positions are partly occupied by water molecules which, however, are not stabilized well and therefore do not contribute significantly to the enthalpic term of complex formation, but to the entropic one. In contrast to water molecules *W2*' and *W3*', the water molecule encircled in blue (Figure 3.4c) is in vdW contact distance to both conformations of **6**, therefore sufficiently stabilized and thus detectable in the crystal structure. Both, the ligand side chain disorder and the disrupted water network may explain the observed enthalpic loss and entropic gain indicated by ITC measurements. As the enthalpic loss is only partly compensated by an entropic gain, the overall difference $\Delta\Delta G_{5/6}$ amounts to +2.8 kJ mol^{-1}.

The P$_2$' substituent of **7** adopts the same binding mode as conformation *A* of **6** (Figure 3.3 and Figure 3.4d). The additional methyl group of **7** sticks out of the binding pocket. The water network pattern adjacent to the P$_2$' substituent of TLN-**7** is also comparable to that of TLN-**6**. However, the additional solvent-exposed methyl group of **7** reshuffles the capping water and provokes a shift of water molecule *W1* for steric reasons (Figure 3.4d). The special stabilization of these two water molecules at their positions in TLN-**6** is underscored by the fact that all complexes populate water molecules at these sites, except for TLN-**1** and TLN-**7**. Their provoked displacement may result in an enthalpic loss of **7** relative to **6**. Interestingly, the pentagonal water network motif on the left rim of TLN-**5** is also found in TLN-**7** at exactly the same location, even though the positions of the stabilizing left terminal methyl groups of **5** and **7** slightly deviate by 0.8 Å. The repetitive occurrence of this pentagonal motif underlines its beneficial energetic state, which overall does not protrude from the thermodynamic profile of TLN-**7**.

Ligand **4** and **8** address the S$_2$' pocket with a *sec*-butyl and a more bulky *neo*-pentyl substituent (Figure 3.3 and Figure 3.7). Although just a single methyl group is added, the

impact on the thermodynamic profile is dramatic: the enthalpy drops by +9.9 kJ mol⁻¹ and the entropy increases by −6.4 kJ mol⁻¹. Overall, the binding free energy $\Delta\Delta G_{4/8}$ is lowered by +3.5 kJ mol⁻¹. This signature presumably originates from an extensive displacement of water molecules. Only the capping water (B factor 47 Å²), *W1*, *W5*, and *W6* are still present with **8**. All water molecules from the left side of the binding pocket are removed from the complex. The water molecules *W2* and *W3* (Figure 3.7) are displaced due to the steric requirements of the additional methyl group. Water molecule *W4* and the one encircled in blue are also repelled from TLN-**8**, even though they would be positioned just outside the Lee-Richards surface of **8**. An explanation for this displacement is likely given by the lower number of neighboring water molecules (*W2* and *W3*) no longer available for hydrogen bonding. The origin of the shift of one water molecule in TLN-**8** relative to TLN-**4**, as indicated by the red arrow, may also be due to the absent H-bond to water molecule *W2*. In TLN-**8**, instead of a displacement, a shift of water molecules *W2* and *W3* to positions outside the Lee-Richards surface of **8** does not likely occur for two reasons: Firstly, the distance between the shifted water molecules would increase and thus weaken H-bond interactions. Secondly, the shifted water molecules *W2* and *W3* would interfere with water molecules observed at the left rim of the complex and shifts of their positions would be enforced.

Ligand **9** exhibits the sterically most demanding P₂' substituent of our inhibitor series. Its phenylethyl group fills the S₂' pocket well, although the crystal structure reveals only a few vdW interactions formed with the protein surface. Similar to **8**, ligand **9** also seems to expel most water molecules from the network formed across the complex surface (Figure 3.3). Its phenylethyl moiety is characterized by three detectable conformations, which deteriorates the stabilization of water molecules, and makes it difficult to resolve their spatial location experimentally. The entropy-driven thermodynamic profile of TLN-**9** perfectly reflects these structural findings.

3.5   Conclusions

The present congeneric series of ligands, gradually modified by hydrophobic P₂' substituents, shows that simple desolvation models attributing a particular $\Delta G_{solvation}$ increment to an attached methyl group are not sufficient to explain the observed differences in the thermodynamic signatures. The attached substituents grow into the hydrophobic bowl-shaped S₂' pocket of thermolysin which is large enough to accommodate all considered moieties ranging from a methyl to a phenylethyl group. It is all the more remarkable that the binding affinity does not improve gradually with the increasing size of the hydrophobic substituent but rather seems optimal for medium-size substituents. As the ligands adopt virtually unchanged binding modes of their basic scaffold, especially the pronounced

modulations of the water network wrapping around the P$_2$' substituent should be responsible for the changes in the thermodynamic profile. Overall, $\Delta\Delta G$ spreads over 7.0 kJ mol^{-1}, $\Delta\Delta H$ varies by 15.8 kJ mol^{-1} and $-T\Delta\Delta S$ by 12.1 kJ mol^{-1}. Three derivatives should be mentioned to underline the extremes: the *sec*-butyl group of **4** differs by only one methyl group from iso-pentyl **5** and *neo*-pentyl **8**. Nonetheless, in TLN-**4** a rather perfect water network wraps around the substituent and is stabilized by favorable water–methyl contacts. The additional methyl group in TLN-**8**, however, creates contacts which are too short for two of the bridging water molecules present in TLN-**4**. In consequence, these water molecules are expelled from the complex and as a result two more water molecules are displaced from the complex due to lacking network stabilization. Accordingly, the *neo*-pentyl derivative **8** loses affinity ($\Delta\Delta G_{4/8} =$ +3.5 kJ mol^{-1}) and a more entropy-driven binding is observed. The difference between TLN-**5** and TLN-**8** is even more impressive: the *iso*-pentyl derivative **5** displays the best binding with the second best enthalpic profile. Shifting a single methyl group to yield the *neo*-pentyl derivative **8** sacrifices 4.4 kJ mol^{-1} in ΔG and turns an enthalpy-driven binder into an entropic one. This difference results from the disruption of a nearly perfect water network which becomes largely repelled from the protein surface.

What can we learn from this study? First of all, the quality of a water network generated locally at the newly created surface of a protein–ligand complex has a significant influence on the binding data of the ligand. Not only the enthalpy/entropy profile can be modulated but more importantly affinity is affected to a surprising extent. Secondly, the optimization strategy of attaching various isomeric alkyl side chains to a given scaffold often regarded as futile, makes perfect sense. The best binder is likely to be the one which creates the most perfect solvation pattern around the formed complex. The solvation/desolvation effect, frequently modeled implicitly by adjusting contributions of buried or exposed surface patches, is not sufficient. Instead, explicit modeling considering the perfection of formed water networks is required.

Admittedly, the present study is a fortunate case, as the experimental structure determinations and the accuracy of the performed ITC titrations allow the data to be analyzed in great detail. We agree, as discussed in previous publications [42, 64], that the analysis approaches the limits of experimental accuracy accomplishable in contemporary protein–ligand structural work. Most likely, routine cases in drug discovery will not disclose the same details as reported here, however, most likely, they are also given there. Nevertheless, such case studies are of utmost importance to unravel the details in operation which make the establishment and understanding of structure–activity relationships so difficult.

Chapter 3

3.6   Experimental Section

3.6.1   Ligand synthesis

The compounds were synthesized as described previously [200].

3.6.2   Crystallization and soaking

Crystals were grown by the sitting drop vapor diffusion method as described earlier with only slight modifications [190, 209]. Thermolysin is commercially available from Calbiochem (EMD Biosciences) with a purity of 62.9%, and used for crystallization without further processing. To yield a concentration of 8 mM, the lyophilized powder was dissolved in pure DMSO and mixed with a vortex mixer for 3 sec. Then, the same volume of crystallization buffer, composed of 100 mM Tris/HCl, pH 7.5 and 3.8 M CsCl, was added to obtain a 4 mM protein solution. The solution was mixed again for 3 sec, followed by centrifugation for 3 min at RCF = 16 000 g. From the clear supernatant, 1 μL was transferred into each well of a crystallization plate. The plate's reservoirs were filled with demineralized water. After at least five days at 18 °C, crystals were obtained and used for ligand soaking. For this purpose, the crystals were transferred into a soaking buffer composed of 100 mM Tris/HCl, pH 7.5, 2 mM $CaCl_2$ as well as 0.2–1 mM of the inhibitors. After soaking for 4–5 h, crystals TLN-**3** to TLN-**9** were transferred into a cryo buffer composed of 10 mM Tris/HCl, pH 7.5, 10 mM $Ca(CH_3COO)_2$, 5% DMSO, and 20% glycerol, incubated for 5 sec and subsequently flash-frozen in liquid nitrogen. For the glycerol-free crystal structure, a cryo buffer composed of 10 mM Tris/HCl, pH 7.5, 10 mM $Ca(CH_3COO)_2$, 5% DMSO, and 20% PEG 400 was applied.

3.6.3   Data collection and processing

The diffraction data of complexes TLN-**3** to TLN-**9** have been collected on the beamline 14.2 operated by the Joint Berlin MX-laboratory at the BESSY II electron storage ring (Berlin-Adlershof, HZB, Germany) [210] with synchrotron radiation at a wavelength of 0.91841 Å on a Rayonix 225 mm CCD detector at a temperature of 100 K. All diffraction images were indexed, processed, and scaled using the *HKL*-2000 package [211]. Data collection and refinement statistics are summarized in Table 3.1.

Structure determination and refinement: The structures of TLN-**3** to TLN-**9** were determined by molecular replacement with the coordinates of a thermolysin search model (PDB accession code 8TLN) [212] using *Phaser* [213] from the *CCP*4 suite [195]. A subset of randomly selected 5% of all reflections were omitted during refinement, but used for the calculation of R_{free}. Model building was performed in *Coot* [214] and refinement using

PHENIX.refine versions 1.8.3_1479 (applied for structures with PDB codes 4MTW and 4MWP) and 1.8.4_1492 (4MXJ, 4MZN, 4N5P, 4N4E, and 4N66) [215]. Rigid body refinement was applied for the first round of refinement, followed by Cartesian simulated annealing. Subsequently, alternate cycles of structural adaptions and refinement were performed until *R* values reached convergence. Temperature factors of all atoms, except for hydrogen atoms, were refined anisotropically. Ligands were built with *MOE* [216] and energetically minimized with *SYBYL* [217]. Ligand restraints were generated with *eLBOW* from the *PHENIX* suite. The models were validated using *MolProbity* [218] as implemented in *PHENIX.refine*.

Alternate conformations of side chains were fitted into well-defined positive F_o–F_c electron densities, even if the resulting $2F_o$–F_c density for the second conformations turned out to be weak. Alternative conformations of side chains were only modeled if their occupancy exceeded 20% after refinement. Only those water molecules which were, after refinement, characterized by a $2F_o$–F_c electron density peak with σ ≥ 1.0 RMSD and a *B* factor not significantly larger than 50 Å² were kept in the structures. Water molecules which displayed negative F_o–F_c electron density after refinement were constrained to an occupancy of 50% and kept as partly occupied water molecules in the structure in case the negative density disappeared after refinement. Water molecules associated with a second side chain conformation were constrained to the occupancy of the respective side chain.

The terminal methyl group of conformation *A* of **6** oriented toward the left (Figure 3.3) was only detectable by a weakly defined positive F_o–F_c electron density and shows a relatively weak $2F_o$–F_c electron density after refinement. The observed positive F_o–F_c electron density, however, is exactly at the position where, according to the geometry of this group, the terminal methyl group should be positioned. Therefore, we decided to place this methyl group.

3.6.4   Isothermal titration calorimetry (ITC)

Measurements were performed with a ITC_{200} instrument from GE Healthcare (Piscataway, New Jersey) as described earlier with slight modifications [190, 209]. Freeze-dried thermolysin was purchased from Calbiochem (EMD Biosciences). The ITC experiments of inhibitors **2**–**9** were performed at 25 °C in a buffer composed of 20 mM HEPES, 200 mM NaSCN, and 2 mM $CaCl_2$ at pH 7.5. Solubility of the hydrophobic thermolysin strongly increases in the presence of NaSCN [180], whereas NaCl and other salts did not sufficiently increase to reach the concentration needed for the ITC measurements. Consequently, NaSCN was used in the buffer as solubilizing agent for thermolysin. However, compared with the 500 mM NaSCN in the buffer used previously [42, 64], the NaSCN concentration

was lowered to 200 mM to decrease its chaotropic effect [163]. Inhibitor solutions with concentrations of 0.4 mM were prepared by directly dissolving the inhibitors in pure buffer. The sample cell contained 200 μL buffer solution with 30 μM thermolysin. The lyophilized powder was freshly dissolved in ITC buffer and centrifuged for 8 min at RCF = 8150 g. Although the lyophilized thermolysin powder contained additional salts, no signal change was detected for samples with and without these salts (removed via dialysis). Therefore, the freeze-dried thermolysin was used without further processing. Measurements were started after the ITC_{200} instrument showed a stable baseline. A first injection with a volume of 0.3 μL inhibitor solution was followed by at least 20 further injections of 1–1.5 μL (injection spacing 180 sec). All measurements were performed in triplicate. The same inhibitor solution was used for all three measurements, whereas thermolysin was freshly prepared before each measurement, to avoid self-digestion of highly concentrated thermolysin samples [194]. Collected data were analyzed using *ORIGIN* 7 SR4 (OriginLab Corporation, Northampton, MA, USA). The initial injection peak was omitted and the areas of the remaining peaks were integrated. The correction of the heats of dilution was conducted by the subtraction of the final, constant injection peak area integrals. The "one set of sites" fitting model (1:1 stoichiometry) was applied, whereby ΔH and K_d were directly obtained [115]. The ΔG and $-T\Delta S$ values were manually calculated using a temperature of 298.15 K.

The measurements for ligands **1** and **2** were already performed previously [42] using a buffer composed of 20 mM HEPES, 500 mM NaSCN, 2 mM $CaCl_2$ at pH 7.5 and 25 °C and slightly different measurement conditions. Ligand **2** was re-measured using the new experimental conditions. Significant deviations between the result of the measurement with the old conditions (ΔG = –34.0±0.3 kJ mol^{-1}, ΔH = –13.0±0.6 kJ mol^{-1}, $-T\Delta S$ = –21.0±0.9 kJ mol^{-1}) and the result of the re-measurement with new conditions (Table 3.2) were observed, although several other ligands tested with both conditions did not show any significant deviations between their measured thermodynamic profiles. The most likely explanation for this is that the values for the enthalpic and the entropic contributions were accidentally interchanged in the results published earlier.

3.6.5   Molecular surface area calculation

The ligands' buried and solvated Connolly surface areas were calculated based on the atomic coordinates obtained by the crystal structures of TLN-**1** to TLN-**9** with the program *MS* [219] applying typical vdW radii for the respective atom types [220] and a probe radius of 1.4 Å. For ligands **6** and **9** with multiple P_2' group conformations, the calculations were performed for the conformation of the respective ligand exhibiting the highest occupancy. Buried and solvated surface areas were calculated for the isolated P_2' groups as well as for the entire ligands (Figure 3.9 and Supporting Information Table S3.2).

3.7   Accession codes

The atomic coordinates and structure factors of thermolysin in complex with ligands **1** to **9** were deposited in the RCSB Protein Data Bank (PDB) [80] with accession codes 3T73 (TLN-**1**), 3T8F (TLN-**2**), 4MXJ (TLN-**3**), 4MTW (TLN-**4**), 4MZN (TLN-**5**), 4N5P (TLN-**6**), 4N4E (TLN-**7**), 4N66 (TLN-**8**), 4MWP (TLN-**9**) and 4OI5 (glycerol-free structure of thermolysin in complex with **7**).

3.8   Abbreviations

TLN, thermolysin; ITC, isothermal titration calorimetry; RCF, relative centrifugal force

Chapter 3

3.9   Acknowledgements

3.10 Supporting Information

3.10.1 Comparison of the glycerol-containing and the glycerol-free crystal structures of ligand 7 in complex with thermolysin

Figure S3.1a shows the water networks spanning around the P_2' groups of the glycerol-free (dark green) and the glycerol-containing (pale brown) crystal structures of **7** in complex with thermolysin. The observed positions of the water molecules around the P_2' substituents of both crystal structures show high agreement. However, certain differences are detectable. Overall, the electron density map of the glycerol-free structure (resolution 1.30 Å, Table S3.1) is less well defined than the one of the glycerol-containing structure (resolution 1.13 Å, Table 3.1). The electron densities are depicted in Figure S3.1c and Figure S3.1d.

3.10.2 Analysis of the positions of the glycerol molecules and the ligands' binding modes

At all six positions of the hydroxy groups of the two glycerol molecules of the glycerol-containing structure close to the active site, electron densities were detectable in the glycerol-free crystal structure and water molecules could be assigned to them. Figure S3.1b shows the thermolysin S_1 pocket with the view onto the superimposed ligands' carboxybenzyl groups. Clearly, all three positions of the glycerol hydroxy groups of the glycerol-containing structure are occupied by water molecules in the glycerol-free crystal structure exhibiting well-defined electron densities. For the glycerol molecule on the right rim of Figure S3.1a and Figure S3.1c, two water molecules were assigned to the densities, while one water molecule occurs at two mutually excluding positions *A* and *B* with a distance of 1.79 Å as shown in Figure S3.1d.

As visible in Figure S3.1b, the carboxybenzyl group of **7** in the glycerol-free crystal structure is positioned closer to the protein surface than the one in the glycerol-containing crystal structure due to the three missing glycerol carbon atoms. As a result of the more deeply buried binding mode of this part of the ligand, only one conformation of the carbamate group is observable in the electron density. However, the remaining ligand parts adopt virtually the same binding modes in both structures, including their P_2' substituents.

At the right upper side of the glycerol-free structure (Figure S3.1d), an unexplained F_o–F_c electron density is observable. A DMSO molecule is present at this position in the glycerol-containing structure. However, in the glycerol-free crystal structure, a DMSO molecule is not unambiguously assignable to this position. Accordingly, we left this density unassigned.

3.10.3 Analysis of the water network patterns adjacent to the P$_2$′ substituents

As visible in Figure S3.1a, water molecules *W1*′ and *W2*′ and the three water molecules encircled in red are not detectable in the glycerol-free crystal structure in the difference electron density. At the hypothetical position of *W2*′ in the glycerol-free crystal-structure, a F_o–F_c density map peak is observable (Figure S3.1d, electron density mesh shown in light green). However, the density is not significantly stronger than noise, therefore we decided against placing a water molecule at this position. One reason, why the density is not sufficient for the placement of a water molecule, may be attributed to the lower resolution of the diffraction data compared to that of the glycerol-containing structure. Considering the remaining structure, no differences between the glycerol-containing and the glycerol-free crystal structures are observable close to the three water molecules indicated in red. These water molecules only form H-bonds to water molecules observable in both complexes but not to the glycerol molecule at the right rim of the complex.

In the glycerol-free crystal-structure, the capping water (*CW*′, Figure S3.1a) is slightly shifted by 0.6 Å (red arrow, Figure S3.1a) and exhibits an increased *B* factor (56 Å^2) compared to the one of the glycerol-containing structure (*B* factor of 38 Å^2). In the upper part next to the ligand and beyond interaction distance to the P$_2$′ group, one water molecule is shifted in space (cyan arrow). This shift likely occurs due to the slightly displaced position of the amide carbonyl oxygen of 0.8 Å (Figure S3.1a and Figure S3.1b).
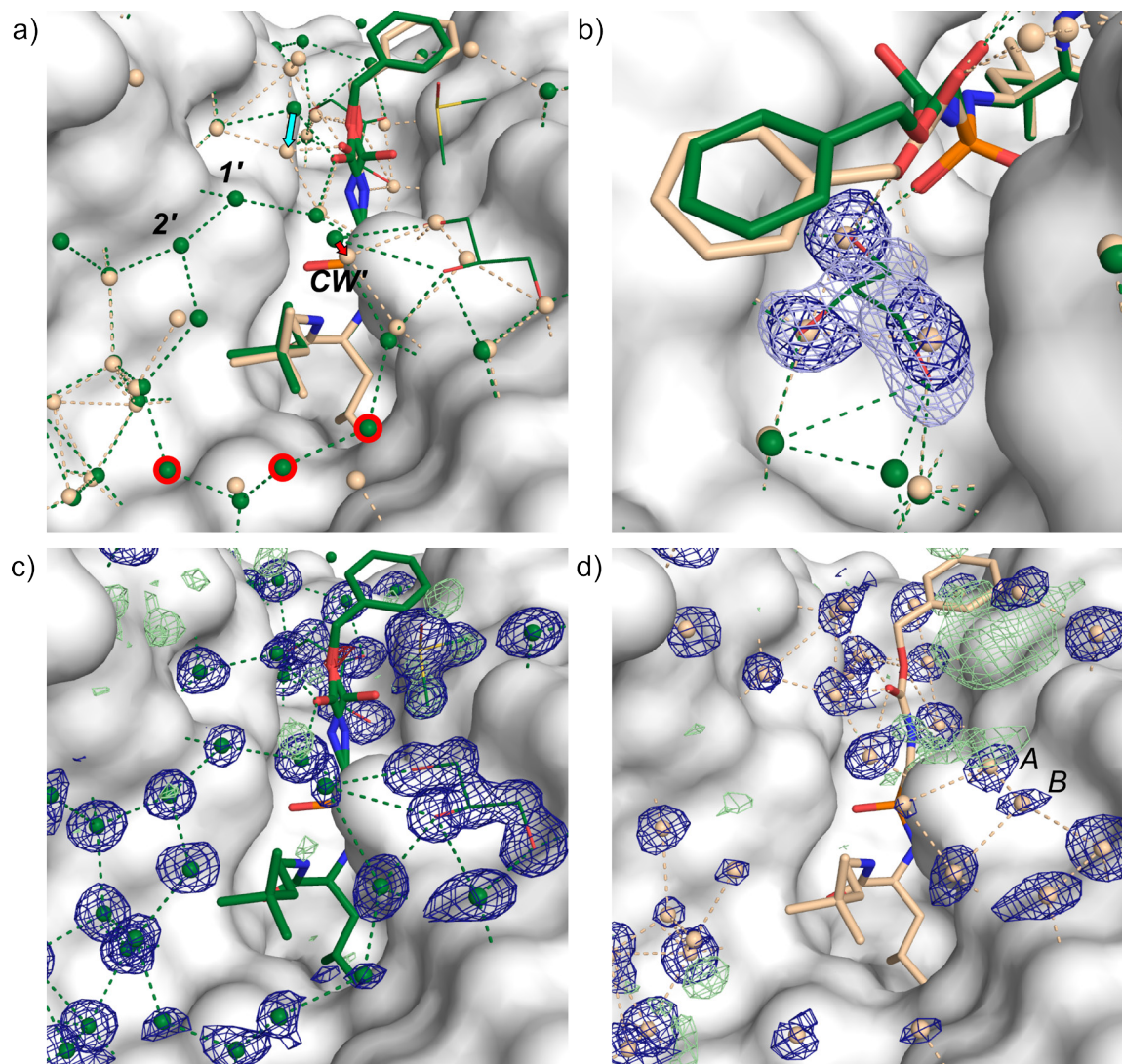
Figure S3.1. Glycerol-containing (dark green) and glycerol-free (pale brown) crystal structures of **7** in complex with thermolysin. The Connolly surface of thermolysin is displayed in white, the ligands' carbon atoms are shown in complex-assigned colors, oxygen atoms are shown in red and nitrogen atoms in blue. Water molecules are represented as spheres in the color of the corresponding ligand and H-bond interaction distances between water molecules are indicated by dashed lines. Important differences in the water networks between the complexes are labeled or encircled and discussed in the text. The F_o–F_c and $2F_o$–F_c electron density maps are shown at a contour level of 0.5 eÅ^{-3}. **a)** and **b)** Superposition of the glycerol-containing and the glycerol-free crystal structures. In **b)**, the $2F_o$–F_c electron densities of the glycerol of the glycerol-containing structure (light blue) and of the three water molecules of the glycerol-free crystal structure (dark blue) are shown. In **c)** and **d)**, the $2F_o$–F_c (blue) and the F_o–F_c (light green) electron density maps are shown within a distance of 10 Å around the ligands. The electron densities of protein and ligand are omitted for clarity.

Table S3.1. Data collection and refinement statistics for the glycerol-free crystal structure of thermolysin in complex with **7**.

	Glycerol-free TLN-**7** (4OI5)
A) Data collection and processing	
Space group	$P6_122$
Unit cell parameters: a, b, c (Å)	92.4, 92.4, 130.3
Matthews coefficient (Å³Da⁻¹)[a]	2.3
Solvent content (%)[a]	47
B) Diffraction data	
Resolution range (Å)	50.00–1.30 (1.32–1.30)
Unique reflections	80778 (3965)
$R(I)_{sym}$ (%)	7.9 (48.7)
Completeness (%)	99.9 (100.0)
Redundancy	5.8 (5.5)
$<I/\sigma(I)>$	22.4 (3.7)
C) Refinement	
Resolution range (Å)	38.16–1.30
Reflections used in refinement (work/free)	80717 (76664/4053)
Final R value for all reflections (work/free) (%)	12.2/14.6
Protein residues	316
Calcium/zinc ions	4/1
Inhibitor atoms	30
Water molecules	469
RMSD from ideality:	
Bond lengths (Å)	0.007
Bond angles (°)	1.1
Ramachandran plot:[b]	
Residues in most favored regions (%)	89.3
Residues in additionally allowed regions (%)	9.6
Residues in generously allowed regions (%)	0.7
Residues in disallowed regions (%)[c]	0.4
Mean B factor (Å²):[d]	
Protein	10.4
Inhibitor	14.1
Water molecules	28.0

Values in parenthesis describe the highest resolution shell. [a] Matthews coefficients and solvent contents were calculated with the Matthews_coef program from the CCP4 suite version 6.3.0 [195]. [b] Ramachandran plots were calculated according to PROCHECK [196]. [c] The Ramachandran outlier Thr26 detectable in every complex structure of TLN is a well-known case in literature [197]. [d] Mean B factors were calculated with MOLEMAN [198].

Table S3.2. Calculated Connolly surface areas of **1** to **9** in complex with thermolysin.[a]

Complex	Entire ligand surface (Å²)[b]	Buried ligand surface (Å²/%)[c]	Entire P₂′ group surface (Å²)[d]	Buried P₂′ group surface (Å²/%)	Solvated P₂′ group surface (Å²/%)
TLN-**1**	370.5	285.0/76.9	47.7	35.8/75.1	11.9/24.9
TLN-**2**	384.2	298.4/77.7	65.1	48.6/74.7	16.5/25.3
TLN-**3**	399.2	309.1/77.4	82.3	61.8/75.1	20.5/24.9
TLN-**4**	413.5	316.6/76.6	99.2	71.6/72.2	27.6/27.8
TLN-**5**	429.4	327.5/76.3	114.7	81.2/70.8	33.5/29.2
TLN-**6**	439.3	329.2/74.9	114.3	77.2/67.6	37.1/32.4
TLN-**7**	449.4	337.4/75.1	124.0	83.2/67.1	40.8/32.9
TLN-**8**	425.4	323.4/76.0	111.8	78.2/69.9	33.6/30.1
TLN-**9**	459.3	326.1/71.0	137.4	74.0/53.9	63.4/46.1

[a] Connolly surface areas were calculated with the program *MS* [219]. [b] Both conformations of the ligands' carbamate groups were considered for the calculation. [c] The thermolysin zinc atom was excluded in the calculation of the buried surface areas. [d] In case of multiple P₂′ group conformations, the one with the highest occupancy was used for the calculation of the surface areas.

3.10.4 Isothermal titration calorimetry (ITC) measurement results

All 27 ITC isotherms which were used for the thermodynamic characterization of the ligands **1**–**9** (Figure 3.8 and Table 3.2 of the main text) and their results for n, K_a, ΔH, and $-T\Delta S$, as calculated by the nonlinear "one set of site" curve fitting model of *ORIGIN*7 SR4, are shown below. As visible from the measurement graphs, the analyzed mean amount of binding sites per thermolysin macromolecule is 0.753±0.04, instead of the theoretical binding site value of 1.0 per macromolecule. Compared to the theoretical value, the lower molar binding ratio n is explainable by partial inactivity of the protein and is seen as a concentration correction factor [132]. However, the experimentally obtained lower binding stoichiometry n has no influence on the dimension of K_a, ΔH, and ΔS. Nevertheless, the binding ratio should be stable throughout all measurements, as strongly varying values indicate problematic measurement conditions, e.g. the use of impure ligands [149]. In contrast to the protein concentration in the cell, the knowledge of the exact concentration and purity of the ligand is critical for the value of K_a, ΔH, and ΔS [149], which can be ensured by accurate weighting and the use of highly pure ligands.

Raw ITC data and binding isotherms

TLN-**1**a

n 0.744±0.00563 Sites
K_a 3.25E5±1.72E4 M⁻¹
ΔH -3603±36.84 cal/mol
ΔS 13.2 cal/mol/K

TLN-**1**b

n 0.782±0.0109 Sites
K_a 4.26E5±4.95E4 M⁻¹
ΔH -3740±70.75 cal/mol
ΔS 13.2 cal/mol/K

TLN-**1**c

n 0.751±0.0177 Sites
K_a 3.74E5±6.96E4 M⁻¹
ΔH -3738±121.7 cal/mol
ΔS 13.0 cal/mol/K

TLN-**2**a

n 0.766±0.00569 Sites
K_a 8.01E5±4.90E4 M⁻¹
ΔH -4866±49.01 cal/mol
ΔS 10.7 cal/mol/K

TLN-**2**b

n 0.776±0.00480 Sites
K_a 8.40E5±4.41E4 M⁻¹
ΔH -4945±41.51 cal/mol
ΔS 10.5 cal/mol/K

TLN-**2**c

n 0.788±0.00445 Sites
K_a 7.65E5±3.53E4 M⁻¹
ΔH -4977±38.05 cal/mol
ΔS 10.2 cal/mol/K

TLN-**3**a

n 0.709±0.00330 Sites
K_a 3.87E6±3.03E5 M⁻¹
ΔH -6313±43.79 cal/mol
ΔS 8.97 cal/mol/K

TLN-**3**b

n 0.723±0.00255 Sites
K_a 3.53E6±2.01E5 M⁻¹
ΔH -6400±33.35 cal/mol
ΔS 8.50 cal/mol/K

TLN-**3**c

n 0.755±0.00311 Sites
K_a 3.61E6±2.49E5 M⁻¹
ΔH -6402±39.20 cal/mol
ΔS 8.54 cal/mol/K

TLN-**4**a

n   0.762±0.00366 Sites
K_a 4.28E6±3.73E5 M⁻¹
ΔH   -5059±36.83 cal/mol
ΔS   13.4 cal/mol/K

TLN-**4**b

n   0.758±0.00385 Sites
K_a 4.10E6±3.71E5 M⁻¹
ΔH   -5038±38.69 cal/mol
ΔS   13.4 cal/mol/K

TLN-**4**c

n   0.690±0.00356 Sites
K_a 4.46E6±4.08E5 M⁻¹
ΔH   -5220±40.64 cal/mol
ΔS   12.9 cal/mol/K

TLN-**5**a

n   0.763±0.00297 Sites
K_a 6.23E6±5.35E5 M⁻¹
ΔH   -5329±32.95 cal/mol
ΔS   13.2 cal/mol/K

TLN-**5**b

n   0.761±0.00329 Sites
K_a 6.90E6±6.93E5 M⁻¹
ΔH   -5277±36.77 cal/mol
ΔS   13.6 cal/mol/K

TLN-**5**c

n   0.738±0.00370 Sites
K_a 6.06E6±6.42E5 M⁻¹
ΔH   -5379±42.31 cal/mol
ΔS   13.0 cal/mol/K

TLN-**6**a

n   0.680±0.00350 Sites
K_a 1.93E6±1.18E5 M⁻¹
ΔH   -3999±25.96 cal/mol
ΔS   15.3 cal/mol/K

TLN-**6**b

n   0.683±0.00376 Sites
K_a 2.17E6±1.50E5 M⁻¹
ΔH   -3869±30.03 cal/mol
ΔS   16.0 cal/mol/K

TLN-**6**c

n   0.694±0.00442 Sites
K_a 1.97E6±1.51E5 M⁻¹
ΔH   -3891±34.71 cal/mol
ΔS   15.7 cal/mol/K

TLN-**7**a

n 0.763±0.00463 Sites
K_a 2.36E6±1.95E5 M^{-1}
ΔH -3394±29.43 cal/mol
ΔS 17.8 cal/mol/K

TLN-**7**b

n 0.762±0.00424 Sites
K_a 2.35E6±1.78E5 M^{-1}
ΔH -3482±27.68 cal/mol
ΔS 17.5 cal/mol/K

TLN-**7**c

n 0.777±0.00526 Sites
K_a 2.46E6±2.34E5 M^{-1}
ΔH -3619±35.17 cal/mol
ΔS 17.1 cal/mol/K

TLN-**8**a

n 0.792±0.00548 Sites
K_a 9.79E5±6.19E4 M^{-1}
ΔH -2790±26.15 cal/mol
ΔS 18.1 cal/mol/K

TLN-**8**b

n 0.729±0.00444 Sites
K_a 1.12E6±6.39E4 M^{-1}
ΔH -2675±22.18 cal/mol
ΔS 18.7 cal/mol/K

TLN-**8**c

n 0.722±0.00543 Sites
K_a 1.06E6±7.26E4 M^{-1}
ΔH -2755±28.21 cal/mol
ΔS 18.3 cal/mol/K

TLN-**9**a

n 0.785±0.00690 Sites
K_a 8.56E5±6.65E4 M^{-1}
ΔH -2644±31.98 cal/mol
ΔS 18.3 cal/mol/K

TLN-**9**b

n 0.843±0.00565 Sites
K_a 8.03E5±4.80E4 M^{-1}
ΔH -2501±23.04 cal/mol
ΔS 18.6 cal/mol/K

TLN-**9**c

n 0.842±0.00560 Sites
K_a 8.19E5±4.77E4 M^{-1}
ΔH -2633±23.78 cal/mol
ΔS 18.2 cal/mol/K

Chapter 3

3.10.5 Calculation of final thermodynamic binding signatures

The changes in the Gibbs' free energy ΔG (kJ mol^{-1}) of the ligands binding to thermolysin were calculated for 298.15 K applying Eq. 3.1:

$$\Delta G \text{ (kJ mol}^{-1}) = -\frac{8.314 \frac{\text{J}}{\text{mol K}} \cdot 298.15 \text{ K} \cdot \ln(K_a)}{1000} \tag{3.1}$$

The enthalpic term was calculated applying Eq. 3.2:

$$\Delta H \text{ (kJ mol}^{-1}) = \frac{\Delta H \frac{\text{cal}}{\text{mol}} \cdot 4.184 \frac{\text{J}}{\text{cal}}}{1000} \tag{3.2}$$

The entropic term was calculated according to Eq. 3.3:

$$-T\Delta S \text{ (kJ mol}^{-1}) = \Delta G - \Delta H \tag{3.3}$$

3.10.6 C values of ITC isotherms

The calculated c values of all ITC isotherms are listed in Table S3.3. The c value is applied as an indicator for the interpretability of an ITC curve and, according to Wiseman et al. [115], should fall into a range between 1 and 1000, whereas also several other, more narrow windows were suggested [132]. As listed in Table S3.3, all c values of the ITC isotherms are between 11 and 158, whereby a reliable analysis of K_a, ΔH, and n is possible.

Table S3.3. *C* values for the ITC isotherms of TLN-**1** to TLN-**9**.

ITC isotherm	c value[a]
TLN-**1**a[b]	11
TLN-**1**b[b]	15
TLN-**1**c[b]	13
TLN-**2**a	18
TLN-**2**b	20
TLN-**2**c	18
TLN-**3**a	82
TLN-**3**b	77
TLN-**3**c	82
TLN-**4**a	98
TLN-**4**b	93
TLN-**4**c	92
TLN-**5**a	143
TLN-**5**b	158
TLN-**5**c	134
TLN-**6**a	39
TLN-**6**b	44
TLN-**6**c	41
TLN-**7**a	54
TLN-**7**b	54
TLN-**7**c	57
TLN-**8**a	23
TLN-**8**b	24
TLN-**8**c	23
TLN-**9**a	20
TLN-**9**b	20
TLN-**9**c	21

[a] *c* value = thermolysin concentration [M] · K_a [M^{-1}] · *n*. [b] Thermolysin concentration = 45 μM (30 μM for the other measurements).

3.10.7 PyMOL file of protein–ligand complexes 1–9

A pre-assembled *PyMOL* [221] file (Figure S3.2) with all protein–ligand complexes from the Figures of the main text is made available online at www.agklebe.de in the Download-Area section and at http://pc1664.pharmazie.uni-marburg.de/water_networks_in_thermolysin_CMC2014.zip.
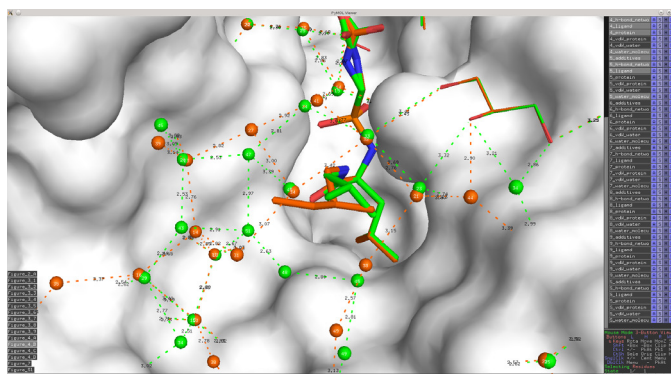
Figure S3.2. *PyMol* user interface with clickable presets at the left hand side of the screen.

Chapter 4

Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands

Stefan G. Krimmer,[1,*] Jonathan Cramer,[1,*] Michael Betz,[1] Veronica Fridh,[2]
Robert Karlsson,[2] Andreas Heine,[1] Gerhard Klebe[1]

*These authors contributed equally

[1]Department of Pharmaceutical Chemistry, University of Marburg,
Marbacher Weg 6, 35032 Marburg, Germany

[2]GE Healthcare Bio-Sciences AB, SE-751 84 Uppsala, Sweden

4.1   Abstract

A previously studied congeneric series of thermolysin inhibitors addressing the solvent-accessible S_2' pocket with different hydrophobic substituents showed modulations of the surface water layers coating the protein-bound inhibitors. Increasing stabilization of water molecules resulted in enthalpically more favorable binding signature, overall enhancing affinity. Based on this observation, we optimized the series by designing tailored P_2' substituents to improve and further stabilize the surface water network. MD simulations were applied to predict the putative water pattern around the bound ligands. Subsequently, the inhibitors were synthesized and characterized by high-resolution crystallography, microcalorimetry and surface plasmon resonance. One of the designed inhibitors established the most pronounced water network of all inhibitors tested so far, composed of several fused water polygons, and showed 50-fold affinity enhancement with respect to the original methylated parent ligand. Notably, the inhibitor forming the most perfect water network also showed significantly prolonged residence time compared to the other tested inhibitors.

4.2   Introduction

Structure-based drug design (SBDD) seeks to optimize ligand binding with respect to a given target protein. Thermodynamic parameters such as ΔG, ΔH, $-T\Delta S$, ΔC_P and of binding kinetic properties, such as kinetic association (k_on) and dissociation (k_off) rate constants, are considered to rationalize and accelerate affinity optimization by a better characterization of the protein–ligand binding process [34, 121, 122, 222–225]. Thermodynamic profiling, at its best based on isothermal titration calorimetry (ITC) [115, 116, 226], is supposed to provide insights into the molecular interactions determining the affinity of a ligand to its target. However, past experience has shown that optimization and prioritization of compounds guided by thermodynamics is difficult, since the enthalpy–entropy profile reflects the binding event as a whole. Hence, the high complexity of this event largely impedes factorization into individual contributions to binding and often only succeeds for congeneric ligand series with minor structural variations [29, 39, 125]. The correlation of structural properties with binding kinetic data is presently poorly understood and reasonable correlations have only been established for a limited number of cases [169]. Furthermore, we still face an incomplete comprehension of the fundamental relationships between thermodynamics, kinetics and molecular interactions [1, 169, 227], which can lead to false predictions made under overly simplified assumptions. Especially, the impact of the versatile and ubiquitously present water molecules is hardly understood. The involvement of water molecules is a major cause for the inherent complexity, particularly arising from the ability of water to establish H-bonds and the resultant tendency to arrange in differently ordered structures [23]. Water molecules can actively mediate H-bonds across the binding interface between protein and ligand and thereby improve affinity [9]. The displacement of well-ordered water molecules from apolar surfaces is discussed as the driving force of the hydrophobic effect (and *not* the formation of hydrophobic interactions), a process of utmost importance for molecular recognition and drug action. According to the so-called "classical" hydrophobic effect [31, 39, 120, 147, 150, 228], the binding of an apolar ligand portion to a hydrophobic protein cavity correlates with an entropic advantage due to the displacement of well-ordered water molecules into the bulk water phase. In contrast, a "non-classical" hydrophobic effect has been defined, which is enthalpy driven. It has been related to a suboptimal hydration of a protein cavity prior to ligand binding [56, 57, 229]. In this case, the enthalpy gain upon binding results from the displacement of orientationally mobile and thus entropically favored water molecules into the bulk phase where they can establish better H-bonds than previously observed in the protein cavity.

Despite the popular binary classification of "classical" and "non-classical", the hydrophobic effect can range from entirely entropy-driven to entirely enthalpy-driven. This is determined by the thermodynamic properties of the water molecules that are involved before and after

binding and by the ligand and the binding site (especially by their molecular shape and polarity), which affect the formation of adjacent water networks [45]. Of utmost importance for the thermodynamic signature is the overall inventory of water molecules with respect to their release into or recruitment from the bulk water phase, and the embedding of water molecules into H-bonding networks of varying completeness and perfection in the protein-bound state. In particular, the way water molecules are able to rearrange and establish H-bonding networks around the newly formed solvent-exposed surface of the protein–ligand complex seems to have a significant impact on the thermodynamic binding signature [42, 64, 65, 104, 203].

In a previous study [104], we structurally and thermodynamically analyzed a series of congeneric thermolysin (TLN) inhibitors with a peptidomimetic parent scaffold (Figure 4.1A) decorated with different hydrophobic P_2' substituents (Figure 4.1B). We selected the zinc metalloprotease TLN from *Bacillus thermoproteolyticus* for our studies [209, 230, 231], as this enzyme has been frequently used as class representative and exhibits excellent crystallographic properties [232, 233]. In addition, TLN is quite rigid because of its high thermal stability, thus reducing structural adaptions and facilitating comparative analyses. The active site of TLN is composed of three subpockets (Figure 4.1A). Firstly, the rather unspecific S_1 pocket, a hydrophobic cavity that recognizes aromatic portions such as Phe. Secondly, the S_1' specificity pocket, a predominantly hydrophobic, deep and narrow pocket, which preferentially accommodates hydrophobic amino acids such as Val, Leu, Ile and Phe, and thirdly the S_2' subsite, a hydrophobic, flat, bowl-shaped pocket, which is well-accessible to bulk water molecules [230]. We selected the S_2' pocket for our studies. Within a previously investigated series [104], we increased the size of the P_2' substituents addressing the well-solvated S_2' pocket systematically from a sole methyl to a phenylethyl substituent as displayed in Figure 4.1B. By detailed ITC analyses (Figure 4.1B), we revealed a difference in binding affinity (K_d: dissociation constant of the thermodynamic equilibrium) of more than one order of magnitude [104], or expressed as standard Gibbs free energy, a maximum $\Delta\Delta G°$ of 7.0±0.4 kJ mol^{-1}. Remarkably, $\Delta\Delta G°$ factored in a huge enthalpy–entropy variation, and indicated pronounced enthalpy–entropy compensation, a phenomenon frequently observed in drug optimization [31, 36, 144, 234]. We could correlate the observed variations in the thermodynamic profiles with crystallographically observed structural changes of the P_2' substituents and, triggered thereby, modulations of the adjacent surface water solvation layer [104]. The $\Delta H°$ contributions appeared to be more favorable (and $-T\Delta S°$ less favorable) where a better-ordered water network was established next to the surface of the newly formed protein–ligand complex. In contrast, $-T\Delta S°$ apparently increased (and $\Delta H°$ decreased) where the first solvation layer next to the bound ligand was unfavorably disrupted. Furthermore, the increasing contribution to desolvate the gradually growing P_2' substituents seemed to
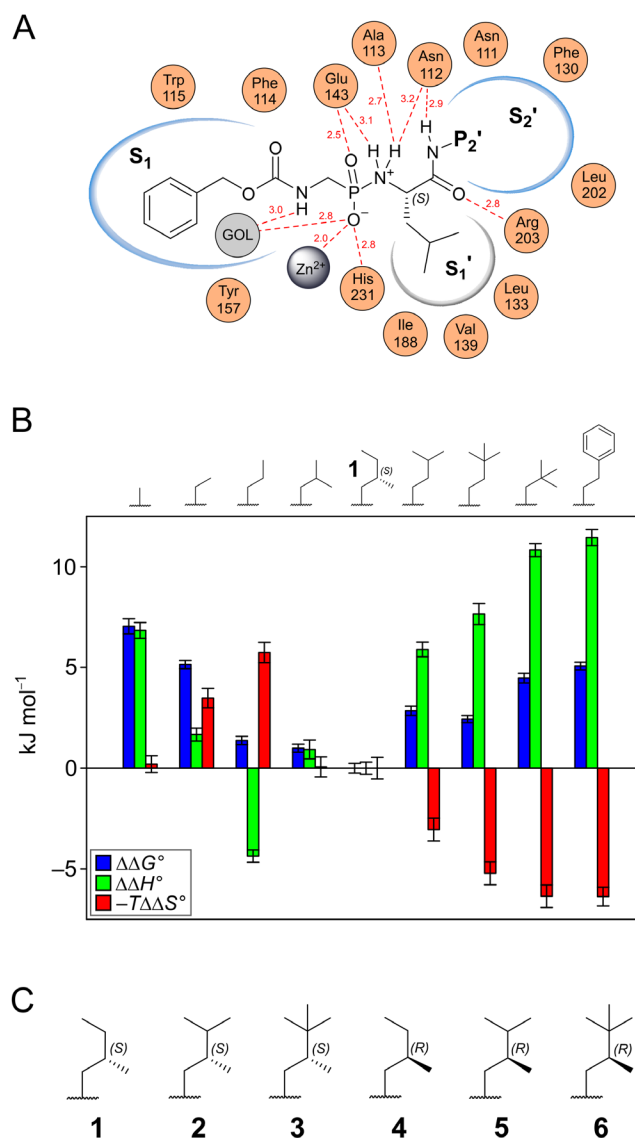
Figure 4.1. (A) Schematic binding mode and protonation state of the parent ligand scaffold in complex with TLN addressing the active site and the S_1, S_1' and S_2' subpockets. Structural components of TLN determinant for the shape of the binding pocket are indicated (GOL = glycerol). Solvent accessibility of sub-pockets is indicated by a blue contour region. H-bonds of at least moderate strength (≤3.2 Å) between parent scaffold and TLN are indicated by red dotted lines, all established to the side-chains of the respective amino acids with the exception of Ala113 (H-bond to backbone carbonyl O). The carboxyl function of Glu143 is protonated while in complex with the ligand and thus can form an H-bond to the double-bonded phosphonamidate O of the ligand [64]. **(B)** Differences in the thermodynamic profiles of the ligands from the previous study [104] were compared to **1** from the latter study. This ligand was also used as a starting point for the design of the ligand series of the current study (**2**–**6**). A positive term resembles a less favorable parameter of $\Delta H°$, $-T\Delta S°$ or $\Delta G°$ relative to ligand **1**, whereas a negative term resembles a more favorable parameter. The chemical structures of the P_2' groups of the ligands are displayed, and the parent scaffold of them is displayed in Figure 4.1A. The graph was prepared with the thermodynamic parameters as analyzed earlier [104]. Mean standard deviations were calculated by the square root of the sum of variances. **(C)** P_2' substituents of phosphonamidate-type TLN inhibitors **1**–**6**. The parent scaffold of the ligands is identical as displayed in Figure 4.1A.

enhance binding entropy, as expected for the "classical" hydrophobic effect of a well-hydrated, apolar cavity [147]. Strikingly, the inhibitor with the highest affinity (ligand **1**, Figure 4.1B) showed both, a pronounced burial of its relatively large P_2' group along with a well-established surface water network wrapping around this substituent in its protein-bound state [104]. This resulted in a well-balanced thermodynamic profile driven by favorable enthalpic and entropic contributions, overall resulting in an increase in binding affinity. Consequently, optimization of the surface water network wrapping around the partly solvent-exposed P_2' substituent appears an useful approach to enhance ligand binding, since the enthalpic gains seem to overall improve affinity.

In the present study, we want to validate this working hypothesis by systematically improving of the surface water network around a newly formed protein–ligand complex to modulate its thermodynamic binding profile and thus increase affinity of a bound ligand. Starting with the best and already fairly well-optimized ligand **1** of our previous series (Figure 4.1B), we designed five additional ligands (**2**–**6**, Figure 4.1C) based on the carboxybenzyl-Gly-(PO₂)-L-Leu-NH₂-P_2' parent scaffold (Figure 4.1A), and attached distinct apolar P_2' substituents to generate differently shaped solvent-exposed surfaces in complex with TLN. Prior to ligand synthesis, we used molecular dynamics (MD) simulations to predict the quality and completeness of the surface water network established around the newly formed complex [235]. The MD simulations suggested the highest completeness and quality for the complex with **3** and lowest for its epimer **6**. As the designed ligands seem to be promising candidates to validate our hypothesis, we synthesized all five to study the established water networks around the formed complexes by X-ray crystallography and thermodynamically by ITC. Furthermore, as we also expected an impact on the binding kinetic properties, we studied the association and dissociation rate constants k_{on} and k_{off} by surface plasmon resonance (SPR).

4.3   Results

4.3.1   Ligand design and solvation pattern prediction using MD simulations

Our design hypothesis to maximize the desolvation of increasingly bulky P_2' substituents along with an energetically favorable ("optimal") surface water network to enhance binding affinity started with the binding pose of **1**, which was already characterized in a previous study (PDB code 4MZN) [104]. This ligand showed a well-established surface water network toward the left rim of the binding pocket including an energetically favorable five-membered water polygon, interconnected by H-bonds (Figure 4.2A, left panel). Deficiencies in the network are suggested on the lower and right-hand side of the S_2' pocket (direction relative to the view of the figure). Here, the *tert*-butyl portion of the ligand with the 2,2-dimethylbutanyl P_2' group (Figure 4.2A, right panel), exhibiting two additional terminal
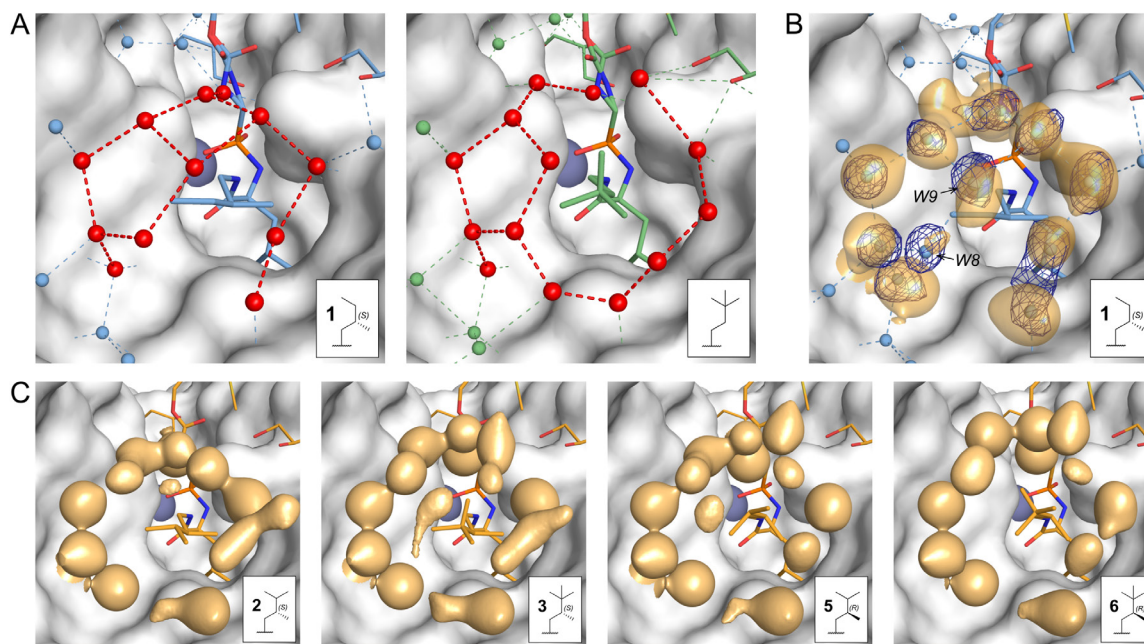
Figure 4.2. Ligand design for optimized surface water network stabilization and evaluation of solvation sites by MD simulations. The inlets of the figures represent the chemical structures of the P$_2$' substituents of the displayed ligands, their parent scaffold is shown in Figure 4.1A. In all panels, the solvent excluded surface of TLN is displayed in white. **(A)** Crystal structure of **1** (blue) and of the TLN ligand with the 2,2-dimethylbutanyl P$_2$' group (green). Water molecules are displayed as spheres, H-bond distances between them as dotted lines. Water molecules from the first solvation layer of the P$_2$' groups are highlighted in red. **(B)** Comparison between the crystal structure of TLN-**1** and the solvation-site predictions by MD simulations. The crystallographically determined binding modes of ligand and additive molecules also used in the MD simulations are displayed as blue stick models with color-coded heteroatoms. Water molecule positions determined in the crystal structure are displayed as blue spheres, and the F_o–F_c omit electron density is displayed as dark blue mesh (contour level 3σ) for the water molecules in the first solvation layer of the P$_2$' group. H-bond distances are indicated as blue dotted lines. Positions of water molecules, which are discussed in the main text, are labeled with identifiers according to Figure 4.4. The yellow, semitransparently contoured regions show computed areas in the first solvation layer of the P$_2$' groups with an occupancy probability by a water molecule of at least 48%. **(C)** Hydration sites of the designed ligands as predicted by MD simulations. The modeled coordinates for ligand and additive molecules used in the MD simulations are displayed as yellow stick models (heteroatoms color-coded). The yellow contoured regions represent areas in the first solvation layer of the P$_2$' groups with an occupancy probability by a water molecule of at least 48%.

methyl groups relative to **1**, stabilized a more complete network in this region, and also established the favorable pentagonal polygon to the left. However, its water network is unfavorably disrupted on top of the *tert*-butyl portion, resulting in an entropically highly favored system with an overall lowered affinity (Figure 4.1B). We therefore envisioned merging the features of both P$_2$' substituents into size-increased **2** (comprising only one of the two additional methyl groups of the *tert*-butyl portion) and **3** (exhibiting both additional

methyl groups of the *tert*-butyl portion). As our design involved the creation of stereogenic centers, we also considered the epimers **5** and **6** of **2** and **3** in our subsequent MD evaluation.

To predict and analyze the pattern of water solvation sites around the designed P$_2$' substituents, we applied our recently introduced MD approach [235] to simulate TLN-**1**, TLN-**2**, TLN-**3**, TLN-**5** and TLN-**6**. Ligand **4** (the epimer of **1**, Figure 4.1C) was not considered for the MD simulations, as this ligand was only synthesized at a late stage of the study with the purpose to complete the congeneric series and to further validate the influence of the rearrangement of water molecules on the thermodynamics of protein–ligand binding. This strategy provided the opportunity to validate our MD protocol on TLN-**1**, as we had already determined a high-resolution crystal structure (1.17 Å) for this complex [104]. MD simulations were run for 20 ns and the spatial positions of water molecules were recorded every 2 ps along the trajectory, from which the propensity of water molecules to occupy the indicated solvation sites was calculated. For the novel complexes, the crystal structure of TLN-**1** was used as a template. During the MD simulations, non-hydrogen atoms and non-water molecules were constrained to their starting coordinates. Similar protocols were applied to model the designed complexes with **2**, **3**, **5** and **6** (for further details, see the Experimental Section). For TLN-**1**, the computed results are superimposed with the difference electron densities of the water molecules found in the crystal structure (Figure 4.2B). The displayed solvation sites encompass a probability greater than 48% to record a water molecule along the trajectory at this site. This contour level was adjusted by visual inspection of the computed map to qualitatively match with the contouring of the crystallographically determined F_o–F_c difference electron density at the commonly applied 3σ level. The results matched convincingly well. Only the site *W8* is predicted as less populated compared to the crystal structure and *W9* was suggested as being slightly displaced by the MD approach. Mutually facing the distributions of the computed solvation sites indicates that TLN-**3** displays the most densely packed and complete surface water network in the series (Figure 4.2C). The simulation of TLN-**5** also assigned a rather elaborate surface water network to this complex, whereas TLN-**2** and in particular TLN-**6** show a gap in the water network capping their P$_2$' substituents.

4.3.2 Stereoselective synthesis of the congeneric phosphonamidate inhibitors 2–6

 Stimulated by the simulation results, we decided to synthesize **2–6**. Compound **4**, the epimer of **1**, was included for reasons of comparison. Scheme 4.1 illustrates representatively the synthesis route to prepare **2–6**. The stereogenic center in the P$_2$' portion of the ligands was synthetically accessible by a strategy employing a chiral oxazolidinone auxiliary. 4-Benzyl

Scheme 4.1. (A) Synthesis of the peptidic P$_2'$ portions of **2–6**. **(B)** Phosphonamidate coupling and deprotection of **34–38**.
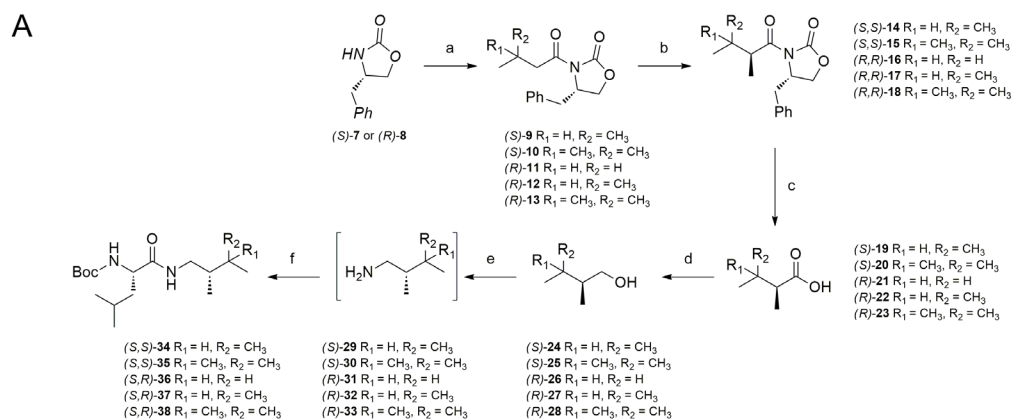
Reagents and conditions: a. n-BuLi, acid chloride, THF, -78 °C to rt, 2 h, 88-95%; b. LDA, MeI, -78 °C to rt, 5 h, 61-82%; c. LiOH, H$_2$O$_2$, H$_2$O, 0 °C to rt, 16 h, 87-100%; d. LiAlH$_4$, Et$_2$O, 0°C to rt, 1 h, 62-95%; e. DEAD, PPh$_3$, TPPA, THF, -15 °C to rt, 48 h; f. EDC, HOBT, DIPEA, Boc-Leu-OH, DCM, rt, overnight, 22-36% over two steps.
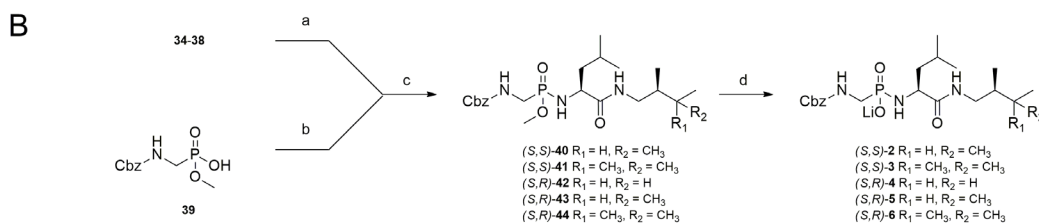
Reagents and conditions: a. HCl, Dioxane, rt, 1 h; b. SOCl$_2$, DCM, 0 °C to rt; c. DIPEA, DCM, 40 °C, overnight; d. LiOH, H$_2$O, rt, 26-37% over two steps.

oxazolidinone (**7** or **8**) was treated with *n*-BuLi followed by the respective acid chlorides. In a diastereoselective enolate alkylation of the resulting *N*-acyloxazolidinones (**9–13**), intermediates **14–18** were synthesized in a diastereomeric ratio of 13:1–16:1. With the exception of **16**, the concentration of the desired diastereomers could be improved to a ratio of >20:1 by recrystallization from cyclohexane. The auxiliary was removed by hydrolysis with LiOH/H$_2$O$_2$. The chiral carboxylic acids **19–23** were subsequently reduced to the corresponding alcohols (**24–28**) with LiAlH$_4$. The conversion to the peptidic intermediates **34–38** was carried out by a multistep procedure involving a Mitsunobu reaction with DPPA followed by an *in situ* Staudinger reduction of the resulting azides. Due to their highly volatile nature, the intermediate amines **29–33** were not isolated. After an aqueous workup procedure, they were reacted with Boc-Leu-OH under standard EDC coupling conditions.

Phosphonic acid monoester **39** was synthesized following a modified literature procedure [236]. For the phosphonamidate coupling reaction, **39** was activated with SOCl$_2$. The peptidic intermediates **34–38** were deprotected using a 4 M solution of HCl in dioxane and reacted with the activated phosphonic acid. The final deprotection of the coupling products **40–44** with aqueous LiOH solution followed by semi-preparative HPLC purification afforded inhibitors **2–6** in high purity.

4.3.3 Structure determination of TLN–2 to TLN–6 by X-ray crystallography

Crystal structures of **2**–**6** (chemical structures in Figure 4.1C) in complex with TLN were collected at high resolutions between 1.16–1.38 Å (Table 4.1). As shown in Figure 4.3, the conformations of the amino acids of their TLN binding sites are highly conserved and superimpose perfectly well, likewise the binding mode of the parent scaffold of all six ligands is virtually identical. The binding mode of this conserved part was already described extensively [64, 194, 237]. In brief, the scaffold coordinates monodentally the zinc ion with its negatively charged phosphonamidate oxygen. The carbamate group is disordered over two conformations with approximately equal occupancy. The *iso*-butyl portion of the ligand's leucine component is buried in the hydrophobic S_1' specificity cavity, a binding motif crucial for achieving high ligand affinity [194]. The S_1 pocket is occupied by a glycerol molecule from the cryobuffer, on top of which the carboxybenzyl moiety of the ligand is positioned. Also picked up from the buffer, a DMSO molecule is binding adjacent to the carbamate group of the ligand. Thus, **1**–**6** differ solely in their P_2' substituents and in the water networks adjacent to them. In the following, the positions of fifteen water molecules *W1*–*W15* (referring to water molecules found at distinct positions in the first solvation layer around the P_2' groups) are described according to the nomenclature and relative to the view angle chosen in Figure 4.4. At the right upper rim of the S_2' pocket, a second glycerol molecule is found (Figure 4.4A–F), which establishes weak H-bonds (distance >3.2 Å) to water molecules *W10* and *W11* in some of the crystal structures. We observed the two glycerol molecules in all 19 crystal structures that we determined in the previous studies [42, 64, 96, 104]. A glycerol molecule is well known to replace three water molecules as a kind of rigidified surrogate in a crystal structure. Nonetheless, to validate and exclude whether these glycerol molecules take any artificial influence on the ligand pose and adjacent water structure, we succeeded to establish alternative crystallization conditions also yielding well-diffracting crystals using PEG400 and methyl-2,4-pentanediol (MPD) instead of glycerol as cryoprotectant. We also collected diffraction data of glycerol-free crystals of TLN in complex with **3**, **5** and **6** in order to validate whether the glycerol molecule exerts any artificial influence on the structural arrangement of the observed water molecules (Figure S4.2, Supporting Information). In summary, in the crystal structures with PEG400 and MPD as cryoprotectant, the positions of the glycerol OH groups are occupied by water molecules, and the established water networks in the S_2' pocket are very similar to the below described complexes with glycerol as cryoprotectant. The minor differences concern only highly mobile water molecules (borderline cases with respect to the placement of water molecules in the refinement model) and the differences observed between the crystals exposed to different cryoprotectants fall into the same range as deviations recognized if diffraction data collected for different crystals of similar protein structures is compared. In the following, we compare

Table 4.1: Data collection and refinement statistics for crystal structures TLN-**2** to TLN-**6** with glycerol as cryoprotectant.[a]

	Complex (PDB code)				
	TLN-**2** (5JT9)	TLN-**3** (5JS3)	TLN-**4** (5JXN)	TLN-**5** (5JVI)	TLN-**6** (5JSS)
(A) Data collection and processing					
Space group	P6$_1$22	P6$_1$22	P6$_1$22	P6$_1$22	P6$_1$22
Unit cell parameters: a, b, c (Å)	92.5, 92.5, 131.2	92.6, 92.6, 131.5	93.0, 93.0, 131.2	92.6, 92.6, 131.2	92.5, 92.5, 131.3
Matthews coefficient (Å^3/Da)[b]	2.4	2.4	2.4	2.4	2.4
Solvent content (%)[b]	48	48	48	48	48
(B) Diffraction data					
Resolution range (Å)	50–1.26 (1.34–1.26)	50–1.16 (1.23–1.16)	50–1.38 (1.46–1.38)	50–1.12 (1.19–1.12)	50–1.19 (1.26–1.19)
Unique reflections	89582 (14202)	114612 (18030)	69138 (10928)	127067 (20268)	106247 (16884)
$R(I)$sym (%)	8.7 (49.4)	6.6 (44.7)	5.3 (49.6)	7.4 (49.6)	9.5 (45.3)
Wilson B factor (Å^2)	8.4	8.6	11.9	8.3	8.3
Completeness (%)	99.9 (99.3)	99.5 (98.3)	99.9 (99.2)	100.0 (99.8)	100.0 (99.7)
Redundancy	15.0 (14.9)	16.1 (15.4)	9.7 (9.7)	19.4 (18.8)	15.1 (15.0)
$<I/\sigma(I)>$	21.8 (5.5)	26.5 (5.8)	26.9 (4.3)	26.1 (6.0)	18.2 (5.2)
(C) Refinement					
Resolution range (Å)	34.20–1.26	40.10–1.16	46.51–1.38	30.36–1.12	40.1–1.19
Reflections used in refinement (work/free)	85102/4480	108881/5731	65681/3457	120713/6354	100934/5313
Final R value for all reflections (work/free) (%)	10.3/12.1	10.3/12.4	11.0/13.5	10.2/11.9	10.3/12.1
Protein residues	316	316	316	316	316
Calcium/zinc ions	4/1	4/1	4/1	4/1	4/1
Inhibitor atoms	30	31	29	30	31
Water molecules	424	428	415	435	424
RMSD from ideality:					
Bond lengths (Å)	0.009	0.01	0.011	0.011	0.011
Bond angles (°)	1.1	1.2	1.2	1.2	1.2
Ramachandran plot:[c]					
Residues in most favored regions (%)	88.1	88.9	88.1	87.8	88.1
Residues in additionally allowed regions (%)	10.7	10	10.7	11.1	10.7
Residues in generously allowed regions (%)	0.7	0.7	0.7	0.7	0.7
Residues in disallowed regions (%)[d]	0.4	0.4	0.4	0.4	0.4
Mean B factors (Å^2):[e]					
Protein non-hydrogen atoms	9.7	9.9	13.0	9.6	9.6
Protein Cα atoms	8.6	8.8	11.7	8.5	8.5
Inhibitor	9.1	9.6	16.1	9.8	10.1
Water molecules	25.8	26.0	30.0	25.5	25.2

[a] The crystal structure of TLN-**1** (PDB code 4MZN, resolution 1.17 Å) was already reported earlier [104]. Statistics of the glycerol-free crystal structures with PEG400 or MPD as cryoprotectant are listed in the Table S4.1 (Supporting Information). Numbers in parentheses represent the values of the highest resolution shells. [b] Matthews coefficients and solvent contents were calculated with the program *Matthews_coef* from the *CCP4* suite [238]. [c] Ramachandran plots were calculated with *PROCHECK* [196]. [d] The Ramachandran outlier is Thr26 as described in literature [197]. [e] Mean B factors were calculated with *MOLEMAN* [198].

Chapter 4
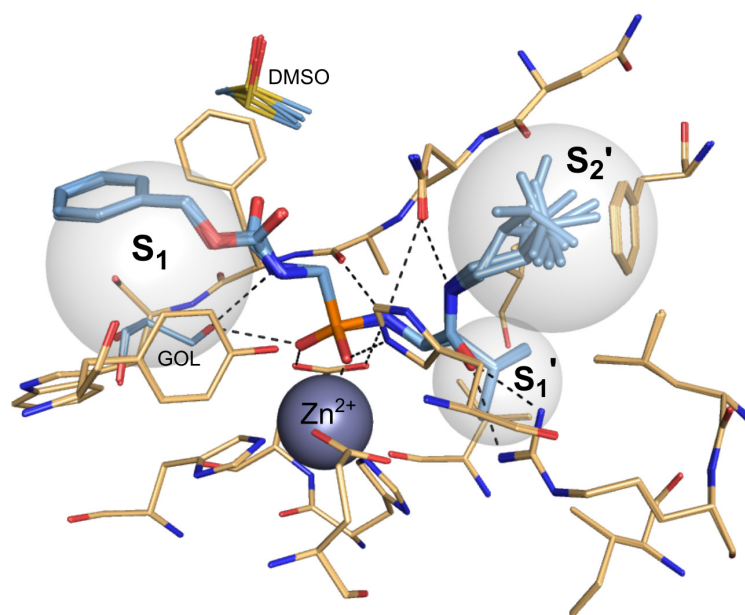
Figure 4.3. Superimposed crystal structures of the TLN–ligand complexes TLN-**1** to TLN-**6**. From all six crystal structures, the ligands, glycerol (GOL) and DMSO molecules are displayed with carbon atoms in light blue (color-coded hetero-atoms), TLN amino acids in beige, and the zinc ion as dark blue sphere. The S_1, S_1' and S_2' subpockets are indicated by semitransparent spheres. Polar interactions of at least moderate strength (≤3.2 Å) are indicated by black dotted lines. All crystal structures were superimposed on TLN-**1** by the alignment of amino acids within 5 Å of **1** (159 heavy atoms). The average RMSD is 0.05±0.02 Å as calculated by *fconv* [239]. As a result of the high rigidity of TLN, structural components of the binding site superimpose almost perfectly.

the crystal structures in terms of observed electron densities in the S_2' pockets (F_o–F_c omit electron densities in Figure 4.4A–F) and refined B factors of the water molecules (B factors of all water molecules from the first solvation layer of TLN-**1** to TLN-**6** displayed as a heatmap in Figure 4.5) and we avoid to discuss only the presence or absence of a water molecule in the refinement model.

4.3.4   Arrangement of water molecules in the S_2' pocket of TLN-1, TLN-2 and TLN-3

The structure of TLN-**1** was published previously [104] and served as starting point for the design of our new congeneric ligand series. Hence, the water networks observed in TLN-**2** to TLN-**6** are modifications of that in TLN-**1**. In TLN-**1**, the ligand places its 2(*S*)-methylbutyl P_2' group into the S_2' pocket (Figure 4.4A). The right cleft of the pocket is addressed by the terminal methyl group of the P_2' portion, whereas the terminal ethyl group is oriented toward the left. In TLN-**1**, *W5–W9* form a five-membered polygon with H-bond distances between 2.8–3.1 Å exhibiting low B factors (Figure 4.5). In total, twelve water molecules are
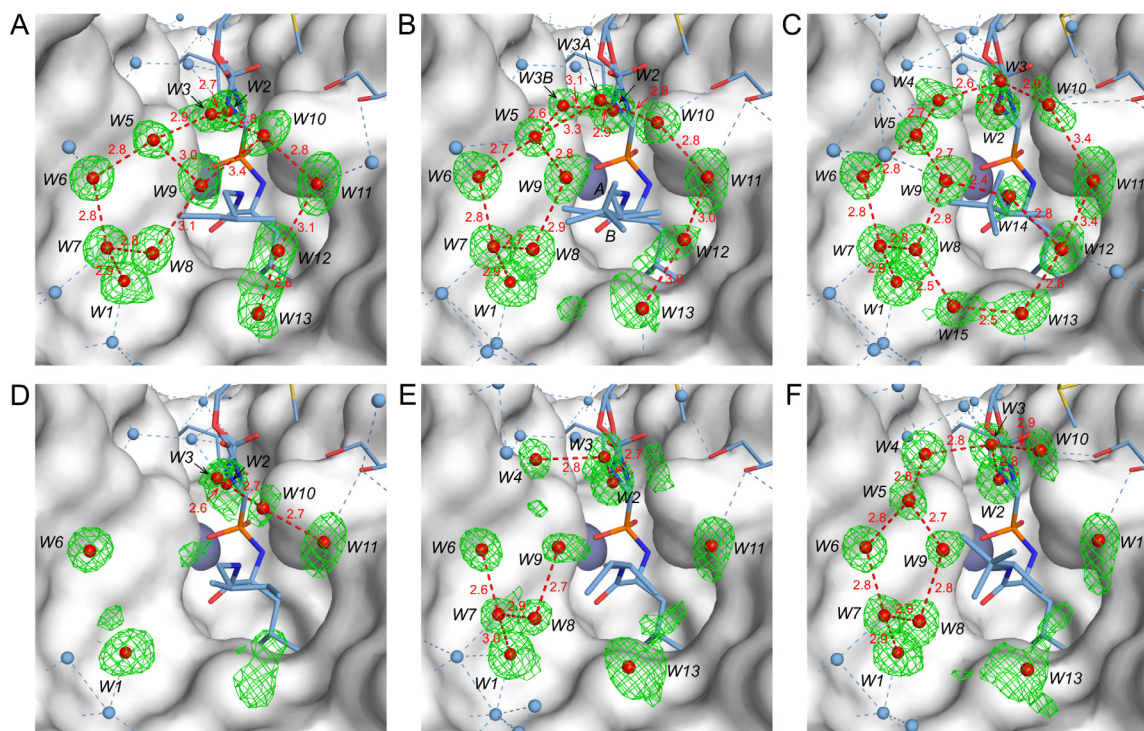
Figure 4.4. Crystal structures of TLN–ligand complexes **(A)** TLN-**1**, **(B)** TLN-**2**, **(C)** TLN-**3**, **(D)** TLN-**4**, **(E)** TLN-**5** and **(F)** TLN-**6**. In each panel, the solvent excluded surface of the respective crystal structure is displayed in white and the bound inhibitor is displayed as stick model (C blue, heteroatoms color-coded). The zinc ion of TLN is indicated as dark blue sphere partly buried by the surface. Water molecules from the first solvation layer of the ligands' P_2' groups are displayed as red spheres and labeled with an identifier (*W1–W15*) which is referred to in the main text; H-bonding distances between them are depicted as red dotted lines (maximum depicted distance: 3.4 Å, without hydrogens). The F_o–F_c omit electron densities of water molecules (green mesh) are displayed at a contour level of 3σ. For clarity reason, water molecules located outside of the first solvation layer of the P_2' groups are displayed as blue spheres with H-bonding distances indicated as blue dotted lines, and their electron densities are omitted. The two conformations of the P_2' group of **2** in complex with TLN (panel B) are labeled with *A* and *B*. F_o–F_c omit electron densities of the TLN-bound ligands are displayed in Figure S4.1 (Supporting Information).

Chapter 4
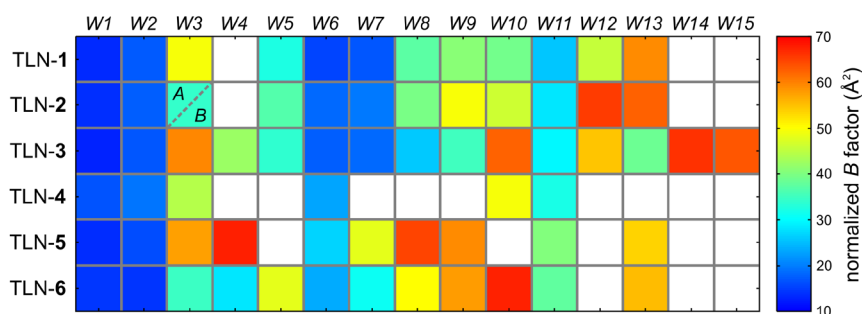
Figure 4.5. Heatmap of normalized *B* factors of water molecules *W1–W15* found in complexes TLN-**1** to TLN-**6**. Chemical structures of **1**–**6** are displayed in Figure 4.1C, water molecule numbering is according to Figure 4.4A–F. The normalized *B* factors are color-coded by a blue-to-red gradient. Individual *B* factor values of water molecules are omitted, since only distinct relative differences between *B* factors are discussed. A white field indicates no sufficient stabilization of the water molecule in the crystal structure for placement in the refinement model. *W3* in TLN-**2** is found in two conformations *A* and *B*. To compensate for differences between the *B* factor scales of different crystal structures as a result of varying crystal quality, *B* factors were normalized on the dataset with the lowest resolution (TLN-**4**, Table 4.1) and calculated as $B_{normalized} = B_{observed} \div B_{average(C\alpha)} \times B_{average(C\alpha)TLN-\mathbf{4}}$, where $B_{observed}$ is the *B* factor of a water molecule as observed in the crystal structure, $B_{average(C\alpha)}$ is the average *B* factor of the Cα atoms of the respective crystal structure and $B_{average(C\alpha)TLN-\mathbf{4}}$ is the average *B* factor of the Cα protein atoms in the crystal structure of TLN-**4**. No significant non-linearity is expected for the *B* factors of the datasets due to the narrow resolution range. Values for experimentally observed and normalized *B* factors are listed in Table S4.2 and Table S4.3 (Supporting Information).

detected in the first solvation layer around the P₂' substituent covering its large, solvent-exposed hydrophobic surface patch (Figure 4.6). Several of the water molecules are anchored to the protein surface via polar interactions: *W1* (Arg203 primary nitrogen, 3.0 Å; His231 carbonyl oxygen, 2.9 Å), *W2* (ligand **1** carboxybenzyl carbonyl O, 2.9 Å; Asn112 amide nitrogen, 3.0 Å), *W6* (Asp226 carboxy oxygen, 2.8 Å), *W11* (Asn111 backbone carbonyl oxygen, 2.8 Å) and *W13* (Tyr193, 3.1 Å). In contrast, *W3, W5, W7, W8, W9, W10* and *W12* are only stabilized by H-bonds to other water molecules or by van der Waals interactions with the apolar surface patch of the P₂' substituent of **1**.

In TLN-**2**, the 2(*S*),3-dimethylbutyl group of **2** orients its P₂' group similarly to **1**, and the additional methyl group is disordered over two positions (Figure 4.4B): conformation *A* (56% occupancy) is sticking out into the solvent (no contacts within hydrophobic interaction distance of ≤4.6 Å), whereas conformation *B* (44% occupancy) is oriented downward, alongside the protein surface. The observed disorder of the P₂' substituent is a result of the shallow, widely open S₂' pocket. Due to the steric requirement of the methyl group in conformation *A*, two distinct, mutually excluding sites are observed for *W3* (occupancies constrained to 50/50 in the refinement model). As a consequence of the reduced occupancy
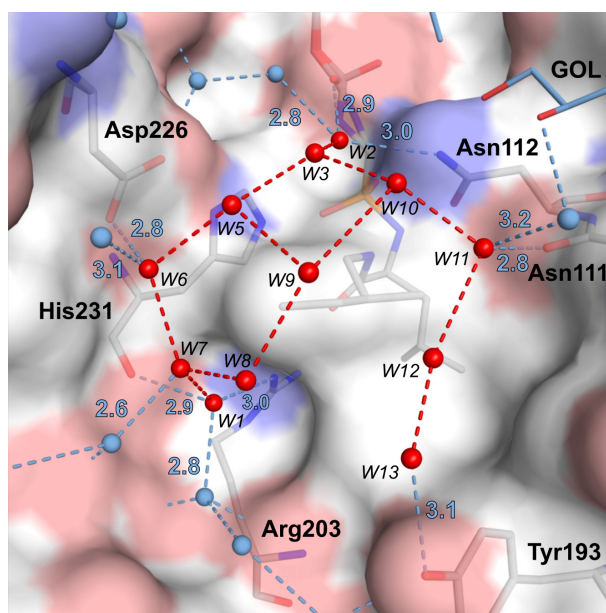
Figure 4.6. Topography of the S_2' pocket of TLN-**1**. The solvent excluded surface of the TLN–ligand complex is shown (color-coded, surface indicated for C in white, O in red, N in blue). Water molecules in the first solvation layer of the P_2' group are displayed as red spheres and labeled with identifiers according to Figure 4.4. Water molecules further remote than the first solvation layer and the glycerol molecule are displayed in blue. The position of **1** and TLN amino acids establishing H-bonds to water molecules of the first solvation layer are indicated by stick models below the surface. H-bonds between the water molecules of the first solvation layer are depicted as red dotted lines, H-bonds established to other water molecules, TLN amino acids or **1** are depicted as blue dotted lines (distances labeled in Å, maximum depicted distance: 3.4 Å).

and owing to a strong correlation of B factors with occupancy [240], the refined B factor for this water molecule (Figure 4.5) has to be regarded with care and will hardly reflect its actual mobility. Furthermore, the distance between $W9$ and $W10$ increases from 3.4 Å in TLN-**1** to 4.5 Å in TLN-**2**, clearly exceeding the maximum distance for an energetically favorable H-bond [201, 241]. The constraint which modifies the water structure between TLN-**1** and TLN-**2** is best visualized by use of the solvent accessible surface areas (SASAs; solvent excluded surface area plus radius of a water molecule) as displayed for the P_2' portions in Figure 4.7: $W3$, $W9$, $W10$, and $W12$ in TLN-**1** would penetrate inside the SASA of TLN-**2**, thus these water molecules in TLN-**2** must be shifted. Because of the expanded water structure, $W12$ and $W13$ become increasingly destabilized (Figure 4.5).

In TLN-**3**, the ligand exhibits a 2(*S*)-3,3-trimethylbutyl substituent (Figure 4.4C), representing the bulkiest P_2' portion in the (*S*)-configurated series. The binding mode of **3** closely resembles that of **2** if both conformations *A* and *B* of TLN-**2** would be merged. Two additional sites for water molecules ($W14$ and $W15$) are refined, resulting in the formation of a six-membered polygon and well-defined electron density (Figure 4.4C) with strongly decreased B factor (Figure 4.5) of the incorporated $W13$. Furthermore, water molecules $W8$
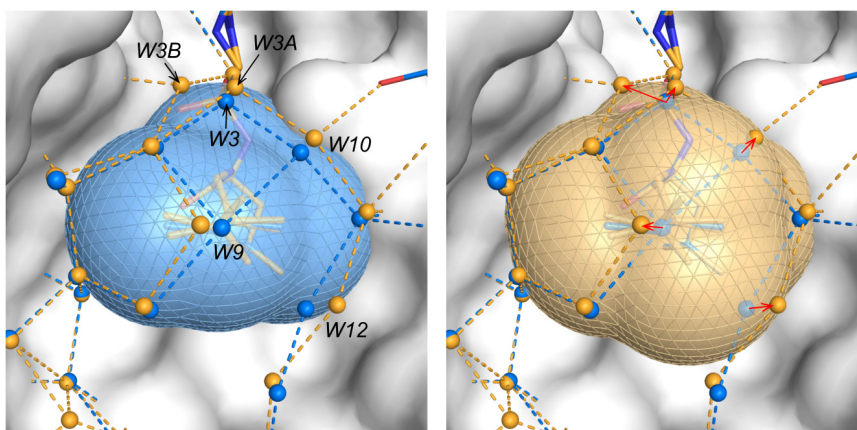
Figure 4.7. Differences in shape and size of the solvent accessible surfaces of TLN-**1** and TLN-**2**. In both images, the crystal structures TLN-**1** (blue) and TLN-**2** (orange) are superimposed. The semitransparent solvent accessible surface of the P_2' group of **1** is displayed in the left panel and of **2** in the right panel. In the right panel, shifts of water molecules observed between crystal structures TLN-**1** and TLN-**2** are indicated by red arrows. For TLN-**2**, both conformations *A* and *B* of the P_2' group of **2** are considered for the displayed solvent accessible surface. The solvent excluded surface of TLN is displayed in white.

and *W9*, both participating in the five and six-membered polygonal water networks in TLN-**3**, are here better stabilized than in TLN-**2** (Figure 4.5). In TLN-**3**, *W10* is shifted distal from the P_2' substituent and is indicated by less electron density along with a higher *B* factor compared to TLN-**2**.

4.3.5 Arrangement of water molecules in the S_2' pocket of TLN-4, TLN-5, and TLN-6

Ligands **4**, **5** and **6** are the epimers of **1**, **2** and **3**. The overall quality of the crystal structure TLN-**4** is slightly lower compared to that of the other five complexes (resolution, R_{work}/R_{free} values, Wilson *B* factor; see Table 4.1). Nevertheless, only about ten water molecules less are observed in TLN-**4** at a total amount of >400. In TLN-**4** (Figure 4.4D), the P_2' portion is flipped over by 180° compared to TLN-**1**: The terminal P_2' ethyl group is oriented toward the right rim, whereas the P_2' methyl group is directed to the left of the S_2' pocket. Only six water molecules are detected in the crystal structure adjacent to the P_2' group mainly stabilized via H-bonds to protein residues.

In TLN-**5** (Figure 4.4E), in contrast to **4**, the P_2' substituent of **5** adopts a conformation analogous to that of the (*S*)-configurated P_2' substituents of **1** to **3**. One methyl group from the terminal *iso*-propyl portion of **5** is not defined in the electron density most likely owing to enhanced mobility. It was therefore not modelled in the structure. The absence of this methyl group is also observed in TLN-**5** with MPD as cryoprotectant (Figure S4.3B, Supporting

Information). The missing of water molecules *W5*, *W10* and *W12* results in an incomplete water network in TLN-**5** (compared to that of the epimeric TLN-bound **2**). In the crystal structure TLN-**5** with MPD as cryoprotectant, *W5* is highly mobile but sufficient electron density is detected to allow placement of a water molecule in the refinement model (Figure S4.2B, Supporting Information). Furthermore, in the structure of TLN-**5** with glycerol as cryoprotectant, some F_o–F_c difference electron density is observed at the positions of *W5* and *W10* (Figure 4.4E), which, however, is too weak to justify placement of a fully populated water molecule in the refinement model of TLN-**5**. In TLN-**6** (Figure 4.4F), *W5* and *W10* were added to the refinement model, but they refined as highly mobile (Figure 4.5). Furthermore, *W10* is missing in the crystal structure of TLN-**6** with MPD as cryoprotectant (Figure S4.2C, Supporting Information). Consequently, the water molecules *W5* and *W10* are highly mobile in TLN-**5** and TLN-**6**, and the local concentration of their electron density is at the borderline for water placement in the refinement model. Thus, the water networks of TLN-**5** and TLN-**6** are highly similar. Overall, in the (*R*)-series a lower amount of water molecules with increased residual mobility (especially of *W5*–*W10*, Figure 4.5) is observed compared to the (*S*)-series.

4.3.6 Thermodynamic signatures of TLN-ligand complex formation measured by ITC

As we recently documented, ITC measurements comparable on the same scale and with minimal error margins can only be obtained if all ligands are studied with the same optimized measurement protocol using the same protein batch [96]. We therefore reevaluated **1** along with **2**–**6** in the present study (for further details see Chapter 4.11.4, Supporting Information). Across the (*S*)-configured series (**1**→**2**→**3**), the binding affinity $\Delta G°$ improved with growing number of methyl groups (Figure 4.8). Remarkably, this effect is determined by an increasingly favorable $–T\Delta S°$, which is only partly compensated by a loss in $\Delta H°$ (slope of $–T\Delta S°$ is steeper than of $\Delta H°$), leaving overall a gain in $\Delta G°$. Interestingly, for the (*R*)-configured series (**4**→**5**→**6**), no affinity enhancement is detected. The mutual compensation of $\Delta H°$ and $–T\Delta S°$ fully nullifies any affinity improvement as the $\Delta H°$ compensation is stronger (slopes of $–T\Delta S°$ and $\Delta H°$ are equal with opposite sign) compared to the (*S*)-series. Accordingly, in the (*S*)-series, a small but significant advantage in $\Delta H°$ is experienced parallel to the growing of the P_2' substituent into the S_2' pocket relative to the (*R*)-series.

To further validate whether the glycerol molecules found in our cryoprotected crystal structures have any distorting effect on the thermodynamic signature, we performed ITC titrations with the addition of different concentrations of glycerol (up to 10%, Figure S4.6 in
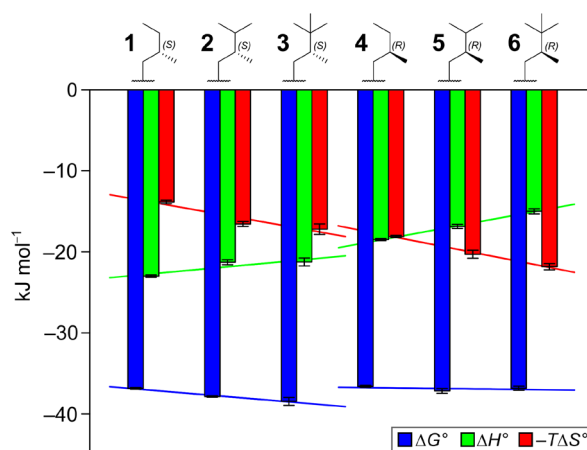
Figure 4.8. Thermodynamic parameters of **1**–**6** upon binding to TLN measured by ITC. The basic scaffold of the displayed P_2' groups is displayed in Figure 4.1A. The columns and the error bars represent the mean values calculated out of three measurements and their standard deviations, respectively. Trendlines of **1**→**3** (*S*) and **4**→**6** (*R*) for $\Delta H°$ (green), $-T\Delta S°$ (red) and $\Delta G°$ (blue) were calculated with *SigmaPlot*. Data values are listed in Table S4.4 (Supporting Information).

the Supporting Information). These titrations revealed a systematical increase of $\Delta H°$ with increasing glycerol concentration paralleled by a compensating decrease of $-T\Delta S°$. Most importantly, the relative difference between $\Delta H°$ and $-T\Delta S°$ remained unchanged, thus no specific effect and only an overall systematic influence of the added glycerol was observed. Comparable systematic influences, for example by the type of salt (NaCl or NaSCN) and its concentration used in the measurement buffer, have been described previously [96]. Similarly, systematic influences on $\Delta H°$ and $-T\Delta S°$ were also observed for the measurement with the addition of different concentrations of DMSO (Figure S4.7, Supporting Information). These findings underscore that ITC data should only be recorded applying highly comparable measurement conditions and evaluated relative to each other in congeneric compound series [96].

4.3.7   Binding kinetics of TLN-ligand complex formation measured by SPR

The binding kinetic parameters of **1**–**3** and **5**–**6** (Figure 4.9) were determined by single-cycle SPR measurements preformed in triplicate for each ligand. Kinetic analysis of the SPR sensorgrams was performed by global analysis of the triplicate data applying a 1:1 binding model [242, 243], which agreed well with individual analysis of the sensorgrams. The relative standard deviations of the individual analysis results were about 30% for k_{on} and 10% for k_{off} (Table S4.8, Supporting Information). Ligand **4** was not tested, as this ligand was synthesized at a later stage of the study. As depicted on the kinetic map (Figure 4.9), **1** showed the fastest k_{off} within the series, whereas the dissociation rates for **2**, **5** and **6** are slower and fall within a

narrow window. Ligand **3** instead shows a significantly prolonged dissociation rate compared to all other members of the series.
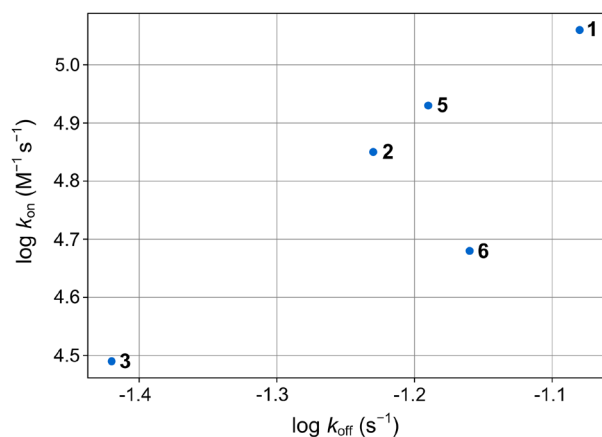
Figure 4.9. Kinetic map (log k_{on} vs. log k_{off}) of **1**–**3**, **5** and **6** as determined by global analysis of the single-cycle SPR measurements performed in triplicate for each ligand. Measurement results and SPR sensorgrams are shown in Table S4.7 and Figure S4.8 (Supporting Information). The results and standard deviations of the kinetic parameters as determined by individual analysis of the SPR sensorgrams are listed in Table S4.8 (Supporting Information).

4.3.8   Buried SASAs of the TLN-ligand complexes

Figure 4.10 shows the computed buried solvent accessible surface areas (SASAs) of **1**–**6** in complex with TLN. One methyl group of the P_2' portion of **5** remained undetected in the
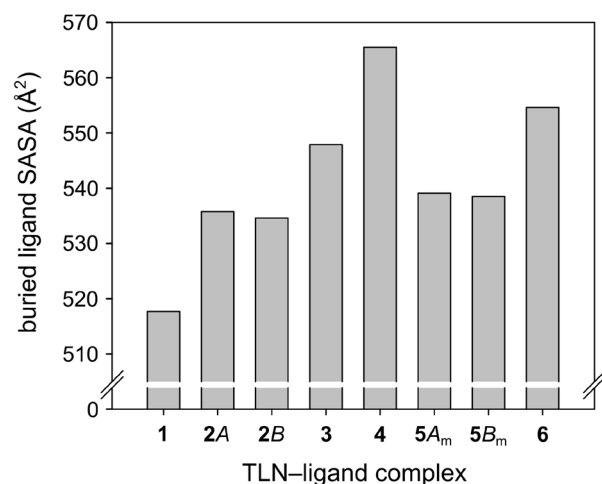
Figure 4.10. Buried solvent accessible surface areas (SASAs) of **1**–**6** while in complex with TLN. The P_2' group of **2** adopts two conformations in the crystal structure (*A* and *B*, Figure 4.4). One methyl group of the P_2' portion of **5** is not visible in the electron density and was thus modeled in two conformations **5A_m** and **5B_m** (Figure S4.9, Supporting Information). Calculated data values are listed in Table S4.9 (Supporting Information).

electron density (Figure 4.4E). Therefore, the missing methyl group was modeled in two conformations A_m and B_m based on the crystal structure (Figure S4.9, Supporting Information) and the buried SASAs of these two conformations were calculated. The SASA buried within the (S)-series increases monotonously from **1→2→3** by approximately 15 Å^2 per added methyl group. TLN-**4** exhibits the largest buried SASA of all six ligands and the buried SASAs of **5** and **6** are slightly larger than those of their respective epimers **2** and **3**.

4.4 Discussion

In the current study, we wanted to validate our hypothesis that ligand binding to an open, rather flat and solvent-exposed binding pocket can be enhanced by optimizing the surface water network wrapping around exposed parts of the bound ligand. We started with the previously characterized peptidomimetic TLN inhibitor **1** [104], and modified its hydrophobic solvent-exposed P$_2$' substituent attached to the parent scaffold. We improved binding affinity by maximizing the desolvation of the increasingly bulky P$_2$' substituents along with an enhancement of the water network wetting the surface of the formed complex. Moreover, we expected that the residence time of the complex could be expanded with increasing quality and perfection of the formed water network. Prior to synthesis, we predicted the putative water pattern around the designed ligands by MD simulations.

In our previous purely descriptive studies, we observed that small changes of the solvent-exposed ligand surface can strongly modulate the stability and complexity of the formed water network adjacent to the bound ligands [42, 64, 104]. Since the spatial arrangement of water molecules across apolar surfaces is governed by a complex architecture and the addition of a sole methyl group can already lead to the unfavorable disruption of the adjacent water network, optimal hydration of a partly solvent-exposed apolar P$_2$' substituent is a challenging design task.

As starting point we chose **1**, which already displayed rather potent inhibition properties but showed local deficiencies in the solvation pattern next to its P$_2$' substituent. This was indicated by a comparison with structures of closely related complexes. By merging features of their P$_2$' substituents with those of **1**, we designed a small series of ligands comprising chiral aliphatic P$_2$' substituents. To predict their impact on the quality of the wetting surface water network, we followed our recently introduced MD approach [235]. To further validate our hypothesis, we considered both chiral orientations, as the simulations suggested significant differences between the stereoisomers. Since the designed ligands exhibited a second stereogenic center at P$_1$', epimeric pairs of ligands resulted. This leads to the disadvantage that differences in the desolvation cost to transfer the corresponding ligands of the epimeric pairs from the bulk solvent phase to the protein pocket cannot be entirely

excluded. However, as the two stereogenic centers are separated by several bonds, we assume quite similar physicochemical properties for the matching epimeric pairs. Across the series, the MD analysis suggested small but significant differences in the completeness of the surface water network formed next to the different P$_2$' substituents rendering the (S)-configurated **3** as most promising candidate of the series. Subsequently, we synthesized the ligands stereoselectively and characterized the complexes formed via crystallography, microcalorimetry and surface plasmon resonance. Concerning the crystal structure analysis, in all cases we obtained diffraction data with very high resolution, also falling into a narrow window (mean: 1.22±0.10 Å, Table 4.1). This is important to reliably compare details of the water structures between the different crystal structures, as deviating resolution can affect such details and will complicate the mutual matching of *B* factors [199].

The parent scaffold of all six ligands adopts virtually the same binding pose (Figure 4.3). Thus, the observed differences between the studied ligands **1–6** most likely originate predominantly from the desolvation differences of the gradually increasing and partly buried P$_2$' substituents and from deviations of the formed surface water networks "wetting" the newly formed complexes. They show varying degrees of completeness and perfection, which in turn, influence the thermodynamic and binding kinetic signature of complex formation. Whereas the fixation of water molecules on the surface of the bound ligand increases binding enthalpy and reduces entropy, in contrast binding entropy is favored and enthalpy lowered by enhanced mobility up to the displacement of water molecules into the bulk water phase. The mobility and occupancy of individual water molecules is indicated by the spatial concentration of the electron densities (Figure 4.4) and the assigned *B* factors (Figure 4.5). A strong fixation of a water molecule results from the embedding into a geometrically rather constrained H-bonding network, also involving the formation of H-bonds to adjacent functional groups of the protein's amino acids.

Although essential in the current series, the arrangement of water molecules is possibly only one component determining the thermodynamic signature of the binding process. The total thermodynamic signature is in any case the sum of many contributions and might be composed of partly compensating or mutually enhancing contributions. Hence, it is even more important that we attempt to only evaluate relative differences of the studied complexes and not their absolute values. Furthermore, ligand binding can be accompanied by global conformational adjustments of the protein partly masking the thermodynamic signature of the local binding event. However, from our experience with the system, global adjustments of TLN are unlikely. The enzyme has proven to be highly rigid and the sole differences introduced between the congeneric ligands are their solvent-exposed P$_2$' groups. Fenley et al. recently evaluated a large-scale MD trajectory of BPTI (bovine pancreatic trypsin inhibitor) and observed remarkable transitions between states of unchanged overall Gibbs free energy but significantly altered enthalpy/entropy inventory [244]. This entropy–enthalpy

transduction might suggest a physical mechanism underlying entropy–enthalpy compensation in such systems. However, we propose that in our congeneric ligand series, where binding occurs to a rigid protein, the ligands always address the same or very similar configurations of TLN.

4.4.1 Prediction of solvation sites by MD simulations and their agreement with crystallographically determined solvation sites

Based on our earlier study on four TLN ligands [42], we developed an MD simulation protocol to predict water networks adjacent to solvent-exposed ligand groups [235]. This protocol correctly reproduced the rupture of a water network between two ligands differing by only one single methyl group [42]. The rupture was responded by a dramatic loss in binding enthalpy ($\Delta\Delta H^{\circ}_{\text{methyl}\rightarrow H}$ = +13.3±0.6 kJ mol^{-1}) and an increase in binding entropy ($-T\Delta\Delta S^{\circ}_{\text{methyl}\rightarrow H}$ = −7.7±0.4 kJ mol^{-1}), overall resulting in a lowered affinity ($\Delta\Delta G^{\circ}_{\text{methyl}\rightarrow H}$ = +5.7±0.3 kJ mol^{-1}). This example underlines the pronounced effect of the ligand-capping water network on the thermodynamic signature.

Importantly, our MD simulation approach does not require any a priori knowledge about water positions but allows the prediction of hydration sites adjacent to protein–ligand complexes from scratch in agreement with experiment. In the present study, the tool was applied to predict TLN-**2,** TLN-**3**, TLN-**5** and TLN-**6**. Its predictive power could be further assessed by simulating TLN-**1**, as its crystal structure had been determined prior to the present study [104]. For TLN-**1**, the computed hydration sites match remarkably well with the difference electron density assigned to water molecules by crystallography (Figure 4.2B). Only the population of *W8* is underestimated, which is in line with our previous observation that water-to-methyl interactions are predicted as too weak by the *AMBER* force field [235].

To estimate whether the designed P$_2$' substituents exhibit higher or lower hydration propensities than those of **1**, the predicted solvation sites were mutually compared. The simulations of TLN-**1** and TLN-**3** suggest some advantages of the latter adjacent to the P$_2$' substituents (Figure 4.2B,C). This should render **3** superior to **1** with respect to affinity, since the enthalpic component of stabilizing the water network is larger and can compensate for the increasing enthalpic cost to (entropical beneficially) desolvate **3** over **1**, as its substituent comprises two additional methyl groups, resulting in a significantly larger buried surface area (Figure 4.10). For TLN-**6**, a possible rupture of the water network in the center of the apolar surface next to the P$_2$' substituent is suggested. Facing the subsequently determined crystal structures with our predictions, TLN-**3** (Figure 4.11A) exhibits the water molecules capping the apolar P$_2$' substituent which are likewise predicted too weak (as *W8* in TLN-**1**). The in the crystal structure well-stabilized *W8* and *W9* are weakly indicated by the MD approach,
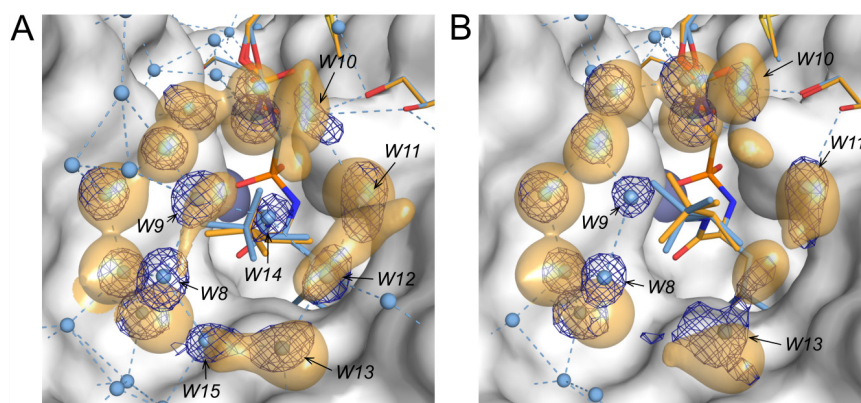
Figure 4.11. Prediction of water solvation sites as calculated by MD simulations in comparison to the crystallographically observed electron densities and refined crystal structure models of **(A)** TLN-**3** and **(B)** TLN-**6**. The modeled coordinates of ligand, glycerol and DMSO molecules used in the MD simulations are displayed as yellow stick models with color-coded heteroatoms. The yellow, semitransparently contoured regions show computed areas in the first solvation layer of the P_2' substituents with an occupancy probability by a water molecule of at least 48%. The crystallographically determined binding modes of ligand and additive molecules are superimposed as blue stick models with color-coded heteroatoms. Water molecule positions determined in the crystal structures are displayed as blue spheres, and the F_o–F_c omit electron density is displayed as dark blue mesh at a contour level of 3σ for the water molecules positioned in the first solvation layer of the P_2' groups. H-bond distances are indicated as blue dotted lines. Positions of water molecules, which are discussed in the main text, are labeled with identifiers according to Figure 4.4. The solvent excluded surface of TLN is shown in white.

and the relatively mobile *W14* (high *B* factor and weak electron density) is not predicted by the computer analysis, at least on the 48% threshold level. *W15* is correctly predicted by a tubular solvation site, which also hosts *W13*. In TLN-**6**, *W8* and *W9* are not predicted by MD, which, however, agrees with the experimental observation that these water molecules are significantly less stabilized in TLN-**6** than in TLN-**1** or TLN-**3** (cf. high *B* factors, Figure 4.5). The shifted position of *W10* was correctly predicted, resulting in a large gap toward *W11* beyond H-bonding distance. Water molecule *W12* is not observed in the electron density, whereas the MD simulation predicts a hydration site at this position. In summary, TLN-**1**, TLN-**3** and TLN-**6** are convincingly predicted on qualitative level and the relative ranking of the epimer complexes TLN-**3** (strong fixation) and TLN-**6** (weak fixation) was correctly assigned.

The predicted solvation sites in TLN-**2** and TLN-**5** differ more strongly from the crystallographically observed electron densities. This results from the disorder of the P_2' substituents of **2** and **5** indicated in the crystal structures, which was difficult to predict as the disorder was not considered in the MD simulations (Figure S4.10, Supporting Information).

4.4.2 Correlation of structural data with thermodynamic signature of complex formation

Our starting ligand **1** has been established as most potent binder from the previously studied series (Figure 4.1B) [104]. We related its superior affinity to an entropically beneficial burial of its surface along with the establishment of an extensive enthalpically favored surface water network. In TLN-**1**, a particularly favorable five-membered water polygon is formed (Figure 4.4A, *W5–W9*), which is well-known for its favorable energetic contribution [245, 246].

Compared to TLN-**1**, the additional methyl group in TLN-**2** creates disorder over two conformational states, perturbing the neighboring water structure (Figure 4.7) and increasing the mobility of the adjacent water molecules: *W3* is observed in two orientations, and *W9*, *W10*, *W12* and *W13* show enhanced *B* factors (Figure 4.5). Consequently, enthalpy decreases from TLN-**1** to TLN-**2** (Figure 4.8). This is overcompensated by a more favorable entropy resulting from the burial of a larger apolar surface area (Figure 4.10) and the enhanced sidechain mobility, which is entropically beneficial. Altogether, the more favorable entropy contribution is only partly compensated by enthalpic losses, resulting overall in a slightly enhanced affinity of **2** over **1** by $\Delta\Delta G°_{1\to2} = -1.0\pm0.1$ kJ mol^{-1}.

The thermodynamic signature of **3** is comparable to that of **2**, with a slightly elevated entropic and virtually unchanged enthalpic term leading to an increased affinity of **3** over **2** (Figure 4.8). The surface water networks of both complexes differ in several regards. In TLN-**3** (Figure 4.4C), a complete six-membered water network polygon (*W8*, *W9*, *W12–W15*) is established adjacent to the five-membered one and is integrated in an eight-membered ring structure (*W3–W5*, *W9–W12*, *W14*). Such fused polygonal water arrangements can be considered as optimal solvation shell to coat the surface of a formed protein–ligand complex in terms of H-bonding and thus inherent enthalpy contributions. Overall, TLN-**3** shows the most perfect water network along with an increased apolar surface burial compared to TLN-**2**, resulting in a slightly superior affinity. This leads to an unchanged enthalpic contribution as desolvation costs are compensated by the formed enthalpically favored water structure. For **3**, however, an entropic benefit remains originating from the burial and desolvation of the additional methyl group ("classical" hydrophobic effect).

The complexes TLN-**5** and TLN-**6** formed with (*R*)-configured ligands also show elaborate surface water networks involving the formation of the stabilizing five-membered water polygon (*W5–W9*). Detailed analysis suggests that the involved water molecules experience much higher *B* factors than in the corresponding complexes with the (*S*)-configured epimers (Figure 4.5). This supposedly less stable arrangement results from the inverted stereochemistry and increases the steric demand of the (*R*)-configured P$_2$' substituents. Furthermore, one terminal methyl group of **5** is not detectable in the crystal structure

suggesting enhanced mobility, likely increasing entropy and reducing the stabilization of the adjacent water network. In contrast to the (S)-series, the (R)-configurated substituents do not enhance affinity (Figure 4.8). The gradual enhancement in $-T\Delta S°$ with growing number of methyl groups is similar within the (S) and (R)-series. However, in the (R)-series the loss in $\Delta H°$ completely nullifies the advantage in $-T\Delta S°$. Thus, overall the affinity enhancement across the (S) series ligands results from an enthalpic advantage of the growing P$_2$' substituents. They achieve more elaborated and energetically improved surface water networks.

Ligand **4** shows an unexpected binding mode, as its P$_2$' group is flipped by 180° compared to the other five ligands (Figure 4.4D). Remarkably, this ligand shows the largest surface burial across the series, significantly higher than that of its epimer **1** (Figure 4.10). The flipped orientation takes considerable impact on the established water structure, showing a much lower amount of recruited water molecules compared to the other TLN complexes. This results in the striking observation that TLN-**4**, even though exhibiting the largest surface burial, shows the lowest affinity across all six ligands (Figure 4.8). This underpins our observation that the sole burial of hydrophobic surface portions of a ligand is clearly not sufficient to explain the binding features. Due to the considerable conformational change of the P$_2$' substituent, a direct comparison of $\Delta H°$ and $-T\Delta S°$ of TLN-**4** with the thermodynamic signatures of the other five ligands is complicated.

4.4.3   Kinetic analysis of the ligands

The overall rather slow association of ligands binding to TLN is likely governed by a large conformational transition (induced fit) of the protein, rendering influences of individual ligands on k_{on} rather insignificant. As a matter of fact, the experimental determination of association rate constants is dependent on the concentration of the studied samples and thus prone to additional experimental uncertainties (e.g. weighting errors or repeated freeze-thaw cycles of the inhibitor solution affecting its concentration). For these two reasons, we refrain from a detailed interpretation of k_{on}. This uncertainty also affects K_d values determined by SPR, and will also afflict a direct comparison with K_d values from ITC measurements. In any case, there might be inherent differences from a theoretical point of view between both techniques involved, making a direct comparison of K_d values taken from both methods difficult. ITC observes a system under thermodynamic equilibrium conditions based on a particular binding model, whereas SPR records a steady state situation in a flow cell using immobilized protein, which does not necessarily relate to the same binding model as different structural states might determine binding kinetics. This may lead to differences in the determined K_d values [247].

Based on the thermodynamic equilibrium K_d values determined by ITC, **3** clearly shows the highest affinity of the series (Figure 4.8), whereas in the SPR measurements the second largest K_d value was determined for this inhibitor (Table S4.7, Supporting Information). Ligand **3** shows, however, a significantly longer residence time compared to all other ligands (Figure 4.9). Since the local interactions of a specific inhibitor conceivably have a higher influence on the dissociation kinetics, we relate this decreased dissociation rate constant of **3** to the formation of the pronounced surface water network caging the hydrophobic P_2' substituent to stabilize the complex and thus prolong residence time of the ligand.

4.5 Conclusion

Drug optimization aims for the tailored optimization of binding parameters to endow a ligand with the required potency, selectivity and binding efficacy. We increasingly recognize that the sole optimization to enhanced binding affinity is not sufficient to render a ligand as most promising candidate for further development. Consequently, additional parameters such as thermodynamic and binding kinetic signatures are consulted to obtain a more detailed view on the binding process. Furthermore, increasing resolution of the crystal structures determined across narrow series of protein–ligand complexes discloses tiny differences in the binding poses, and adaptations of the target protein or modulations of the "wetting" surface water networks. For the medicinal chemist who performs ligand optimization by means of chemical synthesis, it is essential that these modulations, which finally improve the ligand's profile, result from properties of the bound ligand and its partly solvent-exposed substituents. They allow fine-tuning of affinity, enthalpy, entropy and binding kinetic properties, as they are accomplished by well-established medicinal chemistry optimization steps. In the present study, we show that by means of optimizing the composition of a partly exposed, apolar ligand substituent bound to a flat, solvent-exposed binding pocket of a protein, the relevant binding parameters can be fine-tuned using rational design principles. We therefore had to analyze, predict and characterize the substituent's burial and in parallel the quality and perfection of the adjacent formed surface water network, which coats the formed complex. The advantage of this concept is that the chemical adjustments needed to drive the thermodynamic and kinetic parameters into a desired range are performed using the normal toolbox available to medicinal chemists. Our strategy requires the following steps along an iterative design cycle: *(i)* molecular design of the exposed substituent to optimize the pocket burial and adjacent surface water layer, *(ii)* molecular dynamics simulations to validate the proposed surface water network, *(iii)* ligand synthesis, followed by *(iv)* structural, microcalorimetric and binding kinetic characterization of the formed complexes. The Table of Contents Graphic (Figure 4.12) shows the stepwise affinity enhancement across our studied series. The most potent complex TLN-**3** (with the

2(*S*)-3,3-trimethylbutyl P$_2$' substituent, at the far right of the diagram in the Table of Contents graphic) is by about 1.5 orders of magnitude more potent in terms of affinity than the initial purely methylated complex (at the far left of the diagram). As the detailed thermodynamic characterization shows, this only partly results from enhancements of the "classical" hydrophobic effect. Moreover, additionally important, mainly more enthalpy-driven effects result from the optimization of the surface water network that coats TLN-**3** almost perfectly and establishes several fused polygonal water arrangements, which are characterized by a particular stability. Apart from the enhancement of the thermodynamic profile, TLN-**3** shows prolonged residence time that results from a more stable protein–ligand complex. Obviously, the optimized surface water layer captures and holds the ligand more tightly to the protein, thus increasing the barrier for its release. It is possible that this is a rather general concept to modulate binding kinetics, namely by enhancing the interaction of a bound ligand with the adjacent surface water network. Exploitation of this property might allow the medicinal chemist to fine-tune binding kinetic parameters via ligand optimization for many drug targets.
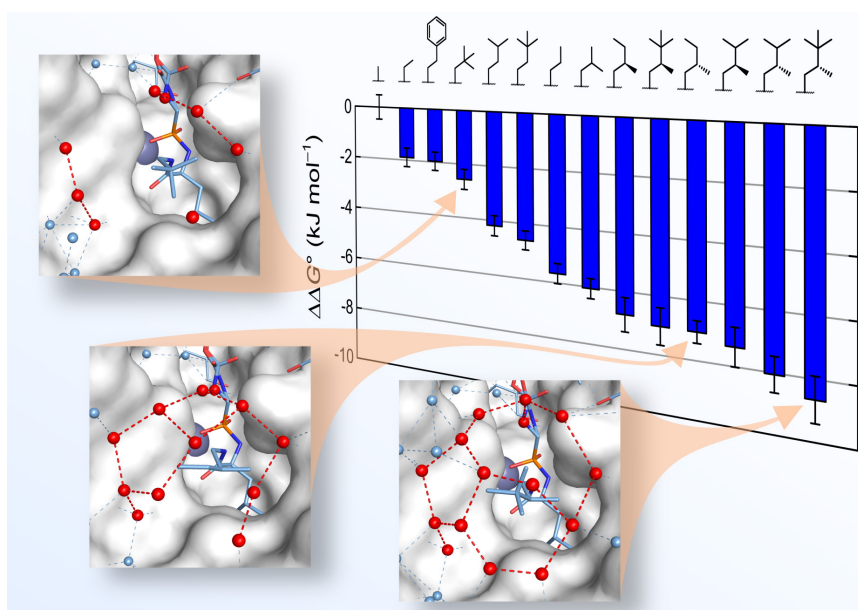
Figure 4.12.

4.6   Experimental Section

4.6.1   Water Network Prediction by Molecular Dynamics Simulations

The crystal structure of TLN-**1** (PDB code 4MZN) was used for modeling of the ligands as well as for the MD simulations. In order to provide a common environment for modeling and simulation, the Cartesian coordinates of protein, zinc ion, and cryobuffer molecules (DMSO, glycerol) were used. The preparation was performed similarly to our protocol described previously [235]. After protonation, all crystallographically observed water molecules were removed. All ligands were modeled within the binding site of TLN-**1**. As template structures for the modelling of the ligands, **1** was used for the (*S*)-configurated **2** and **3**, and the ligand in its complex to TLN from the PDB entry 4MTW [104] was used as a template structure for the (*R*)-configurated **5** and **6**, as they provided suitable exit vectors. Modeling and a subsequent minimization of the S_2' groups was performed using the molecule builder function and the *AMBER99* force field implemented in *MOE* [216]. Atomic charges for the ligands were calculated with the *RESP* methodology [248] based on quantum mechanical calculations obtained by *Gaussian09* [249] at HF/6-31G* level. The MD simulations were performed with the *AMBER14* package [250], using the ff99SB force field and periodic boundary conditions. During all simulation steps, all atoms, except hydrogen atoms and water molecules, were restrained to their coordinates of the crystal structure. In a 20 ns production phase, water molecule positions were recorded every 2 ps. This trajectory was analyzed to calculate the solvation sites using the *VOLMAP* plugin in *VMD* [251]. The protocol is described in detail in our earlier contribution [235].

4.6.2   Ligand Synthesis and Purification

^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on a JEOL ECX-400 or JEOL ECA-500 instrument. All chemical shift values are reported in ppm relative to the non-deuterated solvent signal. An external standard was used for ^{31}P NMR spectra (referenced to: 85% H_3PO_4) and ^{13}C NMR spectra in D_2O (referenced to: trimethylsilyl propanoic acid). ESI-MS spectra were recorded on a Q-Trap 2000 system by Applied Biosystems. For the description of multiplicity the following abbreviations were used: s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, dd = doublet of doublet, m = multiplet, br = broad signal. For high resolution ESI-MS a LTQ-FT Ultra mass spectrometer (Thermo Fischer Scientific) was used. EI-MS analysis was carried out on a Micromass AutoSpec instrument. For HPLC chromatography a Shimadzu LC-20 system equipped with a diode array detector was used. Analytic separations were carried out with a MN Nucleodur 100-5 C18 ec 4.6×250 mm column using a water-acetonitrile gradient. For semi-preparative separations a Water XSelect

CSH C18 10×250 mm column using a water-acetonitrile gradient was used. The purity of all inhibitors used for biophysical experiments was greater than 95%, as determined by HPLC

General procedure for the synthesis of acyloxazolidinones **9–13**: *n*-BuLi (1.2 eq) was slowly added to a solution of **7** or **8** (1.0 eq) in THF at -78 °C under argon. The solution was allowed to warm to room temperature over 30 minutes. The respective acid chloride (1.1 eq) was added to the yellow solution and the reaction mixture was stirred for 2 h at -78 °C. The reaction was quenched with saturated NH$_4$Cl solution and extracted with EtOAc (3×20 mL). The combined organic extracts were washed with brine and dried over MgSO$_4$. The crude product was purified by silica gel column chromatography (cyclohexane/EtOAc 5:1).

(S)-4-Benzyl-3-(3-methylbutanoyl)oxazolidin-2-one (**9**): Compound **9** was synthesized according to the general procedure using *n*-BuLi (2 M in hexanes, 2.4 mL, 6.00 mmol), *(S)*-4-benzyloxazolidin-2-one (**7**, 886 mg, 5.00 mmol) and 3-methylbutanoyl chloride (663 mg, 5.50 mmol). The product was obtained as a colorless oil (1243 mg, 4.76 mmol, 95%). **¹H NMR** (400 MHz, CDCl$_3$) δ = 1.02 (t, *J* = 6.6, 6H), 2.17–2.28 (m, 1H), 2.71–2.76 (m, 1H), 2.79 (dd, *J* = 10.2, 6.0, 1H), 2.89 (dd, *J* = 16.2, 6.7, 1H), 3.31 (dd, *J* = 13.3, 3.3, 1H), 4.12–4.23 (m, 2H), 4.63–4.74 (m, 1H), 7.18–7.39 (m, 5H). **¹³C NMR** (101 MHz, CDCl$_3$) δ = 22.6, 22. 7, 25.1, 38.1, 44.1, 55.3, 66.2, 127.4, 129.1, 129.5, 135. 5, 153.6, 172.8. **MS (ESI+)** *m/z* calculated for C$_{15}$H$_{20}$NO$_3$ [M+H]$^+$: 262.32; found: 262.09.

(S)-4-Benzyl-3-(3,3-dimethylbutanoyl)oxazolidin-2-one (**10**): Compound **10** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 3.4 mL, 8.40 mmol), *(S)*-4-benzyloxazolidin-2-one (**7**, 1240 mg, 7.00 mmol) and 3,3-dimethylbutanoyl chloride (1036 mg, 7.70 mmol). The product was obtained as a colorless oil (1776 mg, 6.45 mmol, 92%). **¹H NMR** (400 MHz, CDCl$_3$) δ = 1.09 (s, 9H), 2.71 (dd, *J* = 13.2, 10.1, 1H), 2.86 (d, *J* = 14.9, 1H), 2.99 (d, *J* = 14.9, 1H), 3.35 (dd, *J* = 13.3, 3.2, 1H), 4.07–4.22 (m, 2H), 4.62–4.78 (m, 1H), 7.16–7.43 (m, 5H). **¹³C NMR** (101 MHz, CDCl$_3$) δ = 29.7, 31.6, 38.2, 46.3, 55.5, 65.9, 127.4, 129.1, 129.5, 135.6, 153.6, 172.0. **MS (ESI+)** *m/z* calculated for C$_{16}$H$_{22}$NO$_3$ [M+H]$^+$: 276.35; found: 276.08.

(R)-4-Benzyl-3-butyryloxazolidin-2-one (**11**): Compound **11** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 6.6 mL, 16.50 mmol), *(R)*-4-benzyloxazolidin-2-one (**8**, 2660 mg, 15.00 mmol) and butyryl chloride (2238 mg, 21.00 mmol). The product was obtained as a colorless oil (3483 mg, 14.08 mmol, 94%). **¹H NMR** (400 MHz, CDCl$_3$) δ = 1.01 (t, *J* = 7.4 Hz, 3H), 1.66–1.80 (m, 2H), 2.77 (dd, *J* = 13.3, 9.7 Hz, 1H), 2.83–3.01 (m, 2H), 3.30 (dd, *J* = 13.3, 3.2 Hz, 1H), 4.14–4.24 (m, 2H), 4.63–4.72 (m, 1H), 7.18–7.36 (m, 5H). **¹³C NMR** (101 MHz, CDCl$_3$) δ = 13.8, 17.8, 37.5, 38.1, 55.3, 66.3, 127.5, 129.1, 129.6, 135.4, 153.6, 173.4. **MS (ESI+)** *m/z* calculated for C$_{14}$H$_{18}$NO$_3$ [M+NH$_4$]$^+$: 265.33; found: 265.19.

(R)-4-Benzyl-3-(3-methylbutanoyl)oxazolidin-2-one (**12**): Compound **12** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 2.8 mL, 7.00 mmol), *(R)*-4-benzyloxazolidin-2-one (**8**, 1050 mg, 5.90 mmol) and 3-methylbutanoyl chloride (787 mg, 6.50 mmol). The product was obtained as a colorless oil (1368 mg, 5.23 mmol, 88%). **¹H NMR** (500 MHz, CDCl₃) δ = 1.01 (dd, *J* = 8.3, 6.7, 2H), 2.15–2.29 (m, 1H), 2.71–2.81 (m, 2H), 2.89 (dd, *J* = 16.2, 6.7, 1H), 3.31 (dd, *J* = 13.3, 3.3, 1H), 4.09–4.23 (m, 2H), 4.63–4.73 (m, 1H), 7.18–7.37 (m, 5H). **¹³C NMR** (126 MHz, CDCl₃) δ = 22.5, 22.7, 25.1, 38.1, 44.1, 55.3, 66.2, 127.4, 129.0, 129.5, 135.4, 153.5, 172.8. **MS (EI)** *m/z* calculated for $C_{15}H_{19}NO_3$ [M]⁺: 261.32; found: 261.

(R)-4-Benzyl-3-(3,3-dimethylbutanoyl)oxazolidin-2-one (**13**): Compound **13** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 3.4 mL, 8.40 mmol), *(R)*-4-benzyloxazolidin-2-one (**8**, 1240 mg, 7.00 mmol) and 3,3-dimethylbutanoyl chloride (1036 mg, 7.70 mmol). The product was obtained as a colorless oil (1876 mg, 6.81 mmol, 97%). **¹H NMR** (500 MHz, CDCl₃) δ = 1.09 (s, 9H), 2.71 (dd, *J* = 13.3, 10.0 Hz, 1H), 2.86 (d, *J* = 14.9 Hz, 1H), 2.99 (d, *J* = 14.9, 1H), 3.34 (dd, *J* = 13.3, 3.3 Hz, 1H), 4.10–4.18 (m, 2H), 4.64–4.73 (m, 1H), 7.21–7.36 (m, 5H). **¹³C NMR** (126 MHz, CDCl₃) δ = 29.7, 31.5, 38.1, 46.2, 55.5, 65.9, 127.4, 129.0, 129.5, 135.6, 153.6, 172.0. **(MS ESI+)** *m/z* calculated for $C_{17}H_{24}NO_3$ [M+H]⁺: 290.38; found: 290.13.

General procedure for the synthesis of acyloxazolidinones **14–18**: A solution of diisopropylamine (1.3 eq) in dry THF under argon was cooled to -78 °C. n-BuLi (1.2 eq) was slowly added to the solution and the reaction mixture was allowed to warm to room temperature over 60 minutes. After cooling to -80 °C the respective oxazolidinone **9–13** (1.0 eq) was added dropwise to the mixture. After 60 minutes MeI (4.0 eq) was added to the solution. The reaction mixture was stirred for 5 h without further cooling. The reaction was quenched with saturated NH₄Cl-solution and extracted with EtOAc (3×20 mL). The combined organic extracts were washed with brine and dried over MgSO₄. The crude reaction product was purified by silica gel column chromatography (cyclohexane/EtOAc 6:1).

(S)-4-Benzyl-3-((S)-2,3-dimethylbutanoyl)oxazolidin-2-one (**14**): Compound **14** was synthesized according to the general procedure using diisopropylamine (591 mg, 5.84 mmol), n-BuLi (2.5 M in hexanes, 2.2 mL, 5.39 mmol), oxazolidinone **9** (1173 mg, 4.49 mmol) and MeI (2550 mg, 17.96 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 892 mg, 3.24 mmol, 68%). **¹H NMR** (400 MHz, CDCl₃) δ = 0.90–0.99 (m, 6H), 1.17 (d, *J* = 6.4 Hz, 3H), 1.94–2.06 (m, 1H), 2.76 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.29 (dd, *J* = 13.3, 3.1 Hz, 1H), 3.54–3.67 (m, 1H), 4.12–4.23 (m, 2H), 4.63–4.74 (m, 1H), 7.19–7.41 (m, 5H). **¹³C NMR** (101 MHz, CDCl₃) δ = 13.9, 18.7, 21.3, 30.7, 37.9, 43.6, 55.6, 66.0, 127.4, 129.0, 129.5, 135.5, 153.2, 177.2. **MS (ESI+)** *m/z* calculated for $C_{15}H_{20}NO_3$ [M+H]⁺: 262.32; found: 262.17.

(S)-4-Benzyl-3-((S)-2,3,3-trimethylbutanoyl)oxazolidin-2-one (**15**): Compound **15** was synthesized according to the general procedure using diisopropylamine (1383 mg, 13.66 mmol), n-BuLi (2.5 M in hexanes, 5.0 mL, 12.50 mmol), oxazolidinone **10** (2895 mg, 10.51 mmol) and MeI (5967 mg, 42.02 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 2486 mg, 8.59 mmol, 82%). **^{1}H NMR** (400 MHz, CDCl$_3$) δ = 1.00 (s, 9H), 1.19 (d, *J* = 7.0 Hz, 3H), 2.76 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.27 (dd, *J* = 13.3, 3.2 Hz, 1H), 3.89 (q, *J* = 7.0 Hz, 1H), 4.15 (d, *J* = 4.8 Hz, 2H), 4.62–4.74 (m, 1H), 7.19–7.42 (m, 5H). **^{13}C NMR** (101 MHz, CDCl$_3$) δ = 13.2, 27.5, 33.7, 37.9, 44.8, 55.7, 65.8, 127.4, 129.0, 129.6, 135.5, 153.5, 176.8. **MS (ESI+)** *m/z* calculated for C$_{17}$H$_{24}$NO$_3$ [M+H]$^+$: 290.38; found: 290.07.

(R)-4-Benzyl-3-((R)-2-methylbutanoyl)oxazolidin-2-one (**16**): Compound **16** was synthesized according to the general procedure using diisopropylamine (1842 mg, 18.20 mmol), n-BuLi (2.5 M in hexanes, 6.7 mL, 16.80 mmol), oxazolidinone **11** (3460 mg, 14.00 mmol) and MeI (7949 mg, 56.00 mmol). The product was obtained as a colorless oil (*dr* 16:1, 2647 mg, 10.13 mmol, 72%). **^{1}H NMR** (400 MHz, CDCl$_3$) δ = 0.94 (t, *J* = 7.4 Hz, 3H), 1.23 (d, *J* = 6.9 Hz, 3H), 1.42–1.54 (m, 1H), 1.72–1.85 (m, 1H), 2.78 (dd, *J* = 13.3, 9.6 Hz, 1H), 3.28 (dd, *J* = 13.3, 3.3 Hz, 1H), 3.60–3.69 (m, 1H), 4.15–4.24 (m, 2H), 4.65–4.72 (m, 1H), 7.20–7.37 (m, 5H). **^{13}C NMR** (101 MHz, CDCl$_3$) δ = 11.8, 17.0, 26.5, 38.0, 39.3, 55.5, 66.1, 127.5, 129.0, 129.6, 135.5, 153.2, 177.3. **MS (ESI+)** *m/z* calculated for C$_{15}$H$_{23}$N$_2$O$_3$ [M+H]$^+$: 279.35; found: 279.24.

(R)-4-Benzyl-3-((R)-2,3-dimethylbutanoyl)oxazolidin-2-one (**17**): Compound **17** was synthesized according to the general procedure using diisopropylamine (658 mg, 6.50 mmol), n-BuLi (2.5 M in hexanes, 2.4 mL, 6.00 mmol), oxazolidinone **12** (1307 mg, 5.00 mmol) and MeI (2839 mg, 20.00 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 2486 mg, 8.59 mmol, 61%). **^{1}H NMR** (500 MHz, CDCl$_3$) δ = 0.93 (d, *J* = 6.8 Hz, 3H), 0.95 (d, *J* = 6.7 Hz, 3H), 1.17 (d, *J* = 6.9 Hz, 3H), 1.94–2.04 (m, 1H), 2.76 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.28 (dd, *J* = 13.4, 3.3 Hz, 1H), 3.59 (quint, *J* = 6.9 Hz, 1H), 4.13–4.22 (m, 2H), 4.61–4.72 (m, 1H), 7.20–7.36 (m, 5H). **^{13}C NMR** (126 MHz, CDCl$_3$) δ = 13.9, 18.6, 21.3, 30.7, 37.9, 43.6, 55.5, 66.0, 127.4, 129.0, 129.5, 135.4, 153.2, 177.1. **MS (ESI+)** *m/z* calculated for C$_{16}$H$_{22}$NO$_3$ [M+H]$^+$: 276.36; found: 276.12.

(R)-4-Benzyl-3-((R)-2,3,3-trimethylbutanoyl)oxazolidin-2-one (**18**): Compound **18** was synthesized according to the general procedure using diisopropylamine (896 mg, 8.85 mmol), n-BuLi (2.5 M in hexanes, 3.3 mL, 8.17 mmol), oxazolidinone **13** (1876 mg, 6.81 mmol) and MeI (3866 mg, 27.24 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 1379 mg, 4.77 mmol, 70%). **^{1}H NMR** (500 MHz, CDCl$_3$) δ = 1.00 (s, 9H), 1.19 (d, *J* = 7.0 Hz, 3H), 2.75 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.27 (dd, *J* = 13.3, 3.2 Hz, 1H), 3.88 (q, *J* = 7.0 Hz, 1H), 4.11–

4.17 (m, 2H), 4.64–4.72 (m, 1H), 7.19–7.36 (m, 5H). **¹³C NMR** (126 MHz, CDCl₃) δ = 13.2, 27.5, 33.6, 37.9, 44.7, 55.6, 65.8, 127.4, 129.0, 129.6, 135.5, 153.5, 176.8. **MS (ESI+)** *m/z* calculated for $C_{12}H_{24}NO_3$ [M+H]⁺: 290.38; found: 290.13.

General procedure for the synthesis of carboxylic acids **19–23**: To a cooled solution of LiOH (2.0 eq) and H_2O_2 (4.0 eq) in water, a solution of acyloxazolidinone **14–18** (1.0 eq) in THF was added. The reaction was stirred at room temperature for 16 h. The solution was extracted with DCM (3×15 mL) and the organic phase was discarded. The aqueous phase was acidified and extracted with DCM (3×15 mL). The combined organic extracts were washed with brine and dried over $MgSO_4$. The crude reaction product was purified by silica gel column chromatography (pentane/Et₂O 5:1).

(S)-2,3-Dimethylbutanoic acid (**19**): Compound **19** was synthesized according to the general procedure using LiOH (184 mg, 4.38 mmol), H_2O_2 (30% in H_2O, 1000 mg, 8.76 mmol) and intermediate **14** (573 mg, 2.19 mmol). The product was obtained as a colorless liquid (254 mg, 2.19 mmol, 100%). **¹H NMR** (400 MHz, CDCl₃) δ = 0.93 (d, *J* = 6.8 Hz, 3H), 0.97 (d, *J* = 6.8 Hz, 3H), 1.13 (d, *J* = 7.0 Hz, 3H), 1.88–2.02 (m, 1H), 2.22–2.33 (m, 1H). **¹³C NMR** (101 MHz, CDCl₃) δ = 13.5, 19.1, 20.8, 30.9, 46.0, 182.8. **MS (ESI-)** *m/z* calculated for $C_6H_{11}O_2$ [M-H]⁻: 115.15; found: 115.15.

(S)-2,3,3-Trimethylbutanoic acid (**20**): Compound **20** was synthesized according to the general procedure using LiOH (203 mg, 4.84 mmol), H_2O_2 (30% in H_2O, 1097 mg, 9.68 mmol) and intermediate **15** (700 mg, 2.42 mmol). The product was obtained as a colorless liquid (313 mg, 2.40 mmol, 99%). **¹H NMR** (500 MHz, CDCl₃) δ = 0.99 (s, 9H), 1.14 (d, *J* = 7.1 Hz, 3H). **¹³C NMR** (126 MHz, CDCl₃) δ = 12.7, 27.6, 32.7, 49.6, 182.5. **MS (ESI-)** *m/z* calculated for $C_7H_{13}O_2$ [M-H]⁻: 129.18; found: 129.16.

(R)-2-Methylbutanoic acid (**21**): Compound **21** was synthesized according to the general procedure using LiOH (722 mg, 17.20 mmol), H_2O_2 (30% in H_2O, 3900 mg, 34.40 mmol) and intermediate **16** (1998 mg, 7.65 mmol). The product was obtained as a colorless liquid (760 mg, 7.44 mmol, 97%). **¹H NMR** (500 MHz, CDCl₃) δ = 0.95 (t, *J* = 7.5 Hz, 3H), 1.18 (d, *J* = 7.0 Hz, 3H), 1.45–1.55 (m, 1H), 1.66–1.76 (m, 1H), 2.36–2.44 (m, 1H). **¹³C NMR** (126 MHz, CDCl₃) δ = 11.7, 16.5, 26.7, 41.0, 183.2. **MS (ESI-):** *m/z* calculated for $C_5H_9O_2$ [M-H]⁻: 101.13; found: 101.11.

(R)-2,3-Dimethylbutanoic acid (**22**): Compound **22** was synthesized according to the general procedure using LiOH (258 mg, 6.14 mmol), H_2O_2 (30% in H_2O, 1393 mg, 12.28 mmol) and intermediate **17** (845 mg, 3.07 mmol). The product was obtained as a colorless liquid (332 mg, 2.86 mmol, 93%). **¹H NMR** (500 MHz, CDCl₃) δ = 0.93 (d, *J* = 6.8 Hz, 3H), 0.96 (d, *J* = 6.8 Hz, 3H), 1.13 (d, *J* = 7.0 Hz, 3H), 1.90–2.01 (m, 1H), 2.21–2.30 (m, 1H). **¹³C NMR** (126

MHz, CDCl$_3$) δ = 13.5, 19.1, 20.8, 30.8, 46.0, 182.8. **MS (ESI-)** *m/z* calculated for C$_6$H$_{11}$O$_2$ [M-H]$^-$: 115.15; found: 115.15.

(R)-2,3,3-Trimethylbutanoic acid (**23**): Compound **23** was synthesized according to the general procedure using LiOH (132 mg, 3.14 mmol), H$_2$O$_2$ (30% in H$_2$O, 713 mg, 6.28 mmol) and intermediate **18** (455 mg, 1.57 mmol). The crude reaction product was purified by silica gel column chromatography (pentane/Et$_2$O 5:1). The product was obtained as a colorless liquid (177 mg, 1.36 mmol, 87%). **¹H NMR** (400 MHz, CDCl$_3$) δ = 0.99 (s, 9H), 1.14 (d, *J* = 7.1 Hz, 3H), 2.24–2.34 (m, 1H). **¹³C NMR** (101 MHz, CDCl$_3$) δ = 12.7, 27.6, 32.7, 49.5, 182.1. **MS (ESI-)** *m/z* calculated for C$_7$H$_{13}$O$_2$ [M-H]$^-$: 129.18; found: 129.22.

General procedure for the synthesis of alcohols **24–28**: A solution of LiAlH$_4$ in Et$_2$O (2.0 eq) was added dropwise to a cooled solution of the respective carboxylic acid **19–23** (1.0 eq) in dry Et$_2$O. After stirring for 1 h the reaction was quenched by the addition of H$_2$O and 1 M HCl. The crude reaction mixture was filtered over Celite and dried over MgSO$_4$.

(S)-2,3-Dimethylbutan-1-ol (**24**): Compound **24** was synthesized according to the general procedure using LiAlH$_4$ (1 M in Et$_2$O, 4.0 mL, 4.02 mmol) and carboxylic acid **19** (233 mg, 2.01 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (161 mg, 1.58 mmol, 78%). **¹H NMR** (400 MHz, CDCl$_3$) δ = 0.83 (d, *J* = 6.8Hz, 3H), 0.86 (d, *J* = 6.9 Hz, 3H), 0.91 (d, *J* = 6.9 Hz, 3H), 1.44–1.54 (m, 1H), 1.64–1.76 (m, 1H), 3.45 (dd, *J* = 10.5, 7.0 Hz, 1H), 3.59 (dd, *J* = 10.6, 5.9 Hz, 1H). **¹³C NMR** (101 MHz, CDCl$_3$) δ = 12.6, 18.1, 20.7, 28.9, 41.5, 66.7.

(S)-2,3,3-Trimethylbutan-1-ol (**25**): Compound **25** was synthesized according to the general procedure using LiAlH$_4$ (1 M in Et$_2$O, 10.0 mL, 10.00 mmol) and carboxylic acid **20** (1118 mg, 8.59 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (950 mg, 8.17 mmol, 95%). **¹H NMR** (400 MHz, CDCl$_3$) δ = 0.88 (s, 9H), 0.94 (d, *J* = 6.8 Hz, 3H), 1.35–1.45 (m, 1H), 3.28–3.36 (m, 1H), 3.82 (dd, *J* = 10.3, 3.7 Hz, 1H). **¹³C NMR** (101 MHz, CDCl$_3$) δ = 12.5, 27.8, 32.2, 45.7, 65.5.

(R)-2-Methylbutan-1-ol (**26**): Compound **26** was synthesized according to the general procedure using LiAlH$_4$ (1 M in Et$_2$O, 7.2 mL, 7.20 mmol, 1.2 eq) and carboxylic acid **23** (613 mg, 6.00 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (493 mg, 5.60 mmol, 93%). **¹H NMR** (500 MHz, CDCl$_3$) δ = 0.89–0.93 (m, 6H), 1.10–1.24 (m, 1H), 1.41–1.49 (m, 1H), 1.51–1.59 (m, 1H), 3.39–3.54 (m, 2H). **¹³C NMR** (126 MHz, CDCl$_3$) δ = 11.5, 16.3, 25.9, 37.5, 68.2.

(R)-2,3-Dimethylbutan-1-ol (**27**): Compound **27** was synthesized according to the general procedure using LiAlH$_4$ (1 M in Et$_2$O, 6.0 mL, 6.00 mmol) and carboxylic acid **22** (326 mg, 2.81 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (176 mg, 1.72 mmol, 61%). **¹H NMR** (500 MHz, CDCl$_3$) δ = 0.83 (d, *J* = 6.8 Hz, 3H), 0.86 (d, *J* = 6.9

Hz, 3H), 0.91 (d, J = 6.9 Hz, 3H), 1.45–1.54 (m, 1H), 1.65–1.73 (m, 1H), 3.44 (dd, J = 10.6, 7.0 Hz, 2H), 3.59 (dd, J = 10.6, 5.9 Hz, 1H). **^{13}C NMR** (126 MHz, CDCl$_3$) δ = 12.6, 18.1, 20.7, 28.9, 41.5, 66.7.

(R)-2,3,3-Trimethylbutan-1-ol (**28**): Compound **28** was synthesized according to the general procedure using LiAlH$_4$ (1 M in Et$_2$O, 2.7 mL, 2.70 mmol) and carboxylic acid **23** (177 mg, 1.36 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (106 mg, 0.91 mmol, 67%). **^{1}H NMR** (500 MHz, CDCl$_3$) δ = 0.88 (s, 9H), 0.95 (d, J = 6.9 Hz, 3H), 1.37–1.45 (m, 1H), 3.33 (dd, J = 10.4, 8.8 Hz, 1H), 3.82 (dd, J = 10.4, 3.9 Hz, 1H). **^{13}C NMR** (126 MHz, CDCl$_3$) δ = 12.5, 27.8, 32.2, 45.8, 65.5.

General procedure for the synthesis of Intermediates **34–38**: PPh$_3$ (3.0 eq) was dissolved in dry THF and cooled to -15 °C. DEAD (1.4 eq) was added to the solution without the temperature exceeding -10 °C. After 30 minutes the respective alcohol was added to the reaction mixture and it was stirred at -15 °C for another 90 minutes. DPPA (1.5 eq) was slowly added to the solution and the reaction was held at -15 °C for further 4 h and then allowed to warm to RT overnight. Water (3 mL) was added to the solution and the mixture was stirred for 48 h. The crude reaction mixture was diluted with EtOAc and extracted with 1 M HCl (3×10 mL). The aqueous phases were combined and the pH was adjusted to 12. The basic mixture was extracted with DCM (3×15 mL). The combined organic phases were washed with brine and dried over K$_2$CO$_3$. Boc-Leu-OH (1.0 eq), EDC (1.3 eq), HOBt (1.3 eq) and DIPEA (3.0 eq) were added to the DCM phase and the mixture was stirred at rt overnight. The solution was diluted with EtOAc and extracted with 1 M HCl (3×10 mL) and sat. NaHCO$_3$ (3×10 mL). The combined organic phases were washed with brine and dried over MgSO$_4$. The crude product was purified by silica gel column chromatography (cyclohexane/EtOAc 7:1).

tert-Butyl ((S)-1-(((S)-2,3-dimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**34**): Intermediate **34** was synthesized according to the general procedure using PPh$_3$ (1141 mg, 4.35 mmol), DEAD (354 mg, 2.03 mmol), alcohol **24** (148 mg, 1.45 mmol), DPPA (600 mg, 2.18 mmol), Boc-Leu-OH (335 mg, 1.45 mmol), EDC (362 mg, 1.89 mmol), HOBt (255 mg, 1.89 mmol), DIPEA (562 mg, 4.35 mmol). The product was obtained as a colorless solid (119 mg, 0.38 mmol, 26%). **^{1}H NMR** (400 MHz, CDCl$_3$) δ = 0.80–0.85 (m, 6H), 0.88–0.95 (m, 9H), 1.43 (s, 9H), 1.46–1.79 (m, 5H), 2.99–3.29 (m, 2H), 3.92–4.15 (m, 1H), 4.84 (br, 1H), 6.12 (br, 1H). **^{13}C NMR** (101 MHz, CDCl$_3$) δ = 13.6, 17.9, 20.5, 22.2, 22.9, 24.8, 28.4, 30.0, 39.0, 41.1, 43.5, 53.2, 80.1, 155.9, 172.6. **MS (ESI+)** m/z calculated for C$_{17}$H$_{35}$N$_2$O$_3$ [M+H]$^+$: 315.48; found: 315.20.

tert-Butyl ((S)-1-(((S)-2,3,3-trimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**35**): Intermediate **35** was synthesized according to the general procedure using PPh$_3$ (2361 mg, 9.00 mmol), DEAD (731 mg, 4.20 mmol), alcohol **25** (345 mg, 3.00 mmol), DPPA (1238 mg, 4.50 mmol), Boc-Leu-OH (694 mg, 3.00 mmol), EDC (748 mg, 3.90 mmol), HOBt (527

mg, 3.90 mmol), DIPEA (1163 mg, 9.00 mmol). The product was obtained as a colorless solid (319 mg, 0.97 mmol, 32%). ^{1}**H NMR** (500 MHz, CDCl$_3$) δ = 0.85 (d, *J* = 6.8 Hz, 3H), 0.89 (s, 9H), 0.91–0.95 (m, 6H), 1.34–1.41 (m, 1H), 1.43 (s, 9H), 1.62–1.72 (m, 2H), 2.89–2.98 (m, 1H), 3.37–3.48 (m, 1H), 4.00–4.10 (m, 1H), 4.87 (br, 1H), 6.07–6.18 (m, 1H). ^{13}**C NMR** (126 MHz, CDCl$_3$) δ = 13.1, 22.2, 23.0, 24.9, 27.5, 28.4, 32.5, 41.1, 41.9, 43.3, 53.3, 80.1, 155.9, 172.6. **MS (ESI+)** *m/z* calculated for C$_{18}$H$_{37}$N$_2$O$_3$ [M+H]$^+$: 329.50; found: 329.19.

tert-Butyl ((S)-4-methyl-1-(((R)-2-methylbutyl)amino)-1-oxopentan-2-yl)carbamate (**36**): Intermediate **36** was synthesized according to the general procedure using PPh$_3$ (2361 mg, 9.00 mmol), DEAD (731 mg, 4.20 mmol), alcohol **26** (264 mg, 3.00 mmol), DPPA (1238 mg, 4.50 mmol), Boc-Leu-OH (694 mg, 3.00 mmol), EDC (748 mg, 3.90 mmol), HOBt (527 mg, 3.90 mmol), DIPEA (1163 mg, 9.00 mmol). The product was obtained as a colorless solid (327 mg, 1.09 mmol, 36%). ^{1}**H NMR** (400 MHz, CDCl$_3$) δ = 0.85–0.96 (m, 12H), 1.08–1.20 (m, 1H), 1.32–1.40 (m, 1H), 1.43 (s, 9H), 1.49–1.61 (m, 2H), 1.61–1.73 (m, 2H), 3.02–3.23 (m, 2H), 3.99–4.11 (m, 1H), 4.84–4.94 (m, 1H), 6.19 (s, 1H). ^{13}**C NMR** (101 MHz, CDCl$_3$) δ = 11.4, 17.2, 22.3, 23.0, 24.9, 27.0, 28.4, 35.0, 41.1, 45.1, 53.3, 80.1, 156.0, 172.7. **MS (ESI+)** *m/z* calculated for C$_{16}$H$_{33}$N$_2$O$_3$ [M+H]$^+$: 301.45; found: 301.28.

tert-Butyl ((S)-1-(((R)-2,3-dimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**37**): Intermediate **37** was synthesized according to the general procedure using PPh$_3$ (1251 mg, 4.77 mmol), DEAD (388 mg, 2.23 mmol), alcohol **27** (162 mg, 1.59 mmol), DPPA (658 mg, 2.39 mmol), Boc-Leu-OH (368 mg, 1.59 mmol), EDC (397 mg, 2.07 mmol), HOBt (280 mg, 2.07 mmol), DIPEA (617 mg, 4.77 mmol). The product was obtained as a colorless solid (136 mg, 0.43 mmol, 22%). ^{1}**H NMR** (500 MHz, CDCl$_3$) δ = 0.81–0.85 (m, 6H), 0.90 (d, *J* = 6.8 Hz, 3H), 0.91–0.95 (m, 6H), 1.43 (s, 9H), 1.45–1.54 (m, 1H), 1.55–1.64 (m, 1H), 1.64–1.75 (m, 2H), 3.02–3.12 (m, 1H), 3.17–3.31 (m, 1H), 3.99–4.08 (m, 1H), 4.87 (br, 1H), 6.16 (br, 1H). ^{13}**C NMR** (126 MHz, CDCl$_3$) δ = 13.59, 17.98, 20.57, 22.28, 22.98, 24.88, 28.41, 30.07, 38.99, 41.13, 43.53, 53.27, 80.08, 155.95, 172.63. **MS (ESI+)** *m/z* calculated for C$_{17}$H$_{35}$N$_2$O$_3$ [M+H]$^+$: 315.48; found: 315.05.

tert-Butyl ((S)-1-(((R)-2,3,3-trimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**38**): Intermediate **38** was synthesized according to the general procedure using PPh$_3$ (716 mg, 2.73 mmol), DEAD (222 mg, 1.27 mmol), alcohol **28** (106 mg, 0.91 mmol), DPPA (377 mg, 1.37 mmol), Boc-Leu-OH (210 mg, 0.91 mmol), EDC (227 mg, 1.18 mmol), HOBt (159 mg, 1.18 mmol), DIPEA (353 mg, 2.73 mmol). The product was obtained as a colorless solid (80 mg, 0.24 mmol, 27%). ^{1}**H NMR** (400 MHz, CDCl$_3$) δ = 0.85 (d, *J* = 6.9 Hz, 3H), 0.88 (s, 9H), 0.90–0.96 (m, 6H), 1.32–1.40 (m, 2H), 1.43 (s, 9H), 1.58–1.75 (m, 2H), 2.88–2.97 (m, 1H), 3.38–3.49 (m, 1H), 3.99–4.14 (m, 1H), 4.86–4.97 (m, 1H), 6.17 (br, 1H). ^{13}**C NMR** (101 MHz, CDCl$_3$) δ = 13.1, 22.3, 23.0, 24.9, 27.5, 28.4, 32.6, 41.1, 42.0, 43.3, 53.3, 80.2, 155.9, 172.6. **MS (ESI+)** *m/z* calculated for C$_{18}$H$_{37}$N$_2$O$_3$ [M+H]$^+$: 329.50; found: 329.25.

General procedure for phosphonamidate coupling and deprotection: $SOCl_2$ (4.5 eq) was dissolved in 5 mL dry DCM under argon and cooled to 0 °C. A solution of benzyl ((hydroxy(methoxy)phosphoryl)methyl)carbamate (3 eq) was added to the solution over 30 minutes. The cooling bath was removed and the reaction was stirred at RT for 3-5 h. All volatile components were removed under reduced pressure. In a second flask, peptidic intermediates **34**–**38** (1 eq) was treated with HCl (4 M in Dioxane,) for 1 h. All volatile components were removed under reduced pressure. The activated phosphonic acid was dissolved in dry DCM. DIPEA and the deprotected leucylamide were added to the solution consecutively. The mixture was heated to 40 °C and stirred overnight. The solution was diluted with EtOAc and extracted with 5% citric acid (3×10 mL), 1 M HCl (1×10 mL) and 1 M NaOH (3×10 mL). The organic phase was washed with brine and dried over $MgSO_4$. The crude product was used in the next step without further purification. The protected phosphonamidate was treated with 3mL of a 0.4 M solution of LiOH in water. If needed, MeCN was added until the solution cleared. The reaction was stirred for 4-5 h at rt. Under ice cooling the pH was adjusted to 8 using 5% AcOH and the solvent was removed under reduced pressure. The residue was dissolved in a small amount of water and purified by semi-preparative HPLC.

Phosphonamidate **2**: Ligand **2** was synthesized according to the general procedure using phosphonic acid monoester **39** (249 mg, 0.96 mmol), $SOCl_2$ (171 mg, 1.44 mmol), intermediate **34** (100 mg, 0.32 mmol), HCl (4 M in Dioxane, 1 mL, 4 mmol) and DIPEA (124 mg, 0.96 mmol). The product was obtained as a colorless solid (45 mg, 0.10 mmol, 31%). **¹H NMR** (500 MHz, D_2O) δ = 0.84 (d, *J* = 6.8 Hz, 3H), 0.87 (d, *J* = 6.7 Hz, 3H), 0.92–0.98 (m, 9H), 1.44–1.52 (m, 1H), 1.52–1.76 (m, 4H), 3.04 (dd, *J* = 13.2, 8.0 Hz, 1H), 3.20 (dd, *J* = 13.4, 6.0 Hz, 1H), 3.26–3.40 (m, 2H), 3.67–3.75 (m, 1H), 5.20 (s, 2H), 7.45–7.56 (m, 5H). **¹³C NMR** (126 MHz, D_2O) δ = 12.7, 17.1, 19.8, 21.5, 22.3, 24.3, 29.5, 38.3, 39.9 (d, *J* = 136.0), 43.5, 43.7, 43.7, 54.3, 67.3, 127.9, 128.5, 128.9, 136.5, 158.3, 177.8. **³¹P NMR** (202 MHz, D_2O) δ = 18.0. **HRMS (ESI+)** calculated for $C_{21}H_{35}N_3O_5P$ [M+H]⁺: 440.2320; found: 440.2323.

Phosphonamidate **3**: Ligand **3** was synthesized according to the general procedure using phosphonic acid monoester **39** (207 mg, 0.80 mmol), $SOCl_2$ (190 mg, 1.60 mmol), intermediate **35** (131 mg, 0.40 mmol), HCl (4 M in Dioxane, 1 mL, 4 mmol) and DIPEA (155 mg, 1.20 mmol). The product was obtained as a colorless solid (68 mg, 0.15 mmol, 37%). **¹H NMR** (500 MHz, D_2O) δ = 0.75 (d, *J* = 6.9 Hz, 3H), 0.81 (s, 9H), 0.82–0.86 (m, 6H), 1.30–1.41 (m, 2H), 1.42–1.52 (m, 1H), 1.53–1.65 (m, 1H), 2.84 (dd, *J* = 12.5, 11.0 Hz, 1H), 3.22 (dd, *J* = 11.8, 5.7 Hz, 2H), 3.27 (dd, *J* = 13.1, 3.4 Hz, 1H), 3.55–3.64 (m, 1H), 5.09 (s, 2H), 7.33–7.44 (m, 5H). **¹³C NMR** (126 MHz, D_2O) δ = 12.4, 21.5, 22.3, 24.3, 26.7, 31.8, 40.0 (d, *J* = 136.3), 42.0, 42.6, 43.7, 43.7, 54.3, 67.2, 127.8, 128.5, 128.9, 136.5, 158.3, 177.8. **³¹P NMR** (202 MHz,

D$_2$O) δ = 18.0. **HRMS (ESI+)** calculated for C$_{22}$H$_{37}$N$_3$O$_5$P: 454.2476 [M+H]$^+$; found: 454.2478.

Phosphonamidate **4**: Ligand **4** was synthesized according to the general procedure using phosphonic acid monoester **39** (117 mg, 0.45 mmol), SOCl$_2$ (81 mg, 0.68 mmol), intermediate **36** (45 mg, 0.15 mmol), HCl (4 M in Dioxane, 1 mL, 4 mmol) and DIPEA (58 mg, 0.45 mmol). The product was obtained as a colorless solid (25 mg, 0.06 mmol, 37%). **^{1}H NMR** (500 MHz, MeOD) δ = 0.85–0.98 (m, 12H), 1.09–1.19 (m, 1H), 1.35–1.48 (m, 2H), 1.52–1.63 (m, 2H), 1.72–1.82 (m, 1H), 2.93–3.04 (m, 1H), 3.09 (dd, *J* = 13.2, 6.3 Hz, 1H), 3.18–3.29 (m, 2H), 3.67–3.76 (m, 1H), 5.04–5.14 (m, 2H), 7.25–7.41 (m, 5H). **^{13}C NMR** (126 MHz, MeOD) δ = 11.6, 17.5, 22.5, 23.5, 25.7, 28.1, 36.1, 41.2, 42.3, 45.3 (d, *J* = 5.3), 46.2, 55.5, 67.6, 128.9, 129.0, 129.4, 138.3, 158.8, 178.2. **^{31}P NMR** (202 MHz, MeOD) δ = 16.4. **HRMS (ESI+)** calculated for C$_{20}$H$_{33}$N$_3$O$_5$P: 426.2163 [M+H]$^+$; found: 426.2166.

Phosphonamidate **5**: Ligand **5** was synthesized according to the general procedure using phosphonic acid monoester **39** (249 mg, 0.96 mmol), SOCl$_2$ (171 mg, 1.44 mmol), intermediate **37** (100 mg, 0.32 mmol), HCl (4 M in Dioxane, 1 mL, 4 mmol) and DIPEA (124 mg, 0.96 mmol). The product was obtained as a colorless solid (37 mg, 0.08 mmol, 26%). **^{1}H NMR** (500 MHz, D$_2$O) δ = 0.84 (d, *J* = 6.8 Hz, 3H), 0.87 (d, *J* = 6.7 Hz, 3H), 0.92–0.98 (m, 9H), 1.44–1.77 (m, 5H), 3.04 (dd, *J* = 13.3, 8.0 Hz, 1H), 3.20 (dd, *J* = 13.3, 6.0 Hz, 1H), 3.27–3.41 (m, 2H), 3.67–3.74 (m, 1H), 5.20 (s, 2H), 7.45–7.55 (m, 5H). **^{13}C NMR** (126 MHz, D$_2$O) δ = 12.7, 17.1, 19.8, 21.5, 22.3, 24.3, 29.5, 38.3, 39.9 (d, *J* = 135.9), 43.5, 43.7, 43.7, 54.3, 67.3, 127.9, 128.5, 128.9, 136.5, 158.3, 177.8. **^{31}P NMR** (202 MHz, D$_2$O) δ = 18.0. **HRMS (ESI+)** calculated for C$_{21}$H$_{35}$N$_3$O$_5$P [M+H]$^+$: 440.2322; found: 440.2323.

Phosphonamidate **6**: Ligand **6** was synthesized according to the general procedure using phosphonic acid monoester **39** (156 mg, 0.60 mmol), SOCl$_2$ (107 mg, 0.90 mmol), intermediate **38** (65 mg, 0.20 mmol), HCl (4 M in Dioxane, 1 mL, 4 mmol) and DIPEA (78 mg, 0.60 mmol). The product was obtained as a colorless solid (29 mg, 0.06 mmol, 31%). **^{1}H NMR** (500 MHz, D$_2$O) δ = 0.75 (d, *J* = 6.9 Hz, 3H), 0.82 (s, 9H), 0.83–0.87 (m, 6H), 1.32–1.41 (m, 2H), 1.43–1.50 (m, 1H), 1.55–1.64 (m, 1H), 2.84 (dd, *J* = 12.9, 10.7 Hz, 1H), 3.18–3.31 (m, 3H), 3.56–3.63 (m, 1H), 5.09 (s, *J* = 16.7 Hz, 2H), 7.33–7.44 (m, 5H). **^{13}C NMR** (101 MHz, D$_2$O) δ = 12.2, 21.3, 22.1, 24.1, 26.5, 31.6, 39.8 (d, *J* = 135.7), 41.8, 42.5, 43.6, 54.2, 67.1, 127.7, 128.4, 128.8, 136.2, 152.8, 172.5. **^{31}P NMR** (162 MHz, D$_2$O) δ = 17.9. **HRMS (ESI+)** calculated for C$_{22}$H$_{37}$N$_3$O$_5$P: 454.2476 [M+H]$^+$; found: 454.2481.

4.6.3   Crystal Preparation and Soaking

TLN crystals were prepared similarly to the procedure as previously described [190]. Lyophilized TLN powder was commercially obtained from Calbiochem (EMD Biosciences). For crystal preparation, 1 mL of demineralized water was pipetted into the reservoir wells of a 24 well sitting drop crystallization plate (Cryschem, Hampton research). An 8 mM protein suspension was prepared by adding TLN powder to 50 µL of pure DMSO. To the resulting suspension, 50 µL of an aqueous solution containing 3.8 M CsCl and 100 mM Tris-HCl, pH 7.5 were added. The TLN powder was completely dissolved by mixing with an Eppendorf pipette, resulting in a clear solution of yellowish color. After centrifugation for 3 min (RCF = 16000 g), 1 µL of clear solution was transferred into each of the protein wells of the crystal plate. Subsequently, the crystal plate was sealed and TLN crystals stopped growing after five days at 18 °C. For ligand soaking, crystals with an obelisk shape were transferred into a soaking solution composed of 100 mM Tris-HCl, pH 7.5, 2 mM $CaCl_2$, 5% DMSO and 1 mM of the respective ligand followed by incubation for 24 h. Afterwards, crystals were transferred into a cryo buffer composed of 10 mM Tris/HCl, pH 7.5, 10 mM $Ca(CH_3COO)_2$, 5% DMSO, 20% glycerol and 1 mM of the respective ligand for 5 s and subsequently flash-frozen in liquid nitrogen. For the glycerol-free crystal structures (Chapter 4.11.2, Supporting Information), either 30% PEG400 or 20% MPD was used as cryoprotectant instead of glycerol.

4.6.4   Data Collection and Refinement

Data collection of datasets TLN-2_{GOL}, TLN-3_{GOL}, TLN-4_{GOL}, TLN-5_{GOL}, TLN-6_{GOL}, TLN-3_{PEG400}, and TLN-6_{MPD} was performed with synchrotron radiation at the BESSY II electron storage ring (Berlin-Adlershof, Germany) operated by the Joint Berlin MX-Laboratory [210] at beamline 14.1 with a wavelength of 0.91841 Å at cryogenic temperature of 100 K on an Dectris Pilatus 6M pixel detector. Dataset TLN-5_{MPD} was collected at Elettra (Trieste, Italy) at beamline XRD1 with a wavelength of 1.0000 Å at cryogenic temperature of 100 K on a Dectris Pilatus 2M pixel detector. Data collection and refinement statistics for the glycerol-containing crystal structures are listed in Table 4.1 and for the glycerol-free crystal structures in Table S4.1 (Supporting Information). Data indexing, integration and scaling was performed with *XDS* [252]. Molecular replacement was performed with *Phaser* [213] from the *CCP*4 software suite version 6.3.0 [238]. A TLN search model from the PDB entry 8TLN was used [212], with flexible side-chains, additives, water molecules and ions removed. Alternating cycles of model building and refinement were performed with *Coot* [253] and *phenix.refine* version 1.10.1-2155 [215], respectively, until the models optimally explained the electron densities and their *R* values reached convergence. For cross-validation, a randomly

chosen 5% of the reflections, which were not applied for model refinement, were used for the calculation of R_{free}. As an initial refinement step, Cartesian simulated annealing was performed (start temperature 5000 K, final temperature 300 K). All macromolecule amino acids were refined with riding hydrogen atoms which were added to the refinement model coordinates with *phenix.refine*. In all structures, *B* factors of all atoms except for hydrogen atoms were refined anisotropically, since this resulted in a significant decrease of the R_{free} value compared to an isotropically refinement of the *B* factors. Ligand molecule building and minimization was performed with *MOE* version 2014.09 [216], and restraints were prepared with *phenix.elbow* [254]. F_o-F_c omit electron densities were created by deletion of parts of the refinement model followed by refinement with *phenix.refine*. The graphical representations of the three dimensional structures were prepared with *PyMOL* [221].

4.6.5 Isothermal Titration Calorimetry Measurements

ITC titrations were performed based on established protocols introducing some modifications [42, 64, 96, 104]. For all measurements, the same Microcal ITC_{200} device (GE Healthcare) was used. Lyophilized powder of native TLN expressed by *Bacillus thermoproteolyticus* was bought from Calbiochem (EMD Biosciences). The protein purity was 60%, additionally containing $Ca(CH_3COO)_2$ and $Na(CH_3COO)$. The powder was weighed with a MX5 balance from Mettler Toledo (readability ± 1 μg) and directly dissolved in measurement buffer without further processing, as the additional salts did not show any interfering heat effects during the measurement [104]. All measurements were performed in a buffer composed of 200 mM NaSCN, 2 mM $CaCl_2$ and 20 mM HEPES, pH 7.5, which was filtered through a 0.22 μm filter and degassed prior to use. A TLN concentration of 30 μM was used in the titration cell, resulting in titration curves described by optimal *c* values in the range of 73 (ligand **4**) to 201 (ligand **3**). The relatively narrow affinity range of the studied ligands made it possible to keep the TLN concentration constant during all measurements, which is important to guarantee a similar activity coefficient of the protein in solution [96]. The ligand solution was prepared by directly weighing the highly pure, freeze dried ligand powder and dissolving it in measurement buffer (without the addition of DMSO). For each measurement, new freeze-dried protein and ligand powder was freshly dissolved in measurement buffer. For highest measurement precision, a 10 injections scheme with an injection volume of 1.3–1.4 μL and a ligand concentration in the syringe of 1 mM was applied [136]. This measurement protocol resulted in injections with strong heat signals exhibiting lower standard deviations compared to a conventional 25 injection scheme. At the end of the titration, a titrant to titrand ratio of at least 2.2 was achieved. After filling of the syringe with ligand solution and prior to injection of the syringe into the measurement cell, a manual 1.0 μL injection was performed outside of the measurement cell in order to adjust the syringe

drive mechanism from an "up" to a "down" movement in order to increase the precision of the first injection [138]. Furthermore, the measurement scheme contained an initial injection of 0.3 µL, which was excluded from data analysis. The measurement settings were adjusted to an initial spacing of 150–200 s, a spacing of 150–160 s between the injections, an injection speed of 2.6–2.8 s (depending on the injection volume), a syringe speed of 1000 rpm, a reference power of 5 µcal s^{-1} and a measurement temperature of 298.15 K. Peak extraction and integration was automatically performed with *NITPIC* version 1.1.2 [155]. The fitting of a 1:1 binding model curve and binding parameter extraction was performed with *SEDPHAT* version 12.1b [177] and plots of the raw data and binding isotherms were prepared with *GUSSI* [255] (Figure S4.5, Supporting Information). For the study of the influence of glycerol on the thermodynamic binding profiles of **2**, **3**, **5** and **6**, buffers composed of 200 mM NaSCN, 2 mM CaCl$_2$, 20 mM HEPES, pH 7.5 and a glycerol concentration of 1%, 2.5%, 5% or 10% were used. For the titrations of **3** and **6** with the addition of DMSO, buffers composed of 200 mM NaSCN, 2 mM CaCl$_2$, 20 mM HEPES, pH 7.5 and a DMSO concentration of 0.065 M or 0.130 M DMSO were used. All measurements were performed in triplicate, from which the mean and the standard deviation were calculated.

4.6.6 Surface Plasmon Resonance Measurements

SPR measurements were performed on a Biacore T200 system (GE Healthcare, Sweden) with analysis and sample compartment temperature set to 25 °C. The binding kinetic assay was developed as a capture assay of biotinylated TLN utilizing the Biotin CAPture Kit (GE Healthcare) combined with single cycle kinetics of the inhibitors. For the biotinylation process, lyophilized powder of native TLN (Calbiochem, see above) was dissolved to approximately 0.5 µg mL^{-1} in 0.1 M sodium borate buffer pH 8.5 supplemented with 2 mM CaCl$_2$ and preincubated with excess molar concentration of a high affinity TLN-inhibitor to protect the binding site from biotinylation. The biotinylation reaction was set up using the EZ-Link Sulfo-NHS-LC-Biotin reagent (Pierce) with a 0.6 biotin/protein molar ratio. Following incubation for 1 h at room temperature, the labelled protein was purified from non-reacted biotin reagents by desalting on a PD-10 column according to the manufacturer's instructions (GE Healthcare) to 50 mM Tris pH 8.0, 100 mM NaSCN, 2 mM CaCl$_2$. Concentration of purified TLN-biotin was determined using the mean value of triplicate absorbance readings at 280 nm on a ND1000 spectrometer (Nanodrop). Aliqous of TLN-biotin were kept at 4 °C for use within the next two days, alternatively flash frozen in liquid nitrogen for storage at -70 °C. Before use, TLN-biotin was defrosted on ice if frozen and centrifuged in a bench top centrifuge at high speed for 5 min at 4 °C.

Sensor Chip CAP was docked in the Biacore and prepared according to the manufacturer's instructions. For each day of SPR measurements, new dilutions of inhibitors from 10 mM

stock solutions in 100% DMSO (stored at -20 °C) and TLN-biotin, were made using freshly prepared buffer. 50 mM Tris pH 8.0, 100 mM NaSCN, 2 mM CaCl$_2$, 2% DMSO was used as sample and running buffer. Runs included three start-up cycles and each analysis cycle contained 4 steps: (1) injection of Biotin CAPture reagent (streptavidin modified with a deoxyriboologonucletide that hybridizes with the complimentary oligonucleotide present on the Sensor Chip CAP) in both reference and active flow cells for 300 s and at 2 µl min^{-1} resulting in response levels of around 3000 RU, (2) injection of 150–200 µg mL^{-1} TLN-biotin at 10 µL min^{-1} for 90–180 s in active flow cell only resulting in capture levels of 800–1200 RU (a capture stabilization time of 300 s was applied for some runs), (3) injection of inhibitor in increasing concentration over reference and active flow cell using a single cycle kinetics procedure with five 120 s injections at 30 µL min^{-1} and a 240 s dissociation time after the final injection, and (4) injection of standard (6 M guanidine-HCl, 0.25 M NaOH) and additional (30% acetonitrile in 0.25 M NaOH) regeneration solutions to remove the Biotin Cap Reagent, TLN-biotin and any bound inhibitor from both flow cells. Inhibitor concentrations in the range of 156–2500 nM were used for **1**, 25–2000 nM for **2** and **5** and 156–10000 nM for **3** and **6**. Blank cycles defined as analysis cycles with buffer only in step (3), were performed first, last and between every inhibitor concentration series. Data were double referenced by first subtraction of reference flow cell and then subtraction of blank cycles. Fitting of data was performed using Biacore T200 evaluation software 2.0, applying a 1:1 binding model compensating for linear drift.

4.6.7 Calculation of Buried Solvent Accessible Surface Areas

The total and buried SASAs of **1**–**6** (Figure 4.10) were calculated with the PISA server from the European Bioinformatics Institute [256]. The buried solvent accessible surface area is defined as the SASA of the ligand which becomes inaccessible to water molecules (radius 1.4 Å) through binding of the ligand to the protein cavity. From the disordered ligand's carbamate group, only the orientation with the carbonyl oxygen directed to the right (relative to the view of Figure 4.4) was considered for the calculation of the surface. Furthermore, the zinc ion, crystallographic additives (glycerol, DMSO) and the crystallographic symmetry mate were considered for the calculations, whereas hydrogen atoms were excluded from the calculation.

4.7 Author Contributions

S.G.K. and J.C. contributed equally to this work. The study was designed by S.G.K., J.C., M.B. and G.K. M.B. performed the MD simulations. J.C. synthesized the ligands. S.G.K. and A.H.

determined the crystal structures. S.G.K. collected the thermodynamic data. V.F. performed the kinetic experiments. Kinetic data was analyzed by V.F. and R.K. Data interpretation was performed by S.G.K, J.C. and G.K. This manuscript was written by S.G.K, J.C. and G.K. with contributions from M.B. and V.F.

4.8   Notes

The authors declare no competing financial interest. All described crystal structures were deposited in the PDB [80] and are available upon article publication under the accession codes 5JT9 (TLN-2_{GOL}), 5JS3 (TLN-3_{GOL}), 5JXN (TLN-4_{GOL}), 5JVI (TLN-5_{GOL}), 5JSS (TLN-6_{GOL}), 5L8P (TLN-3_{PEG400}), 5L41 (TLN-5_{MPD}), and 5L3U (TLN-6_{MPD}).

4.9   Acknowledgements

4.10   Abbreviations used

ITC, isothermal titration calorimetry; SPR, surface plasmon resonance; MD, molecular dynamics; TLN, thermolysin; GOL, glycerol; MPD, methyl-2,4-pentanediol; PEG, polyethylene glycol; SASA, solvent accessible surface area; EDC, 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide; DPPA, diphenyl-phosphoryl azide; DEAD, diethyl azodicarboxylate; HOBt, hydroxybenzotriazole; DIPEA, N,N-diisopropylethylamine; DCM, dichloromethane

4.11 Supporting Information

4.11.1 F_o–F_c omit electron densities of TLN-bound ligands 1–6
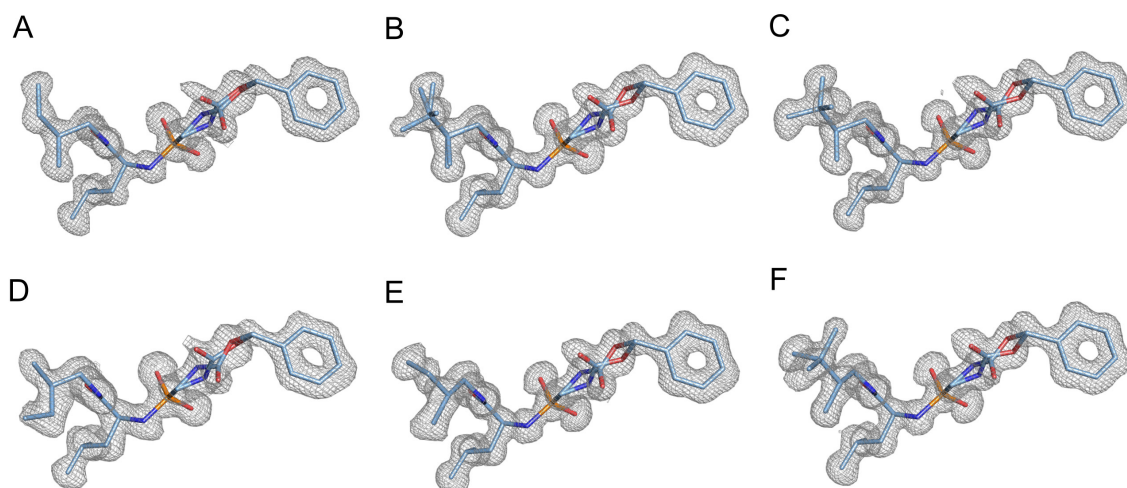
Figure S4.1 (related to Figure 4.4). Crystal structures of TLN bound ligands **(A) 1**, **(B) 2**, **(C) 3**, **(D) 4**, **(E) 5**, and **(F) 6**. Ligands are shown as stick models with carbon atoms in *blue* and color-coded heteroatoms. Their F_o–F_c omit electron densities are displayed as *gray* meshes at a contour level of 3σ. In all six crystal structures, the carbamate group of the bound ligands adopts two conformations. Moreover, the P_2' group of **2** adopts two conformations and one terminal methyl group of the P_2' group of **5** is not detectable in the electron density. The crystal structure of **1** has been published previously [104].

Chapter 4

4.11.2 Glycerol-free crystal structures TLN-3$_{PEG400}$, TLN-5$_{MPD}$ and TLN-6$_{MPD}$
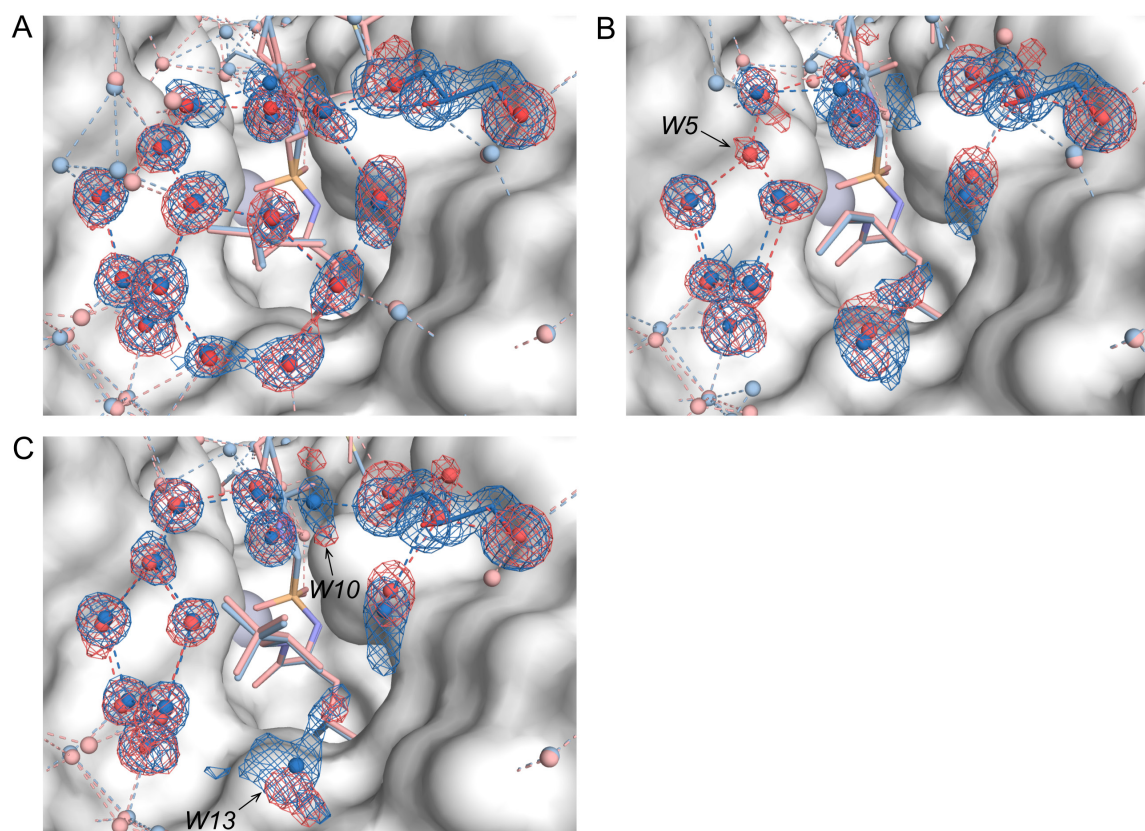
Figure S4.2. Comparison of the positions of the water molecules in the glycerol-containing and glycerol-free crystal structures of **(A)** TLN-**3**, **(B)** TLN-**5** and **(C)** TLN-**6**. Structural elements of the glycerol-containing crystal structures (*blue*) are superimposed on structural elements of the glycerol-free crystal structures (*red*). Ligand and glycerol molecules are displayed as stick models, water molecules as spheres. Distances ≤ 3.4 Å between water molecules are displayed as dotted lines indicating H-bonds. The F_o–F_c omit electron densities are shown as meshes at a contour level of 3σ in colors corresponding to the structures. For clarity reason, structural elements beyond the first solvation layer around all ligand P$_2$' groups and the glycerol molecules are displayed in pale colors and their electron densities are omitted. The solvent excluded surface of the glycerol-containing TLN crystal structure is displayed in *white*. The binding modes of the ligands in the glycerol-containing and glycerol-free crystal are identical and only minor differences in the adjacent water structures are observed. In TLN-**5**$_{MPD}$ (panel B, *red*), water molecule *W5* is observed in the electron density (highly mobile, non-normalized *B* factor of 49 Å^2), whereas the electron density of TLN-**5**$_{GOL}$ (panel B, *blue*) is too weak for the placement of a water molecule at this position in the refinement model. Furthermore, in TLN-**6**$_{MPD}$ (panel C, *red*), the electron density is too weak for the placement of water molecules *W10* and *W13*, whereas in TLN-**6**$_{GOL}$ (panel C, *blue*) these two water molecules are sufficiently stabilized for placement in the refinement model (highly mobile, see Figure 4.5 of the main text). F_o–F_c omit electron densities of the TLN-bound ligands of the glycerol-free crystal structures are displayed in Figure S4.3.
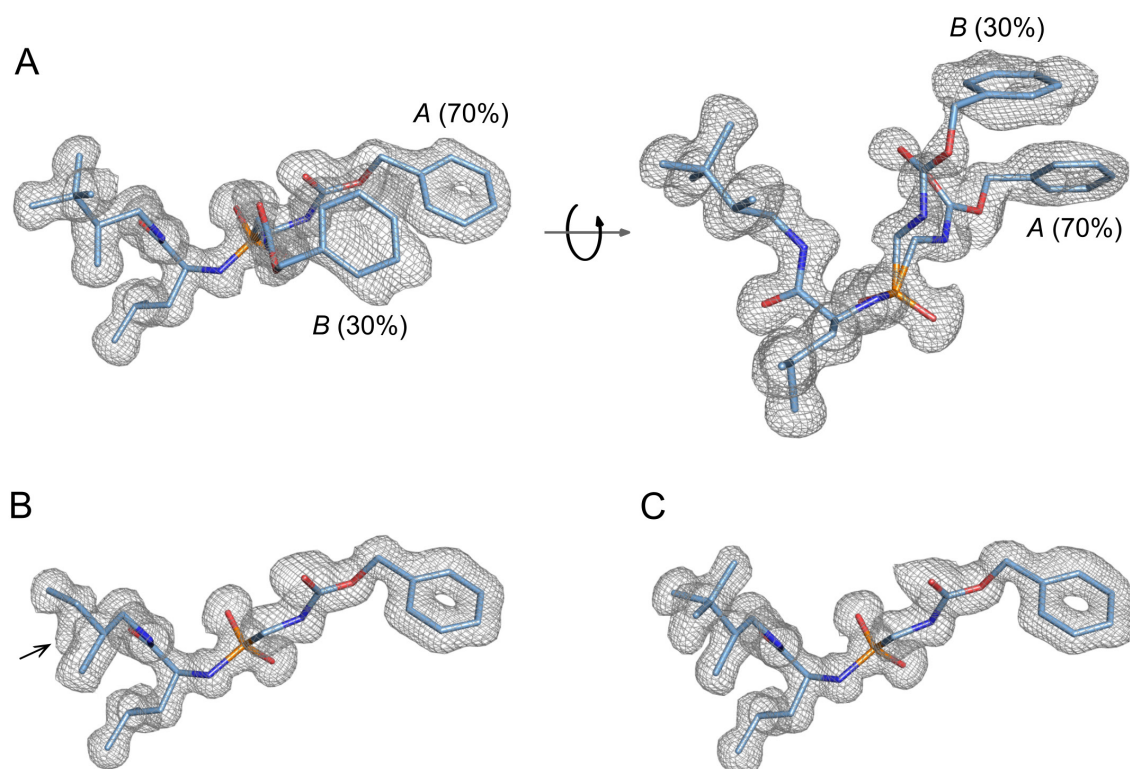
Figure S4.3. Crystal structures of the TLN-bound inhibitors **(A)** TLN-3_{PEG400}, **(B)** TLN-5_{MPD} and **(C)** TLN-6_{MPD}. Ligands are shown as stick models with carbon atoms in *blue* and color-coded heteroatoms. The F_o–F_c omit electron densities are displayed as *gray* meshes at contour levels of 2.5σ for **3** and 3.0σ for **5** and **6**. The carboxybenzyl portion of **3** shows a second conformation B (30% occupancy), occupying the space where a DMSO molecule is found in the other crystal structures (Figure S4.4). After refinement of the model of TLN-**3** against the diffraction data, a positive, featureless F_o–F_c electron density blob remains in position of the $2F_o$–F_c electron density of the phenyl ring of the carboxybenzyl group in conformation B. Since the electron density reflects the average of all conformations that a structural element adopts in the protein crystal, the unexplained F_o–F_c electron density at the position of the phenyl ring most likely originates from a DMSO molecule binding to this site in case **3** adopts conformation A. The electron density of ligand **5** indicates the missing terminal P_2' methyl group by a weak F_o–F_c electron density (black arrow), which, however, is not sufficient for placement of this methyl group in the refinement model.

Chapter 4
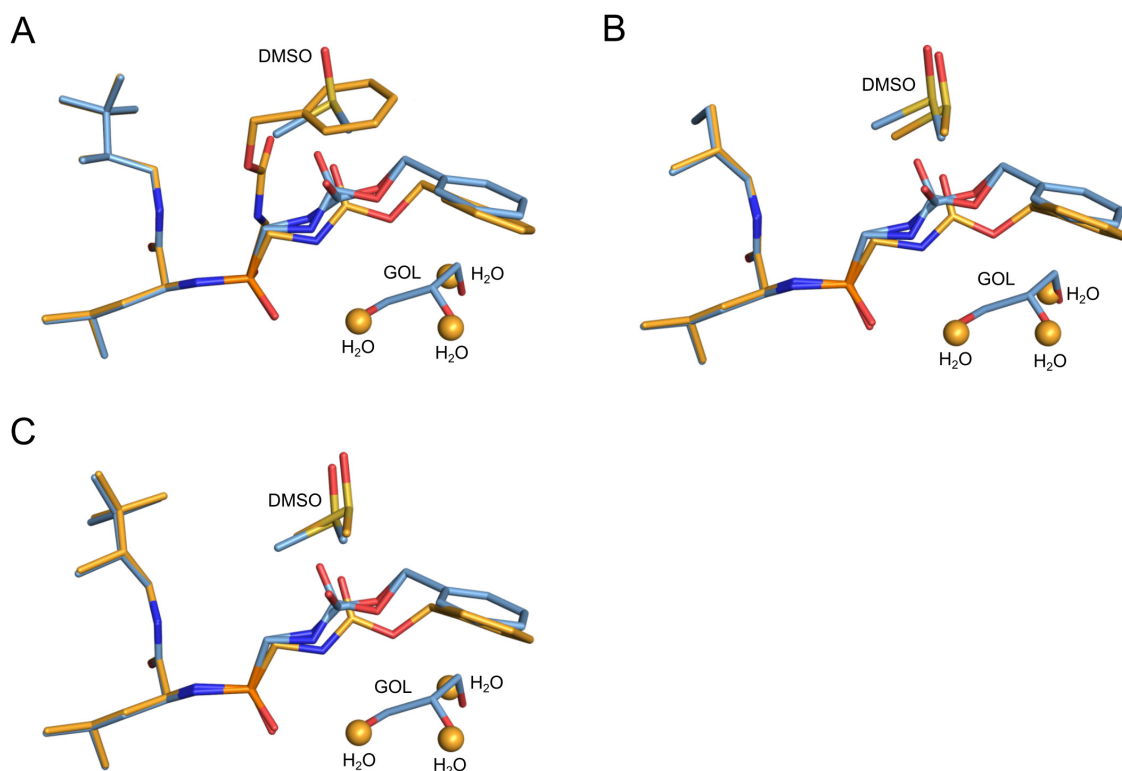
Figure S4.4. Superimposition of crystal structures of **(A)** TLN-**3**$_{GOL}$ (*blue*) and TLN-**3**$_{PEG400}$ (*orange*), **(B)** TLN-**5**$_{GOL}$ (*blue*) and TLN-**5**$_{MPD}$ (*orange*), and **(C)** TLN-**6**$_{GOL}$ (*blue*) and TLN-**6**$_{MPD}$ (*orange*). Heteroatoms are color-coded. The three spheres superimposed on the glycerol (GOL) hydroxyl groups represent water molecules bound to the TLN S$_1$ pocket in the superimposed glycerol-free crystal structure. As a result of the reduced steric requirement of the three water molecules compared to the glycerol molecule, the carboxybenzyl portion of the ligands in the glycerol-free crystal structures is buried more deeply in the S$_1$ pocket of TLN and adopts only a single, fully occupied conformation. Nevertheless, the binding mode of the leucine and P$_2$' portions of the ligands in the glycerol-containing and glycerol-free crystal structures are completely identical and consequently exert similar influences on the water network establishment in the TLN S$_2$' pocket.

Table S4.1. Data collection and refinement statistics for the glycerol-free crystal structures TLN-2_{PEG400}, TLN-5_{MPD} and TLN-6_{MPD}.[a]

	Complex (PDB code)		
	TLN-3_{PEG400} (5L8P)	TLN-5_{MPD} (5L41)	TLN-6_{MPD} (5L3U)
(A) Data collection and processing			
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters			
a, b, c (Å)	92.6, 92.6, 130.4	92.6, 92.6, 130.8	92.8, 92.8, 131.1
Matthews coefficient (Å^3/Da)[b]	2.3	2.3	2.4
Solvent content (%)[b]	47	48	48
(B) Diffraction data			
Resolution range (Å)	50.00–1.29 (1.37–1.29)	50.00–1.25 (1.33–1.25)	50.00–1.23 (1.30–1.23)
Unique reflections	83387 (13109)	91628 (14590)	96812 (15297)
$R(I)_{sym}$ (%)	6.2 (49.4)	8.0 (49.4)	5.5 (47.9)
Wilson B factor (Å^2)	9.6	9.3	9.4
Completeness (%)	99.7 (98.5)	100.0 (99.8)	99.8 (99.0)
Redundancy	12.7 (12.5)	12.8 (12.4)	9.6 (9.7)
$<I/\sigma(I)>$	27.6 (5.5)	19.8 (4.9)	26.8 (4.7)
(C) Refinement			
Resolution range (Å)	40.10–1.29	43.65–1.25	38.42–1.23
Reflections used in refinement (work/free)	79216/4170	87046/4582	91963/4841
Final R value for all reflections (work/free) (%)	10.7/13.3	10.7/13.0	10.4/12.4
Protein residues	316	316	316
Calcium/zinc ions	4/1	4/1	4/1
Inhibitor atoms	31	29	31
Water molecules	423	433	437
RMSD from ideality			
Bond lengths (Å)	0.009	0.011	0.009
Bond angles (°)	1.2	1.1	1.0
Ramachandran plot[c]			
Residues in most favored regions (%)	88.1	89.6	88.5
Residues in additionally allowed regions (%)	10.7	9.3	10.4
Residues in generously allowed regions (%)	0.7	0.7	0.7
Residues in disallowed regions (%)[d]	0.4	0.4	0.4
Mean B factor (Å^2)[e]			
Protein non-hydrogen atoms	10.7	10.6	10.7
Protein Cα atoms	9.5	9.4	9.4
Inhibitor	9.9	11.0	11.5
Water molecules	26.5	26.3	26.5

[a]Numbers in parentheses represent the values of the highest resolution shells. [b]Matthews coefficient and solvent content were calculated with the program *Matthews_coef* from the *CCP*4 suite [238]. [c]Ramachandran plots were calculated with *PROCHECK* [196]. [d]The Ramachandran outlier is Thr26 as described in literature [197]. [e]Mean B factors were calculated with *MOLEMAN* [198].

Chapter 4

4.11.3 Crystallographically determined and normalized *B* factors

Table S4.2 (related to Figure 4.5). Crystallographically determined (non-normalized) *B* factors of *W1–W15* of TLN-**1** to TLN-**6**.

Water molecule ID	*B* factor (Å²)					
	TLN-**1**	TLN-**2**	TLN-**3**	TLN-**4**	TLN-**5**	TLN-**6**
W1	11	10	10	17	10	10
W2	14	13	13	19	12	10
W3	41	25[a]	45	44	42	25
W4	-	-	32	-	49	20
W5	27	27	25	-	-	35
W6	13	13	13	23	19	17
W7	14	14	14	-	35	22
W8	31	29	19	-	47	36
W9	34	36	26	-	43	42
W10	32	34	47	49	-	49
W11	21	21	22	32	29	27
W12	38	48	41	-	-	-
W13	49	46	29	-	39	40
W14	-	-	50	-	-	-
W15	-	-	47	-	-	-

[a] both conformations *A* and *B* of *W3* in TLN-**2** have similar *B* factors.

Table S4.3 (related to Figure 4.5). Normalized *B* factors of *W1–W15* of TLN-**1** to TLN-**6**.

Water molecule ID	normalized *B* factor (Å²)					
	TLN-**1**	TLN-**2**	TLN-**3**	TLN-**4**	TLN-**5**	TLN-**6**
W1	13	14	13	17	13	14
W2	17	17	17	19	16	14
W3	49	34[a]	59	44	57	35
W4	-	-	42	-	68	28
W5	32	37	34	-	-	48
W6	15	18	17	23	27	23
W7	17	20	18	-	48	31
W8	37	39	26	-	65	50
W9	41	49	35	-	59	58
W10	39	46	62	49	-	67
W11	26	28	30	32	40	38
W12	45	65	55	-	-	-
W13	59	62	38	-	53	55
W14	-	-	66	-	-	-
W15	-	-	63	-	-	-

[a] both conformations *A* and *B* of *W3* in TLN-**2** have similar *B* factors.

4.11.4 ITC measurements

Buffer ionization reaction during TLN–ligand complex formation

All ITC measurements were performed in HEPES buffer. Measurements conducted earlier with ligands exhibiting a similar parent scaffold (Figure 4.1A of the main text) revealed the uptake of one proton by Glu143 during the protein–ligand complex formation [64, 96]. The proton is transferred from a buffer molecule, resulting in the ionization of the buffer molecule and a heat of ionization ΔH°_{ion} in addition to the heat signal ΔH°_{bind} from the actual binding event. Thus, the heat signal observed by the ITC experiment (ΔH°_{obs}) is the sum of both, ΔH°_{ion} and ΔH°_{bind}. Since the magnitude of the heat signal ΔH°_{ion} depends on the applied buffer molecule [150], it would be unreasonable to discuss the magnitude of ΔH°_{obs} (and thus also the calculated value of $-T\Delta S^{\circ}$) on an *absolute* scale without prior correction of ΔH°_{ion}. A correction of ΔH°_{ion} is possible by measuring the binding reaction in different buffers showing different heats of ionization [39, 96]. However, in a congeneric series, the discussion of *relative* differences of ΔH°_{obs}, all with an identical contribution of ΔH°_{ion}, is possible and probably even more accurate than the comparison of calculated buffer corrected values due to error propagation. This scenario is given in the current ligand series under investigation [64], as the congeneric ligands exhibit only small changes in the aliphatic portion sticking into the solvent, and consequently exert similar heats of ionization ΔH°_{ion}. As only relative differences between the thermodynamic parameters of **1**–**6** are discussed and not their absolute values, ITC measurements were only performed in one buffer and the heat of ionization was not corrected.

Remeasurement of ligand 1 to guarantee high comparability of the thermodynamic parameters

Ligand **1** was already thermodynamically characterized in the study conducted earlier [104]. However, the protein batch was completely used up for the ITC measurements and therefore it was necessary to perform the ITC measurements of the current study with TLN from a different protein batch. Furthermore, the ITC measurement scheme was optimized in order to improve measurement precision. Moreover, instead of applying *Origin7* from OriginLab for titration curve analysis, we decided to work with the programs *NITPIC* [155] and *SEDPHAT* [177] for a more automated and therefore potentially less user-biased integration and curve fitting procedure. Since ITC measurements can be very sensitive to changing measurement conditions and it is highly recommended to keep measurement conditions similar to achieve high comparability [96], we decided to measure ligand **1** again together with ligands **2**–**6**.

Table S4.4 (related to Figure 4.8). Thermodynamic parameters measured for ligands **1**−**6** by ITC. All measurements were performed in triplicate (a–c) out of which the mean values and the standard deviations were calculated. All parameters are given as rounded numbers. Measurement curves are displayed in Figure S4.5 in order to proof the high data quality of the measurements [96, 257].

Measurement	n	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
TLN-**1**a	1.015	0.368	-36.7	-23.0	-13.7
TLN-**1**b	0.988	0.338	-36.9	-22.8	-14.1
TLN-**1**c	0.952	0.347	-36.9	-23.1	-13.7
	0.985±0.032	**0.351±0.015**	**-36.8±0.1**	**-23.0±0.2**	**-13.8±0.2**
TLN-**2**a	0.990	0.232	-37.9	-21.6	-16.2
TLN-**2**b	0.996	0.242	-37.8	-21.1	-16.6
TLN-**2**c	1.009	0.227	-37.9	-21.1	-16.8
	0.998±0.010	**0.233±0.008**	**-37.9±0.1**	**-21.3±0.3**	**-16.6±0.3**
TLN-**3**a	0.957	0.149	-39.0	-21.1	-17.9
TLN-**3**b	1.029	0.218	-38.0	-20.8	-17.2
TLN-**3**c	0.940	0.189	-38.4	-21.8	-16.6
	0.975±0.047	**0.185±0.035**	**-38.5±0.5**	**-21.2±0.5**	**-17.2±0.6**
TLN-**4**a	0.987	0.409	-36.5	-18.4	-18.0
TLN-**4**b	1.058	0.376	-36.7	-18.4	-18.3
TLN-**4**c	1.021	0.378	-36.7	-18.6	-18.0
	1.022±0.036	**0.388±0.019**	**-36.6±0.1**	**-18.5±0.1**	**-18.1±0.1**
TLN-**5**a	0.947	0.285	-37.4	-16.6	-20.8
TLN-**5**b	0.982	0.291	-37.3	-17.0	-20.3
TLN-**5**c	0.939	0.354	-36.8	-17.0	-19.8
	0.956±0.023	**0.310±0.038**	**-37.2±0.3**	**-16.9±0.3**	**-20.3±0.5**
TLN-**6**a	0.948	0.368	-36.7	-15.3	-21.4
TLN-**6**b	1.005	0.314	-37.1	-15.0	-22.1
TLN-**6**c	0.962	0.380	-36.6	-14.7	-21.9
	0.972±0.030	**0.354±0.035**	**-36.8±0.3**	**-15.0±0.3**	**-21.8±0.4**
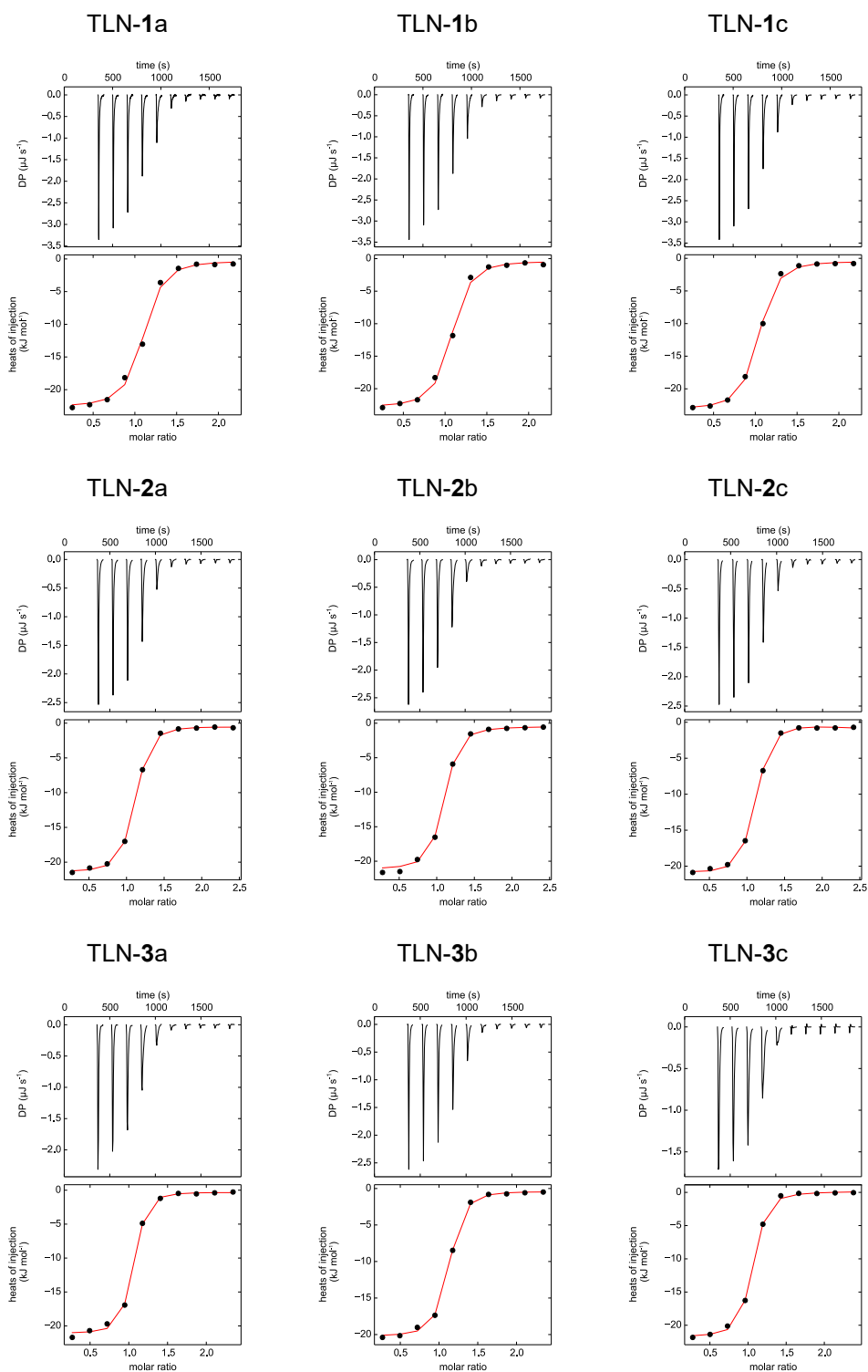
Figure S4.5. ITC measurement data of TLN-**1** to TLN-**6** each performed in triplicate (a–c). The upper panel shows the extracted peaks of the titration curve. The lower panel displays the values of the integrated peaks (*black* dots) as automatically performed by the *NITPIC* algorithm, and, fitted to them, the 1:1 model binding isotherms (*red* curve) as created with *SEDPHAT*.
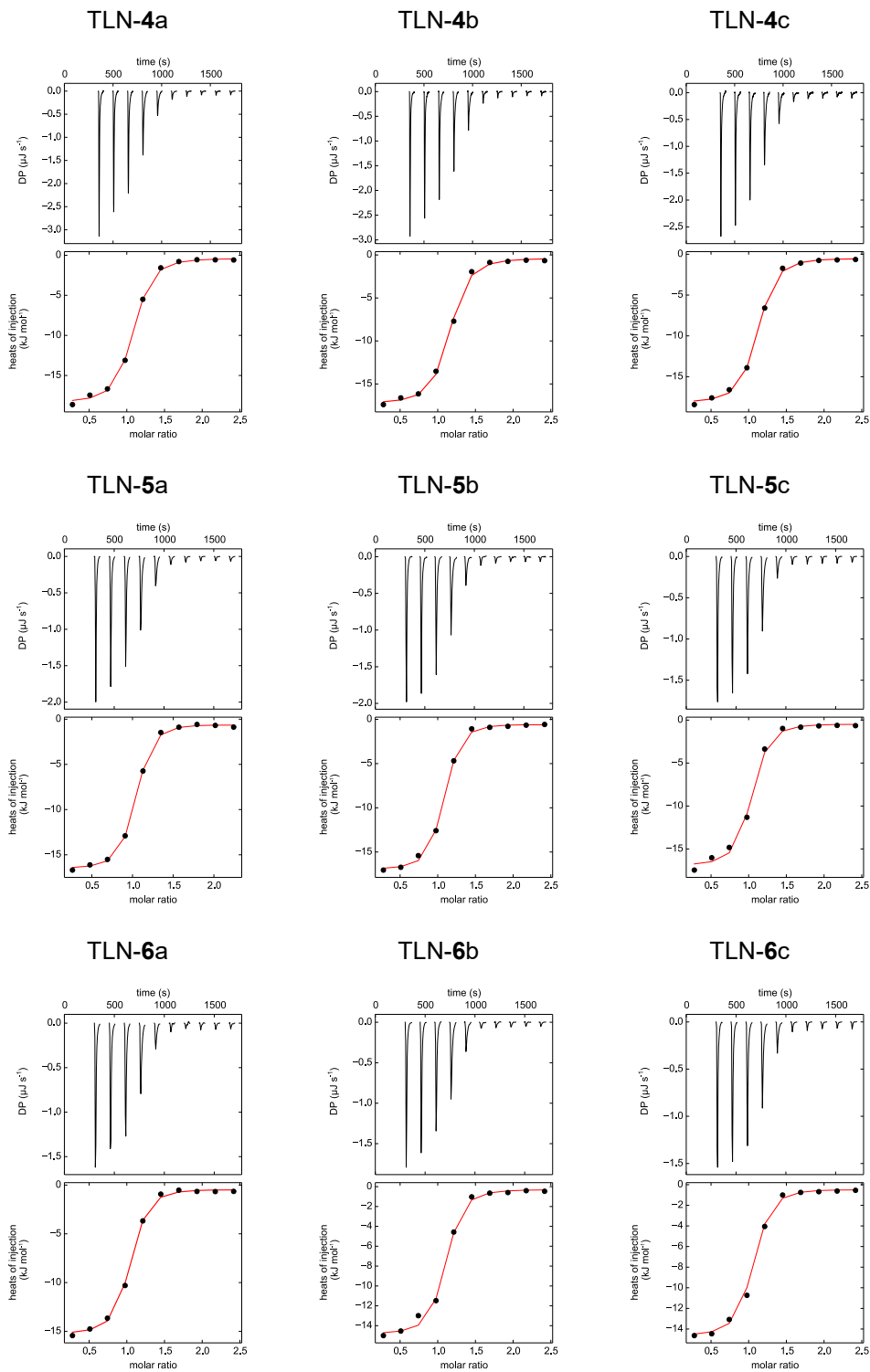
Figure S4.5. (continued)

4.11.5 ITC measurements with the addition of glycerol

For all four tested ligands (**2**, **3**, **5**, and **6**), $\Delta H°$ increases and $-T\Delta S°$ decreases with increasing concentration of glycerol (Figure S4.6). Almost complete compensation of $\Delta H°$ and $-T\Delta S°$ is observed, resulting in virtually unchanged $\Delta G°$. The relative differences of the thermodynamic parameters between the ligands remain constant. For **3** compared to **2**, identical values of $\Delta H°$ and more favorable values of $-T\Delta S°$ and $\Delta G°$ are observed for all measured glycerol concentrations.
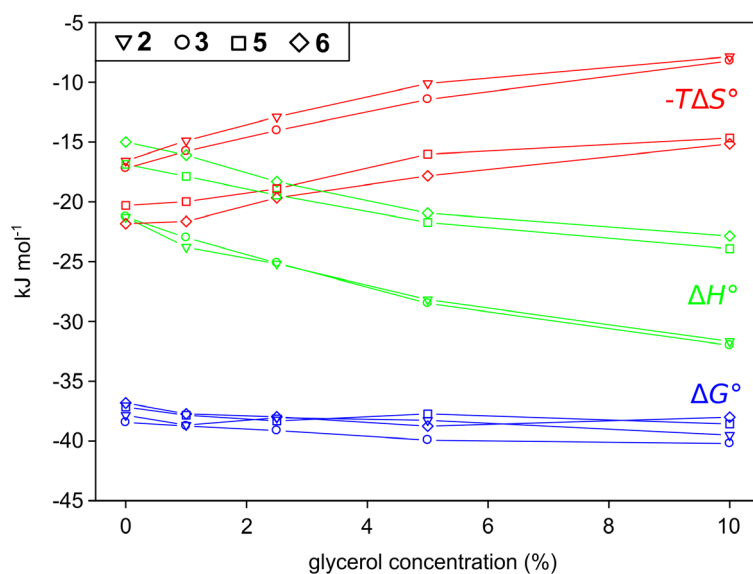
Figure S4.6. ITC measurements of **2**, **3**, **5** and **6** binding to TLN with buffers containing different glycerol concentrations of 0%, 1%, 2.5%, 5% and 10% (v/v). The symbols indicate the different ligands, the colors indicate the thermodynamic parameters $\Delta H°$ (*green*), $-T\Delta S°$ (*red*) and $\Delta G°$ (*blue*). For sake of clarity, standard deviation bars are not shown. All measurements were performed in triplicate. Data values with standard deviations are listen in Table S4.5.

Table S4.5. ITC measurement results of ligands **2**, **3**, **5** and **6** measured with the addition of different concentrations of glycerol. All measurements were performed in triplicate (a–c), from which the mean values and the standard deviations were calculated. All parameters are given as rounded numbers.

Measurement (glycerol conc.)	n	K_d (µM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
TLN-**2**a (1.0%)	1.023	0.188	-38.4	-23.8	-14.6
TLN-**2**b (1.0%)	1.015	0.174	-38.6	-23.6	-15.0
TLN-**2**c (1.0%)	1.080	0.143	-39.1	-24.0	-15.1
1.039±0.035	**0.168±0.023**	**-38.7±0.3**	**-23.8±0.2**	**-14.9±0.3**
TLN-2a (2.5%)	1.087	0.188	-38.4	-25.2	-13.2
TLN-2b (2.5%)	1.060	0.281	-37.4	-25.1	-12.3
TLN-2c (2.5%)	1.034	0.186	-38.4	-25.3	-13.2
1.060±0.027	**0.218±0.054**	**-38.1±0.6**	**-25.2±0.1**	**-12.9±0.5**
TLN-2a (5.0%)	1.037	0.216	-38.0	-28.5	-9.5
TLN-2b (5.0%)	1.014	0.186	-38.4	-28.4	-10.0
TLN-2c (5.0%)	1.025	0.190	-38.4	-27.6	-10.8
1.025±0.012	**0.197±0.016**	**-38.3±0.2**	**-28.2±0.5**	**-10.1±0.6**
TLN-2a (10.0%)	1.023	0.146	-39.0	-31.1	-7.9
TLN-2b (10.0%)	1.066	0.118	-39.5	-31.4	-8.1
TLN-2c (10.0%)	1.104	0.098	-40.0	-32.5	-7.5
1.064±0.041	**0.120±0.024**	**-39.5±0.5**	**-31.7±0.7**	**-7.9±0.3**
TLN-**3**a (1.0%)	1.003	0.154	-38.9	-22.7	-16.2
TLN-**3**b (1.0%)	0.981	0.158	-38.8	-22.9	-16.0
TLN-**3**c (1.0%)	0.963	0.175	-38.6	-23.4	-15.1
0.982±0.020	**0.162±0.011**	**-38.8±0.2**	**-23.0±0.4**	**-15.8±0.5**
TLN-3a (2.5%)	1.012	0.165	-38.7	-24.9	-13.9
TLN-3b (2.5%)	0.990	0.140	-39.1	-25.5	-13.6
TLN-3c (2.5%)	0.991	0.116	-39.6	-25.0	-14.6
0.998±0.012	**0.140±0.025**	**-39.1±0.4**	**-25.1±0.3**	**-14.0±0.5**
TLN-3a (5.0%)	0.965	0.110	-39.7	-28.8	-10.9
TLN-3b (5.0%)	0.962	0.101	-39.9	-28.4	-11.5
TLN-3c (5.0%)	0.973	0.092	-40.2	-28.3	-11.9
0.967±0.006	**0.101±0.009**	**-39.9±0.2**	**-28.5±0.2**	**-11.4±0.5**
TLN-3a (10.0%)	0.976	0.106	-39.8	-31.7	-8.2
TLN-3b (10.0%)	0.914	0.085	-40.4	-31.7	-8.7
TLN-3c (10.0%)	0.927	0.081	-40.5	-32.7	-7.8
0.939±0.033	**0.091±0.013**	**-40.2±0.3**	**-32.0±0.6**	**-8.2±0.5**
TLN-**5**a (1.0%)	1.058	0.189	-38.4	-17.9	-20.5
TLN-**5**b (1.0%)	1.015	0.289	-37.3	-17.7	-19.6
TLN-**5**c (1.0%)	1.053	0.234	-37.8	-18.0	-19.8
1.042±0.024	**0.237±0.050**	**-37.8±0.5**	**-17.9±0.2**	**-20.0±0.4**
TLN-5a (2.5%)	1.042	0.195	-38.3	-19.6	-18.7
TLN-5b (2.5%)	1.016	0.212	-38.1	-18.9	-19.2
TLN-5c (2.5%)	1.054	0.171	-38.6	-19.8	-18.8
1.037±0.019	**0.193±0.021**	**-38.3±0.3**	**-19.4±0.5**	**-18.9±0.3**
TLN-5a (5.0%)	0.936	0.288	-37.3	-22.1	-15.3
TLN-5b (5.0%)	0.933	0.210	-38.1	-21.8	-16.3
TLN-5c (5.0%)	0.991	0.241	-37.8	-21.3	-16.5
0.953±0.033	**0.246±0.039**	**-37.7±0.4**	**-21.7±0.4**	**-16.0±0.7**
TLN-5a (10.0%)	1.067	0.152	-38.9	-24.1	-14.8
TLN-5b (10.0%)	1.063	0.186	-38.4	-24.1	-14.3
TLN-5c (10.0%)	1.042	0.185	-38.4	-23.6	-14.9
1.057±0.013	**0.174±0.019**	**-38.6±0.3**	**-23.9±0.3**	**-14.7±0.3**
TLN-**6**a (1.0%)	0.994	0.231	-37.9	-16.2	-21.7
TLN-**6**b (1.0%)	0.986	0.270	-37.5	-16.0	-21.5
TLN-**6**c (1.0%)	0.996	0.236	-37.8	-16.1	-21.7
0.992±0.005	**0.245±0.021**	**-37.7±0.2**	**-16.1±0.1**	**-21.6±0.2**
TLN-6a (2.5%)	1.020	0.212	-38.1	-18.1	-20.0
TLN-6b (2.5%)	0.996	0.260	-37.6	-18.7	-18.8
TLN-6c (2.5%)	0.981	0.198	-38.3	-18.1	-20.2
0.999±0.020	**0.223±0.032**	**-38.0±0.4**	**-18.3±0.4**	**-19.7±0.7**
TLN-6a (5.0%)	0.955	0.190	-38.4	-20.9	-17.4
TLN-6b (5.0%)	0.939	0.142	-39.1	-21.2	-17.9
TLN-6c (5.0%)	0.953	0.155	-38.9	-20.7	-18.1
0.949±0.009	**0.162±0.025**	**-38.4±0.4**	**-20.9±0.2**	**-17.8±0.4**
TLN-6a (10.0%)	1.065	0.199	-38.2	-23.5	-14.7
TLN-6b (10.0%)	1.018	0.233	-37.9	-22.2	-15.6
TLN-6c (10.0%)	1.012	0.226	-37.9	-22.8	-15.1
1.032±0.018	**0.219±0.018**	**-38.0±0.2**	**-22.9±0.6**	**-15.1±0.5**

4.11.6 ITC measurements with the addition of DMSO
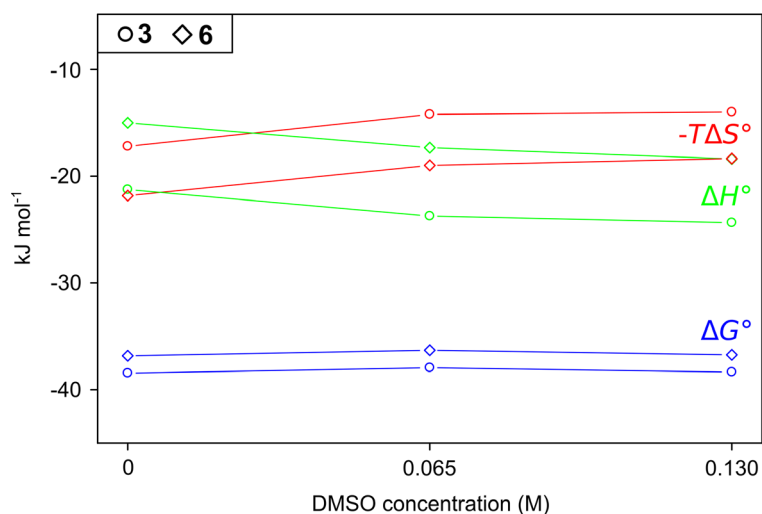
Figure S4.7. ITC measurements of **3** and **6** with the addition of DMSO. The measured DMSO concentrations of 0.065 M and 0.130 M correspond to the molar concentration of glycerol in the solutions with 5% and 10% glycerol (Figure S4.6). Data values are listed in Table S4.6.

Table S4.6. Thermodynamic parameters measured for ligands **3** and **6** by ITC with the addition of DMSO. All measurements were performed in duplicate (a+b) out of which the mean values and the standard deviations were calculated.

Measurement	n	K_d (µM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
TLN-**3**a (0.065 M)	0.907	0.199	-38.2	-24.3	-13.9
TLN-**3**b (0.065 M)	1.027	0.255	-37.6	-23.2	-14.5
	0.967±0.085	**0.227±0.040**	**-37.9±0.4**	**-23.7±0.8**	**-14.2±0.4**
TLN-**3**a (0.130 M)	1.014	0.208	-38.1	-24.3	-13.8
TLN-**3**b (0.130 M)	1.003	0.175	-38.6	-24.4	-14.2
	1.009±0.008	**0.192±0.024**	**-38.4±0.3**	**-24.4±0.1**	**-14.0±0.2**
TLN-**6**a (0.065 M)	1.007	0.459	-36.2	-17.2	-18.9
TLN-**6**b (0.065 M)	1.004	0.414	-36.4	-17.4	-19.0
	1.006±0.002	**0.437±0.032**	**-36.3±0.2**	**-17.3±0.1**	**-19.0±0.0**
TLN-**6**a (0.130 M)	0.992	0.3402518	-36.92	-18.36	-18.56
TLN-**6**b (0.130 M)	1.005	0.3935458	-36.56	-18.42	-18.14
	0.999±0.009	**0.367±0.038**	**-36.7±0.3**	**-18.4±0.0**	**-18.4±0.3**

4.11.7 SPR measurements

Table S4.7 (related to Figure 4.9). SPR measurement results of TLN-ligand complexes TLN-**1**, TLN-**2**, TLN-**3**, TLN-**5** and TLN-**6**. Kinetic analysis was performed by global analysis of single-cycle measurements performed in triplicate (Figure S4.8).

TLN complex	k_{on} (M^{-1}s^{-1})	SE (k_{on})[a]	k_{off} (s^{-1})	SE (k_{off})[a]	K_d (M)	Chi^2 (RU^2)
TLN-**1**	1.14×10^5	6.50×10^2	8.40×10^{-2}	2.70×10^{-4}	7.38×10^{-7}	0.058
TLN-**2**	7.08×10^4	4.10×10^2	5.95×10^{-2}	2.10×10^{-4}	8.41×10^{-7}	0.104
TLN-**3**	3.10×10^4	2.10×10^2	3.78×10^{-2}	1.50×10^{-4}	1.22×10^{-6}	0.257
TLN-**5**	8.58×10^4	7.00×10^2	6.53×10^{-2}	3.20×10^{-4}	7.61×10^{-7}	0.185
TLN-**6**	4.82×10^4	3.30×10^2	6.84×10^{-2}	2.70×10^{-4}	1.42×10^{-6}	0.152

[a] SE = standard error

Table S4.8 (related to Figure 4.9). Kinetic binding parameters of TLN-ligand complexes TLN-**1**, TLN-**2**, TLN-**3**, TLN-**5** and TLN-**6** as determined by individual analysis of the SPR measurements performed in triplicate, from which the mean values and the standard deviations were calculated.

TLN complex	k_{on} (M^{-1}s^{-1})	SD (k_{on})	k_{off} (s^{-1})	SD (k_{off})	K_d (M)	SD (K_d)
TLN-**1**	1.10×10^5	8.06×10^3	7.94×10^{-2}	3.97×10^{-3}	7.24×10^{-7}	8.38×10^{-8}
TLN-**2**	8.25×10^4	3.32×10^4	5.97×10^{-2}	3.07×10^{-3}	8.30×10^{-7}	4.03×10^{-7}
TLN-**3**	2.69×10^4	7.08×10^3	3.87×10^{-2}	5.09×10^{-3}	1.55×10^{-6}	5.95×10^{-7}
TLN-**5**	9.55×10^4	3.32×10^4	6.64×10^{-2}	1.49×10^{-2}	7.55×10^{-7}	2.82×10^{-7}
TLN-**6**	4.00×10^4	1.16×10^4	7.00×10^{-2}	5.98×10^{-3}	1.86×10^{-6}	5.56×10^{-7}
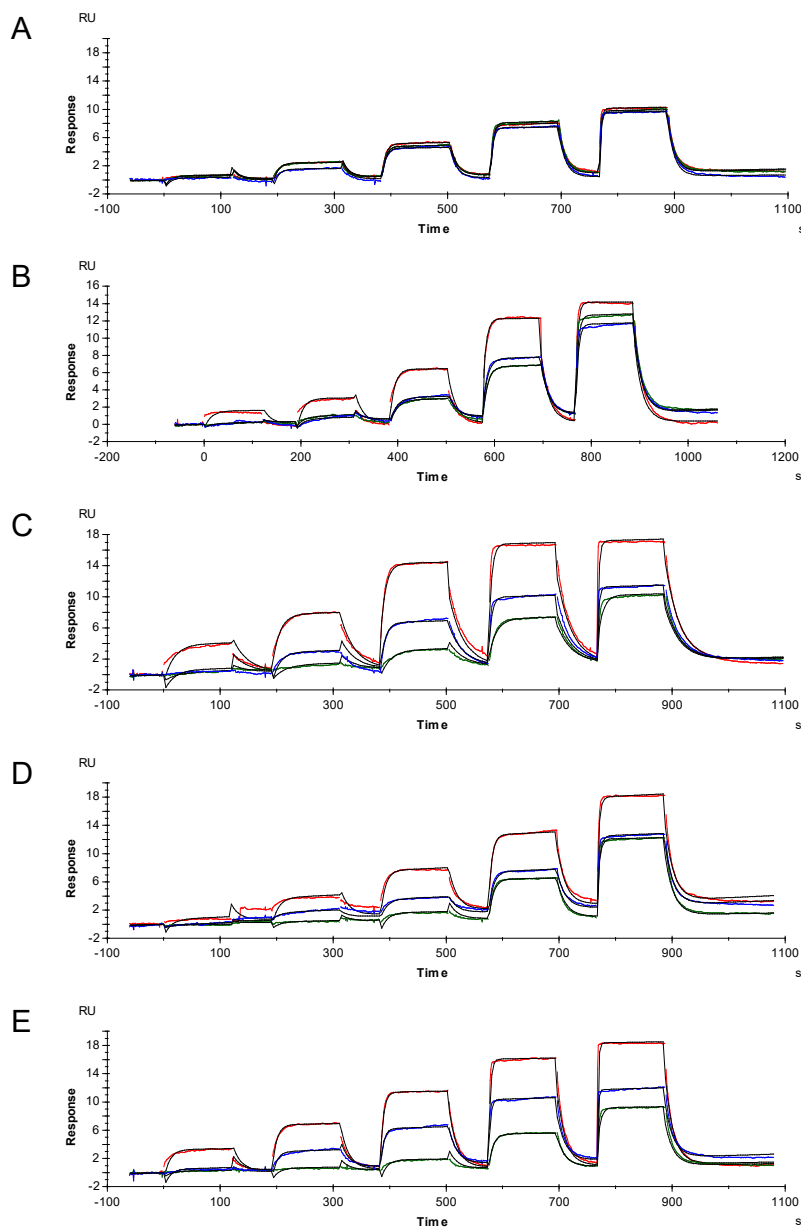
Figure S4.8 (related to Figure 4.9). SPR sensorgrams showing three separate single cycle kinetics runs in five concentrations of each **(A)** TLN-**1**, **(B)** TLN-**2**, **(C)** TLN-**3**, **(D)** TLN-**5**, and **(E)** TLN-**6**. The analyte concentration ranges spanned (A) 156–2500 nM, (B) 25–2500 nM, (C) 156–10000 nM, (D) 25–2500 nM, (E)156–10000 nM. The colouring of sensorgrams represents the concentrations used in each cycle (A) 156, 312, 625, 1250, 2500 nM (*red*, *blue*, *green*), (B) and (D) 25–2500 nM (156, 312, 625, 1250, 2500 nM (*red*), 25, 74, 222, 667, 2000 nM (*blue*, *green*)), (C) and (E) 156, 312, 625, 1250, 2500 nM (*blue*, *green*), 625, 1250, 2500, 5000, 10000 nM (*red*). Global analysis of the triplicate data was performed (*black*) to account for experimental differences between runs such as variation in capture level (800–1200 RU), drift and used analyte concentrations. The concentration range used spanned from <0.2×K_d to >3.0×K_d for all compounds (K_d values determined by SPR are presented in Table S4.7).

4.11.8 Calculation of buried SASAs
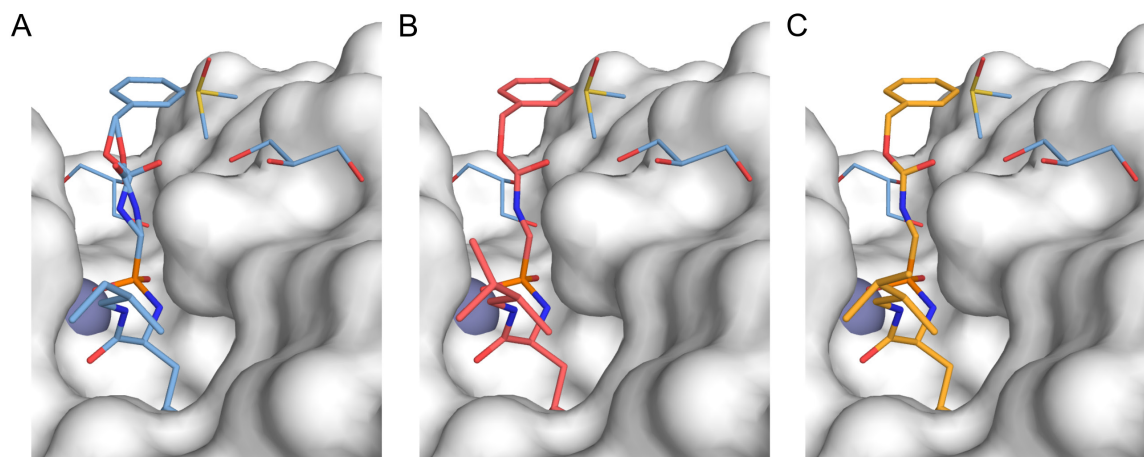
Figure S4.9. Crystal structure TLN-**5** with the missing terminal methyl group of the P$_2'$ portion (panel A, bound ligand in *blue*), and the two modeled conformations **5**A_m (panel B, *red*) and **5**B_m (panel C, *yellow*) of the missing methyl group used for the calculations of the SASAs.

Table S4.9 (related to Figure 4.10). Calculated total and buried solvent accessible surface areas (SASAs) of ligands **1**–**6** in complex with TLN.

Ligand		SASA (Å^2)	
		total	buried
1		702	512
2	conformation *A*	726	536
	conformation *B*	726	535
3		740	548
4		705	566
5	missing P$_2'$ methyl group (Figure S4.9A)	704	520
	methyl group modeled in conformation A_m (Figure S4.9B)	729	539
	methyl group modeled in conformation B_m (Figure S4.9C)	729	539
6		748	555

4.11.9 Comparison MD simulations and crystal structures of TLN-2 and TLN-5
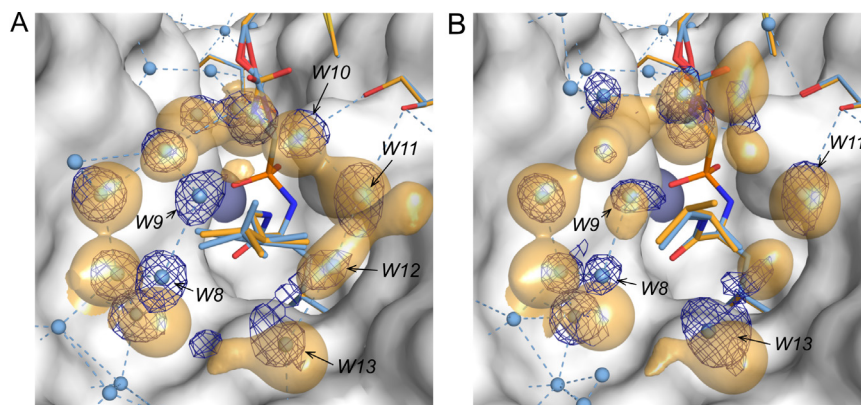
Figure S4.10 (related to Figure 4.11). Prediction of positions of water molecules as calculated by MD simulations in comparison to the crystallographically observed electron densities and refined crystal structure models of **(A)** TLN-**2** and **(B)** TLN-**5**. The modeled coordinates of ligand, glycerol and DMSO molecules used in the MD simulations are displayed as *yellow* stick models with color-coded heteroatoms. The *yellow*, semitransparently contoured regions show computed areas in the first solvation layer of the P$_2$′ groups with an occupancy probability by a water molecule of at least 48%. The crystallographically determined binding modes of ligand and additive molecules are superimposed as *blue* stick models with color-coded heteroatoms. Water molecule positions determined in the crystal structures are displayed as *blue* spheres, and the F_o–F_c omit electron density is displayed as *dark blue* mesh at a contour level of 3σ for the water molecules positioned in the first solvation layer of the P$_2$′ groups. H-bond distances are indicated as *blue* dotted lines. Positions of water molecules, which are discussed in the main text, are labeled with identifiers according to Figure 4.4. The solvent excluded surface of TLN is shown in *white*.

Chapter 5

Elucidating the Origin of Long Residence Time Binding for Inhibitors of the Metalloprotease Thermolysin

Jonathan Cramer,[1,*] Stefan G. Krimmer,[1,*] Veronica Fridh,[1,*] Tobias Wulsdorf,[1] Robert Karlsson,[2] Andreas Heine,[1] Gerhard Klebe[1]

*These authors contributed equally

[1]Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

[2]GE Healthcare Bio-Sciences AB, SE-751 84 Uppsala, Sweden

5.1   Abstract

Kinetic parameters of protein–ligand interactions are progressively acknowledged as valuable information for rational drug discovery. However, a targeted optimization of binding kinetics is not easy to achieve and further systematic studies are necessary to increase the understanding about molecular mechanisms involved. We determined association and dissociation rate constants for 17 inhibitors of the metalloprotease thermolysin by surface plasmon resonance spectroscopy and correlated kinetic data with high-resolution crystal structures in complex with the protein. From the structure–kinetics relationship, we conclude that the strength of interaction with Asn112 correlates with the rate-limiting step of dissociation. This residue is located at the beginning of a β-strand motif that lines the binding cleft and is commonly believed to align a substrate for catalysis. A reduced mobility of the Asn112 sidechain owing to an enhanced engagement in charge-assisted hydrogen bonds prevents the conformational adjustment associated with ligand release and transformation of the enzyme to its open state. This hypothesis is supported by kinetic data of ZF^{P}LA, a known pseudopeptidic inhibitor of thermolysin, which blocks the conformational transition of Asn112. Interference with this retrograde induced-fit mechanism results in variation of the residence time of thermolysin inhibitors by a factor of 74,000. The high conservation of this structural motif within the M4 and M13 metalloprotease families underpins the importance of this feature and has significant implications for drug discovery.

5.2 Introduction

Ultimately, the goal of a drug design project is the development of a clinical candidate that is efficacious *in vivo*. The main surrogate parameter for efficacy in the early stages of development is the affinity of a prospective lead candidate, assessed in an assay working under a thermodynamic equilibrium regime. In many cases, however, equilibrium binding affinity translates poorly to *in vivo* activity. Thus additional parameters such as the binding kinetics of protein–ligand interactions can add further insights and are increasingly appreciated as a valuable indicator for therapeutic efficacy [169, 225, 258]. Recent studies have shown that an increased receptor residence time ($\tau = 1/k_d$) translates into an enhanced pharmacological effect also in an *in vivo* setting [224, 259]. Furthermore, in a study of adenosine A_{2A} receptor antagonists, drug response was more strongly correlated to variations in residence time than affinity under equilibrium conditions (K_D) [260]. Yet, the intentional modification of the kinetic parameters of a given ligand is not easy to achieve, because the underlying molecular determinants are poorly understood and thus a reliable structure–kinetics relationship is difficult to establish. Only in rare cases it is possible to rationalize binding kinetic parameters with respect to variations in the structure of an inhibitor in a mechanistically coherent way [247, 261, 262]. Generally, high molecular weight, pronounced lipophilicity (clogP >5) and enhanced molecular flexibility (number of rotatable bonds >5) are accepted to cause slow dissociation rates [263]. However, since these parameters are likely to correlate with unfavorable pharmacokinetic properties, they are not necessarily the most preferred ones to be considered and improved in a targeted drug optimization process. Computational analyses of buried water clusters in the binding pocket of the adenosine A_{2A} receptor in complex with a series of antagonists have shown that the number and position of thermodynamically unfavorable water molecules correlates with decreasing residence time of the inhibitor [264]. Computational analyses have suggested that desolvation or resolvation of the binding site prior to or after the binding event determine the rate-limiting steps for association or dissociation, respectively [265, 266]. Especially the hydrophobic shielding of buried polar atoms in the protein binding site seems to prevent rapid rehydration and thus prolongs dissociation of a bound ligand [267].

Fluctuations in protein conformation can impose a major influence on the binding kinetic profile of an inhibitor [268, 269]. For enoyl-ACP reductase it has been demonstrated that upon binding of the inhibitor PT70 a loop region that is disordered in the uncomplexed state transforms into an ordered helical structure. This reduces the association and dissociation rates of the binding event significantly [270]. The kinetic selectivity of the marketed drug thiotropium toward the M3 isoform of the muscarinic G protein coupled receptor is believed to originate from differences in the dynamic behavior of the ECL2 region relative to the M2 isoform [271]. Another common motif is the occlusion of the ligand binding site by a

hydrophobic lid [272, 273]. The ability of a ligand to stabilize this closed conformation is reflected in its kinetic profile.

In order to further trace putative mechanisms responsible for extended residence times on the molecular level, we used the metalloprotease thermolysin (TLN) from *B. thermoproteolyticus* as a model system. This enzyme has been successfully consulted in the past as a target surrogate in the development of inhibitors for angiotensin converting enzyme and neutral endopeptidase [233, 274, 275]. Its chemical robustness, excellent crystallographic properties and ready access in large quantities renders TLN an ideal test system to study details of protein–ligand interactions.

The conformational dynamics of TLN and other zinc proteases upon ligand binding have been subject of several studies [212, 276, 277]. It has been hypothesized that TLN, analogous to several other members of the bacterial neutral protease family, undergoes a kind of "hinge-bending" motion about its central α-helix and folds upon the bound substrate. However, experimental evidence for this hypothesis remained elusive for a long time. Approaches that rely on a static crystallographic model have notoriously been hampered by the fact that a dipeptide (Val-Lys), cleaved off from the C-terminus of the protease by autoproteolysis, blocks access to the active site and stabilizes the closed conformation [212]. Exchange of the central zinc ion against other more heavy metal ions finally revealed an alternate conformation that was later confirmed as the putative *apo* form of TLN [276, 277]. Although the overall difference between the two conformers is small (rmsd = 0.62 Å based on 316 C_α atoms) [276], larger conformational changes were observed for the side chains of Met120, Glu143, and Leu144.

In this study, the kinetic binding parameters of 17 closely related TLN inhibitors are analyzed with respect to slight chemical modifications in their partly solvent–exposed P_2' substituents (Figure 5.1A). The ligand series consists of three subsets that can be distinguished according to their C-terminal functionality. Either a charged carboxy group (subset *a*, blue), a neutral carboxamide (subset *b*, green) or an apolar aliphatic substituent (subset *c*, red) were considered. The kinetic profiles are correlated with high-resolution crystal structures of the compounds in complex with TLN (mean resolution across all 17 structures: 1.30±0.15 Å).

5.3 Results

5.3.1 SPR data collection

The kinetic binding parameters of **1–17** were determined by SPR using an assay based on capture of biotinylated TLN as outlined in the experimental section. The sensorgrams
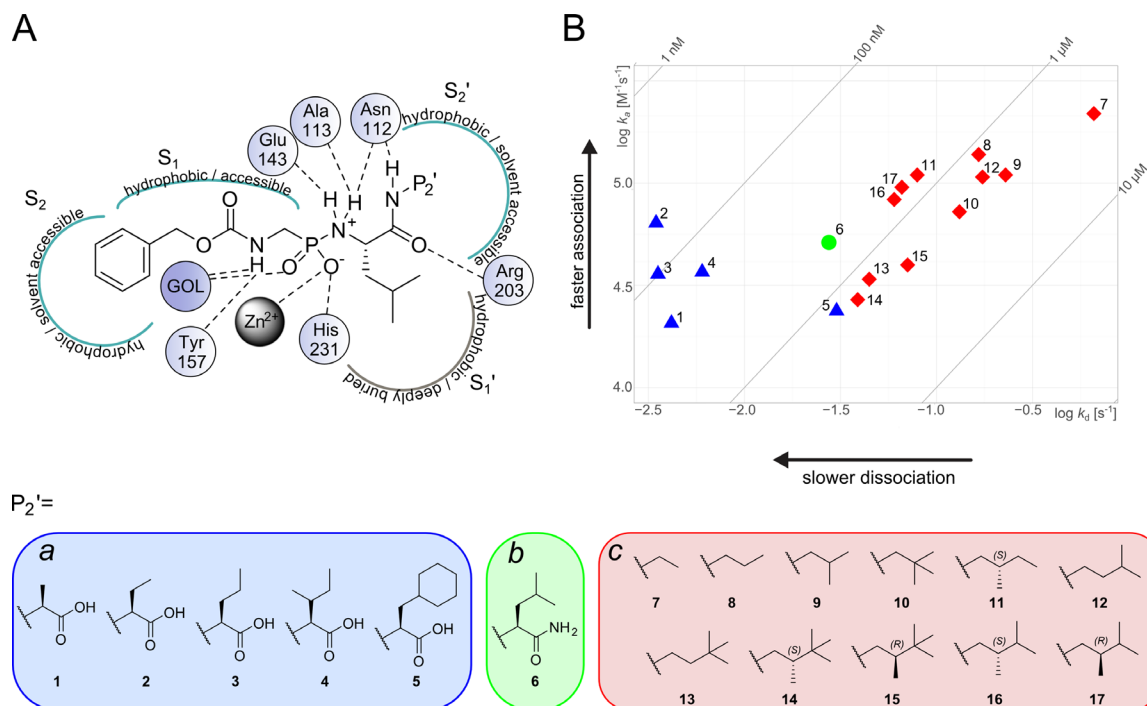
Figure 5.1. Binding kinetic data. **(A)** Schematic representation of the relevant interactions between phosphonopeptide ligands **1**–**17** and TLN. The ligands are categorized in three groups according to the chemical composition of their P_2' groups. Subset *a* (blue): P_2'-carboxy (**1**–**5**), subset *b* (green): P_2'-carboxamide (**6**), subset *c* (red): P_2'-aliphatic (**7**–**17**). **(B)** Kinetic map (log k_a vs. log k_d) of **1**–**17**. Subset *a*: blue triangles, subset *b*: green circle, subset *c*: red diamonds. The diagonal lines indicate regions of equal affinity.

showed clear differences in binding characteristics within and/or between the three investigated series (Supporting Information Figure S5.1).

5.3.2   Kinetic data analysis

Figure 5.1B shows a map of the binding kinetic parameters (log k_a vs. log k_d) of ligands **1**–**17**. The diagonal lines indicate lines of equal affinity (equilibrium dissociation constant $K_D = k_d/k_a$). Overall, the affinity varies within the ligand series from 40 nM to 3.3 μM. The association rate constants across all subsets scatter within one order of magnitude about a mean value of 10^4 M^{-1}s^{-1}. In contrast, the dissociation rate constants display a larger variance between 6.15×10^{-1} s^{-1} and 2.77×10^{-3} s^{-1}. It is apparent that ligands bearing a C-terminal carboxy group (**1**–**5**) show a slower dissociation rate than ligands with only an aliphatic substituent at the P_2' position lacking the acid group (**7**–**17**). Compound **6** with a terminal carboxamide falls in-between both series with a k_d of 4.46×10^{-2} s^{-1}. Within subset *a*, **5** shows a significantly faster dissociation from the TLN binding pocket, with k_d increased by a factor of ten relative to the other ligands of the subset. In subset *c*, **13** and **14** depart toward slower

dissociation and **7** displays the fastest association and dissociation rates compared to the other compounds. The rapid approach to steady state for **7**, resulting in a squared shaped pulse (Supporting Information Figure S5.1), is governed mostly by its fast dissociation rate ($k_d = 6.15 \times 10^{-1}$ s^{-1}). A small subset of the inhibitors (**14–17**) was already analyzed previously in a detailed study including structural, thermodynamic and kinetic properties [66]. While differences in k_a were observed, we mainly focused on an interpretation of dissociation kinetics for this study.

In order to assess the dependence of the binding kinetic data of the phosphonamidates on electrostatic properties, the rate constants of a representative member of each subset were determined at three different ionic strength conditions (Supporting Information Table S5.2). In addition to the standard buffer conditions, NaCl was added in a concentration of 100 mM or 500 mM, respectively. A significant deviation could not be detected in the investigated range of ionic strength.

5.3.3 X-ray crystallography

The high-resolution crystal structures of several ligands in complex with TLN have been reported by us in previous studies [42, 64, 104]. The structures of **5** and **6** were additionally determined in the context of this study (crystallographic information in Supporting Information Table S5.3). In Figure 5.2A, **1** is depicted as a representative example for the binding mode topology of **1–17** within the binding cleft of TLN. The carbamate nitrogen interacts with Tyr157 and a glycerol (GOL) molecule that is picked up from the buffer and bound to the S$_2$ pocket. One oxygen atom of the phosphonamidate moiety binds to the catalytic zinc ion in a monodentate fashion and is further positioned within hydrogen-bond distance of His231. The other oxygen atom interacts with the GOL molecule. The phosphonamidate nitrogen is likely protonated under the applied experimental conditions [237] and interacts with Glu143, Ala113 and Asn112. The hydrophobic leucine side chain of the inhibitors is deeply buried in the S$_1$' pocket. Additionally, a hydrogen-bonding interaction of Arg203 to the leucine carbonyl oxygen is established. The polar groups of the C-terminal ligand portion are in contact with the side chain carboxamide of Asn112 (*vide infra*). The variable hydrophobic P$_2$' moieties are positioned in the shallow, solvent exposed S$_2$' pocket. Figure 5.2B shows a superposition of the binding mode of all investigated ligands. Apart from the chemically deviating P$_2$' substituents, the structures show virtually perfect overlap of the parent inhibitor scaffold.
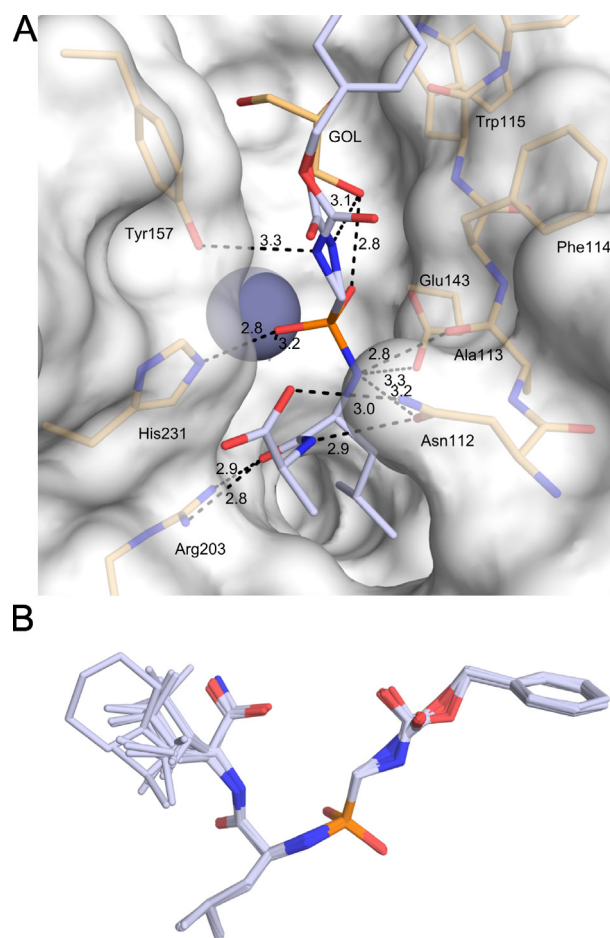
Figure 5.2. General binding mode of phosphonopeptide ligands to TLN. **(A) 1** is shown in blue, protein residues adjacent to the inhibitor are shown in orange. The dashed lines indicate hydrogen bonds between ligand and protein with the distance between the respective heavy atoms annotated in Å. The gray-blue sphere represents the catalytic zinc ion. The solvent-excluded surface of TLN is shown in white. **(B)** Superposition of the binding mode of inhibitors **1**–**17**.

5.4 Discussion

5.4.1 Interactions of inhibitors 1–17 with Asn112

Since all interactions that do not involve the P_2' substituent are geometrically highly conserved between **1**–**17**, the explanation for their deviating k_d values must originate from the interactions of their deviating C-terminal portion. A comparison of the interactions across the three subsets shows that the ligands engage in a deviating hydrogen-bonding pattern with respect to the side chain of Asn112. As shown in Figure 5.3, all inhibitors involve the carboxamide group of this residue into multiple hydrogen bonds. In all cases, the phosphonamidate nitrogen, which is likely protonated under the applied conditions,
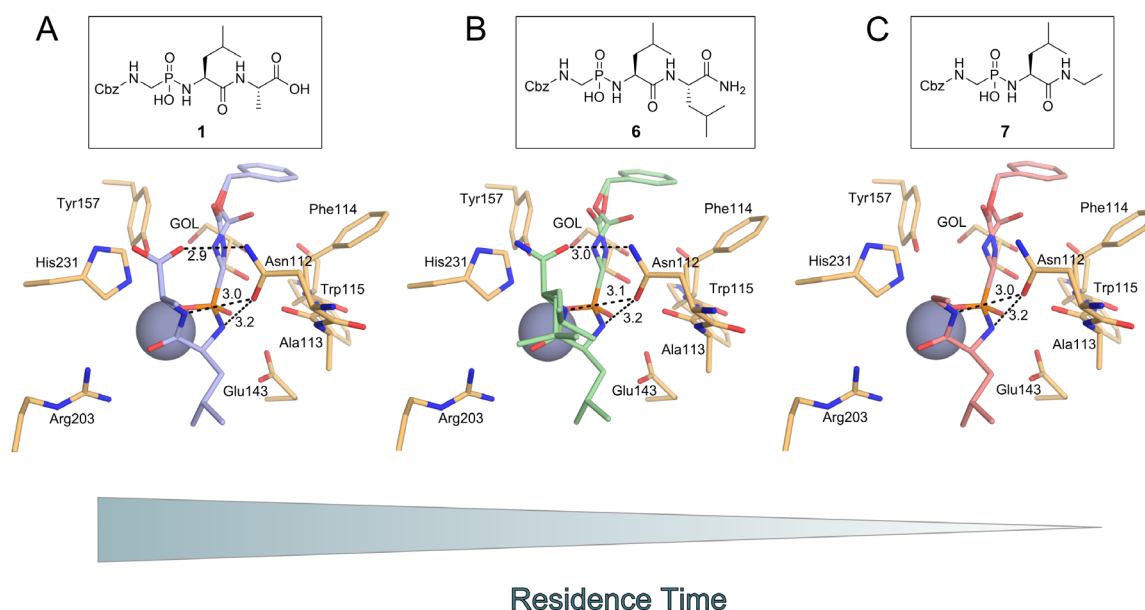
Figure 5.3. Interaction pattern of a representative ligand of each group with Asn112. Residues of the active site are shown in orange. For reasons of clarity, only hydrogen bonds formed by Asn112 are shown. **(A)** The C-terminal amino acid moiety of P_2'-carboxy inhibitor **1** (blue) engages in a bidentate charge-assisted hydrogen bond to Asn112. **(B)** The general binding mode of P_2'-amide inhibitor **6** (green) resembles that of P_2'-carboxy compounds. **(C)** P_2'-aliphatic inhibitor **7** (red) solely interacts with the carbonyl oxygen of Asn112. Bidentate coordination of the C-terminal residue is lost.

interacts with the side chain carbonyl oxygen of Asn112. The secondary amide nitrogen of the P_2' group also forms a hydrogen bond to Asn112. Inhibitors with an aliphatic P_2' group do not show any further interactions with Asn112 (Figure 5.3C). The addition of a C-terminal carboxy group allows further fixation of Asn112 due to a chelating charged-assisted hydrogen bond to the side chain carboxamide nitrogen (Figure 5.3A). The binding mode of **6**, which features a C-terminal primary carboxamide, resembles that of the analogous carboxy compounds (Figure 5.3B). In this case, however, the hydrogen bond between the terminal carboxamide and Asn112 is weakened due to the loss of a formal charge, which potentially induces an electrostatic enhancement in case of **1–5**.

5.4.2 The conformational dynamics of Asn112 and their implications for the kinetics of TLN

Previous discussions concerning the dynamic properties of TLN mainly focused on conformational transformations of Met120, Glu143 and Leu144 [276, 277]. So far, a contribution of Asn112 has not been taken into consideration. Our structural and kinetic results, however, suggest major involvement of this residue in the dissociation step. We

therefore propose the following release mechanism: The rotation of Asn112 is necessary to enable the "hinge-bending" motion, which is required to expel a tightly bound substrate or ligand from the binding pocket. In Figure 5.4A, the structure of TLN in complex with the peptidic inhibitor IMPI (insect metalloprotease inhibitor) from *Galleria mellonella* is shown. IMPI is cleaved by the protease, yet, the hydrolysis product is not released but forms an inhibitory complex with the enzyme [278]. Therefore, its binding mode supposedly resembles a geometry adopted by the substrate after peptide-bond cleavage. The C-terminal
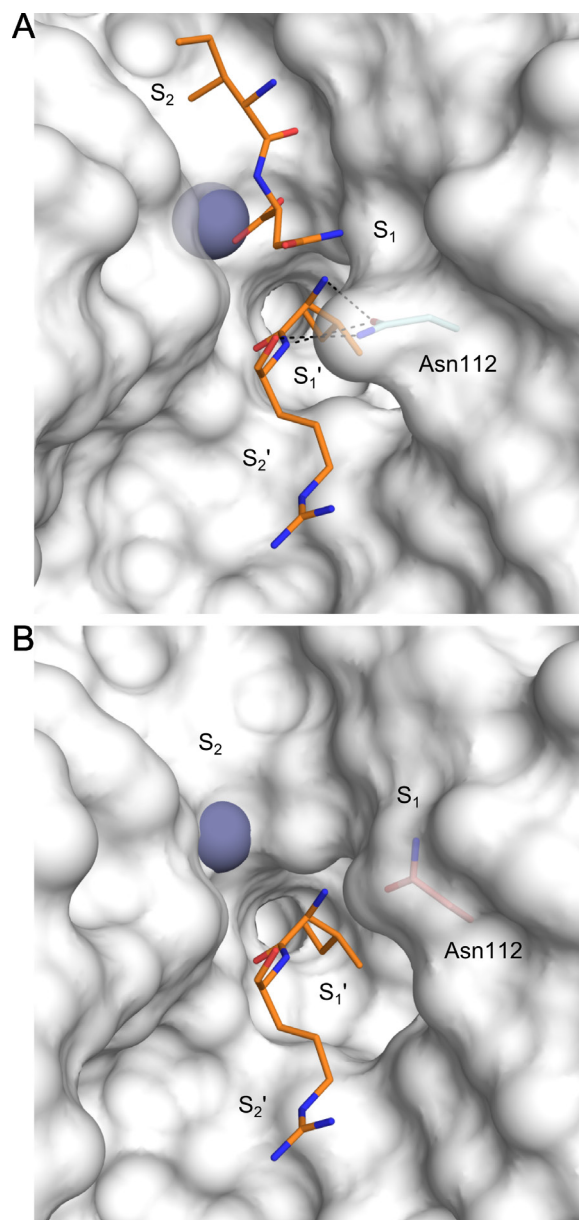
Figure 5.4. Binding mode of the cleaved peptidic product fragments of IMPI (pdb entry: 3ssb). **(A)** Only the four central residues of the "reactive-site" loop are shown in orange. Asn112 in the closed conformation is shown in cyan. **(B)** Superposition of IMPI residues 57–58 with the open conformation of TLN. Asn112 is shown in red.

fragment of IMPI engages in a contact to Asn112 resembling the hydrogen-bonding pattern also observed for our P_2'-carboxy ligands. Obviously, this interaction stabilizes Asn112 in a position, in which its carboxamide side chain forms a lid over the S_1' pocket, augmenting the barrier for the release of the bound substrate product. If the latter complex is superimposed with the open conformation of TLN (Figure 5.4B), it becomes apparent that the outward rotation of Asn112 is involved in an opening of the S_1' specificity pocket, which subsequently allows the release of the cleaved peptide fragment. Our hypothesis supports the assumption that the dissociation of a cleaved substrate follows a sequential process. As a first step, the weakly bound N-terminal peptide (occupying the less specific binding pockets on the unprimed side) has to dissociate from the binding pocket. Once the S_1 pocket is abandoned (and conceivably rehydrated), Asn112 is able to rotate toward this pocket, thus giving access to the well-defined S_1' pocket. This retrograde induced-fit mechanism allows the dissociation of the deeply buried C-terminal product fragment from the hydrophobic cavity. In contrast to this sequential process, a transition-state mimetic inhibitor would have to dissociate from the binding pockets in a concerted way. A pronounced fixation of Asn112 through a strong interaction to the P_2' substituent imposes an additional energetic barrier for the induced-fit step along the escape trajectory [269].

A structural alignment of proteases from the M4 family (TLN-like zinc metalloproteases) reveals that Asn112 is highly conserved across the members of this group (97% consensus in Pfam-A family) [279]. Within the less related M13 family (Neprilysin family), multiple proteases featuring this structural motif are found (79% consensus in Pfam-A family). Figure 5.5 shows the crucial part of the secondary structure alignment of TLN with other members of the M4 and M13 family. Important representatives such as the human endothelin-converting-enzyme (ECE-1) or the human neutral endopeptidase (Neprilysin), associated
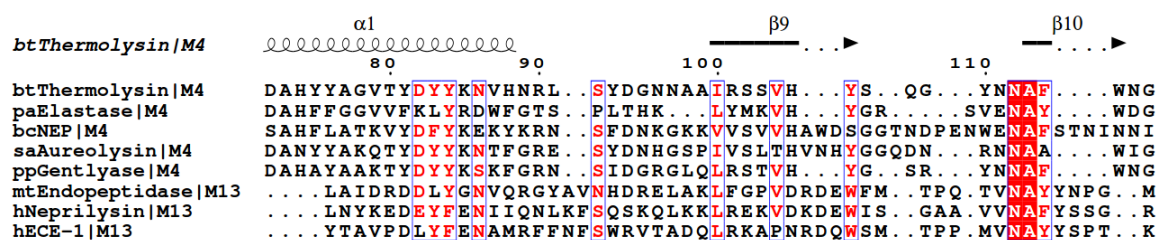
Figure 5.5. A section of the secondary structure alignment of TLN and representative members of the M4 and M13 protease families (alignment of the full sequence in SI). Partially or functionally conserved residues are indicated by red letters, full conservation is highlighted with white letters on red background. Lower case letters indicate the source organism of the respective protein (bt = *B. thermoproteolyticus*, pa = *P. aeruginosa*, bc = *B. cereus*, sa = *S. aureus*, pp = *P. polymyxa*, mt = *M. tuberculosis*, h = human). For the secondary structure alignment the protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm) was used [280]. The results were displayed using the ESPript server (http://espript.ibcp.fr) [281].

with diseases like breast cancer, Alzheimer's disease and heart failure belong to these families [282–284]. Furthermore, it has been shown that N112X mutants (X = A, H, K, R) of TLN lose their catalytic activity [285]. Only the variants N112D and N112E partly retain their proteolytic activity (64% and 19% respectively) [286]. This underpins the crucial role of Asn112 for the substrate turnover of TLN-like proteases.

5.4.3 Structure-kinetics relationship

Based on these considerations, it becomes obvious that the strength of the ligand-to-Asn112 interaction is reflected by the kinetic binding profile. Especially the dissociation rate constant correlates with this property. Members of subset *a* engage in strong, charged-assisted hydrogen bonds between the ligand carboxy and Asn112 carboxamide group. Consequently, these inhibitors show the longest residence times across the series. For **6**, the chelating hydrogen bond is attenuated due to the loss of one formal negative charge. As a result, its k_d corresponds to a three-fold faster release (compared to the mean of **1**–**5**). Total abandonment of this interaction, as given for ligands from subset *c*, results in a thirteen-fold accelerated dissociation (factor between the mean k_d value of **1**–**5** and **7**–**17**).

In a study by Bartlett and Marlowe [287], the binding kinetics of a related inhibitor exhibiting a benzyl group in the P_1 position (ZF^{P}LA, **18**, Figure 5.6) was determined in a photometric inhibition assay. The latter ligand is the most potent TLN inhibitor described in literature, with an affinity of K_D = 68 pM. This high potency originates from a considerably lower dissociation rate resulting in the remarkable residence time of 168 days (k_d = 6.8×10^{-8} s^{-1}). Although inherent differences between the applied photometric method and our SPR analysis to record the kinetic constants hampers a straightforward quantitative comparison, the data of **1**, examined in both studies, suggests that the deviations between both approaches fall maximally into the range of one order of magnitude (for a detailed comparison see Supporting Information Figure S5.2). The exchange of the substitution pattern at P_1 from hydrogen to (*R*)-benzyl resulted in a 74,000-fold deceleration of the dissociation rate constant, based on the photometric inhibition assay (Figure 5.6). This finding is in full agreement with our proposed release mechanism and explains the remarkable kinetic properties of **18**. The binding mode shows some significant deviations from **1**–**17** that all contain a Gly^P-motif in the P_1 position. In the crystal structure of **18**, the Phe^P substituent is positioned in van der Waals distance to Phe114 (Figure 5.7A). The mutual positioning of the two aromatic portions blocks the outward rotation of Asn112. However, as described, this movement is an important prerequisite for the simultaneous release of the bound ligand, while translocating TLN to its open conformation. This hypothesis is

substantiated by a structural superposition of the open conformation of the *apo*-protein and the inhibited complex of **18** (Figure 5.7B).
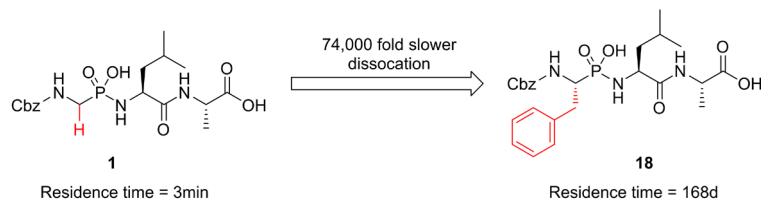
Figure 5.6. Influence of a P$_1$ substitution from hydrogen to *(R)*-benzyl on residence time (residence time for **18** was calculated from k_d determined by Bartlett and Marlowe) [287].
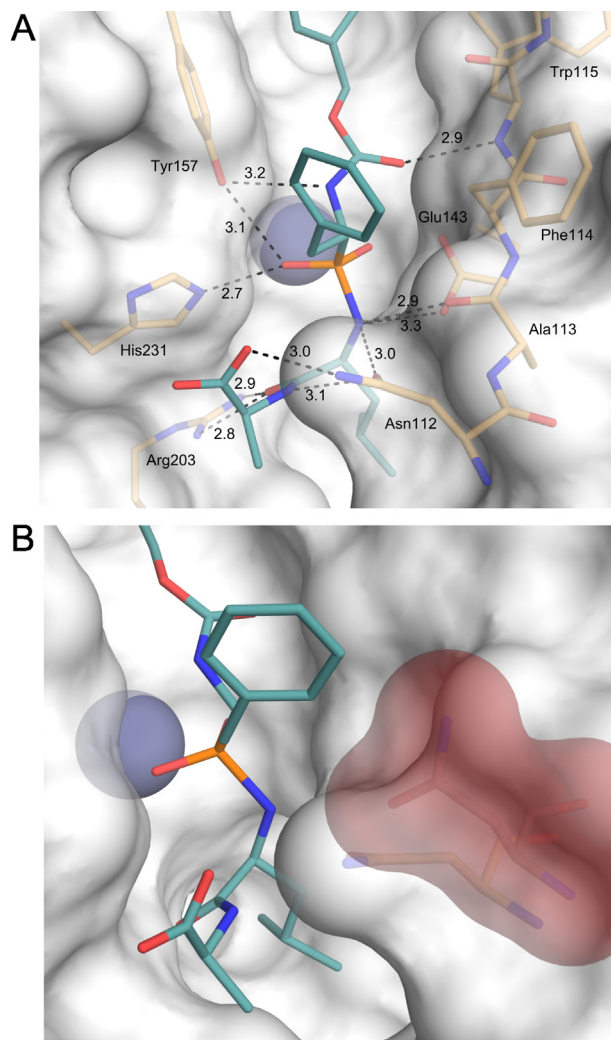
Figure 5.7. Binding mode of ZF^{P}LA (**18**; PDB entry 4TMN). **(A)** The Phe^P substituent is within van der Waals distance to Phe114 and shields the top of the binding cleft. **(B)** Superposition of Asn112 from TLN in the open conformation (red). In the closed conformation, the polar side chain of Asn112 is in close contact to the apolar Phe^P portion of ligand **18**.

Within each subset, binding kinetic parameters are fine-modulated and a more complex picture determines the correlation. As discussed in our previous contribution, the quality and completeness of the surface water network wrapping around ligand-exposed parts of the newly formed protein–ligand complex can take impact on k_d in the range of one order of magnitude [66]. Within subset *a*, **5** is clearly separated from **1**–**4** in the direction of faster dissociation. Since **5** exhibits the largest molecular weight within the series, this result opposes the generally accepted correlation of increasing molecular mass with slower dissociation rate [263]. A closer inspection of the binding modes of **1**–**5** reveals differences in the surface water network adjacent to the partly exposed P$_2$' side chain of the ligands, which provide an explanation (Figure 5.8). Due to the large size of the hydrophobic cyclohexyl
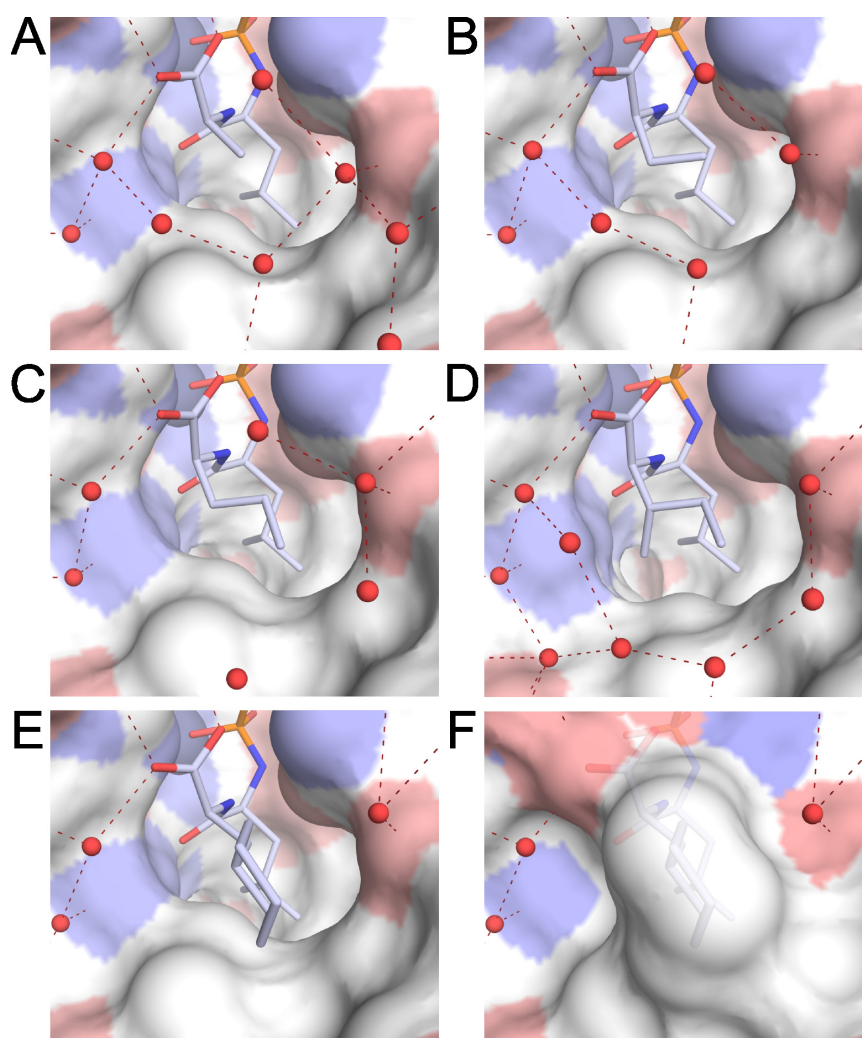
Figure 5.8. Water arrangement around the P$_2$'-portion of ligands. **(A) 1**, **(B) 2**, **(C) 3**, **(D) 4** and **(E) 5** in complex with TLN. **(F)** Representation of the solvent-excluded surface generated from the complex of **5** with TLN. The surface is color-coded according to atom types (red = oxygen, blue = nitrogen, white = carbon). Water molecules are shown as red spheres with red dashes indicating hydrogen bond interactions.

substituent of **5**, the formation of the surface water network remains incomplete and is more strongly perturbed compared to those formed with **1–4**. This destabilization of the surface water network supposedly facilitates dissociation of the bound ligand due to the loss of solvent–solvent interactions that have to be disrupted upon ligand dissociation.

It has been demonstrated for several protein–ligand systems that electrostatic attraction or repulsion can play an important role in binding kinetics [288–291]. While electrostatic factors are generally believed to predominantly influence association through enhanced probability for diffusional collisions or conformational preorganization of the binding partners, minor impact on the dissociation process cannot be neglected. In order to elucidate possible long-range electrostatic effects between the differently charged molecules, we repeated our kinetic experiments in buffers of varying ionic strength. In our experiments, none of the kinetic constants for the TLN-phosphonamidate interactions showed a significant dependence on the ionic strength, regardless of the presence or absence of any nominal charge on the ligand's C-terminus (Supporting Information Table S5.2). We are therefore confident that long-range electrostatic effects do not matter significantly in the current example and the differences in binding kinetics can be traced back to the interaction with Asn112.

5.5   Conclusion

The kinetic characterization of 17 congeneric inhibitors of the metalloprotease TLN by SPR and their correlation with structural data reveals insights into the molecular mechanism of binding. We suggest that the involvement of the Asn112 side chain in hydrogen bonds with the bound ligand provokes a prolonged residence time of the respective inhibitor. Asn112 is crucially involved in the motion transforming the protease from its opened to closed state. We therefore propose that a reinforced interaction to Asn112 by the bound ligand restricts the conformational flexibility of this residue. This deteriorates ligand release from the bound state as the Asn112 side chain has to move out of space to allow access to the S_1' pocket. The activation of this retrograde induced-fit step can be further impeded by augmenting the steric hindrance of the Asn112 movement, as observed in the complex of the highly potent TLN inhibitor ZF^{P}LA (**18**). This ligand exhibits a drastic decrease in the dissociation rate constant by a factor of 74,000 compared to its analog **1** [287]. Additionally, a fine-tuning of the binding kinetic properties results from modulations of the P_2' portion of the ligands that are linked to variations in the structure of the surface water network which wraps around the S_2' pocket. Perturbance of these networks of hydrogen-bonded water molecules may lead to an enhanced dissociation rate.

These results highlight the determining role of changes in protein conformation on the binding kinetics of bound ligands. Therefore, a profound understanding of protein flexibility, experienced during the ligand binding event, is of vital importance for a rational design and optimization of ligand binding kinetic parameters.

5.6   Methods

5.6.1   Surface Plasmon Resonance Measurements

SPR measurements were performed on a Biacore T200 system (GE Healthcare) with analysis and sample compartment temperature set to 25 °C. The binding kinetic assay was developed as a capture assay of biotinylated TLN utilizing the Biotin CAPture Kit (GE Healthcare) combined with single cycle kinetics of the inhibitors as previously described [66]. Briefly, sensor Chip CAP was docked in the Biacore and prepared according to the manufacturer's instructions. For each day of SPR measurements, new dilutions of inhibitors from 10 mM stock solutions in DMSO (stored at –20 °C) and biotinylated TLN, were made using freshly prepared buffer. 50 mM Tris pH 8.0, 100 mM NaSCN, 2 mM $CaCl_2$, 2% DMSO was used as sample and running buffer. Runs included three start-up cycles and each analysis cycle contained 4 steps: (1) injection of Biotin CAPture reagent (streptavidin modified with a deoxyriboologonucletide that hybridizes with the complimentary oligonucleotide present on the Sensor Chip CAP) in both reference and active flow cells for 300 s and at 2 μl/min resulting in response levels of around 3000 RU, (2) injection of 150–200 μg mL^{-1} TLN-biotin at 10 μL/min for 90–180 s in active flow cell only resulting in capture levels of 800–1200 RU (a capture stabilization time of 300 s was applied for some runs), (3) injection of inhibitor in increasing concentration over reference and active flow cell using a single cycle kinetics procedure with five 120 s injections at 30 μL min^{-1} and a 240 s dissociation time after the final injection, and (4) injection of standard (6 M guanidine-HCl, 0.25 M NaOH) and additional (30% acetonitrile in 0.25 M NaOH) regeneration solutions to remove the Biotin Cap Reagent, TLN-biotin and any bound inhibitor from both flow cells. Inhibitors were analyzed in duplicate or triplicate measurements using concentrations in the range of 25–2000 nM for **2**, **4**; 25–2500 nM for **16**, **17**; 156–2500 nM for **1**, **3**, **6**, **11**; 156–10000 nM for **8**, **9**, **14**, **15**, 625–10000 nM for **5**, **10**, **12**, **13** and 1250–20000 nM for **7**. Analysis of **2**, **6** and **8** at higher ionic strength was performed in duplicate measurements at concentrations between 156–2500 nM using the above running buffer supplemented with 100 mM and 500 mM NaCl, respectively. Blank cycles defined as analysis cycles with buffer only in step (3), were performed first, last and between every inhibitor concentration series. Data were double referenced by first subtraction of reference flow cell and then subtraction of blank cycles. Fitting of data was

performed using Biacore T200 evaluation software 3.0, applying a 1:1 binding model compensating for linear drift.

5.6.2 Crystal Preparation and Soaking

Lyophilized TLN powder (Calbiochem) was dissolved in 50 μL DMSO to yield a solution with a concentration of 8 mM, followed by the addition of 50 μL of a solution containing 3.8 M CsCl and 100 mM Tris-HCl (pH 7.5). After centrifugation (3 min, 16,000 g), 1 μL of the clear solution was pipetted into the wells of a 24 well sitting drop crystallization plate (reservoir wells filled with 1 mL demineralized water). The crystal plate was sealed and crystals finished growing after 5 days at 18 °C. Ligands were soaked into the crystals at a concentration of 1 mM for 24 hrs in a buffer composed of 100 mM Tris-HCl (pH 7.5), 2 mM $CaCl_2$, 5% DMSO and subsequently flash-frozen in liquid nitrogen in a buffer composed of 10 mM Tris-HCl (pH 7.5), 10 mM $Ca(CH_3COO)_2$, 5% DMSO, 20% glycerol and 1 mM of the respective ligand.

5.6.3 Data collection, processing, structure determination and refinement

Data collection of the crystal structures TLN-**5** and TLN-**6** (Supporting Information Table S5.3) was performed at BESSY II (Helmholtz-Zentrum Berlin) at the MX-beamline 14.1. Datasets were collected on a Dectris Pilatus 6M pixel detector at a wavelength of 0.91841 Å. *XDS* [252] was used for indexing, integration and scaling of the datasets. The structures were determined by molecular replacement using *Phaser* (version 2.5.0) [213] applying the structure with the PDB code 8TLN [212] as a search model. Subsequently, model refinement (xyz coordinates, individual *B* factors, occupancies) with *Phenix.refine* (version 1.10.1-2155) [215] and model building into σA-weighted maps ($2F_o$–F_c and F_o–F_c) with *Coot* (version 0.7) [214] were performed in alternating cycles until *R* values reached convergence. A randomly chosen subset of 5% of the reflections was excluded from the refinement and used for the calculation of R_{free}. As a first refinement step, Cartesian simulated annealing was performed (default settings). *B* factors for all model atoms (except for hydrogen atoms) were refined anisotropically. Hydrogen atoms (riding model) were added to the amino acids with *Phenix.refine*. Alternative conformations of amino acid side chains and ligand moieties were assigned to the electron density if an occupancy of at least 20% was obtained after refinement. Ligand molecules were modeled with *MOE* [216] and restraints were created with *Phenix.elbow* [254] (**5**) or with the *Grade* web server [292] (**6**).

5.7 Accession Codes

Atomic coordinates and experimental details for the crystal structures of **5** and **6** (PDB codes 5LIF and 5LWD) will be released upon publication.

5.8 Author Contributions

*These authors contributed equally to this work.

5.9 Acknowledgement

Chapter 5

5.10 Supporting Information

5.10.1 SPR measurement

Experimental data from SPR measurements

Table S5.1. SPR derived kinetic data of thermolysin inhibitors **1**–**17**.

Compound	k_a [M^{-1}s^{-1}]	SD k_a [M^{-1}s^{-1}]	log k_a	k_d [s^{-1}]	SD k_d [s^{-1}]	log k_d	K_d [µM]
1	2.38×10^4	3.54×10^3	4.38	5.06×10^{-3}	1.63×10^{-4}	-2.30	0.21
2	6.88×10^4	6.46×10^2	4.84	2.77×10^{-3}	8.49×10^{-5}	-2.56	0.04
3	3.53×10^4	2.40×10^3	4.55	2.87×10^{-3}	3.04×10^{-4}	-2.54	0.08
4	4.94×10^4	6.93×10^3	4.69	7.86×10^{-3}	4.31×10^{-4}	-2.10	0.16
5	3.47×10^4	6.01×10^3	4.54	4.63×10^{-2}	1.02×10^{-3}	-1.33	1.34
6	7.88×10^4	7.07×10^3	4.91	4.46×10^{-2}	3.25×10^{-3}	-1.35	0.57
7	1.88×10^5	1.34×10^4	5.27	6.15×10^{-1}	4.95×10^{-3}	-0.21	3.28
8	1.44×10^5	9.19×10^3	5.16	1.71×10^{-1}	7.07×10^{-4}	-0.77	1.19
9	8.96×10^4	1.41×10^3	4.95	2.34×10^{-1}	5.66×10^{-3}	-0.63	2.61
10	1.01×10^5	1.41×10^3	5.00	2.08×10^{-1}	7.78×10^{-3}	-0.68	2.05
11	1.10×10^5	8.06×10^3	5.04	7.94×10^{-2}	3.97×10^{-3}	-1.10	0.72
12	1.01×10^5	1.27×10^4	5.00	1.82×10^{-1}	5.66×10^{-3}	-0.74	1.80
13	8.65×10^4	4.95×10^2	4.94	1.40×10^{-1}	1.27×10^{-2}	-0.85	1.62
14	2.69×10^4	7.08×10^3	4.43	3.87×10^{-2}	5.09×10^{-3}	-1.41	1.44
15	4.00×10^4	1.16×10^4	4.60	7.00×10^{-2}	5.98×10^{-3}	-1.15	1.75
16	8.25×10^4	3.32×10^2	4.92	5.97×10^{-2}	3.07×10^{-3}	-1.22	0.72
17	9.55×10^4	3.32×10^4	4.98	6.64×10^{-2}	1.49×10^{-2}	-1.18	0.70

Experimental data from SPR measurements at varying ionic strength conditions

Table S5.2. SPR derived kinetic data of selected inhibitors at varying ionic strength conditions.

Compound	Ionic strength[a] [mM]	k_a [M^{-1}s^{-1}]	SD k_a [M^{-1}s^{-1}]	log k_a	k_d [s^{-1}]	SD k_d [s^{-1}]	log k_d	K_d [µM]
2	134	6.88×10^4	3.46×10^3	4.84	2.77×10^{-3}	8.49×10^{-5}	-2.56	0.04
2	234	5.73×10^4	6.06×10^3	4.76	3.29×10^{-3}	2.45×10^{-4}	-2.48	0.06
2	634	4.24×10^4	1.91×10^3	4.63	2.85×10^{-3}	7.25×10^{-4}	-2.55	0.07
6	134	7.88×10^4	7.07×10^3	4.90	4.46×10^{-2}	3.25×10^{-3}	-1.35	0.57
6	234	4.78×10^4	1.19×10^4	4.89	3.38×10^{-2}	3.13×10^{-3}	-1.47	0.43
6	634	5.15×10^4	1.30×10^4	4.71	4.80×10^{-2}	1.85×10^{-3}	-1.32	0.93
8	134	1.44×10^5	9.19×10^3	5.16	1.71×10^{-1}	7.07×10^{-4}	-0.77	1.19
8	234	2.75×10^5	3.00×10^4	5.44	1.93×10^{-1}	3.75×10^{-3}	-0.71	0.70
8	634	2.49×10^5	2.40×10^3	5.40	1.89×10^{-1}	7.64×10^{-3}	-0.72	0.76

[a] Ionic strength was calculated from all ionized buffer components at pH 8.0 as $I = \sum_i c_i \cdot z_i^2$

Sensograms from SPR measurements
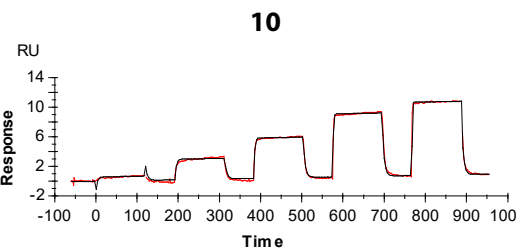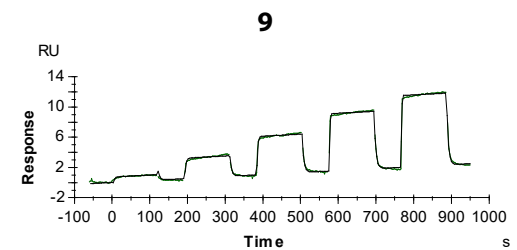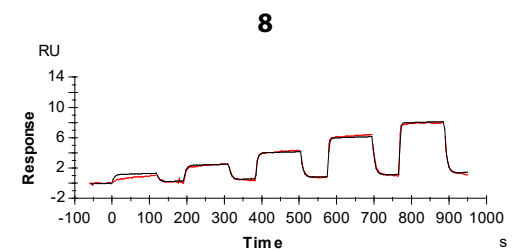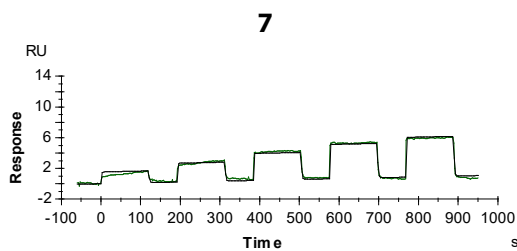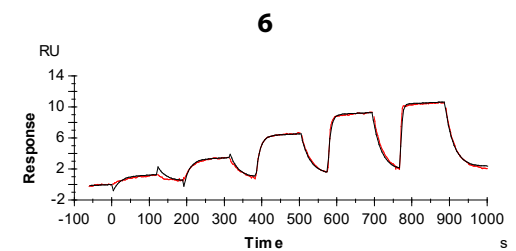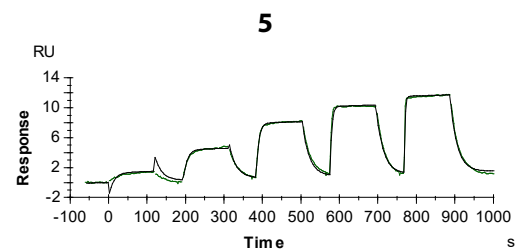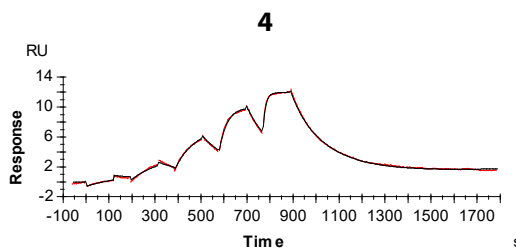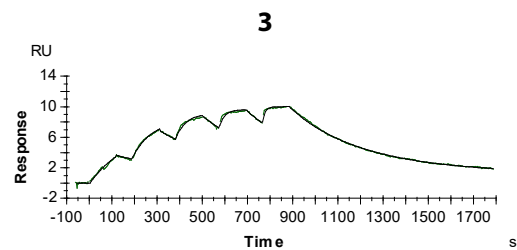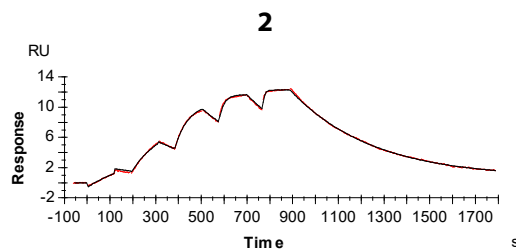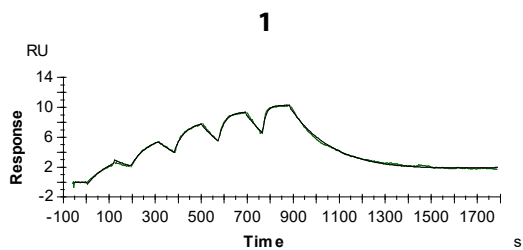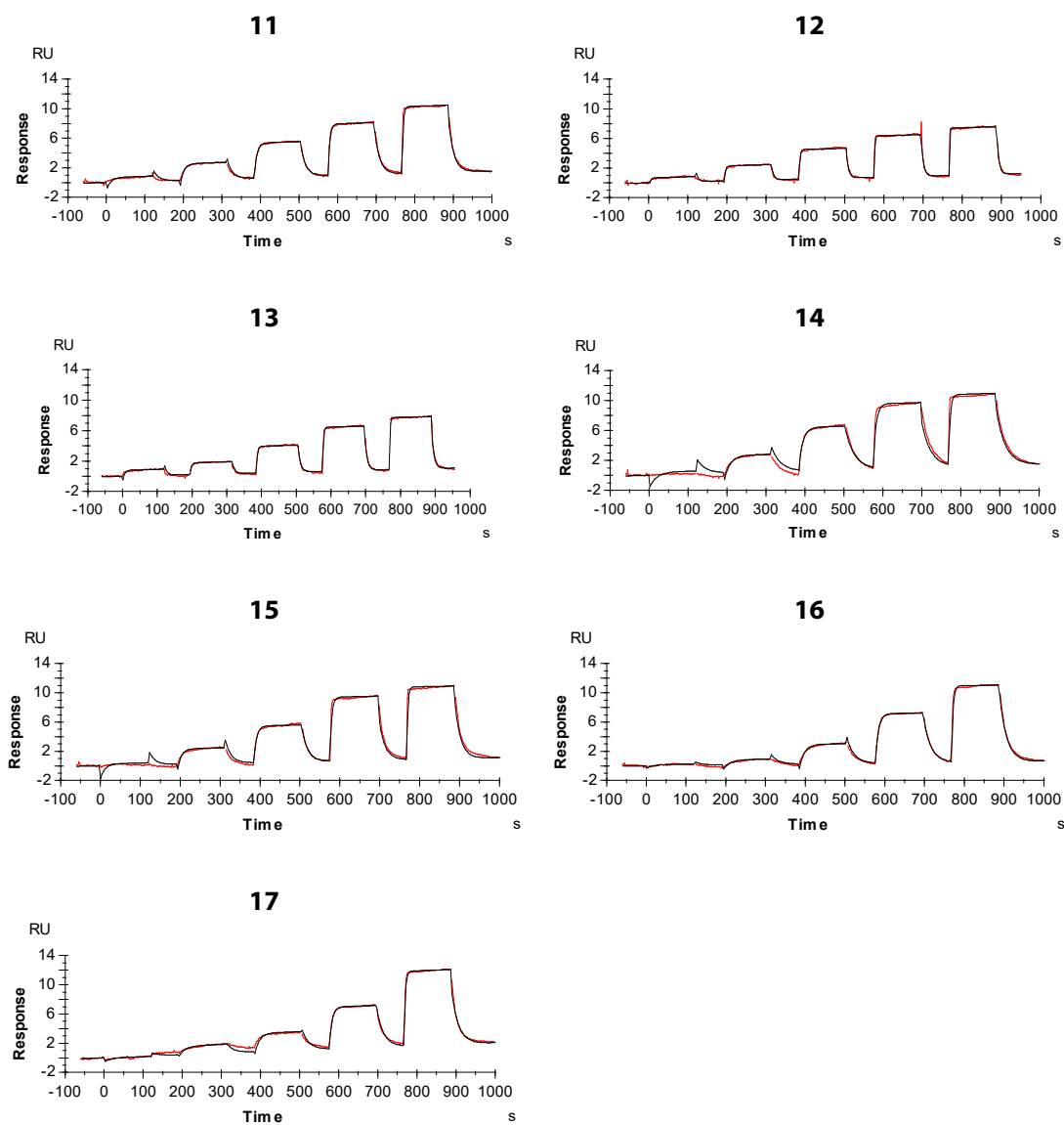
Figure S5.1. SPR sensorgrams showing one representative single cycle kinetics run (green or red) of ligand **1-17**, respectively**,** fitted to a 1:1 kinetic binding model (black). The five concentrations used in the shown sensorgrams were 0.025, 0.074, 0.22, 0.67, 2.0 µM for ligand **2, 4, 16, 17**; 0.16, 0.31, 0.63, 1.25, 2.5 µM for ligand **1, 3, 6, 8, 11**; 0.63, 1.25, 2.5, 5.0, 10.0 µM for ligand **5, 9, 10, 12, 13, 14, 15** and 1.25, 2.5, 5.0, 10.0, 20.0 µM for ligand **7.**

5.10.2 Comparison of kinetic data from SPR and photometric inhibition assay
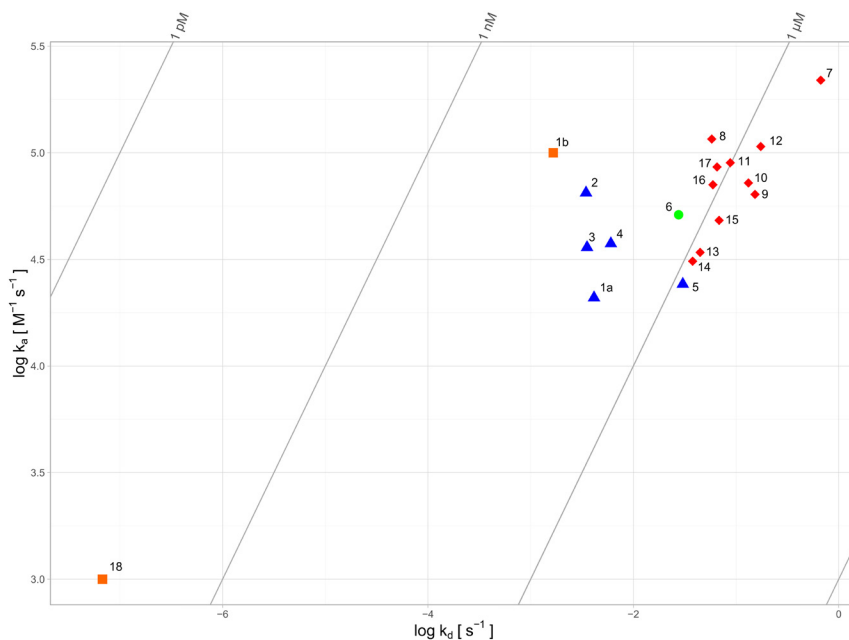
Figure S5.2. Kinetic Map containing binding kinetic data of **1b** and **18** taken from Bartlett and Marlowe (orange blocks) [287]. From the kinetic rate constants of **1a** and **1b** the variation between the inherently different experimental methods (SPR vs. photometric inhibition assay) can be estimated to fall within one order of magnitude.

5.10.3 Crystallographic tables

Table S5.3. Data collection and refinement statistics for crystal structures 5LIF and 5LWD.

	Complex (PDB code)	
	TLN-**5** (5LIF)	TLN-**6** (5LWD)
Data collection and processing		
Space group	$P6_122$	$P6_122$
Unit cell parameters: a, b, c (Å)	91.8, 91.8, 130.0	92.8, 92.8, 130.6
Matthews coefficient (Å^3/Da)[a]	2.3	2.4
Solvent content (%)[a]	46.3	47.7
Resolution range (Å)	50.00-1.31 (1.31–1.39)	50.00-1.23 (1.30–1.23)
Wilson B factor (Å^2)	10.1	9.1
Unique reflections	77996 (12256)	96448 (15229)
R_{sym} (%)	5.7 (49.0)	6.4 (47.3)
Completeness (%)	99.8 (99.0)	99.7 (98.8)
Redundancy	10.6 (10.3)	24.9 (24.6)
$<I/\sigma(I)>$	27.3 (5.2)	36.3 (7.3)
Refinement		
Resolution range (Å)	19.66–1.31	31.75–1.23
Reflections used in refinement (work/free)	74096/3900	91621/4822
R_{cryst} (%)	11.3	10.4
R_{free} (%)	13.6	12.3
Protein residues	316	316
Calcium/zinc ions	4/1	4/1
Inhibitor atoms	35	32
Water molecules	407	401
RMSD from ideality:		
Bond lengths (Å)	0.008	0.010
Bond angles (°)	1.0	1.2
Ramachandran plot (%):[b]		
Residues in most favored regions	87.8	88.5
Residues in additionally allowed regions	11.1	10.4
Residues in generously allowed regions	0.7	0.7
Residues in disallowed regions[c]	0.4	0.4
Mean B factor (Å^2):[d]		
Protein	10.5	10.7
Inhibitor	12.5	10.7
Water molecules	26.0	26.6

The highest resolution shell is described by values in parentheses. [a] Matthews coefficient and solvent content were calculated with the program *Matthews_coef* from the *CCP*4 suite (version 6.3.0) [195]. [b] Ramachandran plots were calculated with *PROCHECK* [196]. [c] The Ramachandran outlier (Thr26) occurs in every structure of TLN and is described in literature [197]. [d] Mean B factors were calculated with *MOLEMAN* [198].

5.10.4 Ligand Synthesis and Purification

Experimental details. ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on a JEOL ECX-400 or JEOL ECA-500 instrument. All chemical shift values are reported in ppm relative to the non-deuterated solvent signal. An external standard was used for ^{31}P NMR spectra (referenced to: 85% H_3PO_4) and ^{13}C NMR spectra in D_2O (referenced to: trimethylsilyl propanoic acid). For the description of multiplicity the following abbreviations were used: s = singlet, m = multiplet. ESI-MS spectra were recorded on a Q-Trap 2000 system by Applied Biosystems. For high resolution ESI-MS a LTQ-FT Ultra mass spectrometer (Thermo Fischer Scientific) was used. For HPLC chromatography a Shimadzu LC-20 system equipped with a diode array detector was used. Analytic separations were carried out with a MN Nucleodur 100-5 C18 ec 4.6×250 mm column using a water-acetonitrile gradient. For semi-preparative separations a Water XSelect CSH C18 10×250 mm column employing a water-acetonitrile gradient was used.

Scheme S5.1. Synthesis of compound **6** from Boc-Leu-OH.
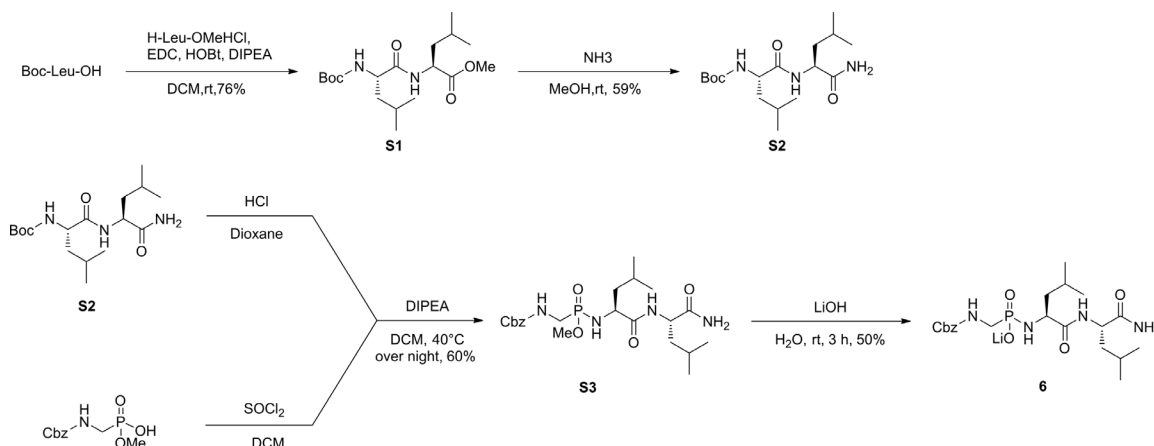
Boc-Leu-Leu-OMe (**S1**). To a suspension of Boc-Leu-OH (347 mg, 1.50 mmol, 1.0 eq), EDC (374 mg, 1.95 mmol, 1.3 eq), HOBt (264 mg, 1.95 mmol, 1.3 eq), and H-Leu-OMe HCl(272 mg, 1.50 mmol, 1.0 eq) in DCM, DIPEA (485 mg, 3.75 mmol, 2.5 eq) was added and the reaction mixture was stirred at rt overnight. The solvent was removed under reduced pressure and the resulting residue was taken up in EtOAc and extracted with 1 M HCl (3×10 mL) and sat. NaHCO$_3$ (3×10 mL). The organic phase was washed with brine and dried over anhydrous MgSO$_4$. The crude product was purified by silica gel chromatography (DCM/MeOH 50:1) to give Boc-Leu-Leu-OMe as a colorless solid (410 mg, 1.14 mmol, 76%). ^{1}H NMR (400 MHz, CDCl$_3$) δ = 0.78 – 1.00 (m, 12H), 1.43 (s, J=9.8, 9H), 1.58 – 1.75 (m, 6H), 3.72 (s, 3H), 4.03 – 4.15 (m, 1H), 4.61 (td, J=8.7, 4.7, 1H), 4.86 (d, J=7.7, 1H), 6.43 (d, J=8.2, 1H). ^{13}C NMR (101 MHz, CDCl$_3$) δ = 21.9, 22.2, 22.9, 24.7, 28.3, 40.9, 41.6, 50.7, 52.4, 53.0, 80.2, 155.8, 172.4, 173.3. **MS (ESI+)** *m/z* calculated for C$_{18}$H$_{35}$N$_2$O$_5$ [M+H]$^+$: 349.49; found: 359.24.

Boc-Leu-Leu-NH$_2$ (**S2**). To 7 mL of a 7 M solution of NH$_3$ in MeOH, Boc-Leu-Leu-OMe (427 mg, 1.19 mmol) was added. The mixture was stirred at rt for 48 h, the solvent was removed under reduced pressure and the crude product was purified by silica gel chromatography (cyclohexane/EtOAc 1:1). The product was isolated as a colorless solid (239 mg, 0.70 mmol, 59%). ^{1}H NMR (400 MHz, CDCl$_3$) δ = 0.85 – 0.99 (m, 12H), 1.43 (s, 9H), 1.45 – 1.83 (m, 6H), 4.06 (bs, 1H), 4.41 – 4.53 (m, 1H), 4.93 (s, 1H), 5.54 (s, 1H), 6.47 – 6.64 (m, 2H). ^{13}C NMR (101 MHz, CDCl$_3$) δ = 21.8, 22.0, 23.0, 23.2, 24.8, 28.4, 40.7, 40.9, 51.3, 53.6, 80.5, 156.1, 175.0, 191.2. **MS (ESI+)** *m/z* calculated for C$_{17}$H$_{34}$N$_3$O$_4$ [M+H]$^+$: 344.48; found: 344.21.

Phosphonamidate S3. SOCl$_2$ (238 mg, 2.00 mmol, 4.0 eq) was dissolved in 5 mL dry DCM under argon and cooled to 0 °C. A solution of benzyl ((hydroxy(methoxy)phosphoryl) methyl)carbamate (233 mg, 0.90 mmol, 1.8 eq) was added to the solution over 30 minutes. The cooling bath was removed and the reaction was stirred at rt for 3–5 h. All volatile components were removed under reduced pressure. In a second flask, Boc-Leu-Leu-NH$_2$ (172 mg, 0.50 mmol, 1.0 eq) was treated with HCl (4 M in dioxane, 0.5 mL, 2 mmol, 4.0 eq) for 1 h. All volatile components were removed under reduced pressure. The activated phosphonic acid was dissolved in dry DCM. DIPEA and the deprotected dipeptide were added to the solution consecutively. The mixture was heated to 40 °C and stirred overnight. The solution was diluted with EtOAc and extracted with 5% citric acid (3×10 mL), 1 M HCl (1×10 mL) and 1 M NaOH (3×10 mL). The organic phase was washed with brine and dried over MgSO$_4$. The product was obtained as a mixture of diastereomers in form of a brown solid (145 mg, 0.30 mmol, 60%). **^{1}H NMR** (400 MHz, CDCl$_3$) δ = 0.71 – 1.06 (m, 12H), 1.16 – 2.27 (m, 6H), 3.40 – 3.82 (m, 5H), 3.81 – 4.00 (m, 1H), 4.50 (s, 1H), 5.00 – 5.23 (m, 2H), 7.25 – 7.49 (m, 5H). **^{13}C NMR** (101 MHz, CDCl$_3$) δ = 21.7, 21.85, 21.90, 22.1, 23.0, 24.5, 24.6, 24.85, 24.94, 37.4, 37.8, 38.7, 39.3, 41.0, 41.1, 43.7, 43.9, 51.3, 51.4, 52.0, 53.4, 53.5, 67.15, 67.24, 128.1, 128.3, 128.6, 136.28, 136.36, 156.8, 157.0, 174.2, 174.3, 175.7, 175.9. **^{31}P NMR**

(162 MHz, CDCl$_3$) δ = 28.6, 29.7. **MS (ESI+)** calculated for C$_{22}$H$_{37}$N$_4$NaO$_6$P: 507.52 [M+Na]$^+$; found: 507.42.

Inhibitor 6: The protected phosphonamidate was treated with 3 mL of a 0.4 M solution of LiOH in water. MeCN was added until the solution cleared. The reaction was stirred for 4 h at rt. Under ice cooling the pH was adjusted to 8 using 5% AcOH and the solvent was removed under reduced pressure. The residue was dissolved in 1 mL of water and purified by semi-preparative HPLC. Lyophilization of fractions containing the pure product afforded a colorless solid (36 mg, 0.08 mmol, 50%). **^{1}H NMR** (500 MHz, D$_2$O) δ = 0.89 – 1.01 (m, 12H), 1.46 – 1.59 (m, 2H), 1.63 – 1.84 (m, 4H), 3.19 – 3.58 (m, 2H), 3.65 – 3.85 (m, 1H), 4.29 – 4.44 (m, 1H), 5.09 – 5.28 (m, 2H), 7.38 – 7.66 (m, 5H). **^{13}C NMR** (101 MHz, D$_2$O) δ = 20.1, 21.0, 22.3, 22.4, 24.2, 24.4, 39.1 (d, J_{CP} = 134.1 Hz), 39.3, 42.9, 51.8, 54.5, 67.2, 127.8, 128.4, 128.8, 136.4, 158.2, 178.1, 178.9. **^{31}P NMR** (202 MHz, D$_2$O) δ = 19.1. **HRMS (ESI+)** calculated for C$_{21}$H$_{35}$N$_4$O$_6$PNa: 493.2186 [M+Na]$^+$; found: 493.2184.

5.10.5 Sequence alignment of representative M4 and M13 proteases

```
btThermolysin|M4

btThermolysin|M4   ..................................................MKKVSTLDLLFVA....IMGVS
paElastase|M4      ..................................................MGNKKEIAIVALTTGFVLTSVT
bcNEP|M4           ..................................................MRKFSRYAFTSMATVTLLSSLT
saAureolysin|M4    ..................................................MRKV...WFSLLGGAMLLGSVA
ppGentlyase|M4     ..................................................MKKV...WFSLLGGAMLLGSVA
mtEndopeptidase|M13 .................................................
hNeprilysin|M13    ............MGKSESQMDITDINTPKPKKK.QRWTPL...EISLSVLVLLLTIIA....VT
hECE-1|M13         MRGVWPPPVSALLSALGMSTYKRATLDEEDLVDSLSEGDAYPNGLQVNFHSPRSGQRCWAARTQVEKRLVVLVVLLAAGLVACLAA


btThermolysin|M4

btThermolysin|M4   ..................................................................................
paElastase|M4      P...AAFAADLIDVSKLPSKAA..QGAPGPVTLQA...............................AVGA
bcNEP|M4           P...YGVGH..............AEEPGMQVEIQED.SFRTGDLTQPSKKAPE.......NVVKDALKENTEHALS.PKQVSA
saAureolysin|M4    P...AALAS..............DTNHKPATSD...I..NFEITQKS.......DAVKA.L...KELPKS...ENVK
ppGentlyase|M4     S...GASAE...............SSVSGPAQLT...P..TFHTEQWKAPSSVSGD.......DIVWSYLNRQKKSLLG...VDSS
mtEndopeptidase|M13 .........MTLAIPSGIDLSHIDADARPQDDLFGHVNGRWLAEHEIPADRATDGAFRSLFDRAETQVRDLIIQASQAGA
hNeprilysin|M13    MIALYATYDDGICKSSDCIKSAARLIQNMDATTEPCTDFFKYACGGWLKRNVIPETSSRYGNFDILRDELEVVLKDVLQEPKT...
hECE-1|M13         LGIQYQTRSPSVCLSEACVSVTSSILSSMDPTVDPCHDFFSYACGGWIKANPVPDGHSRWGTFSNLWEHNQAIIKHLLENSTA...


btThermolysin|M4

btThermolysin|M4   ..................................................................................
paElastase|M4      GGADELKAI....RSTTLPNGKQVTRYEQFH......NGVRVVGE.............AITEVKGPGKSVAAQRSGHF....
bcNEP|M4           ETGVDYKVL....QKRGSYDGTTLLRMQQIY......EGKEVVGH.............QLTAHVDKKGIIKS.ISGAS...
saAureolysin|M4    NHYQDYSVT....DVKTDKKGFTHYTLQPSV......DGVHAPDK.............EVKVHADKSGKVVL.INGDT....
ppGentlyase|M4     SVREQFRIV....DRTSDKSGVSHYRLKQYV......NGIPVYGA.............EQTIHVGKSGEVTS.YLGAV....
mtEndopeptidase|M13 AVGTDAQRIGDLYASFLDEEAVERAGVQPLHDELATIDSAADATELAA...........AL.GTLQRAGVGGGIGV.YVDTDSKDS
hNeprilysin|M13    EDIVAVQKAKALYRSCINESAIDSRGGEPLLKLLPDIYGWPVATENWEQKYGASWTAEKAIAQLNSKYGKKVLINL.FVGTDDKNS
hECE-1|M13         SVSEAERKAQVYYRACMNETRIEELRAKPLMELIERLGGWNITGPWAKDNF......QDTLQVVTAHYRTSPFFSV.YVSADSKNS


btThermolysin|M4

btThermolysin|M4   ..................................................................................
paElastase|M4      VANIA.AD..................................LPGSTTAAV.............SAE...........QVLA
bcNEP|M4           AQNLTQEDLKKPINLSKEE......AKQYIYKKYGNDTKFI.SEPEVKAVIF.............VDEN.....
saAureolysin|M4    DAKKV..................................IAKIANLQP.......AADKAFN..AVKIDKNKAKNL
ppGentlyase|M4     INEDQQEEATQGTT.PKIS......ASEAVYTAYKEAAARIEALPTSDDTISKDAEEPSSVSKDTYAEAANNDKTLSVDKDELSLD
mtEndopeptidase|M13 TRYLVHFT.QSGIGLPDESYYRDEQH.AAVLAAYPGHIARMFGLVHV.......................ES.....RDHAKT
hNeprilysin|M13    VNHVIHID.QPRLGLPSRDYYECTGIYKEACTAYVDFMISVARLIRQE..................ERLPIDENQLALE
hECE-1|M13         NSNVIQVD.QSGLGLPSRDYYLNKTENEKVLTGYLNYMVQLGKLLGGG.................DE.....EAIRPQ


btThermolysin|M4

btThermolysin|M4   ..................................................................................
paElastase|M4      QAKSLKAQG.........................RKT......ENDKVE......LVIRLG...ENNIAQLVYNVSYL
bcNEP|M4           .....................................................................HGQASNAYQVTFA
saAureolysin|M4    QDDVIKENKV...........................EID.............GDSNKYIYNIELI
ppGentlyase|M4     KASVLKDSKIEAVEAEKSS...........IAKIANLQP.......EVDPKA...ELYYYP...KGDDLLLVYVTEVN
mtEndopeptidase|M13 A......DRIVALETKLADAHWDVVKRRDADLGYNLRTFAQLQT......EGAGFDWVSWVTALGSAP...DAMTELVVRQPDYL
hNeprilysin|M13    M......NKVMELEKEIANATAKPEDRNDPMLLYNKMTLAQIQNNFSLEINGKPFSWLNFTNEIMSTVNISITNEEDVVVYAPEYL
hECE-1|M13         M......QQILDFETALANITIPQEKRRDEELIYHKVTAAELQT.....LAPAINWLPFLNTIFYPVE.INESEPIVVYDKEYL


btThermolysin|M4                                                                   β1
                                                                        1      10

btThermolysin|M4   ....................................................ITGTSTVGVGR................
paElastase|M4      IP............GEGLSRPHFVIDAKTGEVLDQWE................GLAHAEAGGPGG................
bcNEP|M4           AT............TPNYVSGTYLVNAHNGEMLKDMVQESSLKISEEYVESVKENKTSNFTSLTGTGK..............
saAureolysin|M4    TV............TPEISHWKVKIDADTGAV..................VEKTNLVKEAAATGTGK..............
ppGentlyase|M4     VL..................................VFQY.DIINEAT.GK..............................
mtEndopeptidase|M13 VTFASLWASVNVEDWKCWARWRLIRARA.P...........WLTRALVAED....FEFYGRTLTGAQQLRDRWKRGVSLVENLM
hNeprilysin|M13    TKLKPILTKYSARDLQNLMSWRFIMDLV..S...........SLSRTYKESR...NAFRKALYGTTSE.TATWRRCANYVNGNM
hECE-1|M13         EQISTLINTTDRCLLNNYMIWNLVRKTS..S...........FLDQRFQDAD.....EKFMEVMYGTKKTCLPRWKFCVSDTENNL


btThermolysin|M4          β2          β3 β4         β5                β6 β7 β8
                  TT                    TT    TT        TT                          
                           20          30       40                    50      60    70

btThermolysin|M4   ....GVLGDQKNINTTY.S.TYYYLQDNT..RGNGIFTYDAKYRTT...............LPGSLWADADNQFFASYDAPAVDA
paElastase|M4      ....NQKIGKYTYGSDY...GPLIVNDRCEMDDGNVITVDMNSSTDD...S.......KTTPFFRFACPTNTYKQVNGA.YSPLNDA
bcNEP|M4           ....DDLGVTRTFGISKQTSGNYALADYT..RGQGIETYDVNYRDIT...N....EESYYPGILATSASTTFN...DPKAVSA
saAureolysin|M4    ....GVLGDTKDININS.IDGGFSLEDLT..HQGKLSAYNFND.QT............GQATLITNEDENFVKDDQRAGVDA
ppGentlyase|M4     ....GVLGDSKSFTTTA.SGSSYQLKDTT..RGNGIVTYTASNRQS.............IPGTLLTDADNVWN...DPAGVDA
mtEndopeptidase|M13 GDAVGKLYVQRHFPPDAKSRIDTLVDQEAYRISISELDWMTPQTRQRALAKLNKFTAKVGYP.IKWRD.......YSK.....
hNeprilysin|M13    ENAVGRLYVEAAFAGESKHVVEDLIAQIREVFIQTLDDLTWMDAETKKRAEEKALAIKERIGYP.DDIVSNDNKLNNEYLE.....
hECE-1|M13         GFALGPMFVKATFAEDSKSIATEIILEIKKAFEESLSTLKWMDEETRKSAKEKADAIYNMIGYP.NFIMDP.KELDKVFND.....
```

Figure S5.3. Secondary structure alignment of thermolysin and proteases from the M4 and M13 family. Only the main chain of thermolysin (Uniprot P00800) was used for the alignment. The sequences of the signal- and propeptide were omitted to conserve canonical numbering throughout the main text. The protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm) was used for secondary structure alignment [280]. The results were displayed using the ESPript server (http://espript.ibcp.fr) [281].

```
                        α1                                                    β9                              β10         β11                              η1
btThermolysin|M4    0000000000000000  →                                                              ═══                     ═══                         000
                          80        90         100           110           120            130
btThermolysin|M4    HYYAGVTYDYYKNVHNRL..SYDGNNAAIRSSVH...YS..QG...YNNAF....WNGSQMVYGDG..D.......GQTFIPLSGG
paElastase|M4       HFFGGVVFKLYRDWFGTS..PLTHK...LYMKVH...YGR.....SVENAY....WDGTAMLFGDG...........ATMFYPLVS
bcNEP|M4            HFLATKVYDFYKEKYKRN..SFDNKGKKVVSVVHAWDSGGTNDPENWENAFSTNINNISMLLYGDP...............MVKA
saAureolysin|M4     NYYAKQTYDYYKNTFGRE..SYDNHGSPIVSLTHVNHYGGQDN...RNNAA....WIGDKMIYGDG..D.......GRTFTNLSGA
ppGentlyase|M4      HAYAAKTYDYYKSKFGRN..SIDGRGLQLRSTVH...YG..SR...YNNAF....WNGSQMTYGDGDGD........GSTFIAFSGD
mtEndopeptidase|M13 ..LAIDRDDLYGNVQRGYAVNHDRELAKLFGPVDRDEWFM..TPQ.TVNAYYNPG..MNEIVFPAAILQPPFFDPQADEAANYGGI
hNeprilysin|M13     ..LNYKEDEYFENIIQNLKFSQSKQLKKLREKVDKDEWIS..GAA.VVNAFYSSG..RNQIVFPAGILQPPFFSAQQSNSLNYGGI
hECE-1|M13          ..YTAVPDLYFENAMRFFNFSWRVTADQLRKAPNRDQWSM..TPP.MVNAYYSPT..KNEIVFPAGILQAPFYTRSSPKALNFGGI

                          α2                              α3                                β12              TT TT  β13
btThermolysin|M4    0000000000000000          0000......0000000000000000                           →                        ─
                          140       150          160           170          180               190             200
btThermolysin|M4    IDVVAHELTHAVTDYTAGLIYQNESG.....AINEAISDIFGTLVEFYAN.K.............NPDWEIGEDVYTPGISGDSLR
paElastase|M4       LDVAAHEVSHGFTEQNSGLIYRGQSG.....GMNEAFSDMAGEAAEFYMR.G.............KNDFLIGYDIKK..GSGALR
bcNEP|M4            FDIAGHEFTHAVTSSESNLEFSGESG.....AINEALSDIMGTAIEKYINNG.............EFNWTIGEQ......TGSTFR
saAureolysin|M4     NDVVAHELTHGVTQETANLEYKDQSG.....ALNESFSDVFGYFV.......D...........DEDFLMGEDVYTPGKEGDALR
ppGentlyase|M4      PDVVGHELTHGVTEYTSNLEYYGESG.....ALNEAFSDVIGNDI.......Q...........RKNWLVGDDIYTPNICGDALR
mtEndopeptidase|M13 GAVIGHEIGHGFDDQGAKYDGDGNLVDWWTDDDRTEFAARTKALIEQYHAYTPRDLVDHPGPPHVQGAFTIGENIGDLGGLSIALL
hNeprilysin|M13     GMVIGHEITHGFDDNGRNFNKDGDLVDWWTQQSASNFKEQSQCMVYQYGNFS....WDLAGGQHLNGINTLGENIADNGGLGQAYR
hECE-1|M13          GVVVGHELTHAFDDQGREYDKDGNLRPWWKNSSVEAFKRQTECMVEQYSNYS....V...NGEPVNGRHTLGENIADNGGLKAAYR

                         η2              η3        α4           α5         β14  β15              α6                  TT    00
btThermolysin|M4    →      000             000       0000    0000000000   →═══════            0000000000000000        TT
                          210             220       230          240       250           260          270           280
btThermolysin|M4    SMSDPAKYG......DPDHYSKRY.TGTQDNGGVHINSGIINKAAYLTSQGGTHYGVSVVGIGRDKLGKIFYRALTQYLTPTSNFS
paElastase|M4       YMDQPSRDGRS.....IDNASQYY.....NGIDVHHSSGVYNRAFYLLANS.........PGWDTRKAFEVFVDANRYYWTATSNYN
bcNEP|M4            NMKNPTAINFSEGLPYPDDYSKYNDLNGEDYGGVHFNSSIINKVAYLMAQGGTHNGVNVNGIGEDRMFDIFYYANTDELNMTSNFS
saAureolysin|M4     SMSNPEQFG.....QPSHMKDYV.YTEKDNGGVHTNSGIPNKAAYNVI.........QAIGKSKSEQIYYRALTEYLTSNSNFK
ppGentlyase|M4      SMSNPTLYD......QPHHYSNLY.KGSSDNGGVHTNSGIINKAYYLLAQGGTFHGVTVNGIGRDAAVQIYYSAFTNYLTSSSDFS
mtEndopeptidase|M13 AY...............................QLSLNGNPAPV...IDGLTG.MQRVFFGWAQIWRTKSRAA
hNeprilysin|M13     AY...............................QNYIKKNGEEKLLPGLDLNH.KQLFFLNFAQVWCGTYRPE
hECE-1|M13          AY...............................QNWVKKNGAEHSLPTLGLTN.NQLFFLGFAQVWCSVRTPE

                           α7             α8
btThermolysin|M4    0000000000000000  0000000000000
                          290        300         310
btThermolysin|M4    QLRAAAVQSATDLYGSTSQEVASVKQAFDAVGVK.............................................
paElastase|M4       SGACGVIRSAQNRN....YSAADVTRAFSTVGVTCPSAL.......................................
bcNEP|M4            ELKLACLKVATNKYGANSIEVQAVQNAFDAAKITGTVKENEKTAENQAPELTVPFTITLRVGDTFDPMRNVKAVDKEDGDLTNKVK
saAureolysin|M4     DCKDALYQAAKDLYDEQTAEQ..VYEAWNEVGVE...........................................
ppGentlyase|M4      NARAAVIQAAKDLYGANSAEATAAAKSFDAVG.............................................
mtEndopeptidase|M13 EAIRRL...AVDPH.........SPPEFRCNGVVRNVDAF........YQAFDVTEDDALFLDPQRRVRIWN.......
hNeprilysin|M13     YAVNSI...KTDVH.........SPGNFRIIGTLQNSAEF........SEAFHCRKN..SYMNPEKKCRVW.......
hECE-1|M13          SSHEGL...ITDPH.........SPSRFRVIGSLSNSKEF........SEHFRCPPG..SPMNPPHKCEVW.......

btThermolysin|M4    .......................................................................................
paElastase|M4       .......................................................................................
bcNEP|M4            HKGDVDTSKPGTYIVEYMVVDSQGRNATATQTVIVEGNGETSDLEPKLTVPVGATIHVGDSFVPMAEVLAIDKEDGDLTSKIKVDG
saAureolysin|M4     .......................................................................................
ppGentlyase|M4      .......................................................................................
mtEndopeptidase|M13 .......................................................................................
hNeprilysin|M13     .......................................................................................
hECE-1|M13          .......................................................................................

btThermolysin|M4    .......................................................................................
paElastase|M4       .......................................................................................
bcNEP|M4            EVDTTKAGTYVLTYTVTDSKGHEVTAKQTVTVKVREEVKNDKPILKVPATTSITEGDQFDPLVGVSATDKEDGDLTSKVAYEGTID
saAureolysin|M4     .......................................................................................
ppGentlyase|M4      .......................................................................................
mtEndopeptidase|M13 .......................................................................................
hNeprilysin|M13     .......................................................................................
hECE-1|M13          .......................................................................................

btThermolysin|M4    .......................................................................................
paElastase|M4       .......................................................................................
bcNEP|M4            TIKAGTYEIIYSVRDSMGNEVNTIQKVLVKDKETSKANGLTNKTNNNNSDKKEYTYKELPKTGVSTTNSAAIGILMIITGTVLTFV
saAureolysin|M4     .......................................................................................
ppGentlyase|M4      .......................................................................................
mtEndopeptidase|M13 .......................................................................................
hNeprilysin|M13     .......................................................................................
hECE-1|M13          .......................................................................................

btThermolysin|M4    ........
paElastase|M4       ........
bcNEP|M4            RKFRKIQK
saAureolysin|M4     ........
ppGentlyase|M4      ........
mtEndopeptidase|M13 ........
hNeprilysin|M13     ........
hECE-1|M13          ........
```
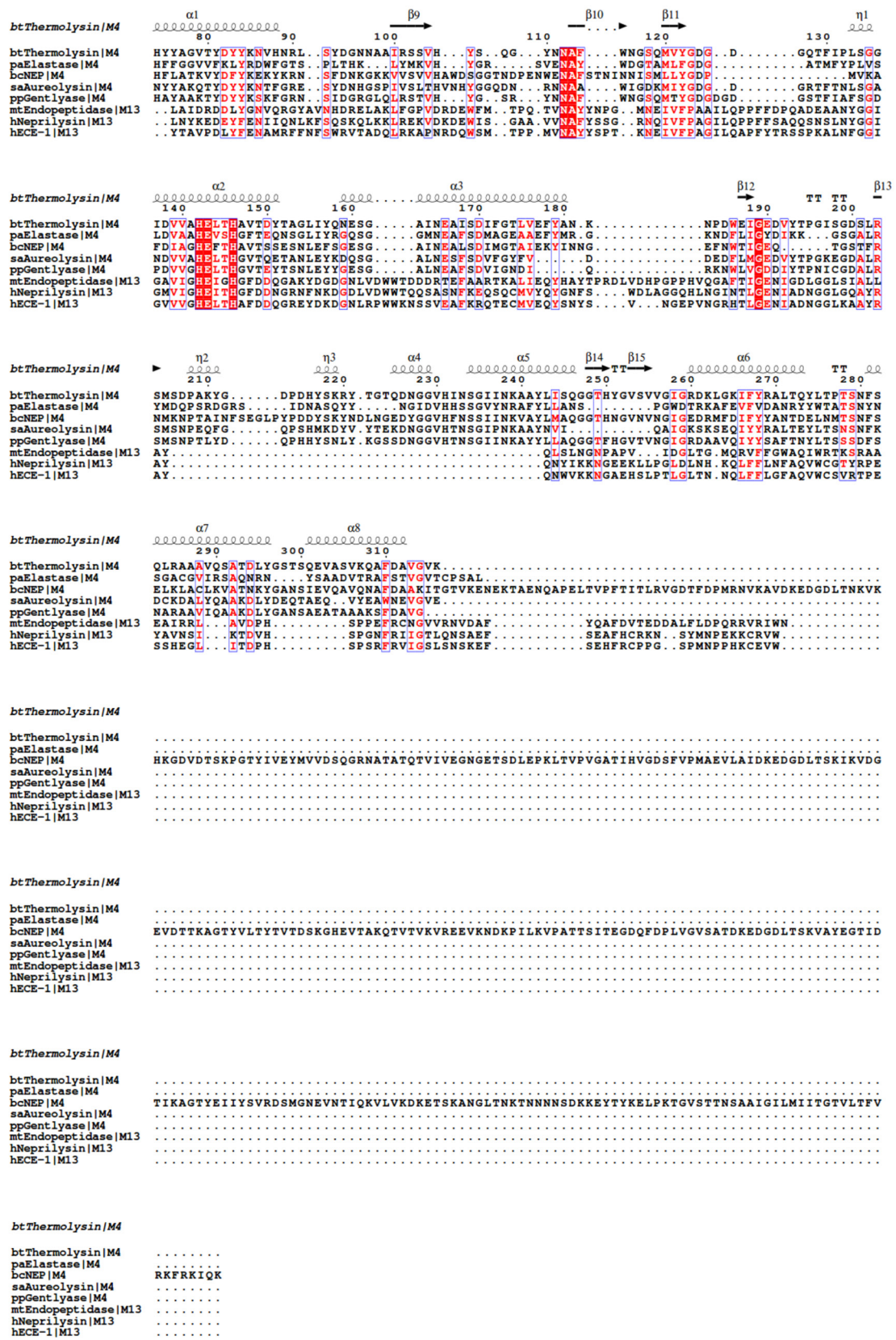
Figure S5.3. (continued)

Chapter 6

How Nothing Boosts Affinity: Hydrophobic Ligand Binding to the Virtually Vacated S_1' Pocket of Thermolysin

Stefan G. Krimmer,[1,*] Jonathan Cramer,[1,*] Johannes Schiebel,[1]
Andreas Heine,[1] Gerhard Klebe[1]

*These authors contributed equally

[1]Department of Pharmaceutical Chemistry, University of Marburg,
Marbacher Weg 6, 35032 Marburg, Germany

6.1   Introduction

The deep, hydrophobic S_1' pocket of thermolysin (TLN), which is large enough to accommodate a P_1' substituent up to a benzyl side chain, has previously been reported by us to be poorly solvated and to contain only disordered water molecules, whose displacement would give rise to a strong enthalpic benefit [194]. In the current contribution, we try to collect more evidence for this poor solvation pattern, maximally composed of disordered water molecules in the S_1' pocket of this metalloprotease, and investigate the origin of the reported enthalpy gain. In this context, we analyze the protein–ligand complex formation of a series of six phosphonamidate-type TLN inhibitors (Figure 6.1A) using X-ray crystallography and isothermal titration calorimetry (ITC). Going from a hydrogen atom to a *iso*-butyl P_1' group (Figure 6.1B), an amazingly strong affinity increase of 26.3 kJ mol^{-1} is observed. For the explanation of the molecular mechanism behind this considerable affinity increase, exact knowledge about the solvation state within the S_1' cavity is required. For TLN-**1**, we determined an experimentally phased electron density map via a zinc multi-wavelength anomalous diffraction (MAD) experiment. The experimentally phased electron density map provides the advantage that no model bias is introduced by phasing of the structure compared to less elaborate strategies such as molecular replacement. Additionally, we attempt to transform the sigma-scaled electron density map to an absolute electron number scale (e$^-$/Å^3) to enable the determination of the exact total electron content within the S_1' cavity. This strategy allows the detection of electron density traces, which can indicate highly mobile, crystallographically disordered water molecules [293]. To characterize the volume and polarity of the S_1' cavity of TLN-**1** experimentally, we used the noble gases xenon and krypton as probes to fill the pocket. Considering the complexes of **1** to **4**, the residual unoccupied volume of the S_1' cavity is gradually reduced, thus decreasing the putative residence volume of a potentially bound mobile water molecule. Inhibitors **5** and **6** even comprise polar functional groups that increase the polarity within the cavity and provide a potential hydrogen-bonding anchor to fix a remaining S_1' water occupant. Both strategies, reducing the residence volume and increasing the polarity within the cavity, should diminish the mobility of a potentially present residual water molecule and increase the concentration of its electrons within the cavity, thus facilitating its detection in a crystal structure.

6.2   Results

6.2.1   Isothermal titration calorimetry

ITC measurements of **2**–**4** were performed by direct titrations (Figure S6.1A–C) and for **1**, **5** and **6** by displacement titrations (Figure S6.1D–F). The thermodynamic profiles of **2**–**4** were
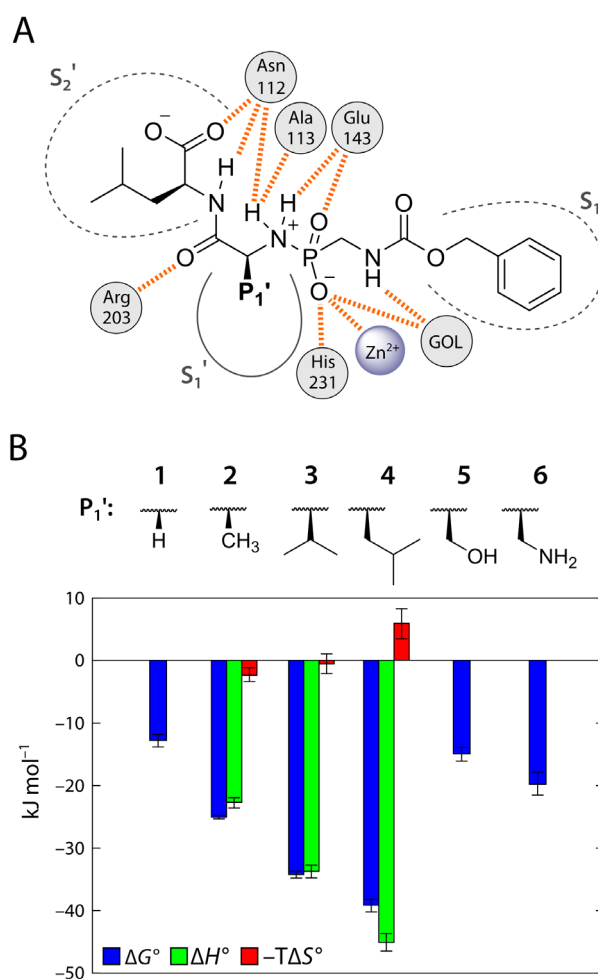
Figure 6.1. Congeneric series of phosphonamidate TLN ligands substituted with different P$_1$′ groups. **(A)** Schematic binding mode of the parent ligand scaffold in complex with TLN. Protein residues and the zinc ion forming hydrogen bonds (orange dashed lines) with the parent scaffold are indicated (GOL = glycerol molecule from the cryo buffer bound to TLN). The S$_1$ and S$_2$′ and pockets of TLN are wide-open and well-accessible to water molecules, whereas the S$_1$′ pocket is deep and apolar. **(B)** P$_1$′ substituents of **1–6** and their thermodynamic binding profiles as determined by ITC. The thermodynamic profiles shown for **2**, **3** and **4** were corrected for the heat of buffer ionization upon complex formation. Error bars represent the 95.4% confidence interval. A reliable determination of the buffer-corrected enthalpy/entropy partitioning of **1**, **5** and **6** (displacement titrations) is impossible and thus in these cases only the standard Gibbs free energy is discussed. Data values are listed in Table S6.1.

corrected for proton transfer reactions upon complex formation via measurement in several buffers showing different ionization enthalpies (Figure S6.2) [150]. Such corrections were not performed for the displacement titrations of **1**, **5** and **6**, because this would have resulted in large experimental uncertainties, rendering the thermodynamic parameters unreliable even for mutual comparison.

The thermodynamic profiles of **1**–**6** are displayed in Figure 6.1B. The affinity strongly increases with the growing hydrophobic size of the P_1' group (**1**→**2**→**3**→**4**) over 4.5 orders of magnitude (expressed in terms of the binding constants as listed in Table S6.1). In particular, the addition of a single methyl group to **1** yielding **2** ($\Delta\Delta G°_{1\to2}$= –12.3 kJ mol^{-1}) and the addition of two methyl groups to **2** yielding **3** ($\Delta\Delta G°_{2\to3}$ = –9.2 kJ mol^{-1}) result in substantial affinity boosts. With respect to the partitioning in $\Delta H°$ and $-T\Delta S°$, the affinity increase from **2**→**3**→**4** is the result of a sharp increase in $\Delta H°$, which is only partially compensated by a decrease in $-T\Delta S°$. The affinities of **5** and **6** fall between those of **1** and **2**. The amino-substituted **6** clearly shows a higher affinity than the hydroxyl-substituted **5**.

6.2.2 Crystal structure analysis

In addition to the already published crystal structures of TLN-**2**, TLN-**3** and TLN-**4** [194], we succeeded in crystallizing TLN-**1**, TLN-**5**, and TLN-**6** (Table 6.1). Furthermore, we also obtained TLN-**1** in complex with xenon (TLN-**1**-Xe) and krypton (TLN-**1**-Kr). Except for the native structure of TLN-**1** that was experimentally phased, all other crystal structures were phased by the standard molecular replacement technique.

6.2.3 Shape, polarity and solvent content analysis within the S_1' cavity and within three internal reference cavities of TLN-1

The structure of TLN-**1** was experimentally phased using a zinc MAD dataset (Table 6.2). The experimental phasing of TLN-**1** without any further least-squares refinement steps resulted in a very clear, high-quality electron density (Supporting Information, Figure S6.3) with an overall figure of merit (FOM) of 0.870. We decided to apply density modification techniques (solvent flattening and histogram matching) on the experimental phases since this resulted in a significantly improved quality of the electron density map (overall FOM increased from 0.763 to 0.870). Since the analyzed cavities are completely buried within the core of TLN and exhibit a narrow shape, the electron density within the cavities is not affected by the applied phase improvement techniques. The experimentally phased density map was put on an absolute electron number scale (see Chapter 6.5.8 for further details), and the total electron content within the S_1' cavity and three internal reference cavities of different solvation state was calculated. The protein model coordinates displayed in Figure 6.2A–E are taken from the experimentally phased, fully refined and deposited model of the native TLN-**1** (PDB code 5M9W), whereas the superimposed electron density map is the experimentally phased electron density without any further model-based refinement. In the following, the cavities and electron density maps are described relative to the view of Figure 6.2B–E.

Table 6.1. X-ray data collection and refinement statistics.[a]

	Complex (PDB code)				
	TLN-**1** (5M9W)	TLN-**1**-Xe (5M69)	TLN-**1**-Kr (5M5F)	TLN-**5** (5LVD)	TLN-**6** (5MA7)
(A) Data collection and processing					
Space group	P6$_1$22	P6$_1$22	P6$_1$22	P6$_1$22	P6$_1$22
Unit cell parameters: a, b, c (Å)	92.5, 92.5, 130.5	92.5 92.5, 129.9	92.4, 92.4, 130.2	92.8, 92.8, 130.4	92.8, 92.8, 129.8
Wavelength (Å)	0.91841	1.40000	0.85940	0.91841	0.91841
Matthews coefficient (Å^3/Da)[b]	2.3	2.3	2.3	2.4	2.3
Solvent content (%)[b]	47	47	47	48	47
(B) Diffraction data					
Resolution range (Å)	50.00–1.21 (1.28–1.21)	50.00–1.44 (1.53–1.44)	50.00–1.33 (1.41–1.33)	50.00–1.25 (1.32–1.25)	50.00–1.30 (1.38–1.30)
Unique reflections	100158 (15893)	59253 (9235)	75575 (11969)	91830 (14553)	80990 (12879)
R_{sym} (%)	6.2 (48.8)	10.2 (48.6)	8.1 (49.9)	6.6 (49.5)	5.3 (49.0)
Wilson B factor (Å^2)	9.9	12.3	10.1	9.4	10.5
Completeness (%)	99.9 (99.5)	99.0 (97.2)	100.0 (99.7)	99.9 (99.3)	99.8 (99.6)
Redundancy	12.9 (12.5)	37.2 (37.1)	20.7 (19.5)	24.9 (24.7)	9.8 (8.7)
$<I/\sigma(I)>$	24.7 (4.9)	28.0 (8.1)	26.0 (6.3)	35.2 (7.0)	26.1 (4.4)
(C) Refinement					
Resolution range (Å)	43.60–1.21	46.24–1.44	43.55–1.33	38.24–1.25	43.67–1.30
Reflections used in refinement (work/free)	95148/5007	56290/2963	71796/3779	87232/4591	76939/4050
Final R value for all reflections (work/free) (%)	10.7/12.7	10.8/14.1	11.1/13.9	10.5/12.5	11.2/13.9
FOM	0.95	0.94	0.94	0.95	0.94
Phase error (°)	9.2	11.0	10.6	9.0	10.9
Protein residues	316	316	316	316	316
Calcium/zinc ions	4/1	4/1	4/1	4/1	4/1
Xenon/krypton atoms	0/0	2/0	0/2	0/0	0/0
Inhibitor atoms	28	28	28	30	30
Water molecules	344	332	384	386	380
RMSD from ideality:					
Bond lengths (Å)	0.012	0.013	0.011	0.009	0.010
Bond angles (°)	1.2	1.2	1.2	1.1	1.2
Ramachandran plot:[c]					
Residues in most favored regions (%)	88.9	89.6	87.8	87.8	88.1
Residues in additionally allowed regions (%)	10.0	9.3	11.1	11.1	11.1
Residues in generously allowed regions (%)	0.7	0.7	0.7	0.7	0.4
Residues in disallowed regions (%)[d]	0.4	0.4	0.4	0.4	0.4
Mean B factors (Å^2):[e]					
Protein non-hydrogen atoms	11.7	13.8	11.3	11.1	12.3
Inhibitor	14.2	20.5	13.3	12.6	19.3
Water molecules	28.1	30.3	26.2	27.2	28.7

[a] Numbers in brackets stand for the highest resolution shells. [b] Calculated using the program *Matthews_coef* from the *CCP4* suite [238]. [c] Calculated using *PROCHECK* [196]. [d] The Ramachandran outlier Thr26 of TLN is described in the literature [197]. [e] Calculated using *MOLEMAN* [198].

Chapter 6

Table 6.2. Zinc MAD dataset of TLN-**1** used for experimental phasing. [a]

	peak	inflection	remote
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: a, b, c (Å)	92.5, 92.5, 131.0	92.6, 92.6, 131.0	92.6, 92.6, 131.0
Wavelength (Å)	1.281960	1.283060	0.918410
Resolution range (Å)	50.00–1.50 (1.59–1.50)	50.00–1.50 (1.59–1.50)	50.00-1.32 (1.40-1.32)
Unique reflections	91983 (11525)	91926 (11423)	137883 (17564)
R_{sym} (%)	4.3 (17.2)	3.6 (14.5)	5.4 (46.0)
Completeness (%)	91.4 (71.0)	91.3 (70.3)	93.2 (73.4)
Redundancy	7.8 (7.5)	7.8 (7.5)	7.9 (7.5)
$<I/\sigma(I)>$	31.0 (9.8)	37.0 (12.1)	24.2 (4.2)

[a] Numbers in brackets stand for the highest resolution shells.

The **S_1' cavity** (Figure 6.2B) comprises a volume of 141 Å^3. The top of the cavity is exclusively formed by apolar amino acid side chains of protein residues in addition to the ligand atoms of the P_1' group and the leucine P_2' portion of **1**. The mid- to lower left part of the cavity is mainly apolar, except for Asp138, that, however, is inaccessible for hydrogen bonding due to its buried geometry. The polar side chains of Glu143, Asp170, and Arg203 describe the right-hand surface portion of the cavity. However, except for the side chain of Glu143, they are all highly involved in saturating polar contacts and thus not available for hydrogen bonding with putative occupants of the S_1' pocket. The total electron content integrated over the entire volume of the S_1' cavity is 6.6 e^-.

Reference cavity 1 (Figure 6.2C) is a highly polar cavity comprising a volume of 59 Å^3. It contains three water molecules that establish multiple hydrogen bonds (2.8–3.2 Å) to adjacent polar functional groups of TLN residues. The observed firm density peaks comprise spherical shapes and have been refined to fully occupied water molecules with low B factors of 11–12 Å^2. The integrated total electron content within this reference cavity 1 is 19.4 e^-.

Reference cavity 2 (Figure 6.2D) opens a volume of 93 Å^3. In the lower part of the pocket, a strong electron density peak originating from a water molecule (refined B factor: 10 Å^2) is detected. This water molecule establishes three hydrogen bonds with Tyr81 (backbone O, 2.9 Å), Arg90 (backbone O, 2.7 Å), and Ser92 (backbone N, 2.8 Å). In contrast, the upper part of the cavity has a highly apolar character. Apart from the described peak assigned to the hydrogen-bonded water molecule, a second, far less intense electron density peak is observed in the center of the large apolar cavity (maximum concentration of 0.62 e^-/Å^3, Figure 6.2D). The distance of this peak to that of the already assigned water molecule is 3.2 Å. The integrated total electron content within reference cavity 2 is 13.4 e^-.
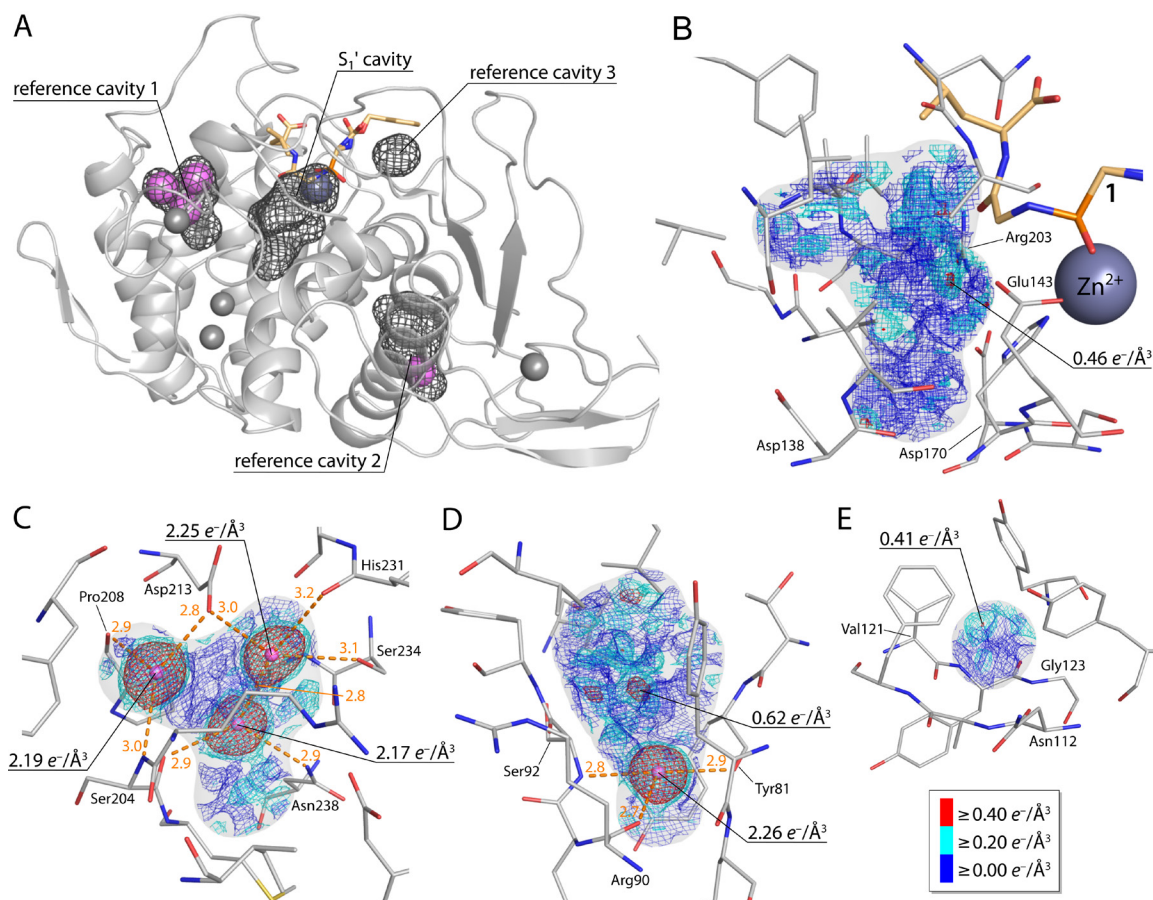
Figure 6.2. Analysis of the experimentally phased electron density map of TLN-**1** within selected cavities. **(A)** Ribbon model of the refined TLN-**1** indicating the solvent excluded surface of the S$_1$' cavity and of the internal reference cavities 1–3 as dark gray mesh. Water molecules within these cavities are shown as magenta spheres. The zinc ion is shown as dark blue sphere and the four calcium ions of TLN are shown as gray spheres. The TLN-bound ligand **1** is indicated as light orange stick model. **(B–E)** Depiction of the experimentally phased absolute-scale electron density map detected within the **(B)** S$_1$' cavity (volume 141 Å^3; total electron content: 6.6 e^-), **(C)** reference cavity 1 (volume 59 Å^3; total electron content: 19.4 e^-), **(D)** reference cavity 2 (volume 93 Å^3; total electron content: 13.4 e^-), and **(E)** reference cavity 3 (volume 16 Å^3; total electron content: 0.9 e^-). The electron density map is displayed as blue, cyan or red mesh indicating three different contour levels in e^-/Å^3. Selected density peaks are labeled with their concentration maximum. Cavity lining TLN residues are shown as thin gray stick models. The solvent excluded surfaces of the cavities are indicated in semi-transparent gray. In panel B, the TLN-bound **1** is shown as light orange stick model, and the zinc ion is shown as dark blue sphere. In panels C and D, water molecules are shown as magenta spheres and hydrogen bond interaction distances as orange dashed lines (labeled in Å). The determined volume and total electron content of the cavities are summarized in Table S6.7 (Supporting Information). The electron density maps of all four cavities are shown at six different contour levels from –0.2 to 0.3 e^-/Å^3 in Figure S6.4 to Figure S6.7.

Chapter 6

Reference cavity 3 (Figure 6.2E) is, apart from the S_1' cavity, the largest cavity which remains non-solvated in the refined structure of TLN-**1**. Its volume of 16 Å³ is significantly smaller than that of the S_1' cavity. Its character is primarily apolar, and it is flanked by polar atoms of Asn112 (backbone O), Val121 (backbone O), and Gly123 (backbone N). The integrated total electron content within reference cavity 3 is 0.9 e^-.

6.2.4 Xenon and krypton derivatization of TLN-1

Pressure derivatization of TLN-**1** with xenon and krypton was performed at 9 and 5 bar, respectively. The position of bound xenon and krypton in the crystal structure can be unambiguously identified in the electron density due to their anomalous scattering properties [294]. To optimize the anomalous signal of xenon, a wavelength of 1.4 Å (f' = 6.3 e^-) was chosen, resulting in a strong anomalous signal without significantly losing scattering power due to the long wavelength and air absorption of the X-ray beam (Table 6.1). In the case of krypton, two datasets of the same crystal were collected directly above and below the K absorption edge of krypton to unambiguously identify bound atoms of the noble gas (Table 6.3). Even though the difference between the wavelengths is only about 0.024 Å, the change of the anomalous signal of krypton is drastic, whereas its F_o–F_c electron density virtually remains unaffected (Supporting Information, Figure S6.8). All datasets of the TLN-**1** noble gas derivatives were collected with high redundancy to maximize the accuracy of the anomalous signal (Table 6.1).

Table 6.3. Datasets of noble gas derivatized TLN-**1** used for the determination of the anomalous signal (Friedels pairs separated).[a]

	TLN-**1**-Kr (above K edge)	TLN-**1**-Kr (below K edge)	TLN-**1**-Xe
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: a, b, c (Å)	92.4, 92.4, 130.2	92.5, 92.5, 130.2	92.5, 92.5, 130.0
Wavelength (Å)	0.859400	0.871500	1.40000
Resolution range (Å)	50.00–1.33 (1.41–1.33)	50.00–1.35 (1.43–1.35)	50.00–1.44 (1.53–1.44)
Unique reflections	142833 (23002)	136781 (22091)	111687 (17774)
R_{sym} (%)	7.8 (48.7)	8.2 (46.6)	9.9 (48.0)
Completeness (%)	99.9 (99.7)	99.9 (99.6)	99.2 (97.6)
Redundancy	11.0 (10.2)	8.6 (8.0)	19.8 (19.3)
$<I/\sigma(I)>$	19.2 (4.6)	15.6 (4.0)	20.9 (5.9)

[a] Numbers in brackets stand for the highest resolution shells.

For both noble gases, binding was observed in the S$_1$' cavity (refined occupancy of xenon: 17%, krypton: 8%) as well as in reference cavity 3 (refined occupancy xenon: 68%, krypton: 20%). The binding of xenon within the upper, highly apolar part of the deeply buried reference pocket 3 has been previously reported by us (unpublished results, PDB code 3LS7). Both types of noble gases populate at the same positions in the S$_1$' cavity as well as in the upper part of reference cavity 3. In the S$_1$' cavity, van der Waals interactions (up to 4.6 Å) are established to the side chains of Val139, His142, Glu143, Ile188, Leu202, Arg203, and the portion of **1** covering the S$_1$' pocket (Figure 6.3). No adaptations of cavity-lining residues of the noble gas derivatized TLN-**1** are observed compared to the native structure of TLN-**1**.
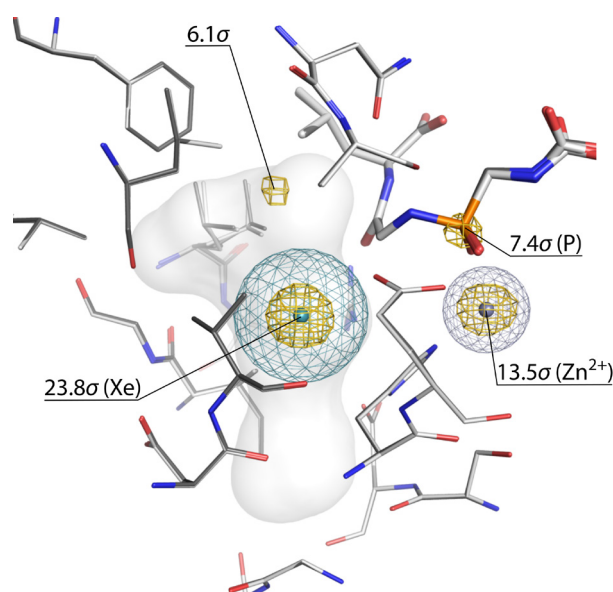
Figure 6.3. Xenon binding site in the S$_1$' cavity of TLN-**1**. The xenon derivatized crystal structure TLN-**1**-Xe (dark gray) is superimposed on the native crystal structure of TLN-**1** (light gray). The center of the bound xenon atom is shown as teal sphere, and its van der Waals radius is indicated as teal mesh. The zinc ion is shown in similar fashion in dark blue. Cavity surface lining TLN residues are shown as thin sticks, the bound **1** is shown as thick stick model. The solvent excluded surface of the S$_1$' cavity is indicated in semi-transparent white. The anomalous map is displayed in gold at a contour level of 5σ. The crystal structure of TLN-**1**-Kr is shown in Figure S6.8.

6.2.5 Comparison between the S$_1$' cavities of TLN-1 to TLN-6

In none of the refined, σ-scaled F_o–F_c electron densities of the six crystal structures TLN-**1** to TLN-**6**, any clearly defined electron density attributable to a bound water molecule could be detected in the S$_1$' cavity. Even in the complexes of TLN with the ligands exhibiting the polar P$_1$' groups (**5** and **6**), the unoccupied part of the cavity remains seemingly empty, even though, as indicated in Figure 6.4, sufficient space to accommodate a water molecule is given.
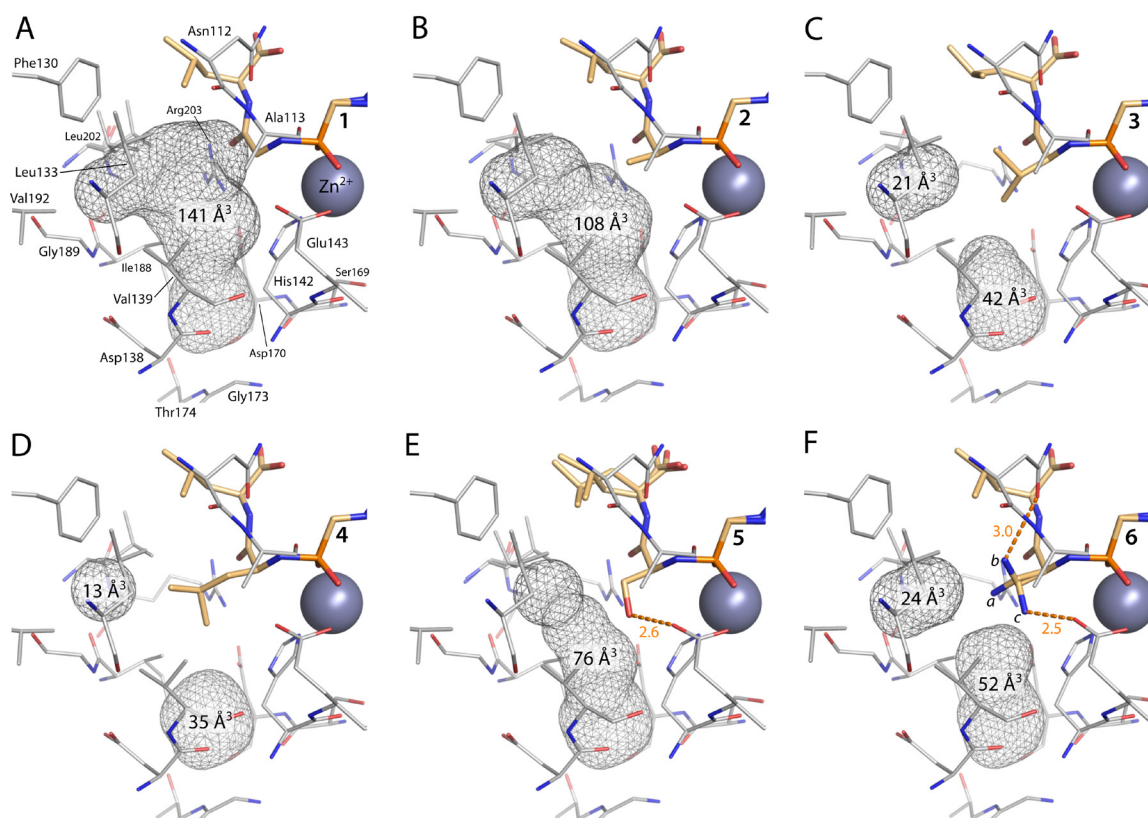
Figure 6.4. S$_1$' cavities of crystal structures TLN-**1** to TLN-**6**. **(A)** TLN-**1**, **(B)** TLN-**2**, **(C)** TLN-**3**, **(D)** TLN-**4**, **(E)** TLN-**5**, **(F)** TLN-**6**. The solvent excluded surfaces of the S$_1$' cavities are displayed as dark gray mesh labeled with their volume in Å^3. Cavity lining amino acids are displayed as gray stick models. Ligands **1**–**6** are shown as thick stick models in light orange. The zinc ion is displayed as a dark blue sphere. Hydrogen bonds formed between TLN and the P$_1$' group of **5** and **6** are indicated as orange dashed lines (labeled in Å). The three conformations of the P$_1$' group of **6** are labeled with *a*–*c*. All three conformations of the latter P$_1$' substituent were considered for the calculation of the residual cavity volume.

Figure 6.4 displays the crystal structure models of TLN-**1** to TLN-**6** and the residual S$_1$' cavities. With growing P$_1$' portion (**1**→**4**), the top part of the cavity is gradually occupied, until only the bottom part of the cavity remains unoccupied (Figure 6.4A–D). Accordingly, the volume of the S$_1$' cavity is gradually decreasing from 141 Å^3 to 48 Å^3 (Figure 6.4). The hydrophobic P$_1$' groups of TLN-**2**, TLN-**3** and TLN-**4** form multiple hydrophobic van der Waals interactions to the pocket-shape determining amino acid side chains of TLN. Whereas the P$_1$' OH group of the serine derivative is found in only one populated conformation forming a hydrogen bond to the side chain of Glu143 (2.6 Å, Figure 6.4E), three conformations (occupancy *a*: 42%, *b*: 27%, and *c*: 31%) can be assigned to the P$_1$' amino function of **6** in the refined model (Figure 6.4F). Conformations *b* and *c* place the amino group within hydrogen-bonding distance to the side chain of Asn112 (3.0 Å) and Glu143 (2.5 Å).

6.3 Discussion

6.3.1 Analysis of the experimentally phased electron density within the S_1' cavity of TLN-1

Even with high-resolution data as in the current case, the analysis of the solvation pattern within a protein cavity by X-ray crystallography is by no means straightforward, since highly mobile ("disordered") or partially occupied water molecules are difficult to detect on first glance. A highly mobile water molecule may be characterized by the lack of a well-defined, sufficiently deep energy minimum on the free energy landscape, resulting in a widely distributed, rather blurred electron density lacking a clearly defined center [295]. Model bias introduced into the electron density via the use of phases transferred from the refined structure of a related complex can obscure the detection of such weak density signals. In particular, the use of the commonly applied molecular replacement as phasing technique can result in a significant impact of model bias [296–298]. Experimentally phased electron densities have the advantage that no model bias is arbitrarily introduced by the application of predefined model phases [293, 299, 300]. Therefore, to reliably detect traces of electron density originating from highly mobile or partly occupied water molecules, we performed an elaborate experimental phasing protocol for TLN-**1**. Even though unbiased phases can be obtained by this procedure, another obstacle arises from the fact that the total electron number within the crystal unit cell F_{000} is impossible to determine experimentally in the case of proteins. In small molecule crystallography, the content of the diffracting unit cell is usually easy to define and thus the number of contributing electrons is clear. In the case of proteins showing a large solvent content particular in the channels passing through the crystal, such an assignment is impossible and thus can only be estimated. In consequence, electron densities in protein crystallography are typically σ-scaled [301], where zero σ corresponds to the average, numerically unknown electron concentration of the unit cell. Even though this step appears very reasonable and pragmatic in usual standard refinements of protein structures, it will make an assignment of an absolute electron number of electrons to a particular integrated volume virtually impossible. Therefore, we attempted to transform the experimentally phased electron density to an absolute electron number scale, where the zero value corresponds to vacuum. Following a similar approach, Liu et al. analyzed a hydrophobic cavity of 134 Å^3 in L99A/M102L T4 lysozyme [300]. In this case, analysis of the experimentally phased electron density on absolute scale discovered a water cluster of three water molecules with an occupancy of approximately 50% within this pocket. In another example, the analysis of an experimentally phased electron density of a central apolar cavity of 40 Å^3 observed in interleukin-1ß settled the debate about the presence of a disordered water molecule in this volume [299, 302–304]. From the data analysis, Quillin et al. concluded that the cavity must indeed be empty [299].

Chapter 6

Reference cavity 3 is considered as an internal control for an entirely empty cavity (Figure 6.2E). This small, apolar cavity of 16 $Å^3$ is due to its size and polarity unsuited to host a water molecule [305–307]. With respect to the calculated volume of 16 $Å^3$ for reference pocket 3 it must be considered that the calculation of the volume that is really available to host a water molecule is non-trivial and is actually further reduced adjacent to apolar residues [308, 309]. In other words, the solvent probe radius adjacent to apolar residues is in fact larger than the commonly applied 1.40 Å. Consequently, the presence of a water molecule in reference cavity 3 can be certainly excluded. The electron concentration peak found within this cavity comprises a concentration maximum of 0.41 $e^-/Å^3$, and the total electron content is 0.9 e^-. We consider this electron content the result of "spillover" of electrons from cavity lining amino acids into the to some degree fuzzy volume of the empty cavity [299, 300], e.g. due to a slight movement of the protein residues. Reference cavities 1 and 2 (Figure 6.2C and D) serve as an internal control for clearly solvated cavities. The calculated total electron content within reference cavity 1 (59 $Å^3$) is 19.4 e^-, a number significantly lower than expected for three fully occupied water molecules (30 e^-). Similar as observed by Liu et al. [300], this is possibly the result of the electron spillover from within the cavity to the outside due to overlapping van der Waals radii of polar cavity-lining atoms and the tightly hydrogen-bonded water molecules. Reference cavity 2 (93 $Å^3$) contains a total electron content of 13.4 e^-. Apart from the unambiguous water density in the lower part of the cavity, a significant amount of blurred electron density with a peak maximum of 0.62 $e^-/Å^3$ is detected in the balloon-like shaped upper, hydrophobic part of the cavity (Figure 6.2D) — which is large enough to host a phenol molecule without that the water molecule in the lower part of the cavity is displaced [191]. The electron concentration maximum is within hydrogen bonding distance (3.2 Å) to the modeled water molecule in the lower part of the cavity and could potentially originate from a second, very low populated water site. Overall, this results in a higher electron content than expected for the single fully occupied water molecule, overcompensating the also in this case expected spillover of electrons of the tightly bound water molecule to the outside of the cavity. Remarkably, the residual electron distribution within the S_1' cavity shows an electron concentration of maximally 0.46 $e^-/Å^3$, a value close to the one found for the empty reference cavity 3 (0.41 $e^-/Å^3$). The total electron content within the S_1' cavity was determined to be 6.6 e^-. If the electron content within the empty reference cavity 3 (0.9 e^-) is extrapolated to the larger volume of the S_1' cavity (141$Å^3$/16 $Å^3$ = 8.8 times larger), this corresponds to 7.9 e^- (0.9 $e^- \times 8.8$), a number higher than the one for the experimentally determined electron content within the S_1' cavity (6.6 e^-). This can be explained by the fact that relative to their volumes reference cavity 3 has a significant larger surface than the S_1' cavity, resulting in a proportionally larger electron "spillover" (*vide supra*). Taking into account these considerations we therefore propose that the TLN-**1** S_1' cavity is indeed a completely unsolvated cavity. We also believe that this is not simply a consequence of the binding of **1**

which seals the pocket from the top. Our complex structures of TLN-**1** with the noble gases clearly demonstrate that the pocket is still well accessible in the crystal by particles as large as a xenon atom, likely due to the residual mobility of the complexes. Thus, smaller water molecules could easily access the S_1' pocket if their penetration would be favorable.

6.3.2 Analysis of TLN-1 in complex with xenon and krypton and of crystal structures TLN-1 to TLN-6

The noble gases xenon and krypton are known to preferentially bind to desolvated, hydrophobic protein cavities through weak van der Waals interactions [310, 311]. These atoms can, therefore, be used as experimental probes to detect such cavities. The S_1' cavity (Figure 6.3) and reference cavity 3 were revealed as noble gas binding sites in TLN-**1**. The fact that both noble gases were found binding to the S_1' cavity even though the derivatization pressure was kept low supports the hypothesis that both cavities do not contain significantly populated solvent molecules, that need to be displaced upon ligand accommodation. The increased occupancy of the gaseous probes in reference cavity 3 compared to the S_1' cavity can be attributed to the deeper burial of this cavity within the apolar interior of TLN. Thus, during the experimental depressurization phase, that needs to be accomplished to transfer the crystal specimen from the pressuring cell to the liquid nitrogen, diffusion of the noble gases is slower. The reduced occupancy of krypton compared to xenon is attributable to the reduced polarizability of the former [312, 313], the lower applied derivatization pressure, and the faster diffusion kinetics of the smaller krypton atoms. Noble gas binding provides a crude estimate about the at least available space within a given cavity, and provides information about an at least detectable electron concentration in a certain volume [312, 314]. For instance, in TLN-**1**-Kr (Figure S6.8), the krypton atom (36 e^-, van der Waals radius 2.0 Å, volume 33.5 Å^3) bound to the S_1' cavity was refined to 8% occupancy. This corresponds to 2.9 electrons (36 e^- × 0.08). Considering the refined temperature factor of 12.5 Å^2 of this krypton atom, krypton will occupy a volume of about 58 Å^3 [295, 314]. Hence, if we transfer this detection limit to our cavity analysis, a putative water molecule occupied as low as 29% (corresponding to 2.9 e^-) and distributed over a volume of 58 Å^3 should be detectable by conventional refinement. If we apply this estimation rule to the S_1' cavities of complexes TLN-**1** to TLN-**6**, it is permissible to conclude that the residual volumes found in TLN-**3** to TLN-**6** (48–76 Å^3, Figure 6.4) must be virtually empty. This is further supported by the observation that the S_1' cavities of TLN-**3**, TLN-**4** and TLN-**6** are divided by the P_1' groups of the bound ligands into two spatially separated cavities, reducing the mobility of a putatively present water molecule even further. Consequently, a potentially present — but nonetheless undetectable — water molecule must be very low occupied to elude crystallographic detection completely. We believe that this minor displacement effect would be insignificant

for the binding event and would hardly influence the overall thermodynamic profile. Thus, for the discussion of the thermodynamic signature, we refer to the empty state of the cavity.

The polar groups of **5** and **6** were introduced into the S_1' pocket with the aim to provide a local anchor for hydrogen bonding for a potentially present and disordered water molecule. If such a water molecule would hypothetically be present in the complexes with the aliphatic ligands, this hydrogen-bonding decoy should result in the stabilization of a disordered water molecule — or possibly in the recruitment of a new water molecule from the bulk water phase — and make it crystallographically detectable [315, 316]. However, no clear electron density peak is observed in the crystal structures of TLN-**5** or TLN-**6**. One reason to be regarded why the polar groups are not available to establish hydrogen bonds to a putative water molecule might arise from the fact that they form hydrogen bonds to protein residues. However, conformation *a* with the highest population of the amino function of **6** orients toward the void of the cavity where it could experience hydrogen bonding (Figure 6.4F). Another reason could be that even though the polarity of the S_1' cavity of TLN-**5** and TLN-**6** increases, the volume of the cavity is reduced, which decreases the probability to find a water molecule within a cavity [306].

6.3.3   Thermodynamic binding profiles of 1–6 as determined by ITC

The affinities of the investigated ligand series fall into the range between milli- to nanomolar binding (Table S6.1). Because of this broad range, it was necessary to apply different measurement protocols to obtain reliable calorimetric data. Ligands **2**, **3** and **4** were measured by direct titration. Due to their low affinities, **1**, **5** and **6** had to be characterized by displacement titrations. The measurement accuracy of the latter ligands is strongly dependent on how accurately the thermodynamic data of the applied reference ligand has been recorded. Any error in the thermodynamic profile of the reference ligand will propagate to the thermodynamic profile of the analyte. Consequently, such measurements usually result in large experimental errors, in particular considering the partitioning of enthalpy and entropy [96]. The affinity of **1** is too low to accurately determine the enthalpy/entropy signature via displacement titration. Furthermore, **5** and **6**, comprising polar P_1' groups, show very likely different changes in their protonation state compared to the ligands with the aliphatic P_1' groups. A superposition of a proton transfer reaction can alter the observed enthalpy of binding and therefore prevent a direct comparability of measured enthalpy values [152]. Unfortunately, an appropriate enthalpy correction is not reliably feasible in a displacement titration scenario due to the resulting highly potentiating errors of the buffer-corrected thermodynamic profiles. Consequently, we decided to refrain from analyzing the

enthalpy/entropy profiles of **1**, **5** and **6**, and instead solely report their affinities. This value is independent of buffer effects and allows accurate comparison with the more potent ligands.

The addition of a single methyl group to the glycine derivative **1** (yielding **2**) results in a more than 100-fold affinity increase (Figure 6.1B and Table S6.1). Adding two further methyl groups to **2** (yielding **3**) results in a further 40-fold affinity increase. Finally, the addition of a further methyl group from **3** to **4** corresponds to a 7-fold affinity increase. Based on a statistical evaluation, a 100-fold affinity increase solely achieved by the addition of one single methyl group has been reported to be a very rare event with a probability of only 0.4% [317]. In the reported cases, the favorable affinity increase resulted from the placement of a methyl group into hydrophobic pockets that the authors assumed to be entirely empty, in combination with an energetically favorable pre-organization of the ligand conformation in solution. This assumption is supported by MD simulations suggesting that the placement of hydrophobic groups into supposedly empty, hydrophobic pockets produces an extraordinarily favorable change of the Gibbs free energy of binding due to the absence of a cavity desolvation step [318, 319]. In the same manner, the extraordinary affinity increase from **1** to **2** (>100-fold) can be attributed to the insertion of a methyl group into a hydrophobic cavity, that is — as experiment confirms — virtually empty. This effect is reduced, but nevertheless still pronounced when further methyl groups are grown into this volume (40-fold affinity increase between **2** and **3**). From **3** to **4**, the comparably low 7-fold affinity increase is the result of a highly favorable enthalpic term, that, however, is partly compensated by unfavorable entropy (Figure 6.1B). Usually, dispersive (van der Waals) interactions are less significant, because such interactions established between protein and ligand are largely canceled out by the required unfavorable disruption of van der Waals interactions between protein and solvent [60]. However, in the case of a vacated pocket that makes a desolvation step obsolete, dispersive interactions formed between protein and ligand upon complexation become determinant in terms of affinity [60]. This is the case in the current study — there is no cost to desolvate the S_1' binding pocket. As a result, the contribution of the established dispersive interactions between protein and ligand to the enthalpy of binding increases with increasing P_1' chain length, and overall strongly affects binding affinity. However, with increasing P_1' chain length, the enthalpic signal is increasingly compensated by a decrease in entropy (Figure 6.1B). This is potentially the result of a loss of conformational degrees of freedom upon complex formation. The low affinities of **5** and **6** can be attributed to the large energetic penalty for the desolvation of their polar functional groups which is not overcompensated by the dispersive interactions with the cavity-lining residues. The hydrogen-bond interactions formed with Glu143 (and Asn112 in the case of **6**) do not suffice to compensate for this loss. Thus, the free energy gain resulting from the establishment of dispersive interactions is largely compensated by the high cost to desolvate **5** and **6**, overall lowering their affinity.

Protein cavities can fulfill essential biological functions, e.g. conformational flexibility, and thus can represent more than "packing defects" [311, 320, 321]. In the case of TLN (and many other metalloproteinases), the S_1' pocket is the most important pocket to discriminate substrates from non-substrates, and thus defines the selectivity profile of the protease. A substrate that exhibits shape complementary and thus fills the hydrophobic S_1' pocket efficiently without requiring a large price for desolvation (e.g. a P_1' leucine, *iso*-leucine or valine side chain) will experience strong dispersive interactions with the protease. It thus gets bound and enzymatically processed. In contrast, if the pre-organized void in the protease is insufficiently filled by an either small or hydrophilic P_1' substrate portion, a pronounced affinity reducing enthalpic penalty results. Thus, the unsolvated state of the TLN S_1' pocket is highly important for the selectivity mechanism of the protease.

An entirely independent proof for the existence of unsolvated protein binding pockets is provided by a current neutron diffraction study [322]. Neutron scattering can reliably differentiate between hydrogen and deuterium. We recently determined the joint X-ray/neutron structures of two trypsin complexes. Crystal growth was performed with the protein in its hydrogen form and crystals were subsequently exposed to fully deuterated buffer over 532 days before data collection. Under such conditions polar hydrogen atoms should exchange by deuterium atoms considering the large excess of deuterons compared to protons. However, as a precondition the polar groups must be accessible to D_2O molecules. In a folded protein, polar hydrogens usually remain only at sites where they are involved in strong hydrogen bonds. Nevertheless, in a sterically accessible pocket of our trypsin complexes we found NH groups where the polar hydrogens are not involved in a hydrogen bond and these NH groups did hardly exchange to ND over the 1.5 years of incubation in D_2O. Only 7% deuterium could be found at these sites. We therefore hypothesize that they hardly experienced any contact with D_2O molecules in the folded protein indicating that these cavities are extremely rarely accessed by water molecules. We believe this is another independent indication that empty and hardly unsolvated pockets exist in folded proteins.

6.4   Conclusion

The discussed experimental data strongly suggest that the S_1' cavity of TLN in complex with ligands **1–6** is not solvated and contains virtually vacuum prior to ligand binding. We found no experimental evidence for the occupation with highly mobile water molecules, and the properties of the cavities (shape, volume, polarity) suggest that the cavity is indeed empty. It seems energetically more favorable to maintain a vacuum than to host one or several water molecules within these cavities. It has been discussed that the generation of vacated pockets in proteins is energetically very costly [323, 324]. Nevertheless, it has to be considered that

the costs for producing such a void have to be afforded during protein-folding and not during ligand or substrate binding.

In the current case, the fact that the TLN S$_1$' cavity is empty takes major consequences on the energetically highly favored accommodation of aliphatic P$_1$' side chains of either substrate or inhibitor molecules. The observed enthalpy driven affinity enhancement with increasing size of the P$_1$' substituent (**1→4**) is mainly a result of the binding of aliphatic groups into a void, where no price for pocket desolvation has to be afforded. The decreasing affinity contribution of a growing side chain can be attributed to the augmenting desolvation penalty of the larger P$_1$' substituents and to a reduction in conformational flexibility. Addition of a polar group entirely destroys binding affinity due to an uncompensated desolvation penalty. Remarkably, the derivative with a P$_1$' benzyl side chain (PDB code 3FV4) that fills the S$_1$' pocket more efficiently than the P$_1$' leucine side chain of **4** is less potent, likely due to the higher desolvation costs for the aromatic side chain [194]. In the case of TLN, the hydration state of the S$_1$' cavity seems to have a pronounced contribution to the selectivity profile of the protease. The remarkable, about 41,000-fold increase in affinity from the introduction of hydrophobic *iso*-butyl group into this pocket also has major implications for medicinal chemistry. The identification of poorly hydrated cavities can represent a valuable strategy to gain overwhelmingly in binding affinity of a prospective drug molecule.

Chapter 6

6.5 Experimental section

6.5.1 Ligand synthesis

The synthesis of **1**–**6** has been reported previously [325].

6.5.2 Crystal preparation and soaking

Crystals were prepared with lyophilized TLN powder commercially obtained from Calbiochem (EMD Biosciences) identical to the procedure as described earlier [66]. For crystal soaking of the low-affinity compounds **1**, **5** and **6**, TLN crystals were transferred into a soaking solution composed of 100 mM Tris-HCl, pH 7.5, 2 mM $CaCl_2$ and 20% DMSO saturated with the respective ligand (ligand precipitate visible), followed by incubation for 24 hrs. Afterward, crystals were flash-frozen in liquid nitrogen after a brief immersion in a cryoprotectant solution saturated with the respective ligand, composed of 10 mM Tris-HCl, pH 7.5, 10 mM $Ca(CH_3COO)_2$, 20%DMSO, and 20% glycerol.

6.5.3 Derivatization of TLN with xenon and krypton

For the noble gas derivatization of the TLN crystals (TLN-**1**-Xe and TLN-**1**-Kr), a pressure cell from Oxford Cryosystems (Long Hanborough, UK) was used [326]. Before derivatization, TLN crystals were soaked with **1** in the above-mentioned soaking buffer for 24 hrs. To protect the crystals from drying out during the pressurization phase, the filter paper of the pressurization cell was drenched with soaking buffer. Subsequently, xenon derivatization was performed at 9 bar for 5 minutes, and derivatization with krypton was carried out at 5 bar for 5 minutes. Derivatization was conducted at relatively low pressure because higher pressure resulted in a strong increase in crystal mosaicity. After the incubation time was up, pressure was quickly released, and TLN crystals were immediately after flash-frozen in liquid nitrogen.

6.5.4 Data collection

Datasets for crystal structures TLN-**1**, TLN-**5**, and TLN-**6** (Table 6.1) were collected at BESSY II (Berlin-Adlershof, Germany) on beamline 14.1 on a Dectris Pilatus 6M pixel detector at 100 K [210, 327]. A three-wavelength MAD dataset of TLN-**1** (Table 6.2) which was applied for the determination of the experimental phases was collected from another crystal at BESSY II on beamline 14.2 on a Rayonix MX-225 pixel detector at 100 K. Datasets for crystal structures TLN-**1**-Xe and TLN-**1**-Kr (Table 6.1) were collected at Elettra (Triest, Italy) on

beamline XRD1 at 100 K on a Dectris Pilatus 2M pixel detector. The dataset for TLN-**1**-Xe was collected at λ = 1.40000 Å. Two datasets were collected for TLN-**1**-Kr from the same crystal (Table 6.3), the first at 0.85940 Å above the krypton *K* absorption edge (that is at 0.86550 Å), against which the structure was refined, and a second dataset at 0.871500 Å below the krypton *K* absorption edge. The second dataset was solely used for analysis of the change of the anomalous signal of krypton to unambiguously identify the bound atom. The anomalous maps TLN-**1**-Xe (Figure 6.3) and TLN-**1**-Kr (Figure S6.8) were created with *ANODE* [328].

6.5.5 Dataset processing and structure determination

All data sets were indexed, integrated and scaled with *XDS* [252]. The phases of TLN-**1**-Xe, TLN-**1**-Kr, TLN-**5**, and TLN-**6** were determined by molecular replacement applying *Phaser* [213] from the *CCP4* suite (version 6.3.0) [238]. The crystal structure of the PDB entry 8TLN was used as molecular replacement search model [212]. The phases for TLN-**1** were determined experimentally by a MAD dataset (Table 6.2) using the intrinsically bound zinc ion of TLN as anomalous scatterer. Experimental phasing was performed applying the programs *SHELXC* (data preparation), *SHELXD* (heavy atom substructure determination) [329] and *SHELXE* (experimental phasing and density modification) [330] as implemented in *HKL2MAP* [331]. The anomalous signal d"/σ of the peak dataset was significant (1.42) to 1.62 Å. Between all datasets the anomalous correlation coefficient was above 30% up to 1.57 Å. The best solution of the *SHELXD* substructure search was CC$_{All}$=71.3 and CC$_{Weak}$= 58.9. The substructure phases calculated by *SHELXE* from the original hand gave a final contrast of 0.56 and a connectivity of 0.81 (inverted hand: contrast 0.35, connectivity 0.66). The experimentally determined phases were combined with the amplitudes of a 1.21 Å native dataset of TLN-**1** (Table 6.1) followed by density modification with the program *DM* [332] from the *CCP4* suite applying solvent flattening and histogram matching. A starting model for conventional refinement was created with *ARP/wARP* [333], where 314 amino acids in a single polypeptide chains (99% sequence coverage) were successfully placed into the electron density map with a resolution of 1.21 Å.

6.5.6 Model building and refinement

Crystal structure model refinement was performed with *phenix.refine* version 1.10.1-2155 [215]. Simulated annealing with default settings was performed as first refinement step. Subsequently, all crystal structure models were refined with riding hydrogen atoms added to protein residues, applying xyz refinement, individual anisotropic *B* factors for all atoms

except for hydrogens, and occupancy refinement. Refinement cycles were intermittent by model building with *Coot* [253]. Ligand building was performed with *MOE* [216], ligand restraints were created with *eLBOW* [254]. Graphical representations of the crystal structure coordinates and electron density maps (automatic ccp4 map normalization turned off) were created with *PyMOL* [221].

6.5.7 Cavity computation and volume calculation

The cavities as displayed in Figure 6.4 were computed with *POVME* [334]. Dummy atoms (DAs) with a radius of 1.40 Å were placed into the cavities (grid spacing 0.2 Å). The radius of 1.40 Å was chosen as it is frequently applied as the radius of a water molecule in the literature [299, 312, 314, 335, 336]. The solvent excluded surfaces of the DA objects representing the cavities were then displayed with *PyMOL*. Since *POVME* can only calculate the solvent accessible volume, the solvent excluded volumes of the DA objects describing the S_1' cavities of TLN-**1** to TLN-**6** (Figure 6.4) were calculated with the program *3V* (radius of the DAs set to 1.40 Å) [337]. The volumes of the S_1' cavity and of reference cavities 1–3 as displayed in Figure 6.2 used for the analysis of the electron density map were created with DA comprising a radius of 1.30 Å. The reduction of the DA radius compared to the DA radius applied for computing the cavities in Figure 6.4 (1.40 Å) was necessary, otherwise the narrow parts of reference cavities 1 and 2, where hydrogen-bonded water molecules are closely bound to TLN residues, would not have been detected.

6.5.8 Placement of the experimentally phased electron density map of TLN-1 on an absolute electron number scale and determination of the total electron content within a cavity

Two mathematical operations are required to transform the experimentally phased, σ-scaled map of TLN-**1** (ccp4 format) characterized by arbitrarily small map voxel values to an absolute electron number density map where zero corresponds to vacuum — thus allowing quantification of the total electron content within a given map volume. First, the values of the map voxels have to be set to the correct scale by applying a scaling coefficient. Subsequently, the still σ-scaled map — where zero corresponds to the average electron concentration of the unit cell — has to be shifted in a way that zero corresponds to vacuum. For that, the average electron concentration (e^-/Å^3) of the unit cell must be added to every map voxel. To have access to an absolute electron number density map as a reference, the σ-scaled density map of the fully refined model of TLN-**1** was transformed to an absolute electron number density map applying the *END* map script developed by Lang et al. [338] running with *phenix.refine*

version 1.8.4-1492. This program computes absolute *electron number density* (*END*) maps from conventionally refined σ-scaled maps. However, it is dependent on a refinement program and relies on refined model phases for the calculation of the electron density map, thus it is not possible to directly use it to transform the experimentally phased σ-scaled density map to an absolute electron number scale without the introduction of model phases. The *END* map script also provided the structure factor F_{000} of TLN-**1** (256668 e^-), which corresponds to the total electron content within the crystal unit cell under consideration of protein model and bulk (disordered) solvent. Thereby, the average electron concentration within the unit cell was accessible by F_{000}/unit cell volume = 256668 e^-/967353 Å^3 = 0.26533 e^-/Å^3.

We decided to use the Zn^{2+} ion (28 e^-) that is intrinsically bound to TLN to derive the scale coefficient. To be independent of the theoretical electron number of the zinc ion and of the necessity of an integration mask covering the entire zinc ion without any electron spillover from inside to the outside of the mask (e.g. due to a slightly disordered zinc) or electron spillover from outside to the inside of the integration mask (e.g. from zinc-complexing residues), we determined the total electron content of a spherical map fragment describing the zinc ion of the refined *END* map. The zinc coordinates of the refined crystal structure TLN-**1** were used for the center of the sphere, and 1.39 Å was used as sphere radius. The total electron content within this map fragment gave 27.3 e^- as calculated with an *in-house* script based on the program *mapman* [339]. Subsequently, an identical integration mask was applied to the σ-scaled experimentally phased electron density for scaling of the latter dataset by applying the program *mapmask* from the *CCP4* suite [238]. The scale coefficient (14.1) was derived in a way that after applying the scale coefficient followed by addition of the average electron concentration (0.26533 e^-/Å^3), integration of the map fragment covering the zinc of the experimentally phased map resulted in exactly 27.3 e^-, just as observed in the map fragment of the refined *END* map. Visually, the experimentally phased electron density map was correctly shifted in a way that zero corresponds to vacuum (Figure S6.4 to Figure S6.7). If an electron concentration of ≥0.0 e^-/Å^3 is displayed (panels C of Figure S6.4 to Figure S6.7), the entire cavities are covered with electron density. In contrast, displaying higher (positive) or lower (negative) electron density concentrations for instance within the empty S$_1$' pocket (Figure S6.4) immediately reduces the visible residual electron density.

For the calculation of the total electron content within a given volume, e.g. the volume of the S$_1$' cavity and reference cavities 1–3, a map fragment was cut out from the entire electron density map applying the program *Coot* via the command "mask map by atom selection", followed by integration over all voxels (including negative values) applying the aforementioned *in-house* script. For the S$_1$' cavity and reference cavities 1-3, the DA objects also used for displaying the contour regions of the cavities as shown in Figure 6.2 were used applying a radius of 1.3 Å.

6.5.9 Measurement of the thermodynamic binding profiles

For the ITC measurements, a Microcal ITC_{200} device from GE Healthcare (Piscataway, New Jersey) was used. Measurements were performed with freeze-dried TLN powder bought from Calbiochem (EMD Biosciences), which was dissolved in buffer directly before measurement without further treatment. Measurement buffers (pH 7.5) were composed of 20 mM ACES, MOPSO, MES, PIPES or cacodylate buffer substance, 500 mM NaSCN, and 2 mM $CaCl_2$. The salt NaSCN was selected because it strongly increases the solubility of TLN [180]. The high concentration of 500 mM was necessary to provide solubility of TLN up to 250 μM that was necessary to enable a direct measurement of the low-affinity ligand **2**. Since different concentrations of NaSCN are known to influence the measured thermodynamic binding parameters [96], this salt concentration was applied for all ligands to guarantee relative comparability of the thermodynamic binding parameters. The thermodynamic profiles of **4** (Table S6.4) and **3** (Table S6.5) were determined by direct titration applying a TLN concentration of 50 μM, resulting in well analyzable isotherms with clear sigmoidal curvatures (Figure S6.1C and B). For the low-affinity compound **2** (Table S6.3), the concentration of TLN was increased to 250 μM, resulting in isotherms described by a *c*-value of 6 (Figure S6.1A). This allowed the experimental determination of the inflection point and the determination of reliable thermodynamic parameters. To determine the heat of ionization associated with the complex formation of TLN and **2**, **3** or **4** (Figure S6.2), measurements were performed in ACES, MOPSO, HEPES, MES, PIPES and cacodylate buffers. The affinities of the weak binding compounds **1**, **5** and **6** were determined by displacement titrations (Figure S6.2D–F and Table S6.6) [141]. 50 μM TLN cacodylate buffer solution was preincubated with different concentrations of weak ligand (**1**: 5, 8, 20 mM; **5**: 5, 10, 15 mM; **6**: 1.5, 2, 2 mM), followed by titration with the reference ligand **4**. Different concentrations of the weak ligand (resulting in titration curves exhibiting different *c*-values) were applied to improve the accuracy of the binding parameter determination by global analysis of the ITC isotherms [179]. Measurement peaks of the raw thermograms were extracted and integrated with *NITPIC* version 1.1.2 [155, 178]. Global analysis of the ITC isotherms was performed with *SEDPHAT* version 12.1b [177, 340]. The thermodynamic profiles of the ligands measured in individual buffers were determined by global analysis applying the model "A+B ↔ AB Hetero-Association". Correction for the heat of ionization of these thermodynamic profiles was performed using the model "A+B ↔ AB Hetero-Association Global Buffer Ionization Enthalpy Analysis". For the displacement titrations, the model "A+B+C ↔ AB+C ↔ AC+B; competing B and C for A" was applied. ITC isotherm graphs were prepared with *GUSSI* [255].

6.6 Abbreviations

TLN, thermolysin; ITC, isothermal titration calorimetry; FOM, figure of merit; DA, dummy atom; ACES, *N*-(2-acetamido)-2-aminoethanesulfonic acid ; MOPSO, 3-morpholino-2-hydroxypropanesulfonic acid; PIPES, piperazine-*N*,*N*-bis(2-ethanesulfonic acid)

6.7 PDB accession codes

Newly reported crystal structures are available in the PDB upon publication of this manuscript (accession codes in brackets): TLN-**1** (5M9W), TLN-**5** (5LVD), TLN-**6** (5MA7), TLN-**1**-Xe (5M69), and TLN-**1**-Kr (5M5F). The crystal structures of TLN-**2** (3FV4), TLN-**3** (3FLF), and TLN-**4** (4H57) were previously reported [194].

6.8 Acknowledgements

6.9   Supporting Information

6.9.1   Isothermal titration calorimetry results

Table S6.1 (related to Figure 6.1). Thermodynamic binding profiles determined for **1**–**6**. The values in parenthesis describe the boundaries of the 95.4% confidence interval. All values are rounded. The thermodynamic profiles of **2**, **3** and **4** were corrected for the heat signal of buffer ionization upon complex formation.

ligand	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
1[a]	5,659 (8,321, 3,762)	-12.8 (-11.9, -13.8)	-	-
2[b]	40.750 (45.793, 36.299)	-25.1 (-24.8, -25.3)	-22.7 (-23.6, -21.9)	-2.4 (-1.2, -3.4)
3[c]	0.992 (1.234, 0.792)	-34.3 (-33.7, -34.8)	-33.7 (-34.8, -32.7)	-0.5 (1.1, -2.1)
4[c]	0.138 (0.201, 0.090)	-39.1 (-38.2, -40.2)	-45.1 (-46.5, -43.7)	5.9 (8.3, 3.5)
5[a]	2,407 (3,652, 1,507)	-14.9 (-13.9, -16.1)	-	-
6[a]	336 (752, 168)	-19.8 (-17.8, -21.5)	-	-

[a] Determined by displacement titration. [b] Direct titration with 250 μM TLN. [c] Direct titration with 50 μM TLN.
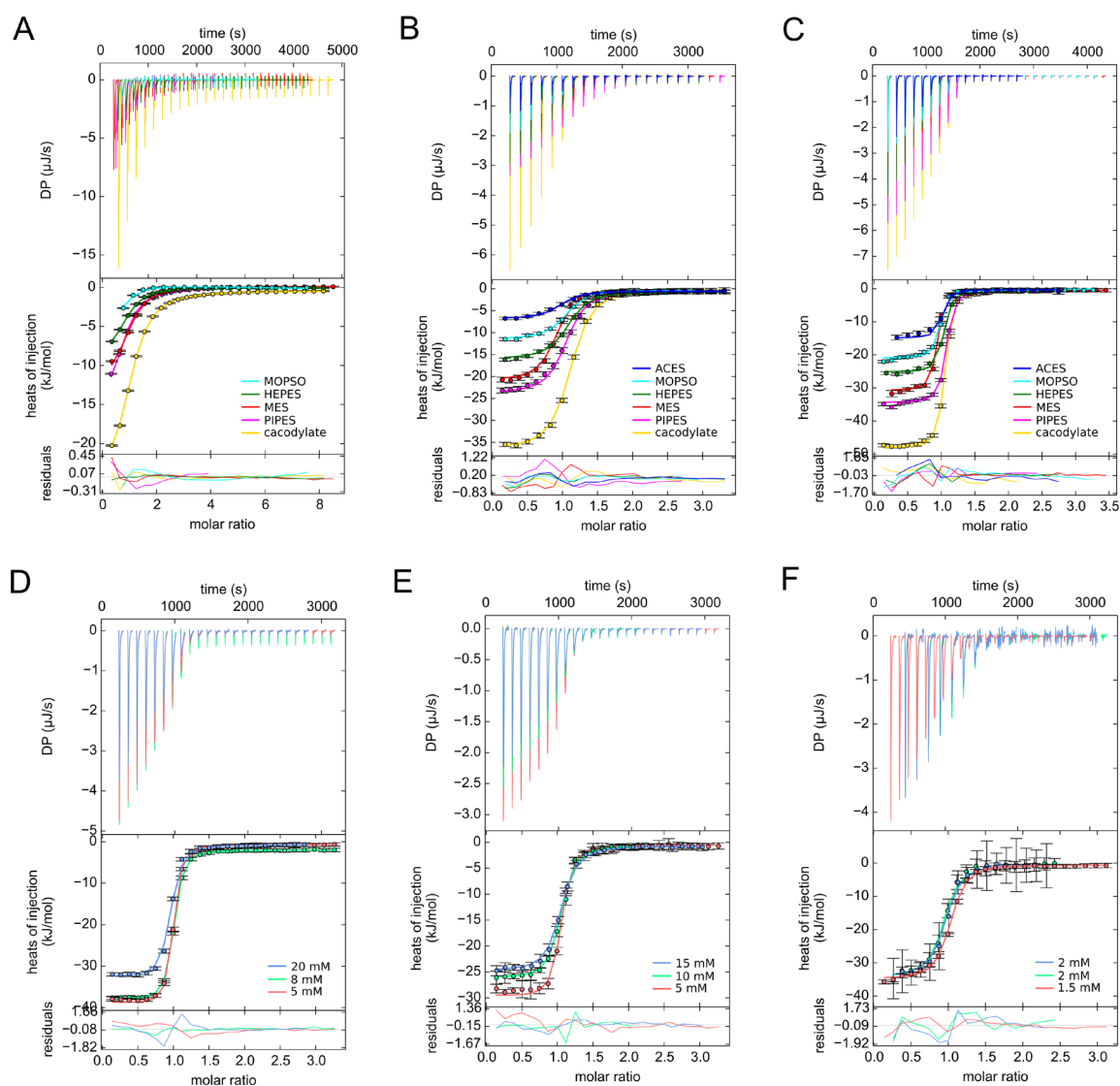
Figure S6.1. ITC thermograms of the **(A)** direct titration of **2**, **(B)** direct titration of **3**, **(C)** direct titration of **4**, **(D)** displacement titration of **1**, **(E)** displacement titration of **5**, **(F)** displacement titration of **6**. The direct titrations shown in panels A–C were performed in different buffers to correct $\Delta H°$ for superimposed buffer ionization reactions and to determine the proton transfer as shown in Figure S6.2. For the displacement titrations shown in panels D–F, the different isotherm colors indicate the different concentrations of weak ligand used for pre-incubation of TLN. For clarity reason always only one thermogram of the measurements performed at least in triplicate is shown. The thermodynamic binding profiles determined for the individual buffers are listed in Table S6.3 (**2**), Table S6.4 (**3**), and Table S6.5 (**4**). The results of the displacement titrations of **1**, **5**, and **6** are listed in Table S6.6.
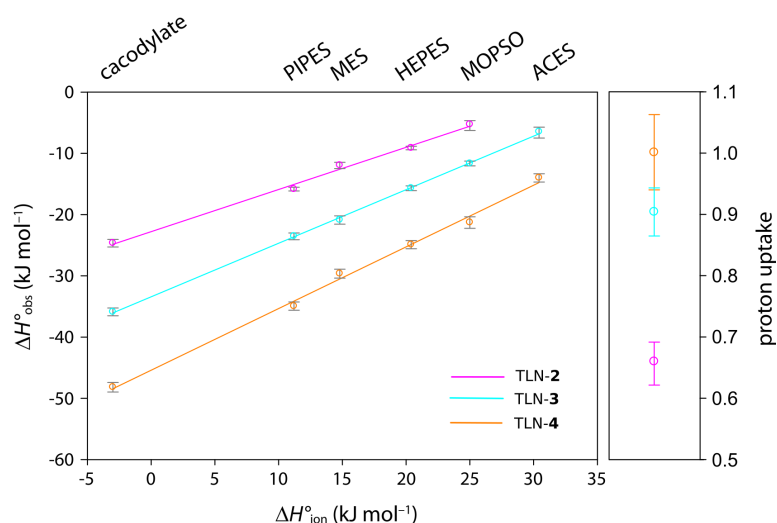
Figure S6.2. Determination of the number of protons transferred and the heat of buffer ionization upon complex formation of TLN-**2**, TLN-**3** and TLN-**4**. In the left diagram, the observed standard enthalpies upon ligand binding $\Delta H°_{obs}$ measured in six different buffers are plotted against the standard heat of ionization $\Delta H°_{ion}$ of the respective buffer [154]. Error bars in gray represent the 95.4% confidence interval of the $\Delta H°_{obs}$ values. Please note that the enthalpic signal of the complex formation of TLN-**2** was not measured in ACES buffer due to a too low exothermic signal. In the right diagram, the number of protons transferred upon complex formation of the three studied reactions is displayed as analyzed by global analysis with *SEDPHAT*. Error bars represent the 95.4% confidence interval. Data values are listed in Table S6.2.

Table S6.2. Proton transfer between buffer and TLN upon complex formation of TLN with ligands **4**, **3** and **2** as analyzed by global analysis with *SEDPHAT*.

Ligand	TLN proton uptake from buffer molecules (minimum and maximum of the 95.4% confidence interval)
4	1.001 (0.940, 1.063)
3	0.904 (0.865, 0.943)
2	0.660 (0.622, 0.692)

Table S6.3. Thermodynamic parameters obtained for **2**. All parameters are listed as rounded values.

buffer	best-fit values global analysis (95.4% confidence interval)				locally fitted concentration correction factor[a]
	K_d (µM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
buffer corrected	40.750 (45.793, 36.299)	-25.1 (-24.8, -25.3)	-22.7 (-23.6, -21.9)	-2.4 (-1.2, -3.4)	
cacodylate	39.604 (43.013, 36.425)	-25.1 (-24.9, -25.3)	-24.7 (-25.3, -24.1)	-0.5 (0.4, -1.3)	1.014, 0.997, 0.971
PIPES	52.301 (54.676, 50.038)	-24.4 (-24.3, -24.5)	-15.8 (-16.1, -15.5)	-8.6 (-8.2, -9.0)	0.731, 0.754, 0.787
MES	35.817 (41.314, 31.060)	-25.4 (-25.0, -25.7)	-11.9 (-12.5, -11.5)	-13.4 (-12.5, -14.3)	0.927, 0.934, 0.951
HEPES	39.526 (43.301, 36.091)	-25.1 (-24.9, -25.4)	-9.1 (-9.4, -8.8)	-16.0 (-15.5, -16.5)	0.800, 0.797, 0.749, 0.796
MOPSO	16.239 (27.511, 9.335)	-27.3 (-26.0, -28.7)	-5.3 (-6.3, -4.7)	-22.1 (-19.8, -24.0)	0.663, 0.561, 0.543
ACES[b]	-	-	-	-	-

[a] The concentration correction factor is in this case equivalent to the n value. The number of active sites per molecule was set to 1, and the concentration of the ligand was considered as accurately known. Therefore, this value corrects for inaccurate protein concentrations, as well as for inactive protein fractions observed in some of the applied buffers. The number of listed concentration correction factors reflects the number of performed measurements. [b]The enthalpic signal in ACES buffer was too low for a reliable determination of the thermodynamic binding profile

Table S6.4. Thermodynamic parameters determined for **3**. All parameters are listed as rounded values.

buffer	best-fit values global analysis (95.4% confidence interval; statistics minimum and maximum)				locally fitted concentration correction factor
	K_d (µM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
buffer corrected	0.992 (1.234, 0.792)	-34.3 (-33.7, -34.8)	-33.7 (-34.8, -32.7)	-0.5 (1.1, -2.1)	
cacodylate	0.929 (1.079, 0.798)	-34.4 (-34.1, -34.8)	-35.9 (-36.5, -35.2)	1.4 (2.5, 0.4)	1.131; 1.065; 1.040
PIPES	1.358 (1.599, 1.148)	-33.5 (-33.1, -33.9)	-23.5 (-24.1, -23.0)	-9.9 (-9.0, -10.9)	0.969; 1.005; 0.992
MES	0.710 (0.937, 0.530)	-35.1 (-34.4, -35.8)	-20.9 (-21.6, -20.2)	-14.2 (-12.8, -15.6)	0.868; 0.835; 0.855
HEPES	1.191 (1.447, 0.975)	-33.8 (-33.3, -34.3)	-15.7 (-16.1, -15.3)	-18.1 (-17.2, -19.0)	1.005; 0.964; 0.956
MOPSO	1.114 (1.427, 0.850)	-34.0 (-33.4, -34.7)	-11.7 (-12.0, -11.3)	-22.3 (-21.3, -23.4)	0.991; 0.984; 0.940
ACES	2.590 (5.013, 1.303)	-31.9 (-30.3, -33.6)	-6.5 (-7.5, -5.7)	-25.4 (-22.7, -27.9)	0.939; 0.782; 0.742

Chapter 6

Table S6.5. Thermodynamic parameters determined for **4**. All parameters are listed as rounded values.

buffer	best-fit values global analysis (95.4% confidence interval; statistics minimum and maximum)				locally fitted concentration correction factor
	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
buffer corrected	0.138 (0.201, 0.090)	-39.1 (-38.2, -40.2)	-45.1 (-46.5, -43.7)	5.9 (8.3, 3.5)	
cacodylate	0.105 (0.141, 0.076)	-39.8 (-39.1, -40.6)	-48.2 (-49.0, -47.4)	8.4 (9.9, 6.8)	0.959; 0.939; 0.954; 0.933
PIPES	0.124 (0.168, 0.089)	-39.4 (-38.7, -40.3)	-34.9 (-35.6, -34.3)	-4.5 (-3.0, -6.0)	1.030; 0.954; 1.020; 0.989
MES	0.119 (0.174, 0.077)	-39.5 (-38.6, -40.6)	-29.6 (-30.4, -28.9)	-9.9 (-8.2, -11.7)	0.895; 0.888; 0.871
HEPES	0.182 (0.267, 0.118)	-38.5 (-37.5, -39.5)	-24.9 (-25.6, -24.2)	-13.6 (-11.9, -15.3)	0.962; 0.961; 0.962
MOPSO	0.207 (0.333, 0.122)	-38.1 (-37.0, -39.5)	-21.3 (-22.3, -20.4)	-16.9 (-14.7, -19.1)	0.928, 0.924, 0.882
ACES	0.279 (0.500, 0.132)	-37.4 (-36.0, -39.3)	-14.0 (-14.7, -13.3)	-23.4 (-21.3, -25.9)	0.942; 0.892; 0.892

Table S6.6. Thermodynamic data of **1**, **5** and **6** as determined by ITC displacement titrations in cacodylate buffer. All values are listed as rounded values.

ligand	best-fit values global analysis (95.4% confidence interval; statistics minimum and maximum)				locally fitted concentration correction factor
	K_d (mM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
1[a]	5.659 (8.321, 3.762)	-12.8 (-11.9, -13.8)	-	-	0.907, 0.944, 0.950
5[a]	2.407 (3.652, 1.507)	-14.9 (-13.9, -16.1)	-	-	1.000, 0.998, 0.981
6[b]	0.336 (0.752, 0.168)	-19.8 (-17.8, -21.5)	-16.9 (-22.7, -13.5)	-2.9 (4.9, -8.0)	0.981, 0.894, 0.916

[a] Thermodynamic parameters $\Delta H°$ and $-T\Delta S°$ could not be determined accurately due to the extremely low affinity of **1** and **5**, and are thus omitted. [b] Thermodynamic parameters of **6** were not corrected for the heats of buffer ionization.

6.9.2 Analysis of the experimentally phased TLN-1
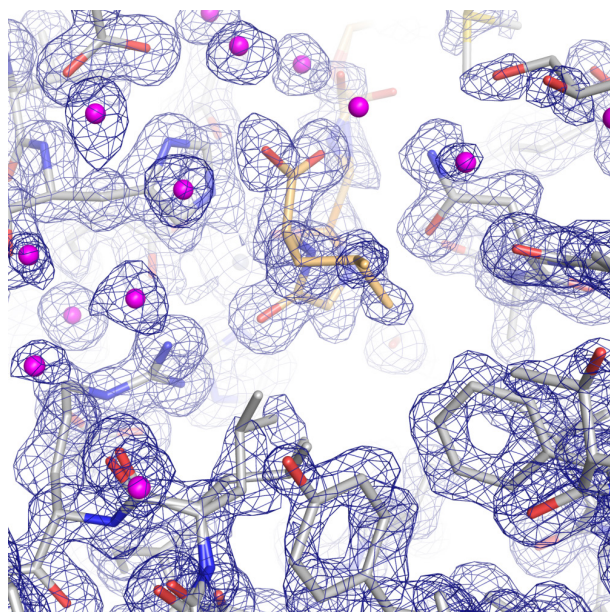
Figure S6.3. Experimentally phased absolute-scale electron density map of TLN-**1** shown as blue mesh at a contour level of 0.7 e$^-$/Å^3. For orientation, the refined model of TLN-**1** is superimposed onto the electron density map. TLN residues are displayed as gray sticks, the bound ligand is displayed as light orange stick model and water molecules are displayed as magenta spheres.

Table S6.7 (related to Figure 6.2). Solvent excluded volume and electron content of selected cavities of TLN-**1**.

cavity	volume (Å^3)	integrated electron content (e$^-$)^a	water molecules in the model
S$_1$'	141	6.6	0
reference cavity 1	59	19.4	3
reference cavity 2	93	13.4	1
reference cavity 3	16	0.9	0

^{a} All integrated electron contents were scaled on the intrinsically bound zinc atom of TLN.
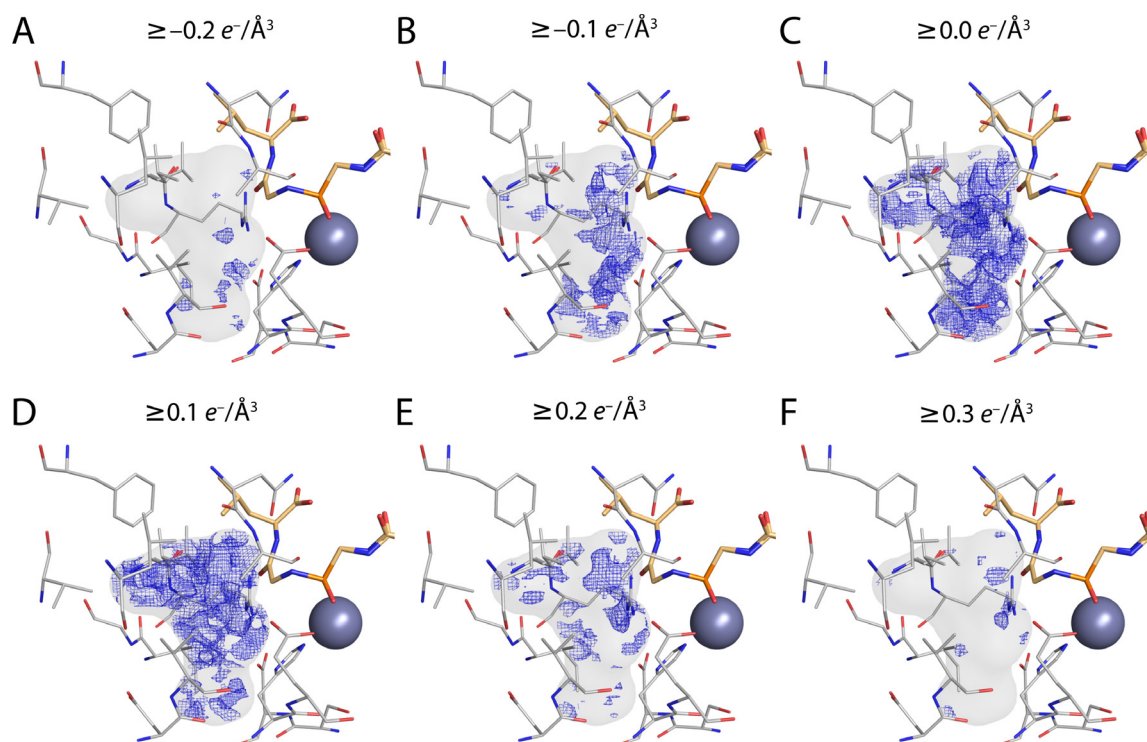
Figure S6.4 (related to Figure 6.2). Absolute-scale electron density map (blue mesh) within the S$_1$' cavity at different contour levels as labeled in panels A–F.
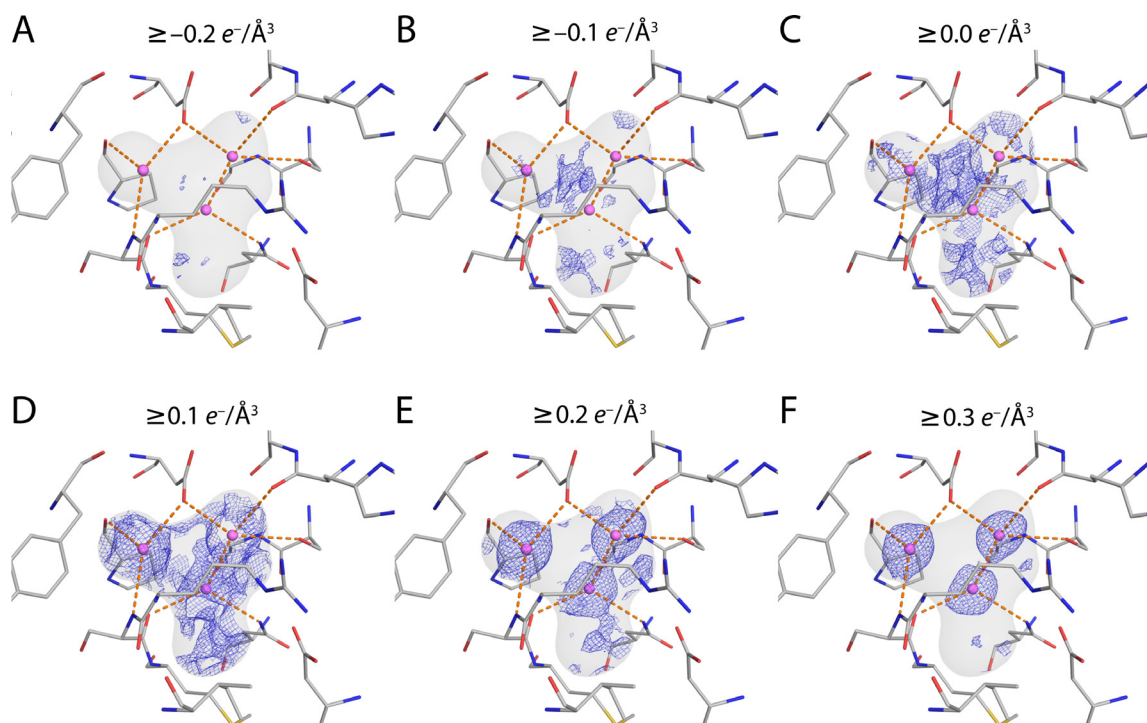
Figure S6.5 (related to Figure 6.2). Absolute-scale electron density map (blue mesh) within the reference cavity 1 at different contour levels as labeled in panels A–F.
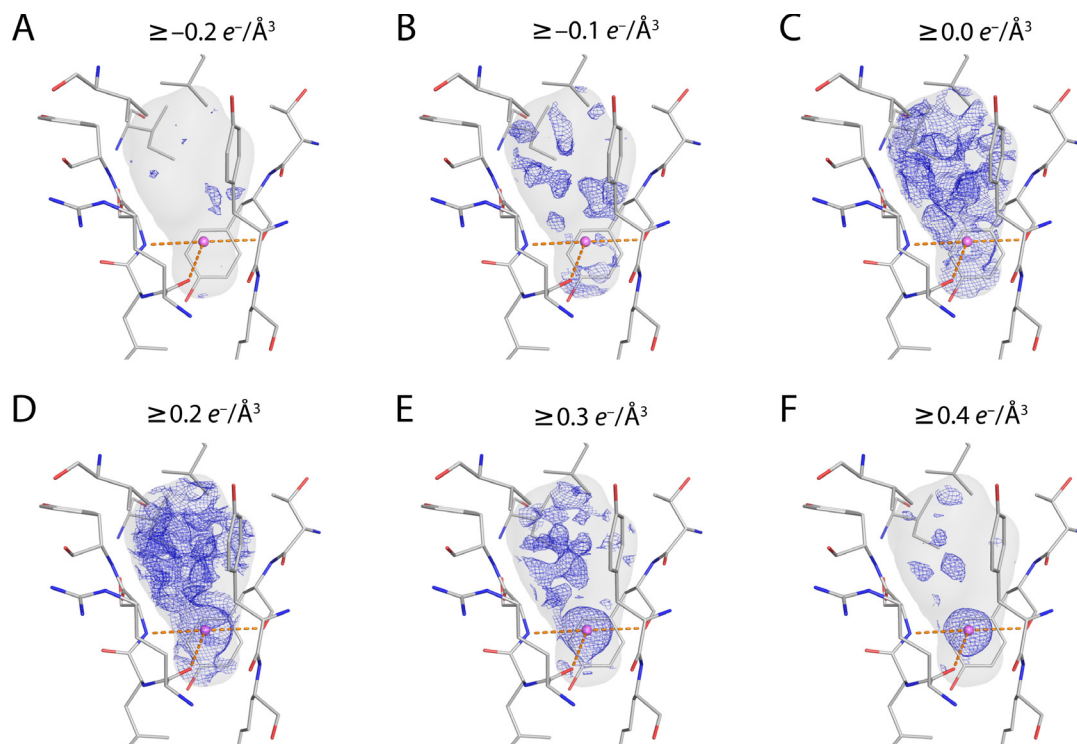
Figure S6.6 (related to Figure 6.2). Absolute-scale electron density map (blue mesh) within the reference cavity 2 at different contour levels as labeled in panels A-F.
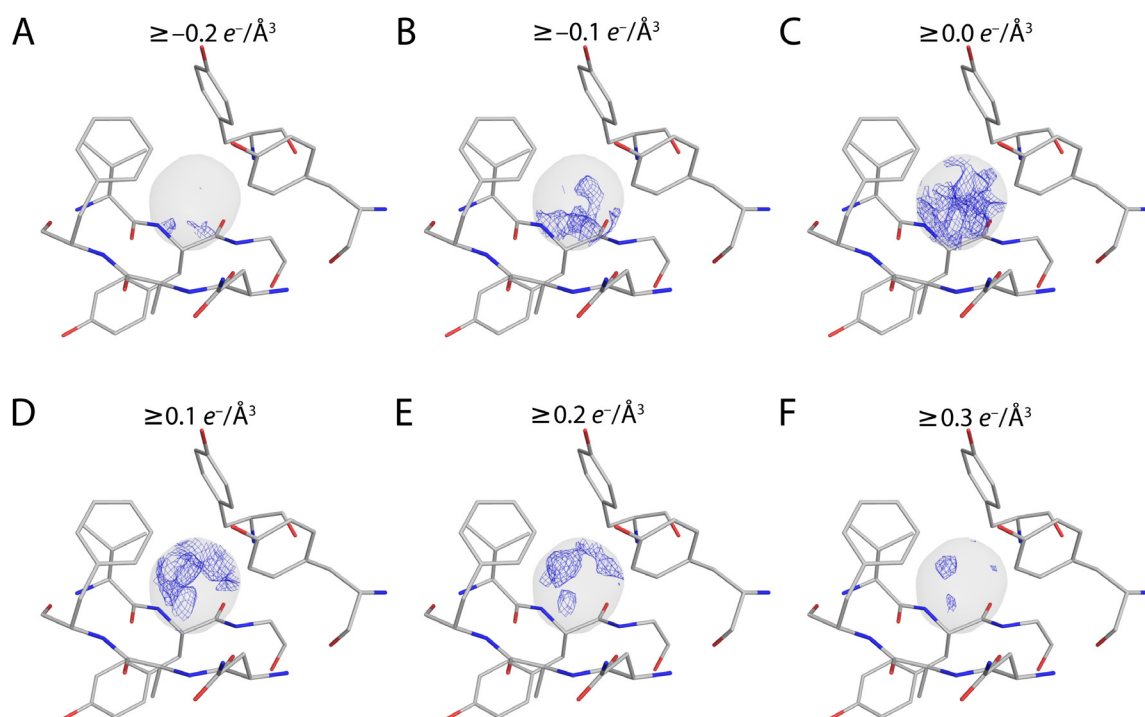
Figure S6.7 (related to Figure 6.2). Absolute-scale electron density map (blue mesh) within the reference cavity 3 at different contour levels as labeled in panels A-F.

Chapter 6

6.9.3 Crystal structure of TLN-1-Kr
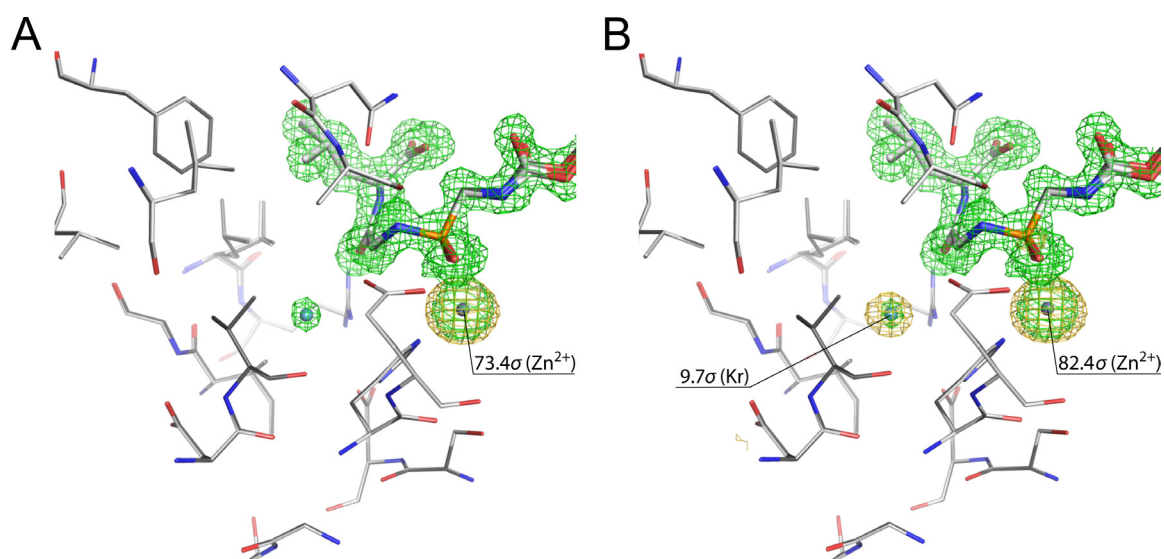
Figure S6.8. Krypton binding site in the S_1' cavity of the crystal structure TLN-**1**-Kr. **(A)** Dataset collected below the krypton absorption edge. **(B)** Dataset collected above the krypton absorption edge. In both panels, the crystal structure of TLN-**1**-Kr (dark gray) is superimposed on the crystal structure of the native TLN-**1** (light gray). The anomalous map is shown in gold at a contour level of 5σ, and the detected peaks are labeled with their intensity maximum. The F_o-F_c omit map of the krypton atom, zinc ion and the ligand is shown in green at a contour level of 3σ

.

Chapter 7

Crystallographic Screening of a 361-Entry Fragment Library against the Aspartic Protease Endothiapepsin

7.1 Introduction to Fragment-Based Lead Discovery

In fragment-based lead discovery (FBLD), targets are screened against libraries of compounds exhibiting a significantly lower molecular weight than drug-like molecules [341, 342]. These so-called "fragments" usually comprise less than 20 non-hydrogen atoms (<300 Da) and their chemical structures follow the "rule of three" [343]. Compared to the high-throughput screening (HTS) of drug-like compounds [344, 345], screening with fragments has several advantages. First, the lower number of atoms strongly reduces the number of theoretically imaginable molecules, making the sampling of the chemical space more efficient [346]. Thus, a fragment library, typically comprising 1,000-5,000 compounds, is about three orders of magnitude smaller than an HTS library [341]. Second, the reduced complexity of the fragments makes it less likely that unfavorable interactions between protein and ligand occur, leading to an increase of the hit rate [341, 347]. Third, fragments often comprise better pharmaceutical properties (e.g. low lipophilicity) and thus, if optimized carefully [348], they result in drugs with likewise improved physical properties [341, 349]. The screening of fragments against a target requires sensitive methods [341, 342], since fragments are only able to establish a few (high-quality) interactions due to their small size, usually resulting in low affinities (0.1-10 mM) [342]. Normally, a biophysical screening cascade is performed, starting with an initial screen applying SPR or thermal shift assay, followed by NMR spectroscopy and subsequently structural elucidation using X-ray crystallography [342]. Even though X-ray crystallography is considered the most powerful screening technique [342], it is usually not considered for primary screening due to its presumed low throughput. After identifying a suitable starting fragment, chemical optimization into a lead structure can be performed by fragment merging, fragment linking, and fragment growing [342].

7.2 Advocacy of X-ray crystallography as primary fragment screening technique

In an initial study, a generally designed 361-entry *in-house* fragment library, not strictly following the Astex's rule of 3 [343], was screened against the aspartic protease endothiapepsin (EP) [350]. To supposedly perform the screening campaign efficiently, a screening cascade comprising a pre-screening by a biochemical cleavage assay followed by a crystallographic screening of the assay-predicted hits was employed. The pre-screening resulted in 55 hits, which were subsequently crystallized. This led to the determination of 11 crystal structures (crystals soaked with mixtures of two fragments). Even though this corresponds to a quite high crystallographic hit rate of 20% (11 out of 55), no correlation between the assay determined affinities of the fragments and the discovered crystallographic hits was observed [350]. This raised the question whether it was meaningful to pre-select

fragments for crystallographic screening according to the assay results and whether other pre-screening assays would determine similar or different hits. Therefore, in a subsequent study [351], the fragment library was screened against EP by several assays, i.e. by a reporter-displacement assay, saturation-transfer NMR, native mass spectrometry, thermophoresis, and by thermal shift assay [351]. Remarkably, the mutual hit overlap between the different techniques was quite poor. Thus, restricting the crystallographic screening to a fragment subset suggested by one or a combination of pre-screening assays as the basis for the determination of optimal crystallographic hits for further chemical optimization has to be questioned. Crystallographic screening only of fragments discovered as hits in pre-screening assays can potentially result in the exclusion of fragments with putatively important binding modes from the crystallographic screening [351]. Furthermore, none of the applied pre-screening techniques provided any information about the binding modes of the fragments, although this is essential for further chemical optimization of a fragment hit [341, 342, 351, 352]. Consequently, it was decided to perform a crystallographic screening of the entire 361-entry fragment library against EP [106].

The crystallographic screening of the entire library was performed by a team of eight crystallographers due to the vast amount of work which accompanies such an endeavor (preparation of individually soaked crystals, collection of the datasets, investigation of the electron density maps for bound fragments, refinement of the crystal structures). Out of the 361 fragments directly screened by crystallography, an exceptionally high hit rate of 20% (71 hits and 86 binding poses) was discovered [106]. Of these hits, 31 were not detected by any of the five pre-screening techniques and an additional 21 hits were determined by only one of the pre-screening techniques [106]. Even though an increase of the crystallographic hit rate could be observed depending on the applied prescreening method (28–55%), this went at the expense of total hits (between 61–94% missed hits) [106]. Consequently, the use of pre-screening techniques for the selection of a fragment subset for a subsequent crystallographic screening can strongly deteriorate the total crystallographic hit rate. Since knowledge of the entire "fragment binding landscape" can be of utmost importance to the success of an FBLD project [106, 108], it is advisable to use X-ray crystallography directly as first (primary) screening technique [106]. Only in cases where the crystallographic screening is difficult to perform (e.g. lack of protein material, poor diffracting crystals, lack of beamtime) can pre-screening techniques be meaningful [106].

Chapter 7

7.3 Phase improvement can be essential to identify weakly bound fragments in the electron density map successfully

In order to identify protein-bound fragments ("hits"), the electron density maps of crystal structure datasets have to be manually inspected for electron densities shaped like the fragments. Even though identifying a hit is admittedly somewhat subjective, weak electron densities increase the chance of overlooking a bound fragment or even make it impossible to unambiguous identify it. The "unmodelled blobs" tool of *Coot* can assist with determining bound fragments characterized by strong density blobs [253]. However, it fails with the discovery of fragments exhibiting weak electron densities. Thus, the detection of bound fragments is mainly a manual task of the experimenter. The first manual inspection of the predominantly only crudely refined 361 crystal structure datasets by the team of eight crystallographers resulted in the discovery of 64 fragment binding poses.

During the refinement of fragment **305** (1.69 Å, Table 7.2) I noticed that the quality of the electron density map visually improved significantly with increasing completeness of the refinement. Whereas the protein-bound fragment was barely recognizable at the beginning of the refinement in the electron density map, it was possible to unambiguously assign the fragment to the electron density after an elaborate refinement was performed. In particular, several cycles of manual addition of water molecules to the model resulted in a significant visual improvement of the electron density map, which increasingly displayed features of the bound fragment. This was accompanied by a decrease of the R_{free} value and a reduction of the phase error. Based on this observation, I concluded that it might be advisable to improve phase and electron density map accuracy before inspecting the map for putatively bound fragments. Otherwise, as in the case of **305**, fragments with weak electron densities could be potentially overlooked. To make the electron density map quality improvement as efficient as possible, I decided to perform a maximum of phase-improving procedures directly in the first refinement cycle with *phenix.refine* after simulated annealing, i.e. automatic placement of water molecules, the addition of riding hydrogen atoms and TLS group ADP or anisotropic ADP refinement, the latter depending on the resolution of the dataset. This strategy — especially the generous automatic placement of water molecules via *phenix.refine* — possibly introduces model bias into the structure. Thus, the structure resulting from this procedure should only be used to identify bound fragments and not as a starting point for further refinement. Fragment **305** in complex with EP is quite difficult to identify in the electron density map after simulated annealing (Figure 7.1A). The subsequent manual placement of 156 water molecule during the first model building cycle only marginally increases the recognizable fragment features of the electron density map (Figure 7.1B). Thus, further manual model building and refinement cycles would be necessary to identify the bound fragment unambiguously. By contrast, the automatic addition of 418 water molecules
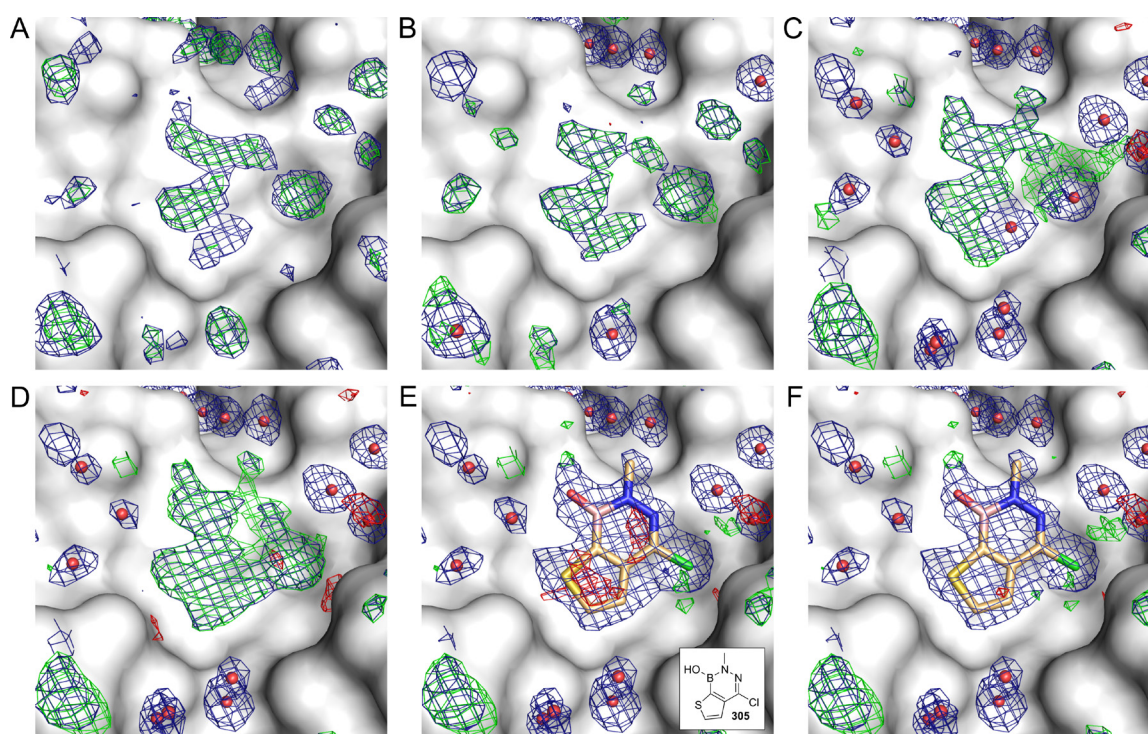
Figure 7.1. Electron density map of **305** in complex with EP after different steps of model building and refinement. In all panels, the molecular surface of EP is displayed in white, the $2mF_o–DF_c$ electron density map is displayed at a contour level of 1σ (blue mesh), and the $mF_o–DF_c$ electron density map at $+3\sigma$ (green mesh) and at $–3\sigma$ (red mesh). **(A)** Electron density map after simulated annealing (R_{work} = 24.2%, R_{free} = 27.3%). **(B)** First model building cycle, manual placement of 156 water molecules (R_{work} = 19.9%, R_{free} = 23.2%). **(C)** Automatic placement of 418 water molecules and addition of riding hydrogen atoms via *phenix.refine* to the structure after simulated annealing displayed in panel A (R_{work} = 17.0%, R_{free} = 20.5%). **(D)** Deletion of two water molecules automatically placed into the electron density of the fragment displayed in panel C. **(E)** Fragment placed into the electron density of panel D and refined with 100% occupancy. Negative difference electron density appears on the refined fragment. **(F)** Occupancy of the fragment displayed in panel E refined to 78%. The negative difference electron density on the fragment disappears.

and the addition of riding hydrogen atoms by *phenix.refine* to the model after simulated annealing results in an electron density map with more visible features of the bound fragment (Figure 7.1C), enabling the identification of the latter. However, to provide a clear view of the electron density of the fragment (Figure 7.1D), two incorrectly placed water molecules have to be deleted. The refined electron density of the fragment appears convincing (Figure 7.1E), even though negative difference electron density appears if the fragment is refined with 100% occupancy. Consequently, the fragment was partially refined to 78% occupancy (Figure 7.1F). The partial occupancy of **305** is the reason for its weak electron density, requiring phase improvement to enable its detection. Since fragments commonly show low affinities, they are often only partially occupied in the crystal structure [76, 353, 354]. Consequently, I

Chapter 7

hypothesized that phase improvement prior to electron density inspection could result in the discovery of further, previously unrecognized hits.

With the aim of performing a second round of manual inspection of phase-improved electron density maps, all datasets were extensively refined by an automatic refinement pipeline script written by Johannes Schiebel (Philipps University of Marburg) [76]. The script solely relies on available crystallographic software; for instance, the command line version of *phenix.refine* is used as refinement software. Several model building and refinement steps are consecutively performed building on each other: 1.) molecular replacement, 2.) cartesian simulated annealing, 3.) xyz coordinate and isotropic ADP refinement, 3.) TLS group ADP refinement, 4.) automatic water placement by *Coot,* followed by automatic identification and deletion of water molecules placed into strong electron density blobs putatively originating from bound fragments, 5.) anisotropic ADP refinement of all atoms excluding water molecules, 6.) anisotropic ADP refinement of all atoms including water molecules, 7.) the addition of riding hydrogen atoms to the protein model and 8.) a second round of automatic water molecule placement and deletion. Each of the steps of the refinement pipeline can be used as starting point for manual structure refinement. After the last step of the refinement pipeline, the phase error was on average reduced by 87% relative to the phase error reduction achieved for fully, manually refined structures [76]. Thus, the crystal structure model at the end of the pipeline is close to convergence and the resulting electron density map is of high quality. Analysis of the individual refinement steps revealed that the strongest impact on the total phase error reduction (68%) comes from the two water molecule placement steps [76]. By default the script uses the water placement algorithm of *Coot*, which places water molecules more conservative than the algorithm of *phenix.refine* (can be used optionally). Thereby, the risk of introducing model bias is reduced. Using *phenix.refine* rather than *Coot* for the automatic water molecule placement only resulted in a slightly stronger reduction of the phase error (90% instead of 87%) [76].

The inspection of the phase-improved electron density maps of the automatically refined crystal structures resulted in the discovery of 22 fragment binding modes in addition to the previously discovered 64 individual binding poses in the only crudely refined structures [76]. Analysis of the additional binding poses revealed that they all exhibit weak electron density features, described by correlation coefficients (CC) between their F_o–F_c and F_c maps of less than 0.7, significantly lower compared to the average CC of 0.79 of the earlier discovered 64 hits [76]. For a fragment to be suited for further chemical optimization, a ligand efficiency (LE) of 1.20 kJ mol^{-1} atom^{-1} is considered a minimum [76, 342, 355]. Whereas the initial hits are characterized by an average LE of 1.22±0.26 kJ mol^{-1} atom^{-1}, the additional hits have an LE of 1.21±0.27 kJ mol^{-1} atom^{-1} [76]. Thus, the LE additional hits only slightly lower than the LE of the initial hits and the affinity of the additional hits is considered sufficiently high to serve as starting points for further chemical optimization. Besides, even if a fragment

comprises an LE below the applied threshold, chemical optimization can turn it into a very potent binder [76, 106]. Furthermore, with respect to assessing the affinity of a fragment hit, it must be considered that the adjacent water structure can have a substantial impact on the affinity and particularly on the thermodynamic binding profile [108, 123, 124]. Clearly, the water structure will experience changes during the chemical optimization process, resulting in a modulation of its impact and thus questioning the significance of the affinity of an initial fragment. In any case, the knowledge about the binding mode of a fragment and its involved functional groups can inspire hit expansion and lead optimization [108]. This underlines the importance of the structural information. In conclusion, overlooking the additionally discovered hits characterized by weak electron densities would have meant losing valuable information and wasting a large amount of already invested resources. The discussed concept to improve the electron density map quality by an elaborate refinement before investigating datasets for putatively bound fragments will possibly also hold for other fragment screening campaigns on other targets and result in maximizing the fragment hit rate. Structure refinement is greatly facilitated and accelerated by the automatic refinement pipeline [76], which should further increase the suitability of X-ray crystallography as primary screening technique [78].

Recently, a new method for the sensitive and less subjective identification of fragment hits called *PanDDA* (Pan-Dataset Density Analysis) has been reported [356]. This approach requires multiple refined datasets of the same crystal system (e. g. crystals of the same protein soaked with different fragments). With these datasets, an average "ground state" electron density map is calculated. Subsequently, the electron density map of a single refined crystal structure dataset is compared to this average "ground state" electron density map. Subsequently, a new electron density map highlighting any disturbances of the "ground state" is calculated, e.g. the binding of a fragment. In combination with the refinement pipeline automatically providing phase-improved, accurate electron density maps, this method could prove valuable in supporting the identification of weakly bound fragments.

7.4 Refinement of fragment-bound crystal structures of EP and description of the binding modes of fragments bound to Asp81 and the S$_1$–S$_3$ pockets

7.4.1 Introductory remarks

Fragments **131**, **158**, **162**, **164**, **171**, **266**, **273**, **285** and **305** were refined by me (Table 7.1 and Table 7.2) and their crystal structures were discussed in two publications [107, 108]. The

Table 7.1. Data collection and refinement statistics of fragment-bound crystal structures of EP.[a]

| | Fragment number (PDB code) | | | | |
	131 (4Y4E)	158 (5J25)	162 (4Y47)	164 (4Y44)	171 (4Y3X)
Data collection and processing					
Wavelength (Å)	0.89500	0.91841	0.89440	0.91841	0.91841
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters					
a, b, c (Å)	45.3, 73.2, 52.8	45.3, 72.9, 52.6	45.3, 72.8, 52.6	45.2, 72.9, 52.6	45.3, 73.0, 52.7
α, β, γ (°)	90.0, 109.4, 90.0	90.0, 109.3, 90.0	90.0, 109.4, 90.0	90.0, 109.0, 90.0	90.0, 109.7, 90.0
Resolution range (Å)	19.18–1.30 (1.38–1.30)	42.74–1.24 (1.31–1.24)	42.72–1.19 (1.26–1.19)	49.69–1.24 (1.31–1.24)	42.70–1.25 (1.33–1.25)
Wilson B factor (Å²)	9.1	9.9	8.6	10.1	8.8
No. of unique reflections	79467 (12595)	86046 (12396)	102282 (15746)	88968 (13388)	87058 (13331)
Multiplicity	4.1 (4.0)	3.2 (3.1)	3.8 (3.3)	3.4 (3.2)	3.8 (3.6)
R_{sym} (%)	6.1 (48.0)	3.0 (24.8)	6.5 (47.3)	3.2 (18.3)	5.9 (40.0)
Completeness (%)	99.4 (97.9)	94.2 (84.2)	98.8 (94.3)	97.5 (91.0)	97.5 (92.7)
$<I/\sigma(I)>$	16.1 (3.8)	21.6 (4.1)	14.5 (3.0)	20.3 (5.3)	13.9 (3.0)
Refinement					
Resolution range (Å)	17.61–1.30	42.74–1.24	36.38–1.19	42.75–1.24	42.70–1.25
No. of reflections used for refinement (work/free)	75469/3973	81741/4303	97150/5113	84517/4449	82703/4353
R_{cryst} (%)	11.4	11.6	13.2	11.2	12.0
R_{free} (%)	13.6	13.7	15.4	12.4	14.2
No. of refined residues	330	330	330	330	330
No. of fragment atoms	13	26	20	10	26
No. of other ligand atoms[b]	20	16	16	10	32
No. of water molecules	296	300	333	282	296
RMSD, bond length (Å)	0.007	0.007	0.006	0.006	0.015
RMSD, bond angles (°)	1.2	1.2	1.2	1.2	1.6
Ramachandran plot (%)[c]					
Most favoured	93.9	92.8	93.9	93.5	93.5
Additionally allowed	6.1	7.2	6.1	6.5	6.5
Generously allowed	0	0	0	0	0
Disallowed	0	0	0	0	0
Mean B values (Å²)[d]					
All atoms	10.5	11.6	10.3	11.8	9.9
Main chain	9.6	10.7	9.5	10.9	9.1
Side chain	11.4	12.4	11.0	12.6	10.7
Fragment atoms	41.6	28.5	15.5	33.1	13.0
Other ligand atoms[b]	27.4	20.5	19.9	17.9	20.7
Water molecules	24.5	24.7	24.0	26.5	24.4

[a] Values in brackets describe the highest resolution shells. [b] Other ligand atoms are from glycerol, PEG, and acetate molecules. [c] Calculated with *PROCHECK* [196]. [d] Calculated with *MOLEMAN* [198].

Table 7.2. Data collection and refinement statistics of fragment-bound crystal structures of EP.[a]

| | Fragment number (PDB code) | | | |
	266 (4Y3S)	273 (4Y3N)	285 (4Y3G)	305 (4Y37)
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters				
a, b, c (Å)	45.3, 73.1, 52.8	45.6, 73.0, 53.0	45.3, 73.1, 52.8	45.4, 73.1, 52.9
a, β, γ (°)	90.0, 109.5, 90.0	90.0, 109.4, 90.0	90.0, 109.5, 90.0	90.0, 109.4, 90.0
Resolution range (Å)	19.78–1.10 (1.17–1.10)	43.01–1.34 (1.42–1.34)	19.80–1.13 (1.20–1.13)	19.21–1.69 (1.79–1.69)
Wilson B factor (Å²)	10.4	12.6	9.8	13.7
No. of unique reflections	127054 (19588)	72919 (11390)	117892 (18348)	36243 (5481)
Multiplicity	3.6 (3.4)	3.6 (3.4)	3.7 (3.6)	3.7 (3.3)
R_{sym} (%)	3.2(48.2)	4.9 (48.8)	6.4 (47.5)	9.9 (48.5)
Completeness (%)	97.0 (93.0)	99.0 (96.2)	97.5 (94.3)	98.4 (93.4)
$<I/\sigma(I)>$	20.5 (2.7)	15.3 (2.3)	11.2 (2.5)	12.1 (2.5)
Refinement				
Resolution range (Å)	19.78–1.10	43.01–1.34	19.80–1.13	19.21–1.69
No. of reflections used for refinement (work/free)	120684/6353	69273/3646	111946/5891	34414/1810
R_{cryst} (%)	12.4	12.4	16.3	18.1
R_{free} (%)	14.1	15.2	18.4	21.7
No. of refined residues	330	330	330	330
No. of fragment atoms	12	14	12	11
No. of other ligand atoms[b]	10	10	6	10
No. of water molecules	273	262	273	268
RMSD, bond length (Å)	0.006	0.007	0.007	0.007
RMSD, bond angles (°)	1.2	1.2	1.2	1.0
Ramachandran plot (%)[c]				
Most favoured	93.5	93.9	93.5	93.1
Additionally allowed	6.5	6.1	6.5	6.9
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B values (Å²)[d]				
All atoms	12.7	14.5	11.3	15.6
Main chain	11.7	13.5	10.5	15.0
Side chain	13.6	15.4	12.1	16.4
Fragment atoms	42.2	24.1	20.1	23.9
Other ligand atoms[b]	17.4	24.0	16.4	27.8
Water molecules	27.3	28.6	24.8	26.9

[a] Values in brackets describe the highest resolution shells. [b] Other ligand atoms are from glycerol, PEG, and acetate molecules. [c] Calculated with *PROCHECK* [196]. [d] Calculated with *MOLEMAN* [198].

Chapter 7

latter two publications also describe experimental details of data collection, structure determination and refinement. However, since the nine fragments exhibit very diverse binding modes and thus it would be difficult to discover and discuss meaningful similarities and differences between their binding modes, fragments **14**, **35**, **39**, **41**, **51**, **54**, **73**, **125**, **158**, **205**, **206**, **207**, **224**, **240**, **261**, **323** and **328** were selected for analysis and mutual comparison, of which only fragment **158** was refined by me. These fragments either bind to Asp81 and/or occupy the S_1–S_3 pockets of EP, as described in the following. In case multiple copies of the same fragment are found simultaneously in complex with EP, only the fragment copies with the aforementioned binding motif are discussed. The descriptions of the binding modes helped to prepare the two publications [107, 108]. Even though the presented fragments do not directly bind to the catalytic dyad of EP and thus are less suited as starting points for chemical optimization, their experimentally determined binding modes provide valuable information concerning how to chemically optimize and expand a given starting structure [107].

7.4.2   Fragments with hydrogen bonds to Asp81

Fragments 14, 41, 51, 125, 158, 261, 323, and 328. Asp81 is addressed by fragments exhibiting hydrogen-bond donor functional groups, mainly represented by amines (Figure 7.2). For fragment **14**, the thiourea's nitrogen next to the methyl group is located within a distance appropriate for the formation of a weak hydrogen bond to Asp81 (3.4 Å). The nitrogen bound to the phenyl ring is acting as hydrogen-bond donor for the formation of an interaction to the OH group of Thr223 (distance as labeld in Figure 7.2). The aromatic moiety of the fragment is placed near the S_3 pocket. Compounds **41** and **158** comprise a close chemical structure and also show an almost identical binding mode (Figure 7.3). Both establish hydrogen bonds between the tertiary nitrogens of their piperidine moieties and Asp81. For **158**, two atoms of the piperidine ring and the methyl group in *para* are not defined in the electron density (Figure 7.2). Moreover, both ligands address Gly80's NH with their carbonyl groups, whereas the amide NH groups of both compounds experience hydrogen bonds with Thr222. Compared to the *apo* structure, in both complexes the carboxyl group of Asp81 is located closer to the fragments' tertiary nitrogens to form more favorable interactions, resulting in an increased distance of Gly80 to the compounds (see Figure 7.2). For **158**, a second fragment molecule is found addressing the S_1 pocket (see Chapter 7.4.6). In the complex of **51** with EP, the compound is not interacting with Asp81 in a direct way. However, a glycerol molecule is located in van der Waals interaction distance between the fragment and the hydrophobic surface of the S_1 pocket. With its central hydroxyl group, the glycerol is in hydrogen bond distance to Asp81 and Ser83. Furthermore, one
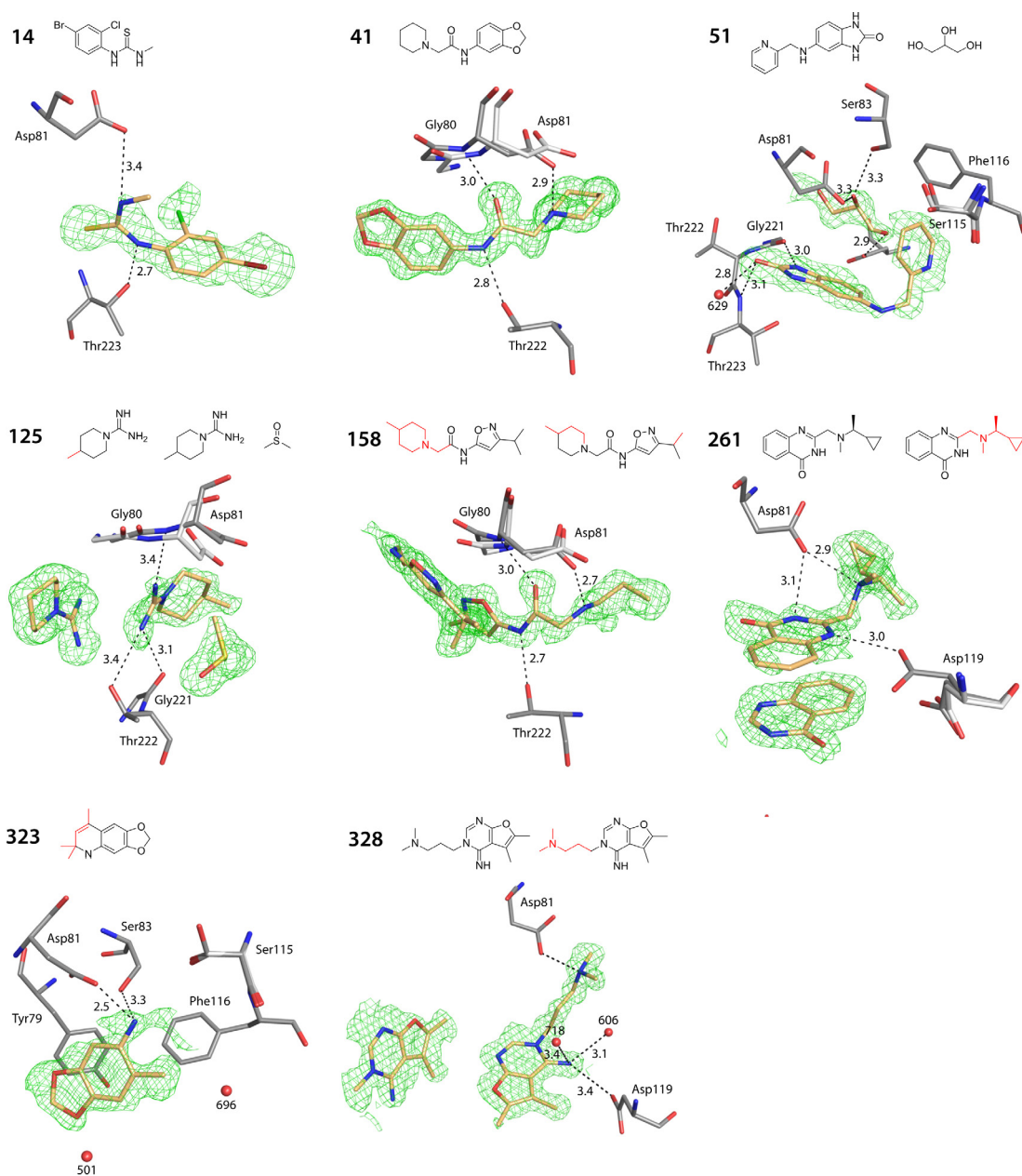
Figure 7.2. Fragments addressing Asp81 with a hydrogen bond. Bound fragment molecules are displayed in orange, protein residues in gray and conformations of residues of the apo structure in light gray (heteroatoms color-coded). Hydrogen-bond interactions are indicated by black dotted lines (distance labeled in Å). F_o–F_c omit electron densities of bound ligands are displayed at a contour level of 2.5σ. All bound fragment molecules and other bound ligand molecules are indicated by 2D chemical structures (red marked atoms indicate insufficient electron density in the crystal structure for atom placement).

Chapter 7
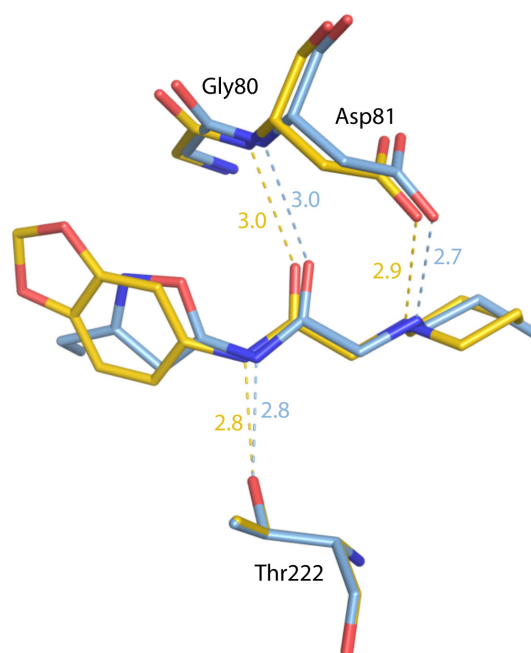
Figure 7.3. Superimposition of fragments **41** (orange) and **158** (blue).

terminal hydroxyl group establishes a hydrogen bond to Asp33. The fragment itself is anchored with its pyridine ring to the protein, performing π-π stacking with Phe116, whereas the imidazolone oxygen forms a hydrogen bond to the backbone NH of Thr223 and to water molecule 629. Moreover, one imidazolone NH interacts with the amide carbonyl of Gly221. The aromatic system is positioned in the S_3 pocket of the active cleft. The guanidine moiety of fragment **125** is in distance to the NH of the Asp81 peptide bond for a long-distance hydrogen-bond interaction (3.4 Å). Compared to its position in the *apo* structure, Asp81 is pushed apart by the fragment due to steric requirements of its 6-membered ring (Figure 7.2), which is located in the S_1 pocket. The second terminal nitrogen of the guanidine group establishes two further hydrogen bonds to Gly221 and the hydroxyl group of Thr222. A second fragment **125** is found in the S_1' pocket of the active site, its methyl group pointing in the opposite direction and placing its guanidine group in van der Waals distance to the guanidine group of the other bound fragment molecule. As an additive from crystal preparation, a DMSO molecule is observed in the structure, located in van der Waals distance in front of the fragment relative to the view direction of Figure 7.2, between the S_2 and S_3 pockets. Three copies of the fragment **261** are found in the crystal structure binding to the active site. One of these molecules establishes hydrogen bonds to Asp81 with its tertiary amino group as well as with its endocyclic amide nitrogen (Figure 7.2). An additional hydrogen bond is formed between the endocyclic amino function and a second, in the *apo* structure not existing conformation of Asp119, refined to an occupancy of 53%. The

fragment's cyclopropyl ring is placed in the S_1 pocket. The aromatic moiety is interacting through π-π stacking with a second bound copy of **261** located in the S_3/S_5 pocket. The quinoline derivative **323** is only partly observed in the electron density (Figure 7.2), with three carbon atoms of the nitrogen containing 6-membered ring missing and additional three carbons next to it. The endocyclic nitrogen is positioned within hydrogen-bonding distance to Asp81 and Ser83, while the dioxole ring reaches into the S_1 pocket. Water molecules 501 and 696 are located within van der Waals distance to the fragment. Compound **328** establishes a hydrogen bond to Asp81 through its tertiary amino group located at the end of the propyl chain. Furthermore, the exocyclic amidine nitrogen is in hydrogen-bonding distance to Asp119 as well as to water molecules 606 and 718. The aromatic ring system of the fragment is located in the S_3/S_5 pocket of the protein. Next to this fragment, a second copy of the molecule is found in the S_4/S_6 pocket, interconnected by van der Waals interactions.

7.4.3   Fragments with hydrogen bonds to Asp81 and Ser115

Fragments 73, 205, 206, 224, and 240. As a common binding motif, five fragments show the formation of hydrogen bonds through their basic functionality with Asp81 in combination with Ser115 as shown in Figure 7.4. For fragments **73** and **205**, this is achieved via their carboximidamide groups. Fragment **73** additionally establishes a hydrogen bond with Ser83, whereas **205** experiences further hydrogen bonds to the backbone oxygen of Ser115 and to water molecule 696. The protein in complex with **205** adopts a second conformation with respect to Ser115 compared to the *apo* state (refined to an occupancy of 34%). Furthermore, the fragment was modeled in two conformations with almost similar occupancy of 45–55%, leading to two variants of the thiophene ring positioned in close proximity to the S_1 pocket. A second molecule of **205** is found in the crystal structure binding to the S_1' pocket. Fragment **206** addresses Asp81 and Ser115 with its primary amino functionality. The furan ring possibly performs a π-π-stacking with the phenyl ring of Phe116, while the trifluoromethyl group is buried in the S_1 pocket and interacts through π-π-stacking with Tyr79. Furthermore, an acetate molecule from the buffer is stabilized adjacent to **206**, forming a hydrogen bond to the amino function of the fragment, which in turn forms a further hydrogen bond to a second conformation of Asp119 (occupancy 55%). Compound **224** and **240** both exhibit a secondary amine, forming hydrogen bonds to Asp81 and Ser115. The indole ring of fragment **224**, which is only partially detectable in the electron density, is aiming toward the solvent and stabilized through lipophilic contacts by the same moiety of a second bound fragment. Ligand **240** establishes, apart from hydrogen bonds to Asp81 and Ser115, a hydrogen bond to water molecule 696. Additionally, a glycerol molecule showing two conformations and a PEG
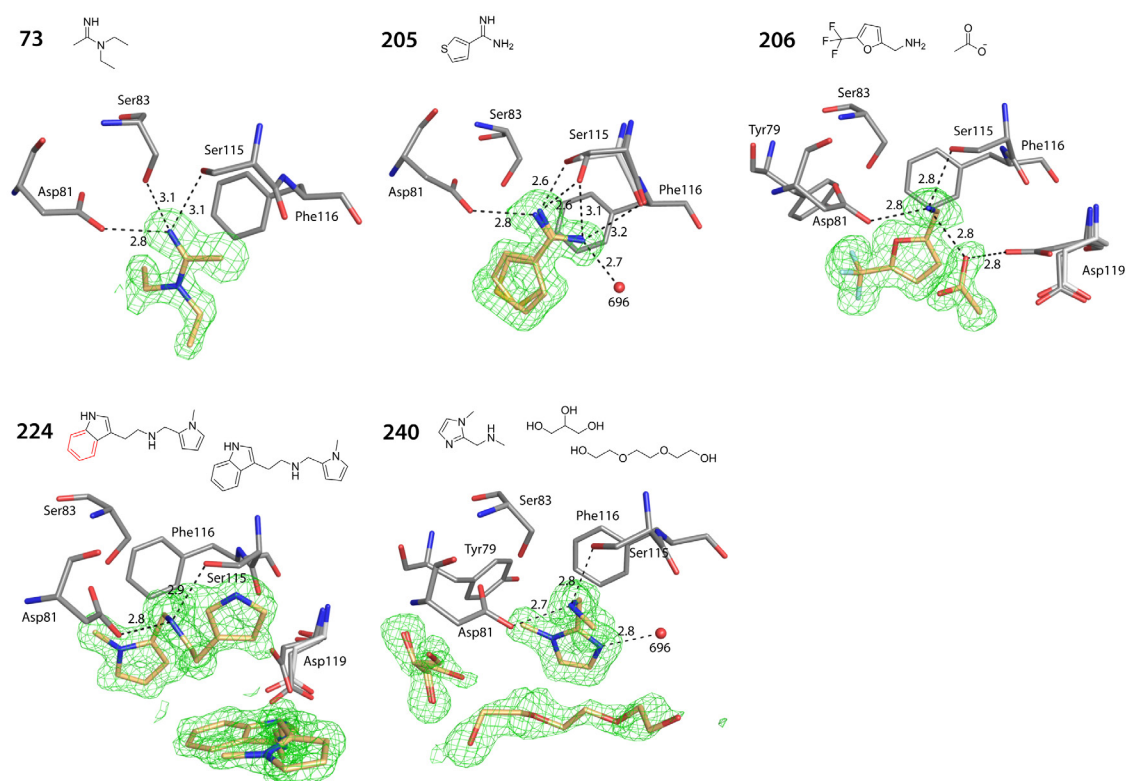
Figure 7.4. Fragments addressing Asp81 and Ser115 by hydrogen bonds. Residues of the *apo* structure are shown in light gray.

molecule are stabilized in van der Waals distance to **240**, which is encased by them in the complex with the protein.

7.4.4 Fragments with hydrophobic aliphatic moieties addressing the S₁ pocket

Fragments 41, 73, 125, 158, and 261. The backside of the S_1 pocket is formed by the phenyl rings of Tyr79, Phe116 and by the isobutyl group of Leu125. Lipophilic aliphatic and cycloaliphatic moieties of several fragments address this hydrophobic pocket (Figure 7.5). Fragments **41**, **125** and **158** bury their piperidine moieties in this pocket, whereas **73** and **261** address it with an ethyl group and a cyclopropyl group, respectively.

7.4.5 Fragments with aromatic moieties buried close to the S₁ pocket

Fragments 39, 205, 206, 224, 240, and 323. The S_1 pocket, characterized by its strict apolarity, hosts aromatic moieties of several fragments. 5-membered aromatic rings almost perfectly superimpose in their binding modes (Figure 7.6, left panel), observed for fragments
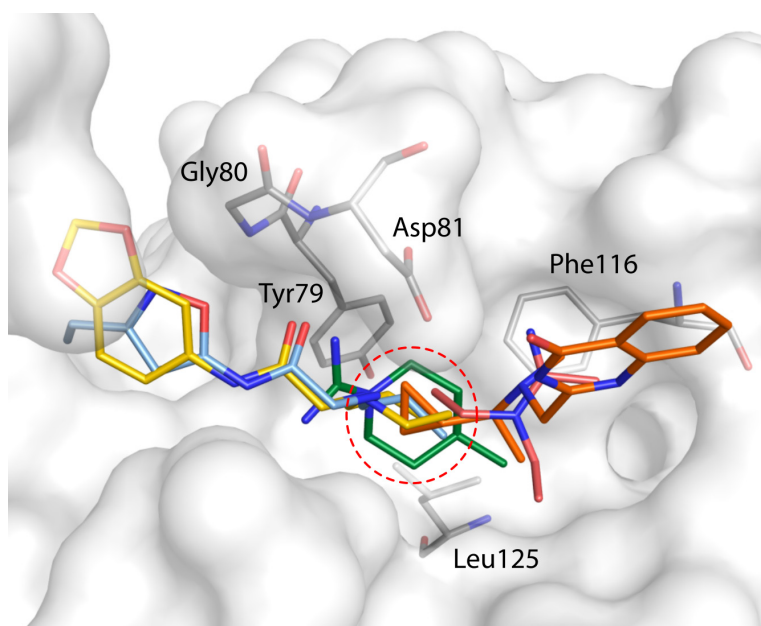
Figure 7.5. Superposition of **41** (yellow), **73** (red), **125** (green), **158** (blue), and **261** (orange). All fragments are positioned with their aliphatic, lipophilic moieties in the hydrophobic S_1 pocket formed by Tyr79, Phe116 and Leu125. The surface of the protein is displayed in white for the complex with fragment **158**.
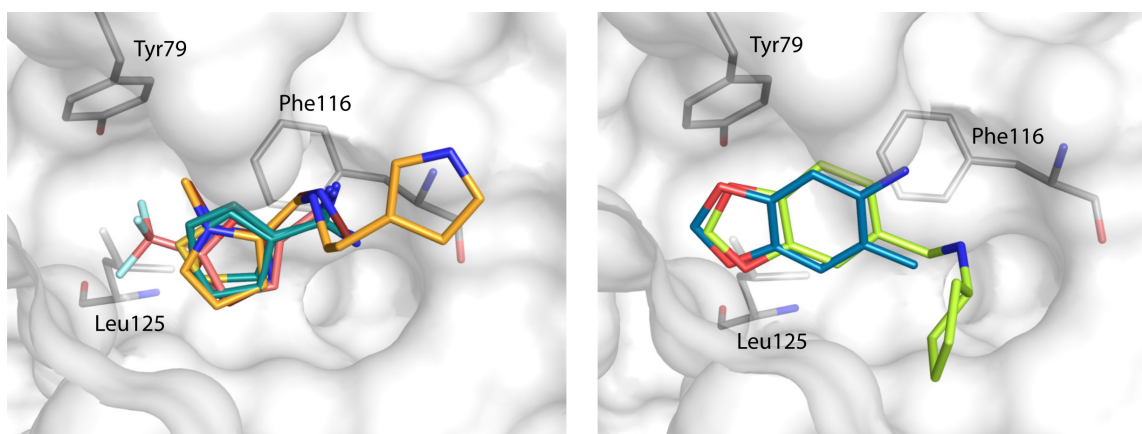
Figure 7.6. Left panel: Fragments with 5-membered aromatic rings located within the S_1 pocket (**205** in teal, **206** in pink, **224** in yellow, **240** in brown). Right panel: Fragments with a benzodioxole moiety placed in the S_1 pocket (**39** in light green, **323** in blue).

Chapter 7

205 (thiophene), **206** (furan), **224** (pyrrole) and **240** (imidazole). The benzodioxol systems of ligands **39** and **323** are positioned almost identically (Figure 7.6, right panel), addressing the S_1 pocket with their aromatic phenyl rings. The dioxole ring of fragment **39** establishes hydrogen bonds to water molecules 511 and 569 (Figure 7.7), the secondary amine is directly bound to Asp119, which undergoes a conformational change compared to its *apo* state, and to water molecule 655, which is further interacting with Ser115. Asp81 slightly shifts due to steric requirements of the phenyl ring.

7.4.6   Fragments addressing the S_2 pocket

Fragments 54, 158, and 207. The S_2 pocket of EP has a lipophilic character, mainly determined by the phenyl ring of Tyr226, the side chains of Ile300, Ile304 and the methyl group of Thr222. Consequently, fragments address this pocket with lipophilic portions (Figure 7.8). One of the two copies of fragment **54** found in the crystal structure buries its difluorophenyl moiety in the S_2 pocket, further stabilized by Gly80 and Asp81 from the flap region of EP. The two neighboring nitrogens of the 6-membered ring, located in the S_1' pocket, are in hydrogen-bond distance of the OH group of Thr222 and to water molecules 529 and 535. Adjacent to the first fragment of **158** (Figure 7.2), a second, in the electron density only partially observed fragment binds to the active site, addressing the S_2 pocket with its *iso*-propyl group (Figure 7.8). The fragment squeezes between the S_2 pocket and the flap
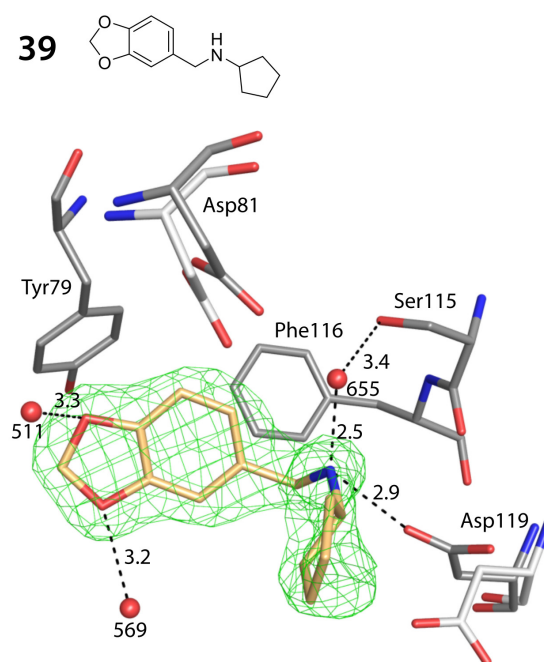

Figure 7.7. Fragment **39** in complex with EP. Residues of the *apo* structure are shown in light gray.
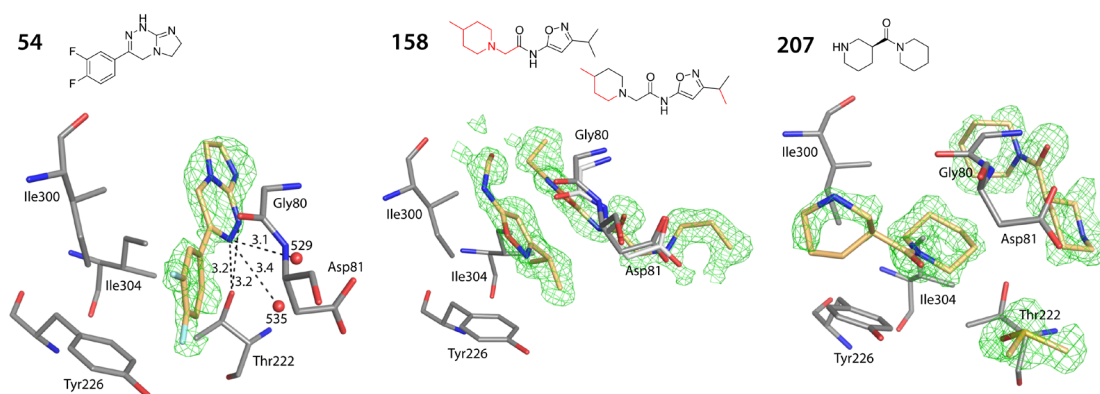
Figure 7.8. Compounds **54**, **158** and **207** addressing the S$_2$ pocket of EP. Residues of the *apo* structure are shown in light gray.

region, with its isoxazole ring even pushing Gly80 and Asp81 slightly apart compared to their *apo* state. The other fragment copy, found binding to the S$_1$/S$_1$' pocket, is positioned in van der Waals distance. One molecule **207**, present three times in the crystal structure, binds with its *N*-cyclic amide in the S$_2$ pocket, again held in position by another bound fragment molecule and the flap region of the protein through van der Waals interactions. The piperidine ring of the fragment is sticking out into the solvent.

7.4.7   Fragments binding only to the S$_3$ pocket

Fragment 35. Two fragments of **35** are found in complex with EP. One of them binds to the S$_3$ pocket, establishing hydrogen bonds through its primary amino function to Asp119, Glu118 and to the backbone oxygen of Ser115 (Figure 7.9). The tertiary nitrogen of the morpholino moiety establishes a hydrogen-bond interaction to Asp119. The ethyl group of the ligand is buried in the apolar back of the pocket, mainly lipophilicly interacting with Phe116. Furthermore, in this complex, several conformational changes of amino acids compared to their *apo* state are detected: For optimal hydrogen-bonding interactions, Asp119 completely undergoes conformational changes, whereas a second conformation is detected for Glu118, refined to an occupancy of 53%. Ser115 is pulled closed to the primary amine for a most favorable interaction distance. Due to steric requirements of the fragment's ethyl group, Asp81 occupies a position further away from the fragment in comparison to the *apo* structure.
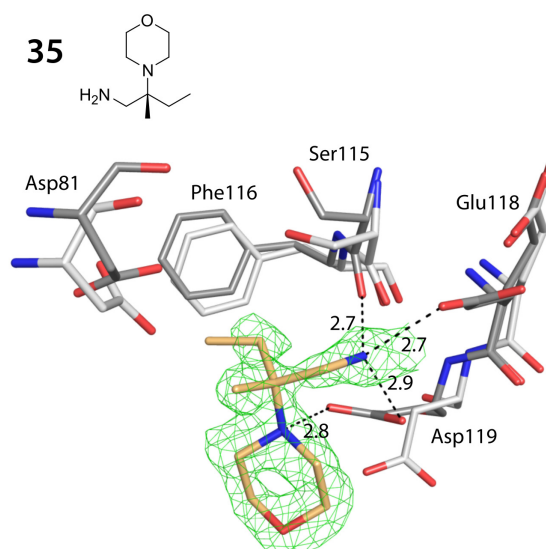
Chapter 7

Figure 7.9. Fragment **35** in complex with EP. Residues of the *apo* structure are shown in light gray.

7.4.8 Fragments with aromatic systems located in the S₃/S₅ pockets

Fragments 14, 54, 224, 261, and 328. The solvent exposed, mainly apolar S_3/S_5 pocket formed by Ile10, Asp15, Asp16, Ile122 and Asp119 is repeatedly occupied mainly by aromatic systems (Figure 7.10). Compounds **14** and **261** address this pocket with their phenyl rings, **224** with an indole moiety, and fragment **328** locates its aromatic 5-6 membered ring system in this region. Fragment **54** places its 6-membered ring of the bicyclic, not aromatic ring system at the same location. Its 5-membered imidazole derivative is pointing into the S_3 pocket, forming a hydrogen bond to Asp119 (Figure 7.11). The amino acid adopts a different conformation compared to its *apo* state, enabling the formation of a hydrogen bond. The difluoro-substituted phenyl ring is placed in van der Waals distance to the amino acids forming the apolar S_4 pocket (Phe280, Phe291 and Leu224).
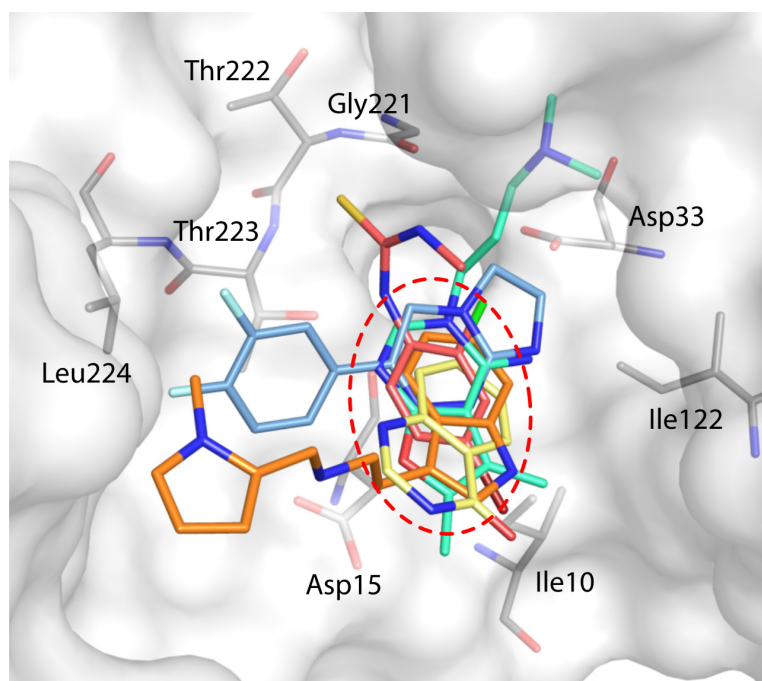
Figure 7.10. Fragments accommodating the S_3/S_5 pocket with an aromatic moiety: **14** (red) **54** (light blue), **224** (orange), **261** (yellow), **328** (light green). The surface of the *apo* structure and adjacent amino acids are shown. Additionally bound fragments and additives are omitted for clarity.
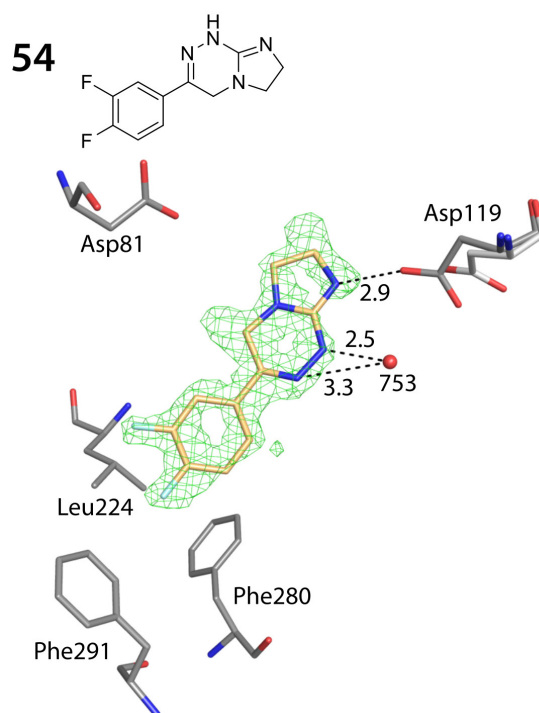

Figure 7.11. Fragment **54** in its complex with EP. Residues of the *apo* structure are shown in light gray.

Chapter 8

**General Discussion
and Conclusion**

8.1 The thermodynamic and kinetic impact of the stabilization of the water structure adjacent to the apolar solvent-exposed surface of thermolysin-bound ligands

The hydrophobic effect is one of the major driving force of most protein–ligand complexation reactions [12, 357]. However, in contrast to the aggregation of simple hydrophobic convex-shaped model systems (e.g. methane) in water [50], the hydrophobic effect in the context of protein–ligand binding cannot be characterized as generally entropy-driven. For instance, the displacement of highly mobile water molecules from a hydrophobic surface upon complex formation results in an enthalpy-driven binding process [57, 58]. The thermodynamic signature of the hydrophobic effect is clearly described by the net change of the thermodynamic properties of all involved water molecules upon complex formation. Since the latter properties can unrestrictedly adopt any enthalpy–entropy ratio [45, 64, 65], the hydrophobic effect can also range from entirely entropy- to entirely enthalpy-driven. Recently, it has been shown that the extent to which the first solvation layer adjacent to the solvent-exposed surface of the protein–ligand complex is stabilized takes a significant impact on the thermodynamic signature of the hydrophobic effect [42, 64]. In the following, the structural, thermodynamic and kinetic characterization of this particular effect is discussed. The presented results have been already published in three publications [66, 104, 105] and are described in great detail in Chapters 3–5.

8.1.1 Thermolysin as a model protein

To study the role of the surface water layer stabilization adjacent to an apolar surface patch on the thermodynamic signature of ligand binding, we selected the zinc metalloprotease thermolysin (TLN) from *Bacillus thermoproteolyticus* as a model system [190, 209, 230]. TLN consists of 316 amino acids, a zinc ion located at the active site that is essential for the catalytic activity [358], and four calcium ions — three located in the C-terminal domain and one located in the N-terminal domain (Figure 8.1). The C-terminal domain, rich in α-helices, is connected via a central α-helix to the N-terminal domain, which is mainly composed of β-sheets (Figure 8.1) [231, 359]. The active cleft of TLN is located on top of the central α-helix (Figure 8.1) [359], and can be divided into three sub-pockets (Figure 8.1) [64, 66, 104, 194, 230]. The deep, apolar S_1' specificity pocket recognizes the side chains of bulky amino acids such as Leu, Ile, and Phe [230]. The unspecific S_2' pocket is predominantly apolar, open and well-solvated. The S_1 pocket is also rather hydrophobic and solvent accessible, and recognizes Phe [64, 194, 230]. A Val-Lys dipeptide binding to the S_1'/S_2' pockets of TLN is frequently observed in "apo" crystals of TLN (Figure 8.1) [212]. This Val-Lys dipeptide was cleaved from the C-terminus of TLN due to a self-digestion of the protein in concentrated
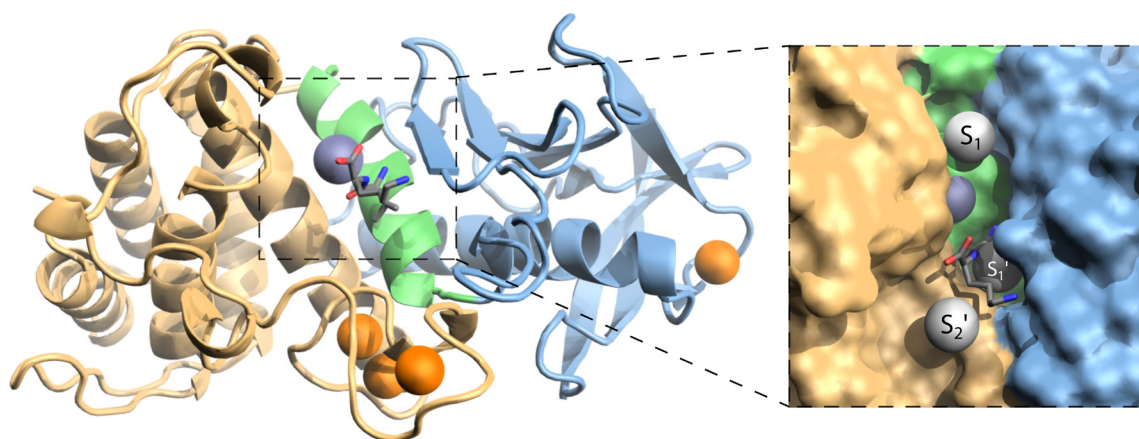
Figure 8.1. The tertiary structure of TLN (PDB code 8TLN) shown as a ribbon model (left). The C-terminal domain is shown in light orange, the N-terminal domain is shown in blue, and the central α-helix is shown in green. The zinc ion is shown as a dark blue sphere, and the four calcium ions are shown as orange spheres. The Val-Lys dipeptide in complex with TLN is shown as a dark gray stick model with color-coded heteroatoms. In the close-up view on the right side, the solvent excluded surface of the active site of TLN is shown and the sub-pockets are indicated as white spheres.

solutions. Consequently, the dipeptide has to be considered an inhibitor that needs to be displaced prior to the binding of other ligands, e.g. during an ITC measurement. Thus, a contribution from this displacement will be included in the measured thermodynamic profile of every ligand. However, since the dipeptide is present in every titration, its contribution cancels out in a relative comparison of thermodynamic binding data (Chapter 2.7). TLN is characterized by a remarkably high thermal stability of $T_{50} = 86.9$ °C (incubation at this temperature for 30 mins reduces the protein activity by 50%) [231, 360]. This is mainly determined by the calcium ion located in the N-terminal domain and the largely α-helical C-terminal domain [231, 361, 362]. As a result of the high thermal stability, TLN also shows high structural rigidity (Figure 4.3). This facilitates the correlation of structural with thermodynamic and binding kinetic data, since structural adaptions of TLN between differently bound inhibitors can be largely excluded (Chapter 2.3). With the flat, apolar, and well-solvated S_2' pocket, TLN features a pocket with optimal properties for the intended studies. In addition, the protein is optimally suited because it is readily commercially available in large quantities (required for the ITC measurements), crystallizes well, and TLN crystals routinely diffract up to a very high resolution (about 1.1 Å). The high diffraction power of TLN crystals is crucial for studying the arrangement of surface water molecules, because the amount of water molecules observed in a crystal structure correlates with the resolution of the dataset [199].

8.1.2 The impact of surface water stabilization on the thermodynamics of ligand binding

The thermodynamic analysis of protein–ligand binding reactions can provide insight into the energetic nature of the affinity-determining interactions of a ligand to its target protein and thus can help to develop a better understanding of the molecular binding mechanism and of the non-covalent interactions involved [66, 363]. Our initial study (Chapter 3) describes the structural and thermodynamical characterization of nine congeneric ligands with respect to the analysis of the impact of surface water rearrangement on the binding properties of the ligands [104]. Since the thermodynamic binding profiles as determined by ITC represent the entire binding event (discussed in Chapter 2.3), congeneric ligand series are essential to extract the energetic contribution of individual ligand portions [39]. The characterized congeneric ligands have an identical parent scaffold, which is substituted at the P_2' position with differently large, strictly apolar portions (Figure 3.1). The high-resolution crystal structures (Table 3.1) revealed that the binding modes of all nine ligands to thermolysin are identical (Figure 3.2a) and that the chemically diverse P_2' group is the only structural difference between their binding modes. Consequently, this enables attributing the differences between their thermodynamic binding profiles to the structural differences induced by the P_2' groups. The ITC measurement protocol was optimized to enable the highest possible level of relative comparability between the thermodynamic profiles of the ligands (discussed in Chapters 2.7, 2.8, 2.10 and 2.11). All measurements were performed by the same operator on the same ITC device, using the same buffer at stable pH and temperature. The same thermolysin batch was used, providing similar protein activity. Furthermore, all ligands were measured with a similar protein concentration, resulting in isotherms described by c-values of 11–158 (Table S3.3), the optimal window for reliable parameter extraction [115, 132, 135]. High ligand purity was confirmed by HPLC and further verified by a stable site parameter n as analyzed by ITC (mean: 0.753±0.04). No correction for the heat of ionization was performed, since it is identical for all nine ligands and thus cancels out in a relative comparison. However, this renders the determined thermodynamic profiles unsuitable for a potential absolute scale comparison with other data (discussed in Chapter 2.8).

The ITC measurements revealed clear differences between the thermodynamic profiles of the nine ligands (Figure 3.8). The maximum difference between enthalpy ($\Delta\Delta H°$=15.8±0.4 kJ mol^{-1}) and entropy ($-T\Delta\Delta S°$=12.1±0.5 kJ mol^{-1}) is much more pronounced than the difference in affinity ($\Delta\Delta G°$=7.0±0.4 kJ mol^{-1}). This is a well-known characteristic of many protein–ligand binding interactions termed enthalpy–entropy compensation (Chapter 2.7) [35, 36]. Remarkably, the calculated buried surface areas of the ligands, as displayed in Figure 3.9, obviously do not correlate directly with the entropy trend. From the first to the fifth

ligand, the buried surface area increases continuously (Figure 3.9), whereas the entropy decreases from the first to the third ligand (Figure 3.8). This rough inverse correlation is contradictory to the "classical" hydrophobic effect (Chapter 1.4.1) [43, 44, 50, 51], where an increase in buried surface area would be responded by an increase in entropy due to desolvation of an increasingly larger apolar surface area [18, 56]. Consequently, we reasoned that other effects must be in operation. Throughout the nine complexes, the water networks adjacent to the P_2' groups undergo considerable transformations (Figure 3.3). This is due to the differently shaped solvent-exposed surfaces established by the protein–ligand complexes. Depending on the topography of these surfaces, water molecules from the first solvation layer have to arrange adjacent to them in differently organized water networks. Figure 3.4 visualizes the Lee-Richards surfaces [207] of selected protein–ligand complexes. This surface visualization can support to understand water rearrangements due to sterical pressure as a result of insufficient contact distances. Similar as described in earlier reports [42, 64], we found a remarkable correlation facing the water networks as observed in the crystal structures with the measured thermodynamic binding profiles of the ligands. Whereas the stabilization and integration of surface water molecules into a hydrogen-bond network favors enthalpy, entropy increases if the mobility of water molecules increases or, in the extreme case, if they are entirely displaced into the bulk water phase (Chapter 3.4). Information about the residual mobility of a water molecule can be derived from a crystal structure as indicated in Figures 3.6, 4.4 and 4.5. Well stabilized water molecules are characterized by a clear, round electron density sphere and a low B factor (Figure 3.6a, b). By contrast, mobile water molecules show weak electron densities and increased B factors (Figure 3.6c, d), or they are not detectable at all in the crystal structure in case of very high mobility. The pronounced water networks observed adjacent to the P_2' groups of the first five ligands of the series (Figure 3.3) are reflected in the measured thermodynamic profiles by a strong enthalpic term (Figure 3.8). By contrast, the last four ligands of the series show strong entropic terms (Figure 3.8), correlating with their weakly stabilized, disrupted water networks (Figure 3.3). One example which impressively depicts how the rearrangement of a single methyl group can already influence the thermodynamic profile is displayed in Figure 8.2. The two compared ligands are substituted with a *iso*-butyl (green) and a *neo*-pentyl (pink) P_2' group, respectively. Thus, they differ only by a single methyl group. The buried surface area of the ligand with the *neo*-pentyl P_2' group is only slightly higher than the buried surface of the ligand with the *iso*-butyl P_2' group (Figure 3.9). However, whereas the ligand with the *iso*-butyl P_2' group stabilizes an elaborate water network adjacent to its P_2' group, the *neo*-pentyl P_2' group fails to stabilize four of these water molecules, resulting in a strong disruption of its adjacent water network. A dramatic loss in enthalpy by $+9.9\pm0.4$ kJ mol^{-1} is the result. However, this loss is largely compensated by a gain in entropy (-6.4 ± 0.5 kJ mol^{-1}).
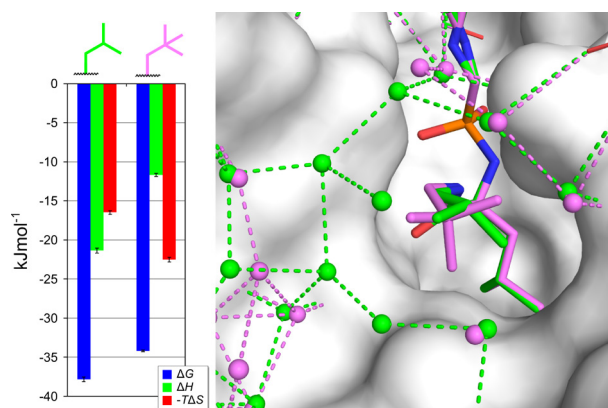
Chapter 8

Figure 8.2. Thermodynamic and structural comparison of two ligands substituted with an *iso*-butyl (green) and a *neo*-pentyl (pink) P$_2'$ group. The diagram shows the thermodynamic profiles of the two ligands determined by ITC. The chemical structure of the parent ligand scaffold is displayed in Figure 3.1. The figure on the right shows the superimposition of the crystal structures of both ligands in complex with thermolysin. Water molecules (spheres) and hydrogen bond distances (dotted lines) are displayed in the color of the respective ligand. The solvent excluded surface of thermolysin is displayed in white. This figure was selected as Table of Contents graphic for publication [104].

Nevertheless, the affinity of the *neo*-pentyl P$_2'$ group substituted ligand is more than half an order of magnitude (+3.5±0.2 kJ mol^{-1}) reduced compared to the *iso*-butyl P$_2'$ group substituted ligand. The three ligands of the series with the highest affinities comprise an *n*-propyl, an *iso*-butyl or an (*S*)-2-methylbutyl P$_2'$ substituent (Figure 3.8). For these ligands, rather perfect water networks are detected adjacent to their P$_2'$ groups (Figure 3.3). Considering the typically assumed radius for a water molecule (1.40 Å), their apolar solvent-exposed surfaces are almost entirely covered (Figure 8.3A–C). The ligand exhibiting overall the highest affinity is substituted with an (*S*)-2-methylbutyl P$_2'$ group (Figure 3.8). The thermodynamic profile of this ligand is characterized by a strong enthalpic as well as a strong entropic term. Remarkably, adjacent to the left rim of the complex (relative to the view direction of Figure 3.3 and Figure 8.3), a pentagonal water network is observed, an arrangement well known for its enthalpically favorable energetic state [245, 246]. Furthermore, Figure 3.9 suggests that the strong entropic term results from a pronounced burial of apolar surface area by this substituent. Consequently, to achieve highest affinities, it seems necessary to maximize desolvation as well as increase the stabilization of the water structure of the first solvation layer. The increase in enthalpy due to the increased water molecule stabilization seems to overcompensate for entropic losses due to the latter, overall improving affinity.

Several conclusions can be drawn from the discussed cases of the first study (Chapter 3) [104]. First of all, it would not have been possible to identify these effects by having only access to the affinity data of the compounds and not to their complete thermodynamic
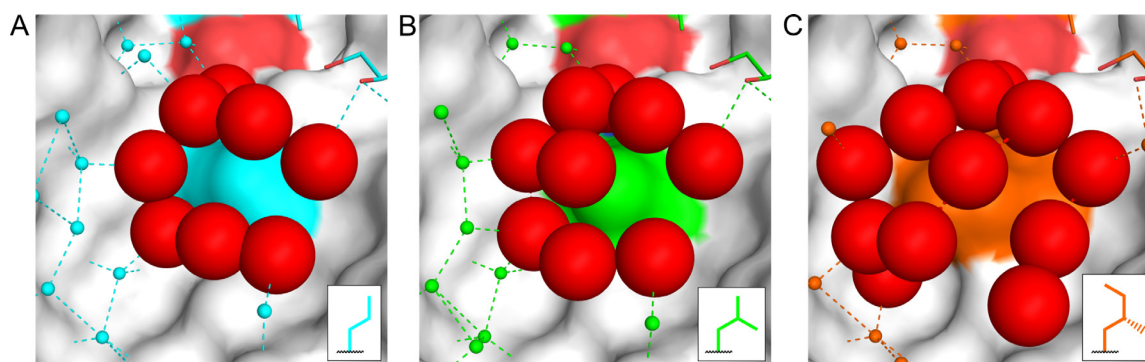
Figure 8.3. Close-up view on the P_2' groups of the three thermolysin-bound ligands exhibiting the highest affinities of the congeneric ligand series of the first study as shown in Figure 3.8. At the P_2' position, the parent scaffold of the ligands (shown in Figure 3.1) is substituted with **(A)** a *n*-propyl group (cyan), **(B)** an *iso*-butyl group (green) and **(C)** an (*S*)-2-methylbutly group (orange). The protein–ligand complexes are displayed as solvent excluded surface representation with the surface of thermolysin in white and the surface of the bound ligand in the color of the respective ligand. Water molecules in van der Waals interaction distance to the P_2' groups of the ligands are displayed as large red spheres with a radius of 1.40 Å. Other water molecules are displayed as small spheres in the color of the respective ligand. Hydrogen-bonding distances are indicated as dotted lines.

binding profiles. The observed changes in affinity are significantly smaller than the changes for enthalpy and entropy, due to a compensation of the latter (Figure 3.8). Furthermore, since affinity data alone do not provide any insights into the "nature" of interactions, it does not allow factorizing into contributions from interface desolvation and from water network stabilization. However, this appears absolutely necessary, since considering only desolvation effects via the buried apolar surface area (Figure 3.9) does not allow to explain the observed affinity trends (Figure 3.8). It must rather be considered that besides desolvation the stabilization of the water networks adjacent to the protein-bound ligands also takes a significant impact on the enthalpy–entropy binding profiles and overall on the affinities of the bound ligands. Modification by a single methyl group can already induce strong differences in the stabilization of the adjacent water structure and thus in the thermodynamic binding profile (Figure 8.2). Consequently, it makes sense to attach different aliphatic side chains to a given scaffold to discover the best binder. For the successful computational prediction of the affinities of such side chain-modified derivatives, it is absolutely necessary to explicitly model the stabilization and contribution of surface water molecules apart from modeling contributions of the apolar surface area desolvation — which, however, does not generally result in an increase in entropy, but rather depends on the individual thermodynamic properties of the displaced water molecules. Only then can a reliable prediction of such cases result. Similar effects are very likely also in operation in other systems. However, data quality (crystal structure resolution, thermodynamic parameter accuracy) and the number of available and comparable test cases are insufficient for their

identification. Thus, studies like the present one applying well-established model systems such as thermolysin hold utmost importance to increase our knowledge about such interactions at a molecular level.

8.1.3 Rational design of a P$_2$′ group with improved water network stabilization properties

Based on the observations and conclusions drawn from the first study (Chapter 3), in a follow-up study (Chapter 4) we attempted to design an aliphatic P$_2$′ group with further improved thermodynamic binding properties [66]. The successful design of such a P$_2$′ group can be considered as validation of our working hypothesis and thus holds special importance. According to our hypothesis, to design a P$_2$′ group with further improved binding properties, a pronounced water network adjacent to its solvent-exposed surface must be increasingly stabilized and in parallel desolvation must be maximized. Initially, two new ligands were designed by "merging" the (S)-2-methylbutyl P$_2$′ group from the ligand with the highest affinity of the initial study (Figure 4.1) with the 2,2-dimethylbutyl P$_2$′ group. The reasoning behind this was to add further methyl groups to the (S)-2-methylbutyl P$_2$′ group at positions where no clash with known water molecule positions would occur and where they help to seal imperfect water–water network contacts across the first solvation layer. The attachment of well-placed methyl groups to the P$_2$′ group of the ligand with the hitherto highest affinity can potentially further increase the stabilization of adjacent water molecules or even recruit new water molecules from the bulk solvent, due to the increased amount of possible van der Waals interactions. Furthermore, by increasing the size of the P$_2$′ group, a larger surface area will be buried upon binding. Both effects should result in a further boost of the ligand affinity.

The superior affinity of the ligand with the (S)-2-methylbutyl P$_2$′ group is the result of a large surface area burial (Figure 3.9) along with the stabilization of an elaborate water network comprising the energetic favorable 5-membered water polygon (Figure 3.3, Figure 4.2A left panel, Figure 4.4A, Figure 4.6). However, the stabilized water network is disrupted at the right rim of the S$_2$′ pocket. By contrast, at this side of the S$_2$′ pocket, the 2,2-dimethylbutyl P$_2$′ group stabilizes a contiguously connected water chain (Figure 3.3 and Figure 4.2A right panel). However, this ligand shows a thermodynamic profile with a sharply increased entropy and lowered enthalpy, overall resulting in a reduced affinity (Figure 4.1B). This supposedly is the result of the disrupted water network capping the ligand's P$_2$′ group (Chapter 3.4). Thus, combining the desirable features of these two P$_2$′ groups could result in a further improved P$_2$′ group resulting in an increased thermodynamic profile and affinity. Initially, two new ligands were designed, the first comprising an (S)-2,3-dimethylbutyl P$_2$′ group and the

second comprising an (S)-2,3,3-trimethylbutyl P$_2$' group. Furthermore, since the P$_2$' groups have a stereocenter, the epimeric (R)-configured ligands also were considered. The epimeric ligand pairs should have very similar desolvation energies. Thus, any changes in the thermodynamic profiles between the epimeric pairs can be attributed predominantly to the established water structures. Overall, the newly designed series comprises six congeneric ligands (Figure 4.1C). To reassure that the synthesis of the six ligands is worth the effort, the putative water networks established adjacent to these P$_2$' groups were predicted by an earlier introduced MD simulation approach (Figure 4.2C) [235]. Given that the crystal structure of the ligand with the (S)-2-methylbutyl P$_2$' group was already known at this point of the study, it was possible to validate the MD simulation predictions by facing the predicted water molecule positions for this P$_2$' group with the crystallographically determined positions. The MD simulation reproduced the known water network very convincingly (Figure 4.2B). Only the stabilization of the two water molecules on top of the apolar P$_2$' group was predicted as too weak. However, this underestimation of water-to-methyl interactions is a known issue of the applied *AMBER* force field [235]. Relative to the predicted water network of the (S)-2-methylbutyl P$_2$' group (Figure 4.2B), the predictions of the water networks adjacent to the (S)-2,3,3-trimethylbutyl P$_2$' group showed a stronger stabilization (Figure 4.2C), whereas for the ligands with the (S)-2,3-dimethylbutyl P$_2$' group and all tested (R)-configured ligands a weaker stabilized water network was predicted (Figure 4.2C). As these results appeared promising and some putatively interesting differences could be expected, we synthesized the five ligands (Scheme 4.1) and performed a structural (Figure 4.4, Table 4.1) and thermodynamical (Figure 4.8) characterization. Furthermore, to analyze whether the surface water network stabilization has any influence on the binding kinetic properties of the ligands (discussed in Chapter 8.1.4), we also performed a kinetic characterization by SPR (Figure 4.9).

The thermodynamic analysis of the new congeneric series revealed that the ligands with the (S)-2,3,3-trimethylbutyl and the (S)-2,3-dimethylbutyl P$_2$' groups indeed have a higher affinity than the ligand exhibiting the (S)-2-methylbutyl P$_2$' group (Figure 4.8). Thus, these two ligands show the highest affinities of all tested congeneric ligands exhibiting the same parent scaffold (as shown in Figure 3.1 and Figure 4.1A) and a strictly apolar P$_2$' group thus far. The enthalpy–entropy partitioning of the latter two ligands revealed a reduced enthalpic term, although this is over-compensated by an increased entropic term, resulting in a gain in affinity. The entropic contribution is also increased for the (R)-configured ligands, although this is completely compensated by a decrease in enthalpy, overall resulting in no significant affinity increase. According to the crystal structure analysis, the similar parent scaffold of all ligands and the arrangement of the amino acids lining the active site of thermolysin superimpose perfectly between the six crystal structures (Figure 4.3). Thus, similar as in the congeneric series of the nine ligands from the first study (Figure 3.1), any

differences observed between the thermodynamic profiles of the six new ligands can be attributed to modulations induced by the P$_2$' groups and the desolvation of the binding interface area. Adjacent to the (*S*)-configurated P$_2$' groups (Figure 4.4A–C), an increasing stabilization of the water networks is observed in the crystal structures. Apart from the electron densities, this stronger stabilization is also reflected in their *B* factors (Figure 4.5). The water network adjacent to the (*S*)-2,3,3-trimethylbutyl P$_2$' group (Figure 4.4C) even comprises multiple water polygons: the already known 5-membered water polygon is connected to a 6-membered polygon and both are integrated into an 8-membered polygon. Such water arrangements can be considered optimal with respect to their enthalpic contribution since water molecules are at least participating in two hydrogen bonds simultaneously (Figure 8.4). In contrast to the (*S*)-configurated ligands, the water networks adjacent to the (*R*)-configurated ligands are disrupted (Figure 4.4D–F) and the *B* factors of these water molecules are significantly increased (Figure 4.5). The reason for this is that the P$_2$' portions in (*R*)-configuration are sterically more demanding and their surface shapes are unsuitable to achieve a pronounced stabilization of an adjacent water network. In both series, aside from the ligand with the (*R*)-2-methylbutyl P$_2$' group — which shows a binding mode flipped by 180° (Figure 4.4D) and thus cannot be compared to the other ligands directly — the buried surface area increases with an increasing number of carbon atoms (Figure 4.10). Thus, in both series, the entropic gain with increasing P$_2$' surface burial can be traced back to an improved desolvation. However, the enthalpic cost for this desolvation process is only over-compensated in the (*S*)-series. Only in this series does the increased stabilization of the water networks adjacent to the P$_2$' substituents compensate for enthalpic losses due to desolvation upon complex formation, overall resulting in a gain in the Gibbs free energy. By contrast, in the (*R*)-series, the gain in entropy due to the increased surface burial is completely compensated by a loss in enthalpy for desolvation. This enthalpic loss is not compensated by the establishment of an enthalpically beneficial water network.

A comparison of the MD simulation predictions with the subsequently determined crystal structures shows that the degree of stabilization of the water networks adjacent to the (*S*)-2,3,3-trimethylbutyl and the (*R*)-2,3,3-trimethylbutyl P$_2$' groups was convincingly predicted on the qualitative level (Figure 4.11). However, as already reported earlier [235] and as also observed for the MD simulation of the water network adjacent to the (*S*)-2-methylbutyl P$_2$' group (Figure 4.2B), the water molecules on top of the apolar patch were predicted as too weak by the applied *AMBER* force field. Furthermore, since the (*S*)-2,3-dimethylbutyl and the (*R*)-2,3-dimethylbutyl P$_2$' groups are disordered in the crystal structures, the MD simulation predictions, which did not consider the disorder, showed reduced predictive power in these cases (Figure S4.10).

The discussed results of the second study (Chapter 4) show that chemically optimizing solvent-exposed apolar ligand portions by increasing their capability to stabilize the adjacent
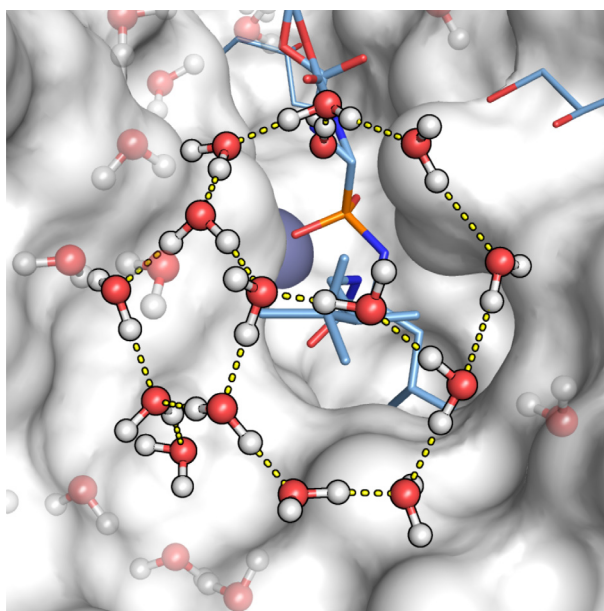
Figure 8.4. Network of the water molecules observed in the crystal structure of thermolysin in complex with the ligand substituted with a (*S*)-2,3,3-trimethylbutyl P_2' group. Hydrogen atoms were added and energy minimized with *MOE* [216]. During the minimization, all heavy atoms from the crystal structure model (PDB code 5JS3) were fixed to their crystallographic coordinates. Water molecules were allowed to freely rotate. The surface of thermolysin is displayed in white, ligand and glycerol molecules are displayed as stick models (carbon in blue, heteroatoms color-coded), and the zinc ion is displayed as dark blue sphere. Water molecules from the first solvation layer adjacent to the P_2' group are displayed as ball and stick models with hydrogen-bond interactions indicated as yellow dotted lines. For clarity reason, water molecules beyond the first solvation layer of the P_2' group are displayed as semitransparent ball and stick models and potential hydrogen bonds between them are not displayed.

water network is indeed a promising design strategy. Thereby, a fine-tuning of the thermodynamic binding properties, as well as the enhancement of the binding affinity, can be achieved. Figure 4.12 displays the differences between the affinities of all tested aliphatic P_2' substituents from the first and the second study (Chapters 3 and 4). Overall, the difference between the weakest (methyl P_2' substituent) and the strongest binder ((*S*)-2,3,3-trimethylbutyl P_2' substituent) is about 1.5 orders of magnitude. As the detailed structural and thermodynamically characterization of the ligands showed, this increase in affinity results from an increased entropic contribution to binding due to increased ligand desolvation ("classical" hydrophobic effect) in combination with an increased enthalpic contribution due to an enhanced stabilization of the water networks adjacent to the solvent-exposed surface of the protein–ligand complexes. Remarkable, adjacent to the (*S*)-2,3,3-trimethylbutyl P_2' group of the ligand with the highest affinity, an almost perfect water network is observed, comprising three partially fused water polygons (Figure 4.4C and Figure 8.4). Apparently, this water network results in an enthalpically favorable component to binding. Prerequisite for the discussed rational optimization of the P_2' group was the

availability of the high-quality data from the first study (Chapter 3). However, also in the case where such high-quality data are lacking — for instance, due to weakly diffracting crystals or due to the lack of protein material — the mentioned MD simulation approach can help to identify the chemical structure of the most suited substituent to be attached to a given scaffold.

8.1.4 The impact of surface water stabilization on the kinetics of ligand binding

The molecular mechanism of protein–ligand binding kinetics is only well understood for a limited number of cases [247, 261–263]. Since the degree of stabilization of the water networks adjacent to a protein–ligand complex can significantly influence the thermodynamic binding profile as well as the affinity (*vide supra*), the stabilization of the adjacent water network can potentially also exert an influence on the binding kinetics of the ligand. Modification of the binding kinetics via the stabilization of the adjacent water network could represent a general mechanism accessible through chemical modification of the ligand, independent of the target protein.

Chapter 5 describes the analysis of the binding kinetics of seventeen congeneric thermolysin inhibitors (Figure 5.1A) [105]. The thermodynamic binding profiles, as well as crystal structures of fifteen of these ligands, were already determined in earlier studies [64, 66, 104]. According to their P_2' substitution, the ligands can be divided into three groups (Figure 5.1). The first group (*a*) comprises five ligands substituted with a carboxy substituent in combination with an aliphatic portion at the P_2' position. The second group (*b*) has only one representative, which is substituted with a carboxamide in combination with an aliphatic portion at P_2'. The third group (*c*) comprises eleven representatives, which are substituted solely with an aliphatic substituent at P_2'. The structural and thermodynamic characterization described in Chapters 3 and 4 were performed with ligands from the latter category. Five of them were already kinetically analyzed in the publication described in Chapter 4 (Figure 4.9). Within this subset of five ligands of group *c*, an indication is given that the degree to which the water network adjacent to the bound ligand is stabilized can have a significant impact on the binding kinetics of the ligand, particularly with respect to the dissociation rate constant. The ligand with the (*S*)-2,3,3-trimethylbutyl P_2' substituent has, compared to the other four ligands of this subset, a significantly prolonged dissociation rate and thus residence time (Figure 4.9). This ligand stabilizes the most pronounced water network (Figure 4.4 C) and overall shows the highest affinity compared to the other ligands (Figure 4.8 and Figure 4.12). The supposed reason for the prolonged residence time is the significantly more stable water network adjacent to this protein-bound ligand, resulting in an additional stabilization of the protein–ligand complex. The energetic barrier to disrupt this water network and the involved

hydrogen bonds is increased relative to the other complexes, resulting in the prolongation of the dissociation rate of this ligand.

Figure 5.1 shows the kinetic map determined for the seventeen ligands. Apparently, the magnitude of the dissociation rate constant k_d correlates with the P_2' substitution of the ligand, whereas the association rate constant k_a is less affected. Mainly differences in the dissociation rate constants were observed between the ligands. This could be the result of a large conformational transition of thermolysin upon ligand binding, rendering the contributions of individual ligands to the rather slow association rate constants less significant. Ligands from the group c substituted only with aliphatic P_2' substituents show the fastest dissociation rates. Ligands from the group a substituted with a charged carboxy portion at P_2' show the slowest dissociation rate constants. The ligand from group b substituted with the neutral carboxamide has a dissociation rate that falls between the other two groups. The superposition of all seventeen crystal structures shows that the binding mode of the parent scaffold of the seventeen congeneric ligands of all three groups a–c perfectly superimposes (Figure 5.2B). Thus, the observed differences in the binding kinetics must be induced by the differences between the P_2' substituents. As shown in Figure 5.3A, a striking difference between the three groups of ligands is their hydrogen-bonding pattern with Asn112. Whereas carboxy-substituted ligands (group a) establish a charge-assisted hydrogen bond with Asn112 (Figure 5.3A), the hydrogen bond formed between the ligand substituent with the neutral P_2' carboxamide (group b) and the residue Asn112 is weakened due to the loss of the charge. Finally, the group of ligands that exhibit solely aliphatic P_2' substituents (group c) cannot establish a hydrogen bond with Asn112 at all. While thermolysin is in complex with a ligand, Asn112 acts as a lid on top of the S_1' pocket, preventing the P_1' portion of the ligand from leaving the binding pocket (Figure 5.2A and Figure 5.4A). For the ligand dissociation to occur, Asn112 needs to provide space via a rotation to the side. Such a rotation of Asn112 takes place in course of a "hinge-bending" motion of thermolysin around its central α-helix (Figure 8.1) [212], resulting in the conformational transition from the closed to the opened state of thermolysin (Figure 5.4B) [276]. The stronger the interaction between the P_2' group of the ligands and the residue Asn112, the more pronounced the fixation of Asn112 and the less likely the rotation occurs. Consequently, as shown in Figure 5.1B, the dissociation rates of the ligands from their protein-bound states decrease from group a (P_2' carboxy) to group b (P_2' carboxamide) to group c (P_2' aliphatic). The importance of Asn112 for the catalytic properties is supported by the high degree of conservation of this residue across the members of the M4 family (Figure 5.5). Our hypothesis is further confirmed by ZF^{P}LA, the most potent thermolysin ligand described thus far (68 pM), exhibiting a residence time of 168 days [287]. This ligand is substituted with a benzyl group at the P_1 position. Exchange of the benzyl group at P_1 with a hydrogen atom increases the dissociation rate by a factor of 74,000 (Figure 5.6). This

amazingly long residence time as the result of the benzyl group attachment, a group that primarily sticks out into the solvent, can be explained by a steric hindrance of Asn112 to rotate from the closed to the opened state (Figure 5.7).

As already described in Chapter 4, the degree of stabilization of the water network adjacent to the protein-bound ligand also seems to have an influence on the binding kinetics of the ligand. However, these effects are less pronounced compared to the discussed kinetic effect achieved through interactions with Asn112. Especially a fine-tuning of the kinetic binding parameters might be achievable by modulating the water network stability. From the ligand group *a* (Figure 5.1A)*,* the ligand comprising a cyclohexyl portion shows a significantly shorter residence time compared to the other ligands of this group (Figure 5.1). Generally, increasing molecular mass is accepted to result in a decrease of the dissociation rate constant [263]. Thus, one would assume that this ligand actually shows a longer residence time than the other ligands, because its molecular weight is the largest of group *a*. The surprisingly increased dissociation rate constant of this ligand can be explained with the water network observed adjacent to it (Figure 5.8). The large cyclohexyl group disrupts the water network, resulting in a less complete water network compared to the other ligands of group *a*. Thus, the barrier for dissociation from the protein-bound state is reduced for this ligand.

The discussed data show that knowledge about the protein conformational flexibility holds utmost importance for the rational modulation of binding kinetic properties. Unfortunately, this will be strongly dependent on the individual characteristics of the system and thus it needs to be specifically investigated for every system. By contrast, the modulation of the stabilization of the water network adjacent to the protein-bound ligand is suggested as a more general concept. However, only fine-tuning of the kinetic binding parameters seems achievable with this approach.

8.2 The dry condition of the S_1' pocket of thermolysin exerts a determinant influence on the thermodynamics of ligand binding

At first glance, the complex formation between the *iso*-butyl P_1' group of the parent scaffold as used in the previously described congeneric ligand series (Figures 3.1, 4.1 and 5.1) and the TLN S_1' pocket appears to be a "textbook" example of the hydrophobic effect. It seems obvious that the complex formation process must be entropy-driven due to a displacement of well-ordered water molecules into the bulk water phase previously arranged along the apolar surface of the TLN S_1' pocket once the ligand's *iso*-butyl P_1' portion enters the pocket. However, the opposite is the case — the complex formation is actually highly enthalpically favorable, as disclosed by the thermodynamic analysis of a congeneric ligand series exhibiting apolar P_1' substituents of growing size (Figure 6.1). Exchanging a hydrogen at P_1' with a

single methyl group resulted in an amazingly strong affinity increase of more than 2 orders of magnitude ($\Delta\Delta G°_{H\rightarrow methyl}$ = −12.3 kJ mol^{-1}). Replacing the hydrogen with a more bulky *iso*-butyl group results in a no less impressive increase in affinity of more than 4 orders of magnitude ($\Delta\Delta G°_{H\rightarrow iso\text{-}butyl}$ = −26.3 kJ mol^{-1}). Going from the methyl to the *iso*-propyl to the *iso*-butyl substituent — for these ligands the thermodynamic partitioning into enthalpy and entropy could be measured by ITC (Figure 6.1B) — showed that increasing the size of the apolar P$_1$' portion is highly enthalpically favored ($\Delta\Delta H°_{methyl\rightarrow iso\text{-}propyl}$ = −11.0 kJ mol^{-1}; $\Delta\Delta H°_{iso\text{-}propyl\rightarrow iso\text{-}butyl}$ = −11.4 kJ mol^{-1}) with a small detrimental entropic contribution ($-T\Delta\Delta S°_{methyl\rightarrow iso\text{-}propyl}$ = +2.1 kJ mol^{-1}; $-T\Delta\Delta S°_{iso\text{-}propyl\rightarrow iso\text{-}butyl}$ = +6.4 kJ mol^{-1}). Clearly, this observation stands in stark contrast to the thermodynamic description of the "classical" hydrophobic effect (see Chapter 1.4.1).

According to crystal structures obtained through conventional crystallography employing electron density maps based on model phases, the S$_1$' cavity appears empty at first sight [194]. However, highly mobile water molecules can easily hide from crystallographic detection due to the lack of a periodic arrangement resulting in a blurred electron density with no clearly defined maximum [194]. The displacement of such a mobile, enthalpically unfavorable water molecule into the bulk water phase could explain the observed enthalpic signal. Model bias, potentially introduced by the use of a predefined model for molecular replacement and imposed during conventional structure refinement, can hamper the detection of weak electron traces resulting from highly mobile water molecules [296–298, 364] Thus, with the aim to elucidate the molecular mechanism behind the extraordinary increase in enthalpy, we determined with the best possible accuracy the solvation pattern within the S$_1$' pocket of TLN. Therefore, we collected a zinc MAD dataset of TLN in complex with the ligand substituted with a hydrogen at P$_1$' (Table 6.2), and performed an experimental phasing of the crystal structure. Applying this strategy, no model bias is introduced [293, 299, 300]. Furthermore, to have access to the total electron content within a certain confinement, we transformed the sigma-scaled electron density map to an absolute electron number scaled map. This allows us to sum up the electron content within a given volume and thereby also trace highly mobile water molecules scattered over a large volume, characterized by a low average electron density level. The remaining five ligands of the series (Figure 6.1) exhibit increasingly bulky or polar P$_1$' substituents, and were designed with the aim to reduce the mobility of a potentially present water molecule — whether by reducing the available volume or by providing a hydrogen bond interaction partner — and, by this, increase the electron concentration locally, making it crystallographically detectable [315, 316].

The S$_1$' cavity of TLN in complex with the P$_1$'-hydrogen substituted ligand opens a volume of 141 Å^3 (Figure 6.4). The experimentally phased, absolute-scale electron density map within the S$_1$' cavity superimposed on the refined model of TLN is shown in Figure 6.2B. Compared to well solvated internal reference cavities (Figure 6.2C and D), no electron concentration

Chapter 8

maximum indicating a water molecule could be observed within the S_1' cavity. In total, 6.6 electrons were detected within the S_1' cavity. An internal reference for an empty cavity — that is clearly empty due to its apolar character and small volume of only 16 $Å^3$ [306, 307] — gave only 0.9 electrons (Figure 6.2E). However, considering that the volume of the S_1' cavity is about 8.8 times larger than the volume of the empty reference cavity, we conclude that the S_1' cavity is free of any water molecules. The detected 6.6 electrons are the result of electrons spilling over from the cavity-lining residues into the volume of the cavity [299, 300].

To further experimentally characterize the S_1' pocket of TLN in complex with the P_1'-hydrogen substituted ligand, we attempted to trap the noble gases xenon and krypton inside the S_1' cavity. These two noble gases are known to preferentially bind to empty, apolar protein cavities through weak van der Waals interactions [310, 311]. Thus, they can serve as probe molecules to experimentally detect such cavities. TLN crystals were derivatized applying an Oxford cryosystems pressure cell [326]. Datasets were collected at wavelengths optimized with respect to maximizing the anomalous signal of the noble gases (Table 6.1). Subsequently, phasing was performed using molecular replacement. It was possible to unambiguously identify the bound noble gases located within the S_1' cavity by their strong anomalous signal (Figure 6.3 and Figure S6.4) [294]. The binding of the noble gases within the S_1' cavity is a further experimental confirmation of the empty solvation state of the cavity. In addition, the occupancy to which the noble gases are populated within the cavity informs about the lower detection limit of an electron concentration within conventionally refined crystal structures [314]. Under consideration of the refined occupancy and the B factor of the krypton atom (8% occupancy, B factor of 12.5 $Å^2$), at least 2.9 electrons scattered over a volume of 58 $Å^3$ are detectable by conventional refinement (for further explanation, see Chapter 6.3.2). Thus, if a water molecule with an occupancy as low as 29% (2.9 electrons) were present in the S_1' cavity of TLN in complex with the ligands exhibiting more bulky P_2' substituents — characterized by a residual volume of the S_1' cavity of about 58 $Å^3$ (Figures 6.4C, D and F) — it should be detectable in the electron density. Since this is not the case, these cavities can be considered empty, or at least as being solvated to an insignificant degree. The two ligands exhibiting polar P_1' side chains (Figures 6.4E, F) can potentially serve as hydrogen bond interaction partners to arrest a putatively present water molecule, and thereby lead to its stabilization and crystallographic detection. However, this attempt also failed to discover a water molecule within the S_1' cavity.

With respect to the described experimental findings, we conclude that the S_1' pocket of thermolysin represents a completely empty ("dry") binding pocket. It is energetically more favorable for the pocket to remain empty than to host a water molecule. This has important implications for the interpretation of the observed enthalpy-driven complex formation process. In agreement with earlier studies [317–319], addressing the empty pocket, where no pocket desolvation step has to be afforded, results in an extraordinarily high affinity gain of

more than four orders of magnitude. This is the result of dispersive interactions between residues of TLN and the P_1' apolar group of the ligand [60]. These interactions become only determinant for the affinity in cases where the desolvation step of the protein binding pocket is missing. Otherwise, the necessity to disrupt the dispersive interactions between water molecules and the binding pocket to vacate the pocket compensates for the dispersive interactions between ligand and protein. In case of the polar P_1' substituted ligands, where a much higher cost of desolvation has to be afforded, the established dispersive interactions are almost completely compensated for, and overall only a slight increase in affinity compared to the P_1'-hydrogen substituted ligand is achieved (Figure 6.1B). With respect to optimizing protein–ligand recognition, the occupancy of dry binding pockets is rewarded with a huge affinity gain. Furthermore, these findings show another facet of the hydrophobic effect. Again, water molecules are determinant for the hydrophobic association — this time, however, their absence is key.

Chapter 8

Appendix

A.1   Additionally refined and deposited crystal structures

Table A.1. Data collection and refinement statistics for the thermolysin–ligand complexes with the PDB codes 5DPE and 5DPF published in [96]. Data collection, processing and refinement were performed as described in Chapter 4.6.4.

	PDB code of crystal structure	
	5DPE	**5DPF**
(A) Data collection and processing		
Space group	$P6_122$	$P6_122$
Unit cell parameters: a, b, c (Å)	92.7, 92.7, 130.3	92.7, 92.7, 130.2
Matthews coefficient (Å^3/Da)^a	47	47
Solvent content (%)^a	2.3	2.3
(B) Diffraction data		
Resolution range (Å)	50-1.34 (1.36-1.34)	50-1.47 (1.50-1.47)
Unique reflections	72715 (3551)	56522 (2772)
$R(I)_{sym}$ (%)	6.3 (46.0)	7.3 (41.1)
Completeness (%)	97.6 (97.1)	99.5 (100.0)
Redundancy	10.3 (7.1)	9.2 (9.1)
$<I/\sigma(I)>$	35.8 (4.4)	29.3 (6.4)
(C) Refinement		
Resolution range (Å)	34.17-1.34	40.16-1.47
Reflections used in refinement (work/free)	72495 (68846/3649)	56294 (53438/2856)
Final R value for all reflections (work/free) (%)	12.7/15.6	11.9/15.2
Protein residues	316	316
Calcium/zinc ions	4/1	4/1
Inhibitor atoms	32	33
Water molecules	393	381
RMSD from ideality:		
Bond lengths (Å)	0.007	0.007
Bond angles (°)	1.183	1.151
Ramachandran plot (%):^{b}		
Residues in most favored regions	88.5	87.8
Residues in additionally allowed regions	10.7	11.5
Residues in generously allowed regions	0.4	0.4
Residues in disallowed regions^c	0.4	0.4
Mean B factor (Å^2):^{d}		
Protein	13.1	13.0
Inhibitor	15.1	13.4
Water molecules	22.7	27.6

Values in brackets describe the highest resolution shell. ^{a} Calculated with the program *Matthews_coef* from *CCP*4 version 6.3.0 [195]. ^{b} Calculated with PROCHECK [196]. ^{c} The Ramachandran outlier is Thr26 as described in literature [197]. ^{d} Calculated with MOLEMAN [198].

Bibliography

[1]   Whitesides GM, Krishnamurthy VM (2005) Designing ligands to bind proteins. *Q Rev Biophys* 38: 385–395.

[2]   Lounnas V, Ritschel T, Kelder J, Mcguire R, Bywater RP, Foloppe N (2013) Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* 5: e201302011.

[3]   Stouch TR (2012) The errors of our ways: Taking account of error in computer-aided drug design to build confidence intervals for our next 25 years. *J Comput Aided Mol Des* 26: 125–134.

[4]   Ballester PJ, Schreyer A, Blundell TL (2014) Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity? *J Chem Inf Model* 54: 944–955.

[5]   Michel J (2014) Current and emerging opportunities for molecular simulations in structure-based drug design. *Phys Chem Chem Phys* 16: 4465–4477.

[6]   Levy Y, Onuchic JN (2006) Water Mediation in Protein Folding and Molecular Recognition. *Annu Rev Biophys Biomol Struct* 35: 389–415.

[7]   Ball P (2011) More than a bystander. *Nature* 478: 467–468.

[8]   Ball P (2008) Water as a biomolecule. *ChemPhysChem* 9: 2677–2685.

[9]   Ladbury JE (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol* 3: 973–980.

[10]   Gibb B (2016) Weird and wonderful water. *Nat Chem* 8: 733–734.

[11]   Francl M (2016) A Brief History of Water. *Nat Chem* 8: 897–898.

[12]   Klebe G (2013) Protein–Ligand Interactions as the Basis for Drug Action. *Drug Design: Methodology, Concepts, and Mode-of-Action* (Springer, Heidelberg, Germany), pp 61–88.

[13]   Marshall GR (2012) Limiting assumptions in structure-based design: binding entropy. *J Comput Aided Mol Des* 26: 3–8.

[14]   Whitesides GM, Snyder PK, Moustakas DT, Mirica KA Designing ligands to bind tightly to proteins. *Physical Biology: From Atoms to Medicine*, ed Zewail AH (Imperial College Press, London, UK), pp 189–215.

[15]   Daze K, Hof F (2016) Molecular Interaction and Recognition. *Encyclopedia of Physical Organic Chemistry 3*, ed Wang Z (John Wiley & Sons), pp 1–51.

[16]   Böhm H-J (2003) Prediction of Non-bonded Interactions in Drug Design. *Protein-Ligand Interactions: From Molecular Recognition to Drug Design*, eds Böhm HJ, Schneider G (WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim), pp 3–20.

[17]   Demchenko AP (2001) Recognition between flexible protein molecules: Induced and assisted folding. *J Mol Recognit* 14: 42–61.

[18]   Homans SW (2007) Dynamics and Thermodynamics of Ligand–Protein Interactions. *Top Curr Chem* 272: 51–82.

[19]   Du X, Li Y, Xia Y-L, Ai S-M, Liang J, Sang P, Ji X-L, Liu S-Q (2016) Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci* 17: E144.

[20]   Bellissent-Funel MC, Hassanali A, Havenith M, Henchman R, Pohl P, Sterpone F, Van Der Spoel D, Xu Y, Garcia AE (2016) Water Determines the Structure and Dynamics of Proteins. *Chem Rev* 116: 7673–7697.

[21]   Chaplin M (2006) Do we underestimate the importance of water in cell biology? *Nat Rev Mol Cell Biol* 7: 861–866.

[22]   Nakasako M (2004) Water-protein interactions from high-resolution protein crystallography. *Philos Trans R Soc Lond B Biol Sci* 359: 1191–1206.

[23]   Ball P (2008) Water as an active constituent in cell biology. *Chem Rev* 108: 74–108.

[24]   Huang C, Wikfeldt KT, Tokushima T, Nordlund D, Harada Y, Bergmann U, Niebuhr M, Weiss TM, Horikawa Y, Leetmaa M, Ljungberg MP, Takahashi O, Lenz A, Ojamäe L, Lyubartsev AP, Shin S, Pettersson LGM, Nilsson A (2009) The inhomogeneous structure of water at ambient conditions. *Proc Natl Acad Sci USA* 106: 15214–15218.

[25]   Gilson MK, Zhou H-X (2007) Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* 36: 21–42.

[26]   Chang CA, Chen W, Gilson MK (2007) Ligand configurational entropy and protein binding. *Proc Natl Acad Sci USA* 104: 1534–1539.

[27]   Rühmann EH, Rupp M, Betz M, Heine A, Klebe G (2016) Boosting Affinity by Correct Ligand Preorganization for the S2 Pocket of Thrombin: A Study by Isothermal Titration Calorimetry, Molecular Dynamics, and High-Resolution Crystal Structures. *ChemMedChem* 11: 309–319.

[28]   Benfield AP, Teresk MG, Plake HR, DeLorbe JE, Millspaugh LE, Martin SF (2006) Ligand preorganization may be accompanied by entropic penalties in protein-ligand interactions. *Angew Chemie - Int Ed* 45: 6830–6835.

[29]   Martin SF, Clements JH (2013) Correlating structure and energetics in protein-ligand interactions: paradigms and paradoxes. *Annu Rev Biochem* 82: 267–293.

[30]   Dunitz JD (1994) The entropic cost of bound water in crystals and biomolecules. *Science* 264: 670.

[31]   Bissantz C, Kuhn B, Stahl M (2010) A medicinal chemist's guide to molecular interactions. *J Med Chem* 53: 5061–5084.

[32]   Amadasi A, Surface A, Spyrakis F, Cozzini P, Mozzarelli A, Kellogg GE (2008) Robust Classification of "Relevant" Water Molecules in Putative Protein Binding Sites. *J Med Chem* 51: 1063–1067.

[33]   Barillari C, Taylor J, Viner R, Essex JW (2007) Classification of Water Molecules in Protein Binding Sites. *J Am Chem Soc* 129: 380–384.

[34]   Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today* 13: 869–874.

[35]   Dunitz JD (1995) Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol* 2: 709–712.

[36]   Chodera JD, Mobley DL (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu Rev Biophys* 42: 121–142.

[37]   Ryde U (2014) A fundamental view of enthalpy–entropy compensation. *Medchemcomm* 5: 1324–1336.

[38]   Rekharsky MV, Mori T, Yang C, Ko YH, Selvapalam N, Kim H, Sobransingh D, Kaifer AE, Liu S, Isaacs L, Chen W, Moghaddam S, Gilson MK, Kim K, Inoue Y (2007) A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy-entropy compensation. *Proc Natl Acad Sci USA* 104: 20737–20742.

[39]   Klebe G (2015) Applying thermodynamic profiling in lead finding and optimization. *Nat Rev Drug Discov* 14: 95–110.

[40]   Baum B, Muley L, Smolinski M, Heine A, Hangauer D, Klebe G (2010) Non-additivity of functional group contributions in protein-ligand binding: A comprehensive study by crystallography and isothermal titration calorimetry. *J Mol Biol* 397: 1042–1054.

[41]   Patel Y, Gillet VJ, Howe T, Pastor J, Oyarzabal J, Willett P (2008) Assessment of Additive/Nonadditive Effects in Structure-Activity Relationships:Implications for Iterative Drug Design. *J Med Chem* 51: 7552–7562.

[42]   Biela A, Betz M, Heine A, Klebe G (2012) Water Makes the Difference: Rearrangement of Water Solvation Layer Triggers Non-additivity of Functional Group Contributions in Protein-Ligand Binding. *ChemMedChem* 7: 1423–1434.

[43]   Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14: 1–63.

[44]   Tanford C (1978) The hydrophobic effect and the organization of living matter. *Science* 200: 1012–1018.

[45]   Snyder PW, Lockett MR, Moustakas DT, Whitesides GM (2014) Is it the shape of the cavity, or the shape of the water in the cavity? *Eur Phys J Spec Top* 223: 853–891.

[46]   Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges* 27: 2985–2993.

[47]   Koshland DE (1958) Application of a Theory of Enzyme Specificity To Protein Synthesis. *Proc Natl Acad Sci USA* 44: 98–104.

[48]   Vogt AD, Di Cera E (2013) Conformational selection is a dominant mechanism of ligand binding. *Biochemistry* 52: 5723–5729.

[49]   Jorgensen WL (1991) Rusting of the lock and key model for protein-ligand binding. *Science* 254: 954–955.

[50]   Frank HS, Evans MW (1945) Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J Chem Phys* 13: 507–532.

[51]   Ross PD, Subramanian S (1981) Thermodynamics of Protein Association Reactions: Forces Contributing to Stability. *Biochemistry* 20: 3096–3102.

[52]   Finney JL, Soper AK (1994) Solvent Structure and Perturbations in Solutions of Chemical and Biological Importance. *Chem Soc Rev* 23: 1–10.

[53]   Buchanan P, Aldiwan N, Soper AK, Creek JL, Koh CA (2005) Decreased structure on dissolving methane in water. *Chem Phys Lett* 415: 89–93.

[54]   Galamba N (2013) Water's Structure around Hydrophobic Solutes and the Iceberg Model. *J Phys Chem B* 117: 2153–2159.

[55]   Meyer EA, Castellano RK, Diedrich F (2003) Ineractions with Aromatic Rings in Chemical and Biological Recognition. *Angew Chem Int Ed* 42: 1210–1250.

[56]   Homans SW (2007) Water, water everywhere - except where it matters? *Drug Discov Today* 12: 534–539.

[57]   Setny P, Baron R, McCammon JA (2010) How can hydrophobic association be enthalpy-driven? *J Chem Theory Comput* 6: 2866–2871.

[58]   Baum B, Mohamed M, Zayed M, Gerlach C, Heine A, Hangauer D, Klebe G (2009) More than a Simple Lipophilic Contact: A Detailed Thermodynamic Analysis of Nonbasic Residues in the S1 Pocket of Thrombin. *J Mol Biol* 390: 56–69.

[59]   Haider K, Wickstrom L, Ramsey S, Gilson MK, Thomas K (2016) Enthalpic Breakdown of Water Structure on Protein Active-Site Surfaces. *J Phys Chem B* 120: 8743–8756.

[60]   Barratt E, Bingham RJ, Warner DJ, Laughton CA, Phillips SE V, Homans SW (2005) Van der Waals Interactions Dominate Ligand-Protein Association in a Protein Binding Site Occluded from Solvent Water. *J Am Chem Soc* 127: 11827–11834.

[61]   Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* 437: 640–647.

[62]   Cheng Y-K, Rossky PJ (1998) Surface topography dependence of biomolecular hydrophobic hydration. *Nature* 392: 696–699.

[63]   Berne BJ, Weeks JD, Zhou RH (2009) Dewetting and Hydrophobic Interaction in Physical and Biological Systems. *Annu Rev Phys Chem* 60: 85–103.

[64]   Biela A, Nasief NN, Betz M, Heine A, Hangauer D, Klebe G (2013) Dissecting the hydrophobic effect on the molecular level: The role of water, enthalpy, and entropy in ligand binding to thermolysin. *Angew Chemie - Int Ed* 52: 1822–1828.

[65]   Snyder PW, Mecinovic J, Moustakas DT, Thomas SW, Harder M, Mack ET, Lockett MR, Heroux A, Sherman W, Whitesides GM (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci USA* 108: 17889–17894.

[66]   Krimmer SG, Cramer J, Betz M, Fridh V, Karlsson R, Heine A, Klebe G (2016) Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein Bound Ligands. *J Med Chem* 59: 10530–10548.

[67]   Renaud J-P, Chung C, Danielson UH, Egner U, Hennig M, Hubbard RE, Nar H (2016) Biophysics in drug discovery: impact, challenges and opportunities. *Nat Rev Drug Discov* 15: 679–698.

[68]   Sanderson MR, Skelly JV eds. (2007) *Macromolecular Crystallography: Conventional and High-Throughput Methods* (Oxford University Press Inc., New York)

[69]   Wlodawer A, Minor W, Dauter Z, Jaskolski M (2013) Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J* 280: 5705–5736.

[70] Shi Y (2014) A glimpse of structural biology through X-ray crystallography. *Cell* 159: 995–1014.

[71] Zheng H, Hou J, Zimmerman MD, Wlodawer A, Minor W (2014) The future of crystallography in drug discovery. *Expert Opin Drug Discov* 9: 125–137.

[72] Brown EN, Ramaswamy S (2007) Quality of protein crystal structures. *Acta Crystallogr Sect D Biol Crystallogr* 63: 941–950.

[73] Bowler MW, Svensson O, Nurizzo D (2016) Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data. *Crystallogr Rev* 22: 223–249.

[74] Mooij WTM, Hartshorn MJ, Tickle IJ, Sharff AJ, Verdonk ML, Jhoti H (2006) Automated protein-ligand crystallography for structure-based drug design. *ChemMedChem* 1: 827–838.

[75] Helliwell JR, Mitchell EP (2015) Synchrotron radiation macromolecular crystallography: science and spin-offs. *IUCrJ* 2: 283–291.

[76] Schiebel J, Krimmer SG, Röwer K, Knörlein A, Wang X, Park AY, Stieler M, Ehrmann FR, Fu K, Radeva N, Krug M, Huschmann FU, Glöckner S, Weiss MS, Mueller U, Klebe G, Heine A (2016) High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* 24: 1398–1409.

[77] Davies TG, Tickle IJ (2012) Fragment Screening Using X-Ray Crystallography. *Top Curr Chem* 317: 33–59.

[78] Amaro RE (2016) Drug Discovery Gets a Boost from Data Science. *Structure* 24: 1225–1226.

[79] Ghosh E, Kumari P, Jaiman D, Shukla AK (2015) Methodological advances: the unsung heroes of the GPCR structural revolution. *Nat Rev Mol Cell Biol* 16: 69–81.

[80] Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *J Mol Biol* 80: 535–542.

[81] Davis AM, Teague SJ, Kleywegt GJ (2003) Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew Chemie - Int Ed* 42: 2718–2736.

[82] Deller MC, Rupp B (2015) Models of protein-ligand crystal structures: Trust, but verify. *J Comput Aided Mol Des* 29: 817–836.

[83] Stanfield R, Pozharski E, Rupp B (2016) Comment on Three X-ray Crystal Structure Papers. *J Immunol* 196: 521–524.

[84] Stanfield R, Pozharski E, Rupp B (2016) Additional Comment on Three X-ray Crystal Structure Papers. *J Immunol* 196: 528–530.

[85] Gutmanas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, Dana JM, Montecelo MAF, van Ginkel G, Gore SP, Haslam P, Hatherley R, Hendrickx PMS, Hirshberg M, Lagerstedt I, Mir S, Mukhopadhyay A, Oldfield TJ, Patwardhan A, et al. (2014) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res* 42: D285–D291.

[86] Kleywegt GJ (2000) Validation of protein crystal structures. *Acta Crystallogr Sect D Biol Crystallogr* 56: 249–265.

[87] Müller P (2006) Hydrogen atoms. *Crystal Structure Refinement: A Crystallographer's Guide to SHELXL*, ed Müller P (Oxford University Press Inc., New York), pp 26–41.

[88]   Howard EI, Sanishvili R, Cachau RE, Mitschler A, Chevrier B, Barth P, Lamour V, Van Zandt M, Sibley E, Bon C, Moras D, Schneider TR, Joachimiak A, Podjarny A (2004) Ultrahigh resolution drug design I: Details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. *Proteins Struct Funct Genet* 55: 792–804.

[89]   Davis AM, St-Gallay SA, Kleywegt GJ (2008) Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discov Today* 13: 831–841.

[90]   Garman E (2003) "Cool" crystals: Macromolecular cryocrystallography and radiation damage. *Curr Opin Struct Biol* 13: 545–551.

[91]   Halle B (2004) Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc Natl Acad Sci USA* 101: 4793–4798.

[92]   Goldflam M, Tarragó T, Gairí M (2012) NMR Studies of Protein–Ligand Interactions. *Protein NMR Techniques*, eds Shekhtman A, Burz DS (Humana Press, New York), pp 233–259.

[93]   Huang H, Melacini G (2006) High-resolution protein hydration NMR experiments: Probing how protein surfaces interact with water and other non-covalent ligands. *Anal Chim Acta* 564: 1–9.

[94]   Yousufuddin M, Bau R (2005) Neutron Diffraction. *Encyclopedia of Inorganic Chemistry*, ed King RB (John Wiley & Sons, Chichester).

[95]   Shu F, Ramakrishnan V, Schoenborn BP (2000) Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin. *Proc Natl Acad Sci USA* 97: 3872–3877.

[96]   Krimmer SG, Klebe G (2015) Thermodynamics of protein-ligand interactions as a reference for computational analysis: How to assess accuracy, reliability and relevance of experimental data. *J Comput Aided Mol Des* 29: 867–883.

[97]   Burnouf D, Ennifar E, Guedich S, Puffer B, Hoffmann G, Bec G, Disdier F, Baltzinger M, Dumas P (2012) kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. *J Am Chem Soc* 134: 559–565.

[98]   Dumas P, Ennifar E, Da Veiga C, Bec G, Palau W, Di Primo C, Pineiro A, Sabin J, Munoz E, Rial J (2016) Extending ITC to Kinetics with kinITC. *Methods Enzymol* 567: 157–180.

[99]   Cooper MA (2002) Optical biosensors in drug discovery. *Nat Rev Drug Discov* 1: 515–528.

[100]   Day YSN, Baird CL, Rich RL, Myszka DG (2002) Direct comparison of binding equilibrium, thermodynamic, and rate constants determined by surface- and solution-based biophysical methods. *Protein Sci* 11: 1017–1025.

[101]   Giannetti AM (2011) From eperimental design to validated hits: A comprehensive walk-through of fragment lead identification using surface plasmon resonance. *Methods Enzymol* 493: 169–218.

[102]   Biacore Sensor Surface Handbook BR-1005-71 Edition AB (GE Healthcare Bio-Sciences AB, Uppsala, Sweden)

[103]   Schuck P, Zhao H (2010) The Role of Mass Transport Limitation and Surface Heterogeneity in the Biophysical Characterization of Macromolecular Binding Processes by SPR Biosensing. *Methods Mol Biol* 627: 15–54.

[104]   Krimmer SG, Betz M, Heine A, Klebe G (2014) Methyl, ethyl, propyl, butyl: futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. *ChemMedChem* 9: 833–846.

[105]   Cramer J, Krimmer SG, Fridh V, Wulsdorf T, Karlsson R, Heine A, Klebe G (2017) Elucidating the Origin of Long Residence Time Binding for Inhibitors of the Metalloprotease Thermolysin. *ACS Chem Biol* 12: 225–233.

[106]   Schiebel J, Radeva N, Krimmer SG, Wang X, Stieler M, Ehrmann FR, Fu K, Metz A, Huschmann FU, Weiss MS, Mueller U, Heine A, Klebe G (2016) Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem Biol* 11: 1693–1701.

[107]   Radeva N, Krimmer SG, Stieler M, Fu K, Wang X, Ehrmann FR, Metz A, Huschmann FU, Weiss MS, Mueller U, Schiebel J, Heine A, Klebe G (2016) Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J Med Chem* 59: 7561–7575.

[108]   Radeva N, Schiebel J, Wang X, Krimmer SG, Fu K, Stieler M, Ehrmann FR, Metz A, Rickmeyer T, Betz M, Winquist J, Park AY, Huschmann FU, Weiss MS, Mueller U, Heine A, Klebe G (2016) Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads To Address a Catalytic Dyad. *J Med Chem* 59: 9743–9759.

[109]   Klebe G (2015) The Use of Thermodynamic and Kinetic Data in Drug Discovery: Decisive Insight or Increasing the Puzzlement? *ChemMedChem* 10: 229–231.

[110]   Kramer C, Lewis R (2012) QSARs, data and error in the modern age of drug discovery. *Curr Top Med Chem* 12: 1896–1902.

[111]   Kramer C, Kalliokoski T, Gedeck P, Vulpetti A (2012) The experimental uncertainty of heterogeneous public K(i) data. *J Med Chem* 55: 5165–5173.

[112]   Kalliokoski T, Kramer C, Vulpetti A (2013) Quality issues with public domain chemogenomics data. *Mol Inform* 32: 898–905.

[113]   Wätzig H, Oltmann-Norden I, Steinicke F, Alhazmi HA, Nachbar M, El-Hady DA, Albishri HM, Baumann K, Exner T, Böckler FM, El Deeb S (2015) Data quality in drug discovery: the role of analytical performance in ligand binding assays. *J Comput Aided Mol Des* 29: 847–865.

[114]   Pethica BA (2015) Misuse of thermodynamics in the interpretation of isothermal titration calorimetry data for ligand binding to proteins. *Anal Biochem* 472: 21–29.

[115]   Wiseman T, Williston S, Brandts JF, Lin L-N (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* 179: 131–137.

[116]   Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* 11: 560–566.

[117]   Thomson JA, Ladbury JE (2004) Isothermal titration calorimetry: a tutorial. *Biocalorimetry 2: Applications of Calorimetry in the Biological Sciences* (John Wiley & Sons, Ltd, Chichester, UK), pp 37–58.

[118]   Perozzo R, Folkers G, Scapozza L (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J Recept Signal Transduct Res* 24: 1–52.

[119]   Freyer MW, Lewis EA (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol* 84: 79–113.

[120]   Biela A, Sielaff F, Terwesten F, Heine A, Steinmetzer T, Klebe G (2012) Ligand binding stepwise disrupts water network in thrombin: Enthalpic and entropic changes reveal classical hydrophobic effect. *J Med Chem* 55: 6094–6110.

[121]   Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat Rev Drug Discov* 9: 23–27.

[122]   Ferenczy GG, Keseru GM (2010) Thermodynamics guided lead discovery and optimization. *Drug Discov Today* 15: 919–932.

[123]   Rühmann E, Betz M, Fricke M, Heine A, Schäfer M, Klebe G (2015) Thermodynamic signatures of fragment binding: validation of direct versus displacement ITC titrations. *Biochim Biophys Acta - Gen Subj* 1850: 647–656.

[124]   Ruehmann E, Betz M, Heine A, Klebe G (2015) Fragments can bind either more enthalpy or entropy-driven: crystal structures and residual hydration pattern suggest why. *J Med Chem* 58: 6960–6971.

[125]   Geschwindner S, Ulander J, Johansson P (2015) Ligand binding thermodynamics in drug discovery: still a hot tip? *J Med Chem* 58: 6321–6335.

[126]   ITC Data Analysis in Origin® Tutorial Guide Version 7.0 (MicroCal LLC Northampton, MA, USA)

[127]   Le VH, Buscaglia R, Chaires JB, Lewis EA (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Anal Biochem* 434: 233–241.

[128]   Microcal ITC200 System User Manual 29017607 AA (GE Healthcare Bio-Sciences AB, Uppsala, Sweden)

[129]   Cubrilovic D, Zenobi R (2013) Influence of dimehylsulfoxide on protein–ligand binding affinities. *Anal Chem* 85: 2724–2730.

[130]   Ghai R, Falconer RJ, Collins BM (2012) Applications of isothermal titration calorimetry in pure and applied research – survey of the literature from 2010. *J Mol Recognit* 25: 32–52.

[131]   Biswas T, Tsodikov O V (2010) An easy-to-use tool for planning and modeling a calorimetric titration. *Anal Biochem* 406: 91–93.

[132]   Broecker J, Vargas C, Keller S (2011) Revisiting the optimal c value for isothermal titration calorimetry. *Anal Biochem* 418: 307–309.

[133]   Turnbull WB (2011) Divided we fall? Studying low-affinity fragments of ligands by ITC. *GE Healthc Life Sci Protoc*.

[134]   Turnbull WB, Daranas AH (2003) On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc* 125: 14859–14866.

[135]   Hansen LD, Fellingham GW, Russell DJ (2011) Simultaneous determination of equilibrium constants and enthalpy changes by titration calorimetry: methods, instruments, and uncertainties. *Anal Biochem* 409: 220–229.

[136]   Tellinghuisen J (2012) Designing isothermal titration calorimetry experiments for the study of 1:1 binding: problems with the "standard protocol." *Anal Biochem* 424: 211–220.

[137]   Tellinghuisen J (2005) Optimizing experimental parameters in isothermal titration calorimetry. *J Phys Chem B* 109: 20027–20035.

[138]   Mizoue LS, Tellinghuisen J (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal Biochem* 326: 125–127.

[139] Tellinghuisen J (2008) Isothermal titration calorimetry at very low c. *Anal Biochem* 373: 395–397.

[140] Tellinghuisen J (2007) Optimizing experimental parameters in isothermal titration calorimetry: variable volume procedures. *J Phys Chem B* 111: 11531–11537.

[141] Zhang YL, Zhang ZY (1998) Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B. *Anal Biochem* 261: 139–148.

[142] Velazquez-Campoy A, Freire E (2006) Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat Protoc* 1: 186–191.

[143] Cornish-Bowden A (2002) Enthalpy-entropy compensation: a phantom phenomenon. *J Biosci* 27: 121–126.

[144] Olsson TSG, Ladbury JE, Pitt WR, Williams MA (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci* 20: 1607–1618.

[145] Reynolds CH, Holloway MK (2011) Thermodynamics of Ligand Binding and Efficiency. *ACS Med Chem Lett* 2: 433–437.

[146] Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35: 198–201.

[147] Olsson TSG, Williams MA, Pitt WR, Ladbury JE (2008) The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. *J Mol Biol* 384: 1002–1017.

[148] Li L, Dantzer JJ, Nowacki J, O'Callaghan BJ, Meroueh SO (2008) PDBcal: a comprehensive dataset for receptor-ligand interactions with three-dimensional structures and binding thermodynamics from isothermal titration calorimetry. *Chem Biol Drug Des* 71: 529–532.

[149] Myszka DG, Abdiche YN, Arisaka F, Byron O, Eisenstein E, Hensley P, Thomson JA, Lombardo CR, Schwarz F, Stafford W, Doyle ML (2003) The ABRF-MIRG'02 Study: Assembly State, Thermodynamic, and Kinetic Analysis of an Enzyme/Inhibitor Interaction. *J Biomol Tech* 14: 247–269.

[150] Jelesarov I, Bosshard HR (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to invesitigate the energetics of biomolecular recognition. *J Mol Recognit* 12: 3–18.

[151] Doyle ML, Louie GL, Dal Monte PR, Sokoloski TD (1995) Tight binding affinities determined from linkage to protons by titration calorimetry. *Methods Enzymol* 259: 183–194.

[152] Baker BM, Murphy KP (1996) Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophys J* 71: 2049–2055.

[153] Parker MH, Lunney EA, Ortwine DF, Pavlovsky AG, Humblet C, Brouillette CG (1999) Analysis of the binding of hydroxamic acid and carboxylic acid inhibitors to the stromelysin-1 (matrix metalloproteinase-3) catalytic domain by isothermal titration calorimetry. *Biochem* 38: 13592–13601.

[154] Goldberg RN, Kishore N, Lennen RM (2002) Thermodynamic quantities for the ionization reaction of buffers. *J Phys Chem Ref Data* 31: 231–370.

[155] Keller S, Vargas C, Zhao H, Piszczek G, Brautigam CA, Schuck P (2012) High-precision isothermal titration calorimetry with automated peak shape analysis. *Anal Chem* 84: 5066–5073.

[156] Baum B, Muley L, Heine A, Smolinski M, Hangauer D, Klebe G (2009) Think twice: understanding the high potency of bis(phenyl)methane inhibitors of thrombin. *J Mol Biol* 391: 552–564.

[157] Grüner S, Neeb M, Barandun LJ, Sielaff F, Hohn C, Kojima S, Steinmetzer T, Diederich F, Klebe G (2014) Impact of protein and ligand impurities on ITC-derived protein-ligand thermodynamics. *Biochim Biophys Acta - Gen Subj* 1840: 2843–2850.

[158] Boström M, Williams DRM, Ninham BW (2003) Specific ion effects: why the properties of lysozyme in salt solutions follow a Hofmeister series. *Biophys J* 85: 686–694.

[159] Xie D, Gulnik S, Collins L, Gustchina E, Suvorov L, Erickson JW (1997) Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurment of the protonation states of the catalytic aspartic acid residues. *Biochem* 36: 16166–16172.

[160] Chu AH, Turner BW, Ackers GK (1984) Effects of protons on the oxygenation-linked subunit assembly in human hemoglobin. *Biochem* 23: 604–617.

[161] Neeb M, Betz M, Heine A, Barandun LJ, Hohn C, Diederich F, Klebe G (2014) Beyond affinity: enthalpy-entropy factorization unravels complexity of a flat structure-activity relationship for inhibition of a tRNA-modifying enzyme. *J Med Chem* 57: 5566–5578.

[162] Gibb CLD, Oertling EE, Velaga S, Gibb BC (2015) Thermodynamic Profiles of Salt Effects on a Host–Guest System: New Insight into the Hofmeister Effect. *J Phys Chem B* 119: 5624–5638.

[163] Gibb CLD, Gibb BC (2011) Anion binding to hydrophobic concavity is central to the salting-in effects of hofmeister chaotropes. *J Am Chem Soc* 133: 7344–7347.

[164] Fox JM, Kang K, Sherman W, Héroux A, Sastry M, Baghbanzadeh M, Lockett MR, Whitesides GM (2015) Interactions between Hofmeister Anions and the Binding Pocket of a Protein. *J Am Chem Soc* 137: 3859–3866.

[165] Spolar RS, Record MT (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263: 777–784.

[166] Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chemie - Int Ed* 41: 2644–2676.

[167] Horn JR, Russell D, Lewis EA, Murphy KP (2001) Van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: are there significant discrepancies? *Biochemistry* 40: 1774–1778.

[168] Mizoue LS, Tellinghuisen J (2004) Calorimetric vs. van't Hoff binding enthalpies from isothermal titration calorimetry: Ba2+-crown ether complexation. *Biophys Chem* 110: 15–24.

[169] Pan AC, Borhani DW, Dror RO, Shaw DE (2013) Molecular determinants of drug-receptor binding kinetics. *Drug Discov Today* 18: 667–673.

[170] Tellinghuisen J, Chodera JD (2011) Systematic errors in isothermal titration calorimetry: concentrations and baselines. *Anal Biochem* 414: 297–299.

[171] Baranauskiene L, Petrikaite V, Matuliene J, Matulis D (2009) Titration calorimetry standards and the precision of isothermal titration calorimetry data. *Int J Mol Sci* 10: 2752–2762.

[172] Wadsö I (2000) Needs for standards in isothermal microcalorimetry. *Thermochim Acta* 347: 73–77.

[173] Wadsö I, Wadsö L (2005) Systematic errors in isothermal micro- and nanocalorimetry. *J Therm Anal Calorim* 82: 553–558.

[174] Hermans J, Barry L (2014) *Equilibrium and kinetics of biological macromolecules* (John Wiley & Sons, Hoboken, New Jersey, USA)

[175] Tellinghuisen J (2004) Volume errors in isothermal titration calorimetry. *Anal Biochem* 333: 405–406.

[176] Good NE, Winget GD, Winter W, Connolly TN, Izawa S, Singh RMM (1966) Hydrogen ion buffers for biological research. *Biochemistry* 5: 467–477.

[177] Houtman JCD, Brown PH, Bowden B, Yamaguchi H, Appella E, Samelson LE, Schuck P (2007) Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein Sci* 16: 30–42.

[178] Scheuermann TH, Brautigam CA (2015) High-precision, automated integration of multiple isothermal titration calorimetric thermograms: New features of NITPIC. *Methods* 76: 87–98.

[179] Zhao H, Piszczek G, Schuck P (2015) SEDPHAT – A platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods* 76: 137–148.

[180] Inouye K, Kuzuya K, Tonomura B (1998) Effect of salts on the solubility of thermolysin: a remarkable increase in the solubility as well as the activity by the addition of salts without aggregation or dispersion of thermolysin. *J Biochem* 123: 847–852.

[181] Wermuth CG (2008) Molecular Variations in Homologous Series: Vinylogues and Benzologues. *The Practice of Medicinal Chemistry*, ed Wermuth CG (Elsevier, Amsterdam, Netherlands), pp 275–289.

[182] Sloan ED (2003) Fundamental principles and applications of natural gas hydrates. *Nature* 426: 353–363.

[183] Teeter MM (1984) Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proc Natl Acad Sci USA* 81: 6014–6018.

[184] Oxtoby NS, Blake AJ, Champness NR, Wilson C (2005) Water superstructures within organic arrays; hydrogen-bonded water sheets, chains and clusters. *Chem - A Eur J* 11: 4643–4654.

[185] Engel CK, Pirard B, Schimanski S, Kirsch R, Habermann J, Klingler O, Schlotte V, Weithmann KU, Wendt KU (2005) Structural basis for the highly selective inhibition of MMP-13. *Chem Biol* 12: 181–189.

[186] Tame JRH, Sleigh SH, Wilkinson AJ, Ladbury JE (1996) The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat Struct Biol* 3: 998–1001.

[187] Wang T, Wade RC (2002) Comparative Binding Energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics. *J Med Chem* 45: 4828–4837.

[188] Matthews BW, Colman PM, Jansonius JN, Titani K, Walsh KA, Neurath H (1972) Structure of Thermolysin. *Nat New Biol* 238: 41–43.

[189] Matthews BW, Weaver LH, Kester WR (1974) The Conformation of Thermolysin. *J Biol Chem* 249: 8030–8044.

[190] Holmes MA, Matthews BW (1982) Structure of thermolysin refined at 1.6 Å resolution. *J Mol Biol* 160: 623–639.

[191] English AC, Groom CR, Hubbard RE (2001) Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng* 14: 47–59.

[192]   Englert L, Silber K, Steuber H, Brass S, Over B, Gerber H-D, Heine A, Diederich WE, Klebe G (2010) Fragment-based lead discovery: Screening and optimizing fragments for thermolysin inhibition. *ChemMedChem* 5: 930–940.

[193]   Behnen J, Köster H, Neudert G, Craan T, Heine A, Klebe G (2012) Experimental and Computational Active Site Mapping as a Starting Point to Fragment-Based Lead Discovery. *ChemMedChem* 7: 248–261.

[194]   Englert L, Biela A, Zayed M, Heine A, Hangauer D, Klebe G (2010) Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: Binding of phosphonamidate to the S1'-pocket of thermolysin. *Biochim Biophys Acta - Gen Subj* 1800: 1192–1202.

[195]   Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 50: 760–763.

[196]   Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26: 283–291.

[197]   Matthews BW (1972) The γ Turn. Evidence for a New Folded Conformation in Proteins. *Macromolecules* 5: 818–819.

[198]   Kleywegt GJ, Zou J-Y (2001) Around O. *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules*, eds Rossmann MG, Arnold E (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp 353–356.

[199]   Carugo O, Bordo D (1999) How many water molecules can be detected by protein crystallography? *Acta Crystallogr Sect D Biol Crystallogr* 55: 479–483.

[200]   Nasief NN, Tan H, Kong J, Hangauer D (2012) Water Mediated Ligand Functional Group Cooperativity: The Contribution of a Methyl Group to Binding Affinity is Enhanced by a COO– Group Through Changes in the Structure and Thermodynamics of the Hydration Waters of Ligand–Thermolysin Complexes. *J Med Chem* 55: 8283–8302.

[201]   Jeffrey GA (1997) *An Introduction to Hydrogen Bonding* (Oxford University Press, Oxford, UK)

[202]   Sharp K (2001) Entropy – enthalpy compensation: Fact or artifact? *Protein Sci* 10: 661-667.

[203]   Breiten B, Lockett MR, Sherman W, Fujita S, Lange H, Bowers CM, Heroux A, Whitesides GM (2013) Water Networks Contribute to Enthalpy / Entropy Compensation in Protein-Ligand Binding. *J Am Chem Soc* 135: 15579–15584.

[204]   Shimokhina N, Bronowska A, Homans SW (2006) Contribution of ligand desolvation to binding thermodynamics in a ligand-protein interaction. *Angew Chemie - Int Ed* 45: 6374–6376.

[205]   Sturtevant JM (1977) Heat capacity and entropy changes in processes involving proteins. *Proc Natl Acad Sci USA* 74: 2236–2240.

[206]   Ford DM (2005) Enthalpy-entropy compensation is not a general feature of weak association. *J Am Chem Soc* 127: 16167–16170.

[207]   Lee B, Richards FM (1971) The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55: 379–400.

[208] Shiao DDF, Sturtevan JM (1969) Calorimetric Investigations of the Binding of Inhibitors to α-Chymotrypsin. I. The Enthalpy of Dilution of α-Chymotrypsin and of Proflavin, and the Enthalpy of Binding of Indole, N-Acetyl-D-tryptophan, and Proflavin to α-Chymotrypsin. *Biochemistry* 8: 4910–4917.

[209] Matthews BW, Jansonius JN, Colman PM, Schoenborn BP, Dupourque D (1972) Three-dimensional Structure of Thermolysin. *Nat New Biol* 238: 37–41.

[210] Mueller U, Darowski N, Fuchs MR, Förster R, Hellmig M, Paithankar KS, Pühringer S, Steffien M, Zocher G, Weiss MS (2012) Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J Synchrotron Radiat* 19: 442–449.

[211] Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276: 307–326.

[212] Holland DR, Tronrud DE, Pley HW, Flaherty KM, Stark W, Jansonius JN, McKay DB, Matthews BW (1992) Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* 31: 11310–11316.

[213] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674.

[214] Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr, Sect D Biol Crystallogr* 60: 2126–2132.

[215] Adams PD, Afonine PV, Gabor B, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66: 213–221.

[216] Molecular Operating Environment (MOE), Chemical Computing Group Inc., Montreal, QC, Canada

[217] SYBYL-X, version 2.0, Tripos International, St. Louis, MO (USA)

[218] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 66: 12–21.

[219] Connolly ML (1983) Analytical molecular surface calculation. *J Appl Crystallogr* 16: 548–558.

[220] Gelin BR, Karplus M (1979) Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry* 18: 1256–1268.

[221] The PyMOL Molecular Graphics System, Schrödinger LLC, New York, NY (USA)

[222] Edink E, Jansen C, Leurs R, De Esch IJP (2010) The heat is on: Thermodynamic analysis in fragment-based drug discovery. *Drug Discov Today Technol* 7: e189–e201.

[223] Freire E (2009) A thermodynamic approach to the affinity optimization of drug candidates. *Chem Biol Drug Des* 74: 468–472.

[224] Lu H, Tonge PJ (2010) Drug-target residence time: critical information for lead optimization. *Curr Opin Chem Biol* 14: 467–474.

[225] Copeland RA (2016) The drug-target residence time model: a 10-year retrospective. *Nat Rev Drug Discov* 15: 87–95.

[226] Holdgate GA, Ward WHJ (2005) Measurements of binding thermodynamics in drug discovery. *Drug Discov Today* 10: 1543–1550.

[227] Chaires JB (2008) Calorimetry and thermodynamics in drug design. *Annu Rev Biophys* 37: 135–151.

[228] Southall NT, Dill KA, Haymet ADJ (2002) A View of the Hydrophobic Effect. *J Phys Chem* 106: 521–533.

[229] Hummer G (2010) Molecular binding: Under water's influence. *Nat Chem* 2: 906–907.

[230] Adekoya OA, Sylte I (2009) The thermolysin family (M4) of enzymes: Therapeutic and biotechnological potential. *Chem Biol Drug Des* 73: 7–16.

[231] Eijsink VGH, Matthews BW, Vriend G (2011) The role of calcium ions in the stability and instability of a thermolysin-like protease. *Protein Sci* 20: 1346–1355.

[232] Petrillo EWJ, Ondetti MA (1982) Angiotensin-converting enzyme inhibitors: medicinal chemistry and biological actions. *Med Res Rev* 2: 1–41.

[233] Holland DR, Barclay PL, Danilewicz JC, Matthews BW, James K (1994) Inhibition of thermolysin and neutral endopeptidase 24.11 by a novel glutaramide derivative: X-ray structure determination of the thermolysin-inhibitor complex. *Biochemistry* 33: 51–56.

[234] Gallicchio E, Kubo MM, Levy RM (1998) Entropy-Enthalpy Compensation in Solvation and Ligand Binding Revisited. *J Am Chem Soc* 120: 4526–4527.

[235] Betz M, Wulsdorf T, Krimmer SG, Klebe G (2015) Impact of Surface Water Layers on Protein-Ligand Binding: How Well Are Experimental Data Reproduced by Molecular Dynamics Simulations in a Thermolysin Test Case. *J Chem Inf Model* 56: 223–233.

[236] Guo J, Huang W, Scanlan TS (1994) Kinetic and Mechanistic Characterization of an Efficient Hydrolytic Antibody: Evidence for the Formation of an Acyl Intermediate. *J Am Chem Soc* 116: 6062–6069.

[237] Holden HM, Tronrud DE, Monzingo AF, Weaver LH, Matthews BW (1987) Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphonamidate transition-state analogues. *Biochemistry* 26: 8542–8553.

[238] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67: 235–242.

[239] Neudert G, Klebe G (2011) fconv: format conversion, manipulation and feature computation of molecular data. *Bioinformatics* 27: 1021–1022.

[240] Carugo O (1999) Correlation between occupancy and B factor of water molecules in protein crystal structures. *Protein Eng* 12: 1021–1024.

[241] Grabowski SJ (2006) *Hydrogen Bonding–New Insights* (Springer, Heidelberg, Germany)

[242] Roden LD, Myszka DG (1996) Global Analysis of a Macromolecular Interaction Measured on BIAcore. *Biochem Biophys Res Commun* 225: 1073–1077.

[243] Karlsson R, Fält A (1997) Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. *J Immunol Methods* 200: 121–133.

[244] Fenley AT, Muddana HS, Gilson MK (2012) Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding. *Proc Natl Acad Sci USA* 109: 20006–200011.

[245] Vaitheeswaran S, Yin H, Rasaiah JC, Hummer G (2004) Water clusters in nonpolar cavities. *Proc Natl Acad Sci USA* 101: 17002–17005.

[246] Lee J, Kim SH (2009) Water polygons in high-resolution protein crystal structures. *Protein Sci* 18: 1370–1376.

[247] Winquist J, Geschwindner S, Xue Y, Gustavsson L, Musil D, Deinum J, Danielson UH (2013) Identification of Structural–Kinetic and Structural–Thermodynamic Relationships for Thrombin Inhibitors. *Biochemistry* 52: 613–626.

[248] Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J Phys Chem* 97: 10269–10280.

[249] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, et al. (2009) *Gaussian09, Revision D.01* (Gaussian Inc., Wallingford, CT, USA)

[250] Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE, Darden TA, Duke RE, Gohlke H, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossváry I, Kovalenko A, Lee TS, LeGrand S, et al. AMBER 14 (2014), University of California, San Francisco.

[251] Humphrey W, Dalke A, Schulten K (1996) VMD: Visual Molecular Dynamics. *J Mol Graph* 14: 33–38.

[252] Kabsch W (2010) XDS. *Acta Crystallogr Sect D Biol Crystallogr* 66: 125–132.

[253] Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* 66: 486–501.

[254] Moriarty NW, Grosse-Kunstleve RW, Adams PD (2009) Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr, Sect D Biol Crystallogr* 65: 1074–1080.

[255] Brautigam CA (2015) Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol* 562: 109–133.

[256] Krissinel E, Henrick K (2007) Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol* 372: 774–797.

[257] Falconer RJ (2016) Applications of isothermal titration calorimetry - the research and technical developments from 2011 to 2015. *J Mol Recognit* 29: 504–515.

[258] Schoop A, Dey F (2015) On-rate based optimization of structure - kinetic relationship-surfing the kinetic map. *Drug Discov today Technol* 17: 9–15.

[259] Walkup GK, You Z, Ross PL, Allen EKH, Daryaee F, Hale MR, O'Donnell J, Ehmann DE, Schuck VJA, Buurman ET, Choy AL, Hajec L, Murphy-Benenato K, Marone V, Patey SA, Grosser LA, Johnstone M, Walker SG, Tonge PJ, et al. (2015) Translating slow-binding inhibition kinetics into cellular and in vivo effects. *Nat Chem Biol* 11: 416–423.

[260]   Guo D, Heitman LH, IJzerman AP (2015) The Role of Target Binding Kinetics in Drug Discovery. *ChemMedChem* 10: 1793–1796.

[261]   Segala E, Guo D, Cheng RKY, Bortolato A, Deflorian F, Doré AS, Errey JC, Heitman LH, IJzerman AP, Marshall FH, Cooke RM, Dore AS, Errey JC, Heitman LH, IJzerman AP, Marshall FH, Cooke RM (2016) Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J Med Chem* 59: 6470–6479.

[262]   Chang A, Schiebel J, Yu W, Bommineni GR, Pan P, Baxter M V., Khanna A, Sotriffer CA, Kisker C, Tonge PJ (2013) Rational optimization of drug-target residence time: insights from inhibitor binding to the Staphylococcus aureus FabI enzyme-product complex. *Biochemistry* 52: 4217–4228.

[263]   Miller DC, Lunn G, Jones P, Sabnis Y, Davies NL, Driscoll P (2012) Investigation of the effect of molecular properties on the binding kinetics of a ligand to its biological target. *Medchemcomm* 3: 449–452.

[264]   Bortolato A, Tehan BG, Bodnarchuk MS, Essex JW, Mason JS (2013) Water Network Perturbation in Ligand Binding: Adenosine A2A Antagonists as a Case Study. *J Chem Inf Model* 53: 1700–1713.

[265]   Mondal J, Friesner RA, Berne BJ (2014) Role of Desolvation in Thermodynamics and Kinetics of Ligand Binding to a Kinase. *J Chem Theory Comput* 10: 5696–5705.

[266]   Pearlstein RA, Sherman W, Abel R (2013) Contributions of water transfer energy to protein-ligand association and dissociation barriers: Watermap analysis of a series of p38a MAP kinase inhibitors. *Proteins Struct Funct Bioinforma* 81: 1509–1526.

[267]   Schmidtke P, Luque FJ, Murray JB, Barril X (2011) Shielded Hydrogen Bonds as Structural Determinants of Binding Kinetics: Application in Drug Design. *J Am Chem Soc* 133: 18903–18910.

[268]   Carroll MJ, Mauldin RV, Gromova AV, Singleton SF, Collins EJ, Lee AL (2012) Evidence for dynamics in proteins as a mechanism for ligand dissociation. *Nat Chem Biol* 8: 246–252.

[269]   Copeland RA (2011) Conformational adaptation in drug-target interactions and residence time. *Futur Med Chem* 3: 1491–1501.

[270]   Luckner SR, Liu N, am Ende CW, Tonge PJ, Kisker C (2010) A Slow, Tight Binding Inhibitor of InhA, the Enoyl-Acyl Carrier Protein Reductase from Mycobacterium tuberculosis. *J Biol Chem* 285: 14330–14337.

[271]   Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482: 552–556.

[272]   Liu Y, Stoll VS, Richardson PL, Saldivar A, Klaus JL, Molla A, Kohlbrenner W, Kati WM (2004) Hepatitis C NS3 protease inhibition by peptidyl-alpha-ketoamide inhibitors: kinetic mechanism and structure. *Arch Biochem Biophys* 421: 207–216.

[273]   Garvey EP, Schwartz B, Gartland MJ, Lang S, Halsey W, Sathe G, Carter HL, Weaver KL (2009) Potent inhibitors of HIV-1 integrase display a two-step, slow-binding inhibition mechanism which is absent in a drug-resistant T66I/M154I mutant. *Biochemistry* 48: 1644–1653.

[274]   Bohacek R, de Lombaert S, McMartin C, Priestle J, Grütter M (1996) Three-Dimensional Models of ACE and NEP Inhibitors and Their Use in the Design of Potent Dual ACE/NEP Inhibitors. *J Am Chem Soc* 118: 8231–8249.

[275]   Tiraboschi G, Jullian N, Thery V, Antonczak S, Fournie-Zaluski M-C, Roques BP (1999) A three-dimensional construction of the active site (region 507-749) of human neutral endopeptidase (EC.3.4.24.11). *Protein Eng, Des Sel* 12: 141–149.

[276]   Hausrath AC, Matthews BW (2002) Thermolysin in the absence of substrate has an open conformation. *Acta Crystallogr Sect D Biol Crystallogr* 58: 1002–1007.

[277]   Holland DR, Hausrath AC, Juers D, Matthews BW (1995) Structural analysis of zinc substitutions in the active site of thermolysin. *Protein Sci* 4: 1955–1965.

[278]   Arolas JL, Botelho TO, Vilcinskas A, Gomis-Ruth FX (2011) Structural evidence for standard-mechanism inhibition in metallopeptidases from a complex poised to resynthesize a peptide bond. *Angew Chemie - Int Ed* 50: 10357–10360.

[279]   Sonnhammer ELL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28: 405–420.

[280]   Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr Sect D Biol Crystallogr* 60: 2256–2268.

[281]   Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42: W320–W324.

[282]   McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Solomon SD, Swedberg K, Zile MR (2014) Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med* 371: 993–1004.

[283]   Smollich M, Gotte M, Yip GW, Yong E-S, Kersting C, Fischgrabe J, Radke I, Kiesel L, Wulfing P (2007) On the role of endothelin-converting enzyme-1 (ECE-1) and neprilysin in human breast cancer. *Breast Cancer Res Treat* 106: 361–369.

[284]   Funalot B, Ouimet T, Claperon A, Fallet C, Delacourte A, Epelbaum J, Subkowski T, Leonard N, Codron V, David J-P, Amouyel P, Schwartz J-C, Helbecque N (2004) Endothelin-converting enzyme-1 is expressed in human cerebral cortex and protects against Alzheimer's disease. *Mol Psychiatry* 9: 1122–1128.

[285]   Yasukawa K, Kusano M, Inouye K (2007) A new method for the extracellular production of recombinant thermolysin by co-expressing the mature sequence and pro-sequence in Escherichia coli. *Protein Eng, Des Sel* 20: 375–383.

[286]   Kusano M, Yasukawa K, Hashida Y, Inouye K (2006) Engineering of the pH-dependence of thermolysin activity as examined by site-directed mutagenesis of Asn112 located at the active site of thermolysin. *J Biochem* 139: 1017–1023.

[287]   Bartlett PA, Marlowe CK (1987) Possible role for water dissociation in the slow binding of phosphorus-containing transition-state-analog inhibitors of thermolysin. *Biochemistry* 26: 8553–8561.

[288]   Fedosova NU, Champeil P, Esmann M (2002) Nucleotide Binding to Na,K-ATPase: The Role of Electrostatic Interactions. *Biochemistry* 41: 1267–1273.

[289]   Ledvina PS, Tsai AL, Wang Z, Koehl E, Quiocho FA (1998) Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: equilibrium, kinetic, and crystallographic studies. *Protein Sci* 7: 2550–2559.

[290]   Radić Z, Kirchhoff PD, Quinn DM, McCammon JA, Taylor P (1997) Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase. Distinctions between active center ligands and fasciculin. *J Biol Chem* 272: 23265–23277.

[291]   Sinha N, Mohan S, Lipschultz CA, Smith-Gill SJ (2002) Differences in Electrostatic Properties at Antibody–Antigen Binding Sites: Implications for Specificity and Cross-Reactivity. *Biophys J* 83: 2946–2968.

[292]   Smart OS, Womack TO, Sharff A, Flensburg C, Keller P, Paciorek W, Vonrhein C, Bricogne G (2011) *Grade (http://www.globalphasing.com)* (Global Phasing Ltd., Cambridge, UK)

[293]   Burling FT, Weis WI, Flaherty KM, Brünger AT (1996) Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* 271: 72–77.

[294]   Schiltz M, Fourme R, Prangé T (2003) Use of Noble Gases Xenon and Krypton as Heavy Atoms in Protein Structure Determination. *Methods Enzymol* 374: 83–119.

[295]   Levitt M, Park BH (1993) Water: Now you see it, now you don't. *Structure* 1: 223–226.

[296]   Read RJ (2001) Model phases: probabilities, bias and maps. *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules*, eds Rossmann MG, Arnold E (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp 325–331.

[297]   Brünger AT, Rice LM (1997) Crystallographic refinement by simulated annealing: Methods and applications. *Methods Enzymol* 277: 243–269.

[298]   Hodel A, Kim S-H, Brünger AT (1992) Model bias in macromolecular crystal structures. *Acta Cryst A* 48: 851–858.

[299]   Quillin ML, Wingfield PT, Matthews BW (2006) Determination of solvent content in cavities in IL-1beta using experimentally phased electron density. *Proc Natl Acad Sci USA* 103: 19749–19753.

[300]   Liu L, Quillin ML, Matthews BW (2008) Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc Natl Acad Sci USA* 105: 14406–14411.

[301]   Urzhumtsev A, Afonine PV, Lunin VY, Terwilliger TC, Adams PD (2014) Metrics for comparison of crystallographic maps. *Acta Crystallogr Sect D Biol Crystallogr* 70: 2593–2606.

[302]   Ernst JA, Clubb RT, Zhou HX, Gronenborn AM, Clore GM (1995) Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* 267: 1813–1817.

[303]   Yu B, Blaber M, Gronenborn AM, Clore GM, Caspar DLD (1999) Disordered water within a hydrophobic protein cavity visualized by x-ray crystallography. *Proc Natl Acad Sci USA* 96: 103–108.

[304]   Matthews BW, Morton AG, Dahlquist FW (1995) Use of NMR to Detect Water Within Nonpolar Protein Cavities. *Science* 270: 1847–1849.

[305]   Sonavane S, Chakrabarti P (2008) Cavities and atomic packing in protein structures and interfaces. *PLoS Comput Biol* 4: e1000188.

[306]   Hubbard SJ, Gross KH, Argos P (1994) Intramolecular Cavities in Globular Proteins. *Protein Eng* 7: 613–626.

[307]   Wolfenden R, Radzicka A (1994) On the probability of finding a water molecule in a nonpolar cavity. *Science* 265: 936–937.

[308] Bhat S, Purisima EO (2006) Molecular surface generation using a variable-radius solvent probe. *Proteins Struct Funct Genet* 62: 244–261.

[309] Li AJ-J, Nussinov R (1998) A set of van der Waals and Coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins Struct Funct Genet* 32: 111–127.

[310] Prangé T, Schiltz M, Pernot L, Colloc'h N, Longhi S, Bourguet W, Fourme R (1998) Exploring hydrophobic sites in proteins with xenon or krypton. *Proteins Struct Funct Genet* 30: 61–73.

[311] Tanwar AS, Goyal VD, Choudhary D, Panjikar S, Anand R (2013) Importance of hydrophobic cavities in allosteric regulation of formylglycinamide synthetase: Insight from xenon trapping and statistical coupling analysis. *PLoS One* 8: e77781.

[312] Quillin ML, Breyer WA, Griswold IJ, Matthews BW (2000) Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme. *J Mol Biol* 302: 955–977.

[313] Schlitz M, Shepard W, Fourme R, Prange T, de la Fortelle E, Bricogne G (1997) High-Pressure Krypton Gas and Statistical Heavy-Atom Refinement: a Successful Combination of Tools for Macromolecular Structure Determination. *Acta Crystallogr D Biol Crystallogr* 53: 78–92.

[314] Matthews BW, Liu L (2009) A review about nothing: Are apolar cavities in proteins really empty? *Protein Sci* 18: 494–502.

[315] Adamek DH, Guerrero L, Blaber M, Caspar DLD (2005) Structural and energetic consequences of mutations in a solvated hydrophobic cavity. *J Mol Biol* 346: 307–318.

[316] Xu J, Baase WA, Quillin ML, Baldwin EP, Matthews BW (2001) Structural and thermodynamic analysis of the binding of solvent at internal sites in T4 lysozyme. *Protein Sci* 10: 1067–1078.

[317] Leung CS, Leung SSF, Tirado-Rives J, Jorgensen WL (2012) Methyl Effects on Protein–Ligand Binding. *J Med Chem* 55: 4489–4500.

[318] Young T, Abel R, Kim B, Berne BJ, Friesner RA (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc Natl Acad Sci USA* 104: 808–813.

[319] Wang L, Berne BJ, Friesner RA (2011) Ligand binding to protein-binding pockets with wet and dry regions. *Proc Natl Acad Sci USA* 108: 1326–1330.

[320] Lee C, Maeng JS, Kocher JP, Lee B, Yu MH (2001) Cavities of alpha(1)-antitrypsin that play structural and functional roles. *Protein Sci* 10: 1446–1453.

[321] Vallone B, Brunori M (2004) Roles for holes: are cavities in proteins mere packing defects? *Ital J Biochem* 53: 46–53.

[322] Schiebel J, Gaspari R, Sandner A, Ngo K, Gerber H-D, Cavalli A, Ostermann A, Heine A, Klebe G (2017) Charges Shift Protonation: Ultimately, Neutron Diffraction Discloses that Aniline and 2-Aminopyridine Become Protonated Upon Binding to Trypsin. *Angew Chemie - Int Ed*. doi:10.1002/anie.201701038R1

[323] Eriksson AE, Baase WA, Zhang X-J, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255: 178–183.

[324] Kellis JT, Nyberg K, Fersht AR (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28: 4914–4922.

[325] Cramer J, Klebe G (2017) An Allyl Protection and Improved Purification Strategy Enables the Synthesis of Functionalized Phosphonamidate Peptides. *Synthesis (Stuttg).* doi:10.1055/s-0036-1588393

[326] Oxford Cryosystems (1999) Portable xenon pressure chamber. *Acta Crystallogr Sect D Biol Crystallogr* 55: 724.

[327] Mueller U, Förster R, Hellmig M, Huschmann FU, Kastner A, Malecki P, Pühringer S, Röwer M, Sparta K, Steffien M, Ühlein M, Wilk P, Weiss MS (2015) The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *Eur Phys J Plus* 130: 141.

[328] Thorn A, Sheldrick GM (2011) ANODE: Anomalous and heavy-atom density calculation. *J Appl Crystallogr* 44: 1285–1287.

[329] Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr Sect D Biol Crystallogr* 58: 1772–1779.

[330] Sheldrick GM (2007) A short history of SHELX. *Acta Crystallogr Sect A Found Crystallogr* 64: 112–122.

[331] Pape T, Schneider TR (2004) HKL2MAP : a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Crystallogr* 37: 843–844.

[332] Cowtan K (1994) "dm": an automated procedure for phase improvement by density modification. *Jt CCP4 ESF-EACBM Newsl Protein Crystallogr* 31: 34–38.

[333] Lamzin VS, Wilson KS (1993) Automated refinement of protein models. *Acta Crystallogr D Biol Crystallogr* 49: 129–147.

[334] Durrant JD, Votapka L, Amaro RE, Sørensen J, Amaro RE (2014) POVME 2.0: An enhanced tool for determining pocket shape and volume characteristics. *J Chem Theory Comput* 10: 5047–5056.

[335] Qvist J, Davidovic M, Hamelberg D, Halle B (2008) A dry ligand-binding cavity in a solvated protein. *Proc Natl Acad Sci USA* 105: 6296–6301.

[336] Gerstein M, Richards FM, Chapman MS, Connolly ML (2001) Protein surfaces and volumes: measurement and use. *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules*, eds Rossmann MG, Arnold E (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp 531–545.

[337] Voss NR, Gerstein M (2010) 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res* 38: 1–8.

[338] Lang PT, Holton JM, Fraser JS, Alber T (2014) Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc Natl Acad Sci USA* 111: 237–242.

[339] Kleywegt GJ, Jones TA (1996) xdlMAPMAN and xdlDATAMAN - Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr Sect D Biol Crystallogr* 52: 826–828.

[340] Brautigam CA, Zhao H, Vargas C, Keller S, Schuck P (2016) Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nat Protoc* 11: 882–894.

[341] Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H (2016) Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov* 15: 605–619.

[342] Scott DE, Coyne AG, Hudson SA, Abell C (2012) Fragment based approaches in drug discovery and chemical biology. *Biochemistry* 51: 4990–5003.

[343] Congreve M, Carr R, Murray C, Jhoti H (2003) A "Rule of Three" for fragment-based lead discovery? *Drug Discov Today* 8: 876–877.

[344] Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, Green DVS, Hertzberg RP, Janzen WP, Paslay JW, Schopfer U, Sittampalam GS (2011) Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* 10: 188–195.

[345] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv Drug Deliv Rev* 46: 3–26.

[346] Barker A, Kettle JG, Nowak T, Pease JE (2013) Expanding medicinal chemistry space. *Drug Discov Today* 18: 298–304.

[347] Leach AR, Hann MM (2011) Molecular complexity and fragment-based drug discovery: Ten years on. *Curr Opin Chem Biol* 15: 489–496.

[348] Leeson PD, St-Gallay SA (2011) The influence of the "organizational factor" on compound quality in drug discovery. *Nat Rev Drug Discov* 10: 749–765.

[349] Hann MM (2011) Molecular obesity, potency and other addictions in drug discovery. *Medchemcomm* 2: 349–355.

[350] Köster H, Craan T, Brass S, Herhaus C, Zentgraf M, Neumann L, Heine A, Klebe G (2011) A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J Med Chem* 54: 7784–7796.

[351] Schiebel J, Radeva N, Köster H, Metz A, Krotzky T, Kuhnert M, Diederich WE, Heine A, Neumann L, Atmanene C, Roecklin D, Vivat-Hannah V, Renaud JP, Meinecke R, Schlinck N, Sitte A, Popp F, Zeeb M, Klebe G (2015) One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* 10: 1511–1521.

[352] Rees DC, Congreve, M, Murray CW, Carr R (2004) Fragment-based lead discovery. *Nat Rev Drug Discov* 3: 660–672.

[353] Tiefenbrunn T, Forli S, Gonzalez A, Tsai Y, Elder JH, Olson AJ, Stout CD (2014) Crystallographic Fragment-Based Drug Discovery : Use of a Brominated Fragment Library Targeting HIV Protease. *Chem Biol Drug Des* 83: 141–148.

[354] Echols N, Moriarty NW, Klei HE, Afonine PV, Bunkoczi G, Headd JJ, McCoy AJ, Oeffner RD, Read RJ, Adams PD, Terwilliger TC, Adams PD (2014) Automating crystallographic structure solution and refinement of protein – ligand complexes. *Acta Crystallogr D Biol Crystallogr* 70: 144–154.

[355] Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today* 10: 464–469.

[356] Pearce ANM, Bradley AR, Collins P, Krojer T, Nowak RP, Talon R, Marsden BD, Kelm S, Shi J, Deane CM, Von Delft F (2016) A Multi-Crystal Method for Extracting Obscured Signal from Crystallographic Electron Density. *bioRxiv*. doi:10.1101/073411

[357] Bronowska AK (2011) Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design. *Thermodynamics - Interaction Studies - Solids, Liquids and Gases*, ed Moreno-Pirajan JC (InTechOpen), pp 1–49.

[358] Matthews BW (1988) Structural basis of the action of thermolysin and related zinc peptidases. *Acc Chem Res* 21: 333–340.

[359] Inouye K, Kusano M, Hashida Y, Minoda M, Yasukawa K (2007) Engineering, expression, purification, and production of recombinant thermolysin. *Biotechnol Annu Rev* 13: 43–64.

[360] Eijsink VGH, Veltman OR, Aukema W, Vriend G, Venema G (1995) Structural determinants of the stability of thermolysin-like proteinases. *Nat Struct Biol* 2: 374–379.

[361] Veltman OR, Vriend G, Berendsen HJC, Van Den Burg B, Venema G, Eijsink VGH (1998) A single calcium binding site is crucial for the calcium-dependent thermal stability of thermolysin-like proteases. *Biochemistry* 37: 5312–5319.

[362] Conejero-Lara F, De Filippis V, Fontana A, Mateo PL (1994) The thermodynamics of the unfolding of an isolated protein subdomain. The 255-316 C-terminal fragment of thermolysin. *FEBS Lett* 344: 154–156.

[363] Doyle ML (1997) Caracterization of Binding Interactions by Isothermal Titration Calorimetry. *Curr Opin Biotechnol* 8: 31–35.

[364] Read RJ (1997) Model phases: Probabilities and bias. *Methods Enzymol* 277: 110–128.

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor **Prof. Gerhard Klebe** for his constant scientific support, which was invaluable for the success of this thesis. His persistent encouragement and enthusiasm about my results was very inspiring. I am grateful to him for the provided opportunities to present my research at international conferences and for supporting my scientific career.

I am very thankful to **Prof. Andreas Heine** for introducing me to macromolecular X-ray crystallography, and the great amount of time he invested in me. I want to thank him for the many insightful discussions and for always having an open door for questions, his collaboration in several projects and for co-examining my thesis.

I want to thank all my internal and external collaborators, especially **Jonathan Cramer** and **Michael Betz** for working so closely and fruitfully together with me. Large parts of this thesis were done in collaboration with them. Further thanks go to **Tobias Wulsdorf**, **Veronica Fridh**, **Robert Karlsson**, **Prof. David Hangauer**, **Nader Nasief**, and the whole **endothiapepsin group**, especially **Johannes Schiebel** and **Nedyalka Radeva**. A special thank you goes to *dott.ssa* **Nicole Bertoletti** for helping me out numerous times — sometimes late at night — and for keeping up the good mood. I am grateful to **Alexander Metz**, **Corey Taylor** and **Mahalia Lepage** for critical proofreading of my manuscripts. I want to thank **Felix Terwesten** and **Phong Nguyen** for their efforts and time invested into the maintenance of the computer network. I am grateful to **Lydia Hartleben** for her administrative support and to **Christian Sohn** for his technical help.

I want to thank my friends and colleagues from the **Klebe** and **Kolb labs** for providing helping hands for all kinds of problems, creating a welcoming working atmosphere and a great time also apart from science. In addition to the people already mentioned above, I especially want to acknowledge **Manuel Neeb**, **Florent Chevillard**, **Timo Krotzky**, **Martin Stieler**, **Thomas Rickmeyer**, **Chris Rechlin**, **Barbara Wienen-Schmidt**, **Kan Fu and Felix Gut**.

My parents have always provided me with the greatest support possible, during good and not-so-good times, for which I am deeply grateful.

Curriculum Vitae

Personal Details

Name	Stefan Günter Krimmer
Date of birth	October 27th, 1984
Place of Birth	Villingen-Schwenningen, Germany
Nationality	German

Education

04/2017 (expected)	**PhD (Dr. rer. nat.) in Life Sciences** Institute of Pharmaceutical Chemistry, Philipps University of Marburg, Germany
08/2011	**State Examination, Board License as Pharmacist in Germany**
04/2011	**MS (Diploma) Degree in Pharmacy** Martin-Luther University of Halle-Wittenberg, Germany
2005–2009	**Study of Pharmacy** Philipps University of Marburg

Research Experience

since 08/2011	**Dissertation** Drug design group of Prof. Dr. Gerhard Klebe, Institute of Pharmaceutical Chemistry, Philipps University of Marburg, Germany
01/2010–07/2010	**Master Thesis** Group of Prof. Dr. Jindřich Kopeček, Department of Pharmaceutics and Pharmaceutical Chemistry, University of Utah, Salt Lake City, USA Thesis title: "Synthesis and *in vitro* Characterization of PCL–PHPMA Polymeric Micelles for Drug Delivery"
03/2009	**Student Internship** Research group of Prof. Dr. Gerhard Klebe, Philipps University of Marburg, Germany
08/2008–09/2008	**Student Internship** Division of Medical Biotechnology, Paul Ehrlich Institute in Langen, Germany

Professional Experience

08/2011–11/2016	**Teaching Assistant**, Philipps University of Marburg, Germany
08/2010–01/2011	**Pharmacist Intern**, *"Mohren-Apotheke"* drugstore in Konstanz, Germany
07/2004–03/2005	**Compulsory Community Servant** of the emergency rescue service of the German Red Cross in Villingen-Schwenningen, Germany

Journal Publications

12. Cramer J, <u>Krimmer SG</u>, Heine A, Klebe G. **Paying the Price of Solvation in Solvent-Exposed Protein Pockets: Impact of Solubilizing Polar Groups on Affinity and Binding Thermodynamics in a Series of Thermolysin Inhibitors.**
manuscript in preparation

11. <u>Krimmer SG</u>,* Cramer J,* Schiebel J, Heine A, Klebe G. **How Nothing Boosts Affinity: Hydrophobic Ligand Binding to the Virtually Vacated S_1' Pocket of Thermolysin.**
manuscript in preparation

10. Cramer J,* <u>Krimmer SG</u>,* Fridh V,* Wulsdorf T, Karlsson R, Heine A, Klebe G. **Elucidating the Origin of Long Residence Time Binding for Inhibitors of the Metalloprotease Thermolysin.**
ACS Chemical Biology **2017**, 12: 225–233

9. <u>Krimmer SG</u>,* Cramer J,* Betz M, Fridh V, Karlsson R, Heine A, Klebe G. **Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein Bound Ligands.**
Journal of Medicinal Chemistry **2016**, 59: 10530–10548

8. Radeva N, Schiebel J, Wang X, <u>Krimmer SG</u>, Fu K, Stieler M, Ehrmann FR, Metz A, Rickmeyer T, Betz M, Winquist J, Park AY, Huschmann FU, Weiss MS, Mueller U, Heine A, Klebe G. **Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads to Address a Catalytic Dyad.**
Journal of Medicinal Chemistry **2016**, 59: 9743–9759

7. Radeva N, <u>Krimmer SG</u>, Stieler M, Fu K, Wang X, Ehrmann FR, Metz A, Huschmann FU, Weiss MS, Mueller U, Schiebel J, Heine A, Klebe G. **Experimental Active-Site Mapping by Fragments - Hot Spots Remote from the Catalytic Center of Endothiapepsin.**
Journal of Medicinal Chemistry **2016**, 59: 7561–7575

6. Schiebel J, <u>Krimmer SG</u>, Röwer K, Knörlein A, Wang X, Park AY, Stieler M, Ehrmann FR, Fu K, Radeva N, Krug M, Huschmann FU, Glöckner S, Weiss MS, Mueller U, Klebe G, Heine A. **High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits.**
Structure **2016**, 24: 1398–1409

5. Schiebel J, Radeva N, <u>Krimmer SG</u>, Wang X, Stieler M, Ehrmann FR, Fu K, Metz A, Huschmann FU, Weiss MS, Mueller U, Heine A, Klebe G. **Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study.**
ACS Chemical Biology **2016**, 11: 1693–1701

4. Betz M, Wulsdorf T, <u>Krimmer SG</u>, Klebe G. **Impact of Surface Water Layers on Protein-Ligand Binding: How Well Are Experimental Data Reproduced by Molecular Dynamics Simulations in a Thermolysin Test Case.**
Journal of Chemical Information and Modeling **2016**, 56: 223–233

3. <u>Krimmer SG</u>, Klebe G. **Thermodynamics of protein–ligand interactions as a reference for computational analysis: how to assess accuracy, reliability and relevance of experimental data.**
Journal of Computer-Aided Molecular Design **2015**, 29: 867–883

2. <u>Krimmer SG</u>, Betz M, Heine A, Klebe G. **Methyl, Ethyl, Propyl, Butyl: Futile but not for Water, as the Correlation of Structure and Thermodynamic Signature Shows in a Congeneric Series of Thermolysin Inhibitors.**
ChemMedChem **2014**, 9: 833–846

1. <u>Krimmer SG</u>, Pan H, Liu J, Yang J, Kopeček J. **Synthesis and Characterization of Poly(ε-caprolactone)-*block*-Poly[*N*-(2-hydroxypropyl)methacrylamide] Micelles for Drug Delivery.**
Macromolecular Bioscience **2011**, 11: 1041–1051

---

* These authors contributed equally

Oral Presentations

o ***German Pharmaceutical Society (DPhG) annual meeting,* October 4–7, 2016 in Munich, Germany**
15 min talk on "Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein Bound Ligands".

o ***International Workshop – New Approaches in Drug Design & Discovery,* March 24–27, 2014 in Rauischholzhausen, Germany**
30 min talk on "Influence of water molecules on protein–ligand binding: Correlation of structure and thermodynamic signature".

o ***247th ACS National Meeting,* March 16–20, 2014 in Dallas, Texas, USA**
30 min talk on "Influence of water molecules on protein–ligand binding: Correlation of structure and thermodynamic signature".

o ***Developments in Protein Interaction Analysis (DiPIA),* November 4–7, 2012 in Beijing, China**
15 min talk on "The Hydrophobic Effect on Molecular Level: Influence of Water on Protein–Ligand Binding".

Poster Presentations

o ***Seventh Joint BER II and BESSY II User Meeting,* December 9–10, 2015 in Berlin**
Poster on „Coating Bound Ligands by Tailored Water Networks to Improve Inhibitor Binding Profile for Thermolysin".

o ***Fourth Joint BER II and BESSY II User Meeting,* December 12–14, 2012 in Berlin**
Poster on "Investigation of the Thermodynamic and Kinetic Influence of Water Molecules on Protein–Ligand Binding".

Stipends

2012   Travel stipend of the German Academic Exchange Service (DAAD) to the DiPIA conference in Beijing, China

Erklärung

Ich versichere, dass ich meine Dissertation

"Systematic Correlation of Structural, Thermodynamic and Residual Solvation Properties of Hydrophobic Substituents in Hydrophobic Pockets Using Thermolysin as a Case Study"

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen sind Zitate als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den . . . .

(Stefan G. Krimmer)