The Role of Water in Protein-Ligand Binding

A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich

der Philipps-Universität Marburg

vorgelegt von

Adam Jozef Biela

aus

Groß Strehlitz

Marburg/Lahn 2012

Vom Fachbereich Pharmazie

der Philipps-Universität Marburg als Dissertation am _____
angenommen.

Erstgutachter: Prof. Dr. Klebe

Zweitgutachter: Prof. Dr. Steinmetzer

Tag der mündlichen Prüfung am _____

Die Untersuchungen zum vorliegenden Thema wurden auf Anregung von Prof. Dr. G. Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Januar 2008 bis Januar 2012 durchgeführt.

Table of Content

1 Einleitung/Introduction

1.1 Drug Discovery

Krankheiten begleiten die Menschheit seit ihren Anfängen. Aber wie heilt man eine Krankheit, wie findet man die richtige Therapie? Am Anfang der Menschheitsgeschichte bediente man sich der Natur, indem man ohne jegliche Vorkenntnisse über die Ursache der Krankheit verschiedene Pflanzen, Pflanzenteile oder deren Extrakte nicht immer in rationaler Weise einsetzte und schaute ob eine Wirkung erzielt werden kann. Dieser Prozess war also nicht rational sondern beruhte auf dem Testen vorhandener Substanzen. Insbesondere im Laufe des letzten Jahrhunderts gab es schließlich enorme Fortschritte auf dem Gebiet der Medizin, wodurch es nun möglich wurde die Ursachen einer Krankheit aufzuklären. Nun konnte man auf molekularer Ebene erkennen, was die Krankheit auslöste. Meistens handelt es sich hierbei um Proteine, die aus verschiedensten Gründen einen pathogenen Mechanismus im Körper einleiten. Die Bandbreite wie diese körpereigenen Moleküle einen schädlichen Einfluss ausüben, ist breit gestreut: sie können erblich bedingt nicht richtig funktionieren (Morbus Gaucher, Mukoviszidose), sind in zu niedriger (Laktoseintoleranz) oder zu hoher Konzentration (Hyperinsulinämie) vorhanden oder die Aktivität wird durch körpereigene Substanzen derart falsch reguliert, dass eine schädliche Dauer- oder Inaktivität (Depression, Bluthochdruck) die Folge ist. Aber auch körperfremde Proteine aus Bakterien oder Viren wurden schnell als Zielscheibe entdeckt um den Fremdorganismus (Gyrase, HIV-Protease) selbst zu beseitigen.

Im Vergleich zu den Anfängen der Therapiesuche war es also oft möglich die Ursache einer Erkrankung auf ein einzelnes Proteinmolekül zurückzuführen. Die Folge dieser Erkenntnis war dann ein bedingungsloser Focus der Arzneistoffentwicklung auf das in irgendeiner Art außer Kontrolle geratene Protein. Wurde ein körperfremdes Protein als überlebenswichtig für den Fremdorganismus eingestuft, wurde es gehemmt (HIV-Protease-Inhibitoren), oder war ein körpereigenes Protein falsch reguliert, fand man entsprechende Substanzen, die wieder für eine ausgewogene Regulierung des Proteins sorgten (Selektive-Serotonin-Reuptake-Inhibitoren). Dieses ist stark vereinfacht die Herangehensweise bei der Suche nach

neuen Therapien im 20. Jahrhundert. Die Chemie war dann auch längst, parallel zu den Fortschritten aus der Medizin, in der Lage mithilfe der *Parallelsynthese* Unmengen von Substanzen zu synthetisieren, die alle potentielle Wirkstoffkandidaten darstellen.

Obwohl nun das Wissen über die Pathophysiologie einer Krankheit vorhanden war, waren die Probleme dieselben wie vor Jahrhunderten. Wie findet man die richtige Therapie? Man ersetzte zwar die damaligen pflanzlichen und tierischen Substanzgemische durch eine einzelne Substanz, aber die Suche nach der Richtigen ist und bleibt eine der größten Herausforderung auf dem Gebiet der Arzneistoffentwicklung. Eine der ersten vielversprechendsten Ansätze entwickelte sich in den späten Achtzigern und basierte auf einer altbekannten Idee: man testete nicht unbedingt nach rationalen Konzepten, sondern untersuchte einfach die vorhandenen Substanzen auf deren Wirkung gegen das gewünschte Zielmolekül. In diesem als *High-throughput screening (HTS)* bezeichneten Verfahren werden im Schnelldurchlauf ganze Bibliotheken aus Hunderttausenden von Substanzen auf deren Aktivität hin geprüft. Die Treffer aus dem HTS bieten schließlich erste Gerüststrukturen (Leitstruktur) für die weitere Optimierung. Obwohl diese Strategie einleuchtend erscheint und sehr erfolgreich war in den letzten beiden Jahrzehnten, sind Nachteile bei diesem Verfahren nicht von der Hand zu weisen. Ein zentrales Problem ist die eingeschränkte Größe der Bibliothek, die auch in den größten bekannten Fällen nicht mehr als 10^6-10^7 Substanzen enthält. Die Anzahl der synthetisierbaren Verbindungen wurde dagegen auf 10^{60} geschätzt.[1,2] Dies macht deutlich, dass nur ein Bruchteil an möglichen Verbindungen überhaupt getestet werden kann und ein Großteil von potentiell aktiven Substanzen nicht berücksichtigt wird. Darüber hinaus ist die Trefferrate oft viel zu niedrig[3] und die gefundenen Strukturen sind weit davon entfernt, optimale Arzneistoffkandidaten zu werden, da wichtige physikochemische und pharmakokinetische Gesichtspunkte im HTS völlig außer Acht gelassen werden. Dieses muss dann im aufwendigen Optimierungsprozess korrigiert werden, wo schließlich Löslichkeit, Stabilität, Bioverfügbarkeit, Permeabilität und Bindungsaffinität der Leitstruktur verfeinert werden.

Ein anderer vielversprechender Ansatz ist das *fragment-based drug discovery (FBDD)*, das vor über 25 Jahren begründet wurde.[4] Es beruht auf der Idee, dass im ersten Schritt nur Substanzen berücksichtigt werden, die nicht größer als 160 Da sind. Durch diese Vorgabe wird die Anzahl synthetisierbarer Verbindungen drastisch reduziert. Der gesamte Bereich wird nun auf 10^7 geschätzt[2] mit der Konsequenz, dass auch kleine Bibliotheken aus 10^3-10^4

Substanzen einen Großteil des Machbaren abdecken. Auch hier stellen die gefundenen Fragmente erste Leitstrukturen dar. Die Vorteile gegenüber dem HTS liegen aber auf der Hand: weniger Substanzen werden getestet, die Optimierung der Treffer ist wegen der weit weniger komplexen Strukturen leicht möglich und man kann früh im Optimierungsprozess auf die physikochemischen Parameter Einfluss nehmen, während dies oft bei Leitstrukturen aus dem HTS wegen der Komplexität und Größe der Moleküle schwer gelingt.

Das computergestützte Drug Design verzichtet dagegen auf experimentelle Verfahren und versucht die Aktivität von Strukturen so gut es geht *in silico* vorherzusagen. Man unterscheidet hier ligand- und strukturbasierte Ansätze. Ist keine Kristallstruktur vom Zielprotein vorhanden, bedient man sich dem Ligand-basierten Verfahren. Hier wird aus bisher entdeckten aktiven Liganden nach gemeinsamen Strukturmerkmalen gesucht und daraus ein Pharmakophor-Modell abgeleitet. Man definiert also ein Minimum an Merkmalen, die für eine ausreichende Inhibition vorhanden sein müssen. Im strukturbasierten Design muss die 3D-Struktur des Proteins dagegen aufgeklärt und die Bindetasche eindeutig definiert sein. Mit diesem Wissen ist es nun möglich eine virtuelle, theoretisch unbegrenzte Substanzbibliothek mithilfe des Docking zu screenen. Die gefundenen Liganden werden schließlich mit einer Scoring-Funktion bewertet, um eine Art Ranking aufzustellen. Die Scoring-Funktion ist in diesem Verfahren ein kritischer Punkt, da sie darüber entscheidet, welche Liganden beim Docking als sinnvoll betrachtet werden. Hier können mehrere Scoring-Funktionen ausgewählt werden, die auf total unterschiedlichen Konzepten basieren. Die wissensbasierte Scoring-Funktion bedient sich statistisch ermittelter Potentiale, die aus Beobachtungen von großen Strukturdatenbanken (CSD, PDB) abgeleitet wurden. Dabei bildet das Vorkommen distanzabhängiger Kontaktgeometrien die Grundlage für das Potential. Zu dieser Klasse gehören DrugScore[(5)] und SuperStar[(6)]. GoldScore dagegen verwendet theoretische Kraftfelder um die Pose eines Liganden zu bewerten.[(7)] Dieser Ansatz ermittelt die Stärke jeder Interaktion zwischen dem Liganden und dem Protein. Die Summe aller Beiträge im Komplex spiegelt dann die Bindungsaffinität des Liganden wider. Hier wird also auf rein theoretischer Basis versucht die unterschiedlichen Bindungsbeiträge zu quantifizieren.

Eine zuverlässige Vorhersage ist nur bedingt mit Scoring-Funktionen möglich, weshalb eine experimentelle Evaluation immer notwendig ist. Daher ist die Weiterentwicklung von Scoring-Funktionen ein stets aktuelles Forschungsgebiet. Gerade experimentelle

1. Einleitung

Bindungsstudien über einzelne Bindungsaspekte wie z.B. Wasser-Einfluss, Kooperativität oder Enthalpie-Entropie-Kompensation liefern oftmals entscheidende Erkenntnisse, die in die Scoring-Funktion einfließen und sie zuverlässiger machen.

1.2 Bindungsprozess/Binding process

Wie oben beschrieben ist der Bindungsprozess durch viele komplexe Phänomene gekennzeichnet, welche verantwortlich dafür sind, wieso Scoring-Funktionen oftmals Defizite aufweisen die Bindungsaffinität richtig abzuschätzen. Trotz dieser Komplexität ist es aber möglich, alle beobachteten Vorgänge innerhalb des Bindungsprozesses auf drei Bereiche zurückzuführen: die *shape complementarity*, die Wasserstoffbrücken und die Van-der-Waals Wechselwirkungen.

1.2.1 Shape complementarity

Die *shape complementarity* beschreibt eine der Grundvoraussetzung für die Bindung zwischen einem Liganden und einem Zielprotein. Nur zueinander komplementäre Systeme, die räumlich zueinander passen, können miteinander interagieren. Diese Beobachtung geht auf Emil Fischer zurück, der es hypothetisch am Beispiel der Bindung zwischen Substrat und Protein beschrieben hatte. Dieser Vorgang ist wegen der Analogie zu einem Schlüssel, der nur für ein Schloss konzipiert wurde, als Schlüssel-Schloss-Prinzip bekannt. Mit der Zeit erkannte man, dass weder das Protein noch der Ligand starre Systeme darstellen. Die Proteinbindetasche zeigt oftmals adaptive Eigenschaften (Induced fit), um mit dem Liganden bestmöglich zu interagieren. Die starre Vorstellung von Emil Fischer wurde daher zunehmend von einem Induced-Fit-Konzept abgelöst, das beiden Bindungspartnern einen gewissen Grad an Flexibilität erlaubt.

1.2.2 Wasserstoffbrücken/Hydrogen bonds

Durch nichtkovalente Interaktionen, wie die Wasserstoffbrücken und die Van-der-Waals Wechselwirkungen, lassen sich alle Bindungsvorgänge, sei es vor oder nach der Komplexierung, beschreiben. Besonders die Rolle der Interaktion im Komplex ist allgegenwärtig, während der Einfluss der Interaktion vor der Komplexbildung oftmals unterschätzt wird, wie nachfolgend beschrieben.

Eine neue Wasserstoffbrücke im Komplex wird nämlich meistens mit einem positiven Beitrag zur Bindungsaffinität assoziiert. Aus thermodynamischer Sicht resultiert aus jeder neuen Wasserstoffbrücke ein bestimmter enthalpischer Beitrag ($\Delta H_{H\text{-}Brücke}$). Ob dieser Beitrag jedoch einen günstigen Effekt auf die Bindungsaffinität ($\Delta G_{Bindung}$) hat, kann man leider mit großer Sicherheit nicht vorhersagen. Oftmals wird nämlich übersehen, dass der resultierende enthalpische Term ($\Delta H_{H\text{-}Brücke}$) sich aus der Summe von zwei enthalpischen Beiträgen ergibt: $\Delta H_{H\text{-}Bindung}$ aus der neu gebildeten Interaktion im Komplex (exothermer Prozess) und das $\Delta H_{H\text{-}Desol.}$ aus der Desolvatation der beteiligten Gruppen (endothermer Prozess). Nur wenn die Distanz und die Orientierung der neuen Wasserstoffbrücke im Komplex optimal sind, überwiegt der exotherme Term, und es wirkt sich günstig auf $\Delta G_{Bindung}$ aus. Im Umkehrschluss bedeutet dies, dass eine nichtoptimale Interaktion sogar einen ungünstigen Einfluss auf die Bindungsaffinität ausübt, da der endotherme Term der Desolvatation nicht kompensiert wird und somit $\Delta G_{Bindung}$ erniedrigt wird. Die freie Energie einer Wasserstoffbrücke ($\Delta G_{H\text{-}Brücke}$) setzt sich jedoch zusätzlich auch aus einem entropischen Term zusammen, was die Sache noch komplizierter macht. Die Enthalpie-Entropie-Kompensation[8,9] zeigt eindrucksvoll, dass auch eine optimal gesetzte Wasserstoffbrücke eine Verbesserung der Bindungsaffinität nicht garantiert, da der enthalpische Beitrag nahezu vollständig durch einen ungünstigen entropischen Term kompensiert wird. Die Ursache dieser Kompensation liegt darin, dass durch eine neue Interaktion die adressierten Proteinbereiche stärker fixiert werden. Ein vorher in der Apoform leicht ungeordneter Rest wird geordneter und der entsprechende Verlust in der Entropie ($- T\Delta S_{H\text{-}Protein}$) macht die günstige Enthalpie einer Wasserstoffbrücke zunichte. Ein Ansatz dies zu erfassen, ist die Berücksichtigung der Flexibilität eines Proteinrestes, der durch eine neue Wasserstoffbrücke adressiert wird. Anhand von kristallographischen B-Werten und MD-Simulationen kann man diese Flexibilität bewerten und entscheiden, wie aussichtsreich eine neue Wasserstoffbrücke im Hinblick auf die Bindungsaffinität ist. Alternativ besteht das Konzept zwei Wasserstoffbrücken einzuführen, die denselben Rest adressieren. Die erste würde für den entropischen Beitrag zahlen, während aus der zweiten Bindung kein nachteiliger entropischer Term resultiert, da eine Interaktion mit einer schon vorfixierten Region stattfindet.[10]

Es ist daher sehr komplex exakt festzulegen, welchen Wert die freie Energie einer einzelnen Wasserstoffbrücke ($\Delta G_{H\text{-}Brücke}$) besitzt. Mehrere experimentelle Versuche wurden trotzdem

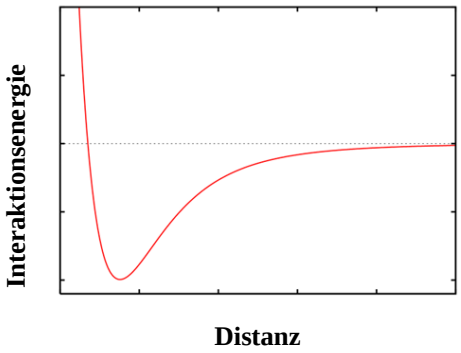
unternommen, diesen Beitrag abzuschätzen. Fersht *et al.*[11] untersuchten die tyrosyl-t-RNA-Synthase im Komplex mit ihrem Substrat. Die Reste, die mit dem Substrat in Form einer Wasserstoffbrücke interagierten, wurden systematisch mutiert. Die anschließende kinetische Untersuchung der Interaktion quantifizierte die Verschlechterung der Bindungsaffinität. So konnte der positive Beitrag einer neutralen Wasserstoffbrücke auf 2,1 - 6,3 kJ/mol abgeschätzt werden. Dieser Wert konnte durch eine Studie von Williams *et al.*[12] nahezu identisch mit einem anderen Verfahren reproduziert werden. Hier untersuchte man die Interaktion von Peptiden aus dem D-Ala-D-Ala-Terminus des Peptidoglykans mit Vancomycin. Die Bindung von leicht modifizierten Peptiden zu Vancomycin wurde anschließend mithilfe der NMR-Spektroskopie untersucht. In dieser Arbeit konnte der Wert von Fersht *et al.*[11] reproduziert werden.

1.2.3 Van-der-Waals-Wechselwirkung/Van-der-Waals interactions

Eine weitere nichtkovalente Interaktion in Protein-Ligand-Komplexen ist die aus enthalpischer Sicht weitaus schwächere Van-der-Waals-Wechselwirkung. Diese Kraft ist nicht gerichtet und wird abhängig vom Dipolcharakter der beteiligten Molekülpartner in drei Bereiche untergliedert. Sind zwei permanente Dipole vorhanden, wird die anziehende Kraft zwischen diesen als Keesom-Wechselwirkung bezeichnet. Ein Beispiel dafür ist das Chlorwasserstoff-Molekül. Das partiell negativ geladene Chloratom interagiert mit dem partiell positiv geladenen Wasserstoff. Die Interaktion zwischen einem permanenten Dipol und einem polarisierbaren Molekül wird dagegen als Debye-Wechselwirkung bezeichnet. Das Molekül mit dem Dipolmoment ist hier in der Lage einen Dipol im unpolaren Molekül zu induzieren. Ein Beispiel für diese Van-der-Waals-Wechselwirkung ist der Komplex aus Chlorwasserstoff und Argon.[13] Das teils negativ geladene Chlor verformt die Elektronenhülle des Argons, da sich beide Elektronenhüllen bei ausreichender Annäherung abstoßen. Auf diese Weise kommt es zu einer asymmetrischen Elektronenverteilung im Argon. Das teils positive Wasserstoff kann analog zum Chlor ebenfalls ein Dipolmoment im Argon induzieren, da es die Elektronen nun anzieht. Die weitaus bekannteste Van-der-Waals-Wechselwirkung ist die London-Kraft. Hier interagieren zwei unpolare Moleküle miteinander, die kein permanentes Dipolmoment aufweisen. Während der Interaktion von unpolaren Molekülen kommen sich die unpolarisierten Elektronenhüllen näher, was energetisch kein günstiger Zustand darstellt. Die Folge ist die spontane Induktion von

Dipolmomenten auf beiden Seiten, um die ungünstige Nähe von Elektronenhüllen zu vermeiden. Die Energie dieser Wechselwirkung hängt stark vom Abstand zwischen den unpolaren Atomen ab. In Abbildung 1 ist die Abhängigkeit der Energie einer Van-der-Waals-Wechselwirkung von der Distanz gezeigt.

Abbildung 1: Die Abhängigkeit der Energie der Van-der-Waals-Wechselwirkung von der Distanz.



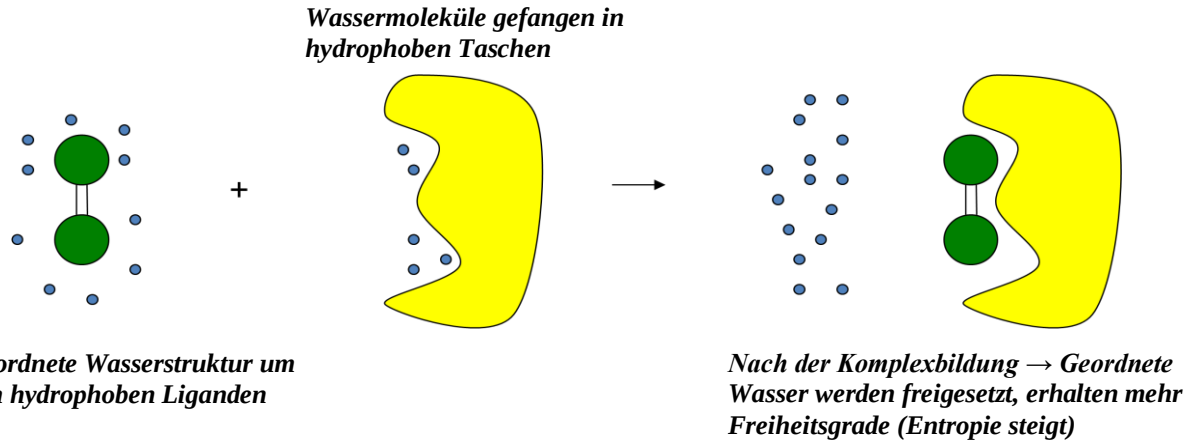
Im Minimum dieser Kurve existiert die größte Anziehung zwischen unpolaren Atomen. Wird der Abstand größer, nähert sich die Energie langsam dem Wert 0 an. Wird der Kontakt aber näher als die typische Van-der-Waals-Distanz, fällt die Energie sprunghaft ab und schließlich überwiegen abstoßende Kräfte. Die Distanz für einen optimalen Van-der-Waals-Kontakt zwischen zwei Kohlenstoffen wird in einem Protein-Ligand-Komplex auf 3,6-3.9 Å geschätzt.

Verglichen mit den beiden anderen ist die London-Kraft am schwächsten, wenn nur die Energie der Interaktion betrachtet wird. Die Betrachtungen werden zusätzlich durch das Faktum verkompliziert, dass sich alle in einem biologischen System ablaufenden Prozesse in Wasser abspielen. Wie bei der Wasserstoffbrücke muss man auch hier berücksichtigen, dass vor der eigentlichen Bindung das polare Molekül desolvatisiert werden muss. Der energetische Beitrag aus der Debye-Kraft kann jedoch die hohe Desolvatationsenergie nicht kompensieren. Die Folge ist, dass es, wie bei einer schlecht platzierten Wasserstoffbrücke, wegen Desolvatationseffekten insgesamt zu einer Verschlechterung der Bindungsaffinität kommt. Die hydrophobe Optimierung von Arzneistoffkandidaten ist daher auf die London-Kräfte angewiesen.

Wenn die enthalpischen Beiträge aus der London-Kraft sehr niedrig ausfallen und gerade mal ausreichen für die Desolvatation der hydrophoben Bindungspartner, ist die Frage berechtigt, welche treibenden Kräfte hinter dieser Interaktion stecken. Vor Jahrzehnten wurde schließlich postuliert, dass der entscheidende Beitrag aus der Verdrängung von geordneten

Wasserstrukturen aus hydrophoben Bereichen der Bindungspartner resultiert.[14] Die Wassermoleküle können vor der Bindung mit einem hydrophoben Rest keine Wasserstoffbrücken eingehen. Die Folge davon ist, dass sich die Wassermoleküle reorganisieren und die Wasserstoffbrücken zu benachbarten Wassermolekülen gestärkt werden. Dieses erhöht natürlich den Grad der Ordnung an der Grenzfläche zwischen der Wasserphase und dem hydrophoben Bereich. Die beschriebenen Wasserstrukturen um hydrophobe Bereiche werden in der Literatur als „Eisberge“ oder „Clathrate“ bezeichnet.[15,16] Bei der Bindung werden diese energetisch ungünstigen Wassermoleküle wieder in die Wasserphase freigesetzt. Die Konsequenz ist eine Erhöhung der Unordnung des gesamten Systems, da die Wassermoleküle wieder ihre Freiheitsgrade gewinnen (Abbildung 2). Die hydrophobe Wechselwirkung gilt daher von jeher als getrieben von einer Reorganisation der beteiligten Wassermoleküle.

Abbildung. 2: Schematische Darstellung des hydrophoben Effekts. In grün wird der Ligand und in gelb das Protein mit hydrophoben Taschen gezeigt. Als kleine blaue Kugeln sind die Wassermoleküle wiedergegeben.



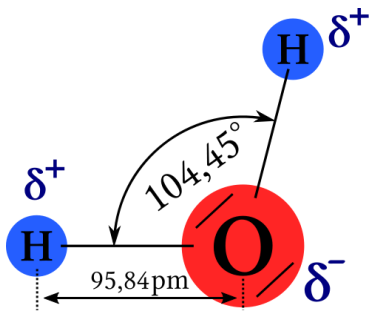
1.3 Wasser in Protein-Ligand-Interaktionen/Water in protein-ligand interaction

Obwohl das Wassermolekül auf den ersten Blick einfach erscheint, ist es in seinen chemischen und strukturellen Eigenschaften sehr speziell. Einzigartig ist, dass es als einziges Molekül in der Natur bei den für biologische Systeme relevanten Temperaturen in allen drei Aggregatzuständen vorkommt. Wasser hat bei 4°C seine größte Dichte und das Volumen steigt beim Gefrieren sprunghaft an (Dichteanomalie). Einzig Quecksilber besitzt unter den

Flüssigkeiten eine höhere Oberflächenspannung als Wasser. Es besitzt außerdem eine der höchsten Wärmekapazitäten aller bekannten Flüssigkeiten (75,4 $Jmol^{-1}K^{-1}$).

Strukturell besitzt Wasser zwei Wasserstoffatome, die mit einem Sauerstoffatom verbunden sind (Abbildung 3). Auffällig ist, dass der Winkel zwischen den OH-Bindungen (104,5 °) vom erwarteten Wert für einen Tetraeder (109,5 °) abweicht. Die freien Elektronenpaare des Sauerstoffs nehmen im Molekül einen erhöhten Raumbedarf ein, was den Winkel zwischen den OH-Bindungen ein wenig zusammenstaucht.

Abbildung 3: Darstellung der chemischen Struktur des Wassers (Bild erstellt von Patrick Emil-Zörner).



Wegen der hohen Elektronegativität des Sauerstoffs ist das Wassermolekül stark polarisiert. Der Sauerstoff zieht die Elektronen des Wasserstoffs zu sich und es resultiert ein starkes Dipolmoment (1,8 Debye). Wasser kann daher in Protein-Ligand-Komplexen sowohl als Akzeptor als auch Donator für Wasserstoffbrücken dienen. Insgesamt kann ein Wassermolekül theoretisch vier Wasserstoffbrücken ausbilden. Zusammengenommen muss man feststellen, dass Wasser vor allem für biologische Systeme die wichtigste Flüssigkeit auf der Erde darstellt, ist aber zugleich auch eine der kompliziertesten.

Wegen der vielen komplexen Eigenschaften des Wassers wird nun verständlich, warum im strukturbasierten Design oftmals die Effekte von Wasser vernachlässigt werden. Fakt ist jedoch, dass jeder beobachtete Bindungsprozess im Wasser stattfindet. Eine Vernachlässigung der Wassereffekte ist in jeder Hinsicht naiv, da es das gesamte System auf den schon gebildeten Komplex reduziert. Wenn man das gesamte System betrachtet, sind beide Bindungspartner, Protein und Ligand, auch vor der Bindung komplett von Wasser umgeben. Bevor es zur Bindung zwischen Ligand und Protein kommen kann, müssen die Wasserhüllen beider Bindungspartner abgestreift werden. Diese Desolvatationsenergien müssen berücksichtigt werden, da sie die Bindungsaffinität entscheidend beeinflussen. Polare oder geladene funktionelle Gruppen haben z.B. viel höhere Desolvatationsbeiträge als

neutrale Liganden.[17] Der Beitrag von polaren Gruppen im Liganden könnte daher nachteilig sein, wenn keine vergleichbaren Interaktionen im Komplex stattfinden. Die Abschätzung dieser Energien ist mit verschiedenen Solvatationsmodellen möglich. Implizite Modelle behandeln das Wasser als ein Kontinuum, das den Liganden umgibt, während explizite Modelle einzelne Wassermoleküle betrachten, die mit dem Liganden interagieren. COSMO[18] oder SM8[19] sind Beispiele für implizite Modelle und SPC/E[20] für ein explizites Modell, das oft in MD-Simulationen benutzt wird. Oftmals unterscheiden sich diese Werte jedoch sehr stark, abhängig davon welches Modell benutzt wird. Das zeigt, dass diese Energien nicht zuverlässig genug vorhergesagt werden können. Die Desolvatation von Proteinseite her spielt natürlich auch eine Rolle, aber die Abschätzung dieser Energien ist noch komplexer[21] und experimentelle Ansätze existieren nur für Spezialfälle[22].

Eine andere entscheidende Rolle spielt Wasser direkt im Komplex. Grundsätzlich gilt die Annahme, dass vergrabene Wassermoleküle im Komplex einen negativen Bindungsbeitrag liefern, da das Wassermolekül durch die Fixierung Freiheitsgrade verliert. Im strukturbasierten Design wird daher versucht gefundene Wasser im Komplex zu verdrängen um einen entropischen Vorteil zu erhalten.[23,24] Oft erkennt man jedoch auch keine Verbesserung in der Bindungsaffinität wegen einer Enthalpie/Entropie-Kompensation.[25] Teils sind die Wassermoleküle so stark am Protein gebunden, dass der entropische Vorteil nicht ausreicht um die Energie zu kompensieren, die benötigt wird, um diese Wassermoleküle aus der Bindetasche zu entfernen. Eine Abschätzung darüber wie stark ein Wassermolekül am Protein gebunden ist, ist ebenfalls schwierig, da stark gebundene Wasser in unterschiedlichen Kristallformen oftmals nicht konserviert sind.[26]

Andererseits existieren Ansätze, die zeigen, dass auch ein Einschluss von Wassermolekülen einen günstigen Effekt auf die Bindungsaffinität haben könnte.[27] Die Idee dahinter ist simpel und überzeugend. Wenn ein Ligand eine Bindetasche komplett ausfüllt, kann eine wasservermittelte Interaktion zu Proteinresten die Bindetasche ausdehnen. So entstehen zusätzliche Wasserstoffbrücken, die sich günstig auf den enthalpischen Term auswirken. Ein Problem ist jedoch die Unkenntnis darüber, wie und wo ein Wassermolekül platziert werden muss, um zu garantieren, dass die freiwerdende Enthalpie auch ausreicht den entropischen Nachteil gewinnbringend wettzumachen. Eine zuverlässige Vorhersage ist hier nicht möglich. Auch das mangelnde Verständnis über die Enthalpie/Entropie-Kompensation bei der Wasserverdrängung macht eine Vorhersage über die Auswirkung auf die Bindungsaffinität

nahezu unmöglich. Aus Gründen der Einfachheit werden im Drug Design derartige Einflüsse von Seiten des Wassers her größtenteils vernachlässigt. Andererseits existieren viel zu wenige Studien, die auf die Problematik von Wasser vor und nach der Komplexbildung eingehen. Wegen der allgemein schwachen Datenlage ist eine umfassende Berücksichtigung von Wasser im Drug Design bisher nicht möglich, muss aber in der Zukunft verstärkt berücksichtigt werden.

1.4 Zielsetzung/Aims of the thesis

Die Arbeit ist in fünf Teile untergliedert. Der Focus dieser Arbeit liegt auf der thermodynamischen und kristallographischen Untersuchung von Wassereffekten bei der Interaktion zwischen Protein und Ligand. Die thermodynamische Charakterisierung der Interaktion wurde mittels *Isothermaler Titrationskalorimetrie* (ITC) durchgeführt. Die erste Serie war durch eine schwache Bindungsaffinität im mikromolaren Bereich gekennzeichnet. Leider ist dieser Affinitätsbereich schwer zugänglich für eine thermodynamische Charakterisierung. Daher besteht das erste Kapitel dieser Arbeit in der Etablierung und Validierung einer ITC-Methode, die es erlaubt, thermodynamische Daten mit einer hohen Genauigkeit aus einer mikromolaren Inhibitorserie zu gewinnen. Die Methode basiert auf einem Verdrängungsansatz, der von Sigurskjold[28] für äußerst starke Liganden im pikomolaren Bereich beschrieben wurde.

Im zweiten Teil wird diese schwach affine Serie dann mit der validierten Verdrängungs-ITC thermodynamisch untersucht. Diese Serie wurde entworfen und synthetisiert um die Bindungsbeiträge zu erfassen, die sich aus der Desolvatation der polaren S1 Tasche von Thrombin ergeben. Anschließend wurden Desolvatationsenergien aller untersuchten Liganden mit quantenmechanischen Berechnungen abgeschätzt, um deren Auswirkungen auf das thermodynamische Profil zu diskutieren.

Der dritte Teil handelt über die Rolle von Wasser bei der hydrophoben Wechselwirkung. Obwohl als gesichert gilt, dass die treibenden Kräfte hinter dem hydrophoben Effekt durch die Verdrängung von Wassermolekülen aus hydrophoben Bereichen resultieren, wurden in der Vergangenheit zahlreiche Beispiele[29-31] entdeckt, die nicht in dieses Konzept passen, da die Interaktion trotz hydrophoben Bindungspartnern durch einen unerklärlich hohen enthalpischen Term dominiert wird. Die letzten beiden Berichte gehen davon aus, dass eine unzureichende Solvatation der hydrophoben Bindetasche die Ursache für das beobachtete

Enthalpiesignal darstellt.[32,33] Basierend auf dieser Hypothese wurde eine Serie konzipiert, in der systematisch der Rest variiert ist, der die hydrophobe S3/4 Tasche von Thrombin adressiert. Diese Tasche gilt zwar als hydrophob, ist aber dafür ausreichend solvatisiert. Ziel war es nun zu zeigen, ob die Adressierung einer gut solvatisierten hydrophoben Bindetasche mit hydrophoben Liganden den klassischen hydrophoben Effekt zeigt.

Die vierte Studie beschäftigt sich mit der hydrophoben S_2' Tasche von Thermolysin. Die benachbarte schwach solvatisierte S_1' Tasche lieferte in einer früheren Arbeit ein enthalpisches Signal[33], das ebenfalls mit der schwachen Solvatation der untersuchten Proteintasche erklärt wird. Um die Solvatations-Hypothese zur enthalpisch getriebenen hydrophoben Wechselwirkung indirekt zu bestätigen, wurde eine hydrophobe Ligandenserie für die Interaktion mit der S_2' Tasche konzipiert und thermodynamisch charakterisiert. Eine kristallographische Untersuchung der Thermolysin-Komplexe wurde schließlich durchgeführt um die Auswirkung auf die adressierten Wassermolekülen zu dokumentieren.

Die fünfte Studie befasst sich mit dem kooperativen Effekt, der in einer Arbeit von Baum *et al.*[34] entdeckt wurde. Eine kinetische Untersuchung von zwei Serien zeigte dort, dass eine Kooperativität zwischen der hydrophoben Interaktion zur S3/4 Tasche und der Stärke einer Wasserstoffbrücke in Thrombin existiert. Basierend auf diesen Ergebnissen wurden fünf neue Serien synthetisiert, die alle mit denselben Resten die hydrophobe S3/4 Tasche von Thrombin adressieren. Die Serien unterscheiden sich lediglich in dem Bereich, der sich in der Nähe der Wasserstoffbrücke im β-Faltblatt befindet. Ziel war es, mit unterschiedlichen Substitutionen einen Einfluss auf die Stärke dieser Wasserstoffbrücke auszuüben und zu evaluieren, ob die hydrophobe Wechselwirkung tatsächlich mit der Stärke der Wasserstoffbrücke korreliert. Zusätzlich wurde auf einem reduzierten Ligandgerüst mit quantenchemischen Rechnungen überprüft, ob die Stärke der Wasserstoffbrücke tatsächlich durch die eingeführten Reste beeinflusst wird.

1.5 Referenzen/References

1 Bohacek R.S., McMartin C. & Guida W.C. (1996) The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev.* **16**, 3-50.

2 Fink T. & Reymond J.L (2007) Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million

Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery. *J. Chem. Inf. Model.* **47**, 342–353.

3 Schuffenhauer A., Ruedisser S., Marzinzik A.L., Jahnke W., Blommers M. *et al.* (2005) Library design for fragment based screening. *Curr Top Med Chem.* **5**, 751-762.

4 Jencks W.P. (1981) On the attribution and additivity of binding energies. *Proc Natl Acad Sci* **78**, 4046-4050.

5 Gohlke H., Hendlich M. & Klebe G. (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol.* **295**, 337-356.

6 Verdonk M.L., Cole J.C. & Taylor R. (1999) SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J Mol Biol.* **289**, 1093-1108.

7 Jones G., Willett P., Glen R.C., Leach A.R. & Taylor R. (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol.* **267**, 727-748.

8 Lumry R. & Rajender S. (1970) Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers.* **10**, 1125-227.

9 Sharp K. (2001) Entropy-enthalpy compensation: fact or artifact? *Protein Sci.* **3**, 661-667.

10 Ward W.H., Holdgate G.A., Rowsell S., McLean E.G., Pauptit R.A. *et al.* (1999) Kinetic and structural characteristics of the inhibition of enoyl (acyl carrier protein) reductase by triclosan. *Biochemistry.* **38**, 12514-12525.

11 Fersht A.R., Shi J.P., Knill-Jones J., Lowe D.M., Wilkinson A.J. *et al.* (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature.* **314**, 235-238.

12 Williams H., Searle M.S., Mackay J.P., Gerhard U. & Maplestone R.A. (1993) Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. *PNAS* **90**, 1172-1178.

13 Blustin P.H. (1978) A Floating Gaussian Orbital calculation on argon hydrochloride (Ar • HCl). *Theoret. Chim. Acta* **47**, 249–257.

14 Tanford C. (1978) The hydrophobic effect and the organization of living matter. *Science*. **200,** 1012-1018.

15 Némethy G. & Scheraga H.A. (1962) Structure of Water and Hydrophobic Bonding in Proteins. II. Model for the Thermodynamic Properties of Aqueous Solutions of Hydrocarbons *J. Chem. Phys.* **36**, 3401-3417.

16 Tanford C. (1980) The hydrophobic Effect: Formation of Micelles and Biological Membranes, *John Wiley and Sons*

17 Wang J. Wang W. Huo S., Lee M. & Kollman P.A. (2001) Solvation Model Based on Weighted Solvent Accessible Surface Area. *J. Phys. Chem. B* **105**, 5055–5067.

18 Klamt A. & Schüürmann G. (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc.*, *Perkin Trans. 2*, 799–805.

19 Cramer C.J. & Truhlar D.G. (2008) A Universal Approach to Solvation Modeling. *Acc Chem Res.* **41**, 760-768.

20 Berendsen H.J.C., Grigera J.R. & Straatsma T.P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271.

21 Hamelberg D. & McCammon J.A. (2004) Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. *J Am Chem Soc.* **126**, 7683-7689.

22 Holdgate G.A., Tunnicliffe A., Ward W.H., Weston S.A., Rosenbrock G. *et al.* (1997) The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. *Biochemistry.* **36**, 9663-9673.

23 Lam P.Y., Jadhav P.K., Eyermann C.J., Hodge C.N., Ru Y. *et al.* (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science.* **263**, 380-384.

24 Watson K.A., Mitchell E.P., Johnson L.N., Son J.C., Bichard C.J. *et al.* (1994) Design of inhibitors of glycogen phosphorylase: a study of alpha- and beta-C-glucosides and 1-thio-beta-D-glucose compounds. *Biochemistry*. **33**, 5745-5758.

25 Mikol V., Papageorgiou C. & Borer X. (1995) The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J Med Chem.* **38**, 3361-3367.

26 Otting G., Liepinsh E. & Wuthrich K. (1991) Protein hydration in aqueous solution. *Science* **254**, 974-980.

27 Wang H. & Ben-Naim A. (1996) A possible involvement of solvent-induced interactions in drug design. *J Med Chem.* **39**, 1531-1539.

28 Sigurskjold B.W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* **277**, 260-266.

29 Smithrud D.B., Wyman T.B. & Diederich F. (1991) Enthalpically driven cyclophane arene inclusion complexation: solvent-dependent calorimetric studies. *J. Am. Chem. Soc.* **113**, 5420-5426.

30 Eftink M.R. & Harrison J.C. (1981) Calorimetric studies of p-nitrophenol binding to α- and β cyclodextrin. *Bioorg. Chem.* **10**, 388-398.

31 Bertrand G.L, Faulkner Jr. J.R., Han S.M. & Armstrong D.W. (1989) Substituent effects on the binding of phenols to cyclodextrins in aqueous solution. *J. Phys. Chem.* **93**, 6863-6867.

32 Bingham R.J., Findlay J.B., Hsieh S.Y., Kalverda A.P., Kjellberg A. *et al.* (2004) Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J. Am. Chem. Soc.* **126**, 1675-1681.

33 Englert L., Biela A., Zayed M., Heine A., Hangauer D. *et al.* (2010) Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: binding of phosphonamidate to the S_1'-pocket of thermolysin. *Biochim Biophys Acta.* **1800**, 1192-1202.

34 Baum, B., Muley, L., Smolinski, M., Heine, A., Hangauer, D. *et al.* (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.

2 Validation of ITC displacement titration and implications to thermodynamic fragment characterization

2.1 Abstract

Fragment-based lead discovery (FBLD) has matured in recent years into a very propitious strategy in drug development giving multiple promising starting points to generate drug candidates. Usually, a fragment screening campaign provides an impressive series of favorable hits to be considered for hit-to-lead evaluation. It has recently been suggested that first hit candidates with largest enthalpic efficiency are the most promising candidates for further optimization. Therefore, in particular for fragments the information about the enthalpic binding contribution is of utmost importance to take an optimal decision. Unfortunately, reliable thermodynamic data are difficult to collect for fragments and they are strongly affected by large errors due to the exceptionally weak binding affinities. In consequence, the thermodynamic parameter enthalpy can hardly be considered as prime selection criterion. We suggest a displacement ITC approach to determine the thermodynamic signature of weak binding fragments with high accuracy and reliability. To validate the reliability and scope of this new technique K_d values between 132 µM and 12 pM have been determined for a series of thrombin inhibitors and faced with independent kinetically recorded K_i data. As insufficient protein solubility can also hamper determination of thermodynamic data we showed that the displacement approach can also be used to circumvent poor solubility of the protein. It is shown that significantly reduced protein concentrations reveal equally reliable data using the suggested displacement technique instead of direct titrations.

2.2 Introduction

Once a novel protein has been discovered and validated as putative drug target for therapeutic intervention, usually high-throughput screening is applied to discover initial hits as prospective candidates for further development. Huge compound libraries of small molecules are enumerated and provide pharmaceutical industry in prospective cases with

large numbers of potential hits, mostly with affinities in one- to two-digit micromolar range.[1] Considering the molecular weight of these hits rather disappointing ligand binding efficiencies (affinity per unit molecular mass) become apparent.[2,3] In consequence, screening has been expanded towards increasingly smaller candidates, so-called fragments which leave sufficient room for optimization. Also this approach unravels large samples of promising hits for further development.[4] At this point, a crucial decision has to be made as to which of these fragments should be taken to the next level of development. As this step will allocate major resources, it requires reliable and relevant parameters to support this decision process. Recently, the use of thermodynamic data collected by isothermal titration calorimetry (ITC) was suggested as a promising additional indicator.[5,6] With respect to ligand optimization, it appears advisable to start with a hit of pronounced enthalpic binding characteristics as strategies to improve the entropic component of binding free energy are more obvious to achieve.[7] In consequence, it was proposed to start in particular with a fragment hit showing the largest enthalpic efficiency. ITC is the straight-forward-to-use biophysical technique to record the required data. However, its general sensitivity ranges from low micromolar to one-digit nanomolar binders, thus hardly applicable to weak binding fragments.

Sigurskjold[8] first introduced theoretically the concept of displacement titration to expand the above-mentioned dynamic range of ITC. By applying elaborate measurement protocols the range of sensitivity could be expanded towards high-affinity binding using this approach. In such cases, the protein solution is saturated with a weak-binding ligand followed by the titration of a competitive high affine ligand. The latter will successively displace the weak-binding ligand from the binding site. In this concept the binding affinity of the highly potent ligand is artifically lowered to a range where it can be measured by ITC. This displacement strategy has been tested experimentally to extremely strong HIV protease inhibitors with affinities in picomolar range[9] and a detailed measurement protocol for strong binders is available.[10] With respect to drug development, the thermodynamic profile of highly potent and fully optimized ligands is interesting for retrospective analysis and to elucidate structure-function relations but not applicable in the acute process of selecting the most promising fragments for optimization.

The study of fragment hits proved to be very difficult as these show mostly low binding affinities. Especially, the determination of accurate association constants of these low

affinity systems would mean a huge contribution in order to understand why fragments actually bind. Common methods for interaction analysis like surface plasmon resonance or kinetic assay are well known to have problems to generate reliable data in the high micro to millimolar range. ITC is another technique which can be applied to address this affinity range by some modifications in the experimental procedure. Unfortunately, considering the low molecular weight of fragments we would expect a scenario where the interaction between fragment and protein is, in addition to the weak binding affinity, linked with a low enthalpy signal. This makes the ITC experiment even more complex and material-consuming as the ITC titration must be carried out at a higher protein concentration to realize sufficient heat changes. The high protein concentration in a direct titration is crucial either because the amount of available protein material is limited or the insufficient solubility of the biomolecule makes ITC experiments at high concentrations impossible. Additionally, the titration curves recorded under these conditions are difficult to analyze because they lack sigmoidal shape (Figure 1). As the inflection point (stoichiometry) in this mode is poorly defined, the enthalpy of binding can only be extracted by fixing the stoichiometry parameter during the fitting procedure. Another approach how to deal with low affinity systems is to carry out low c-value titrations. Considering the importance of consulting reliable thermodynamic signatures of initial weak binding hits with respect to lead optimization, more reliable measuring protocols to record accurate microcalorimetric data of two-digit micromolar to even millimolar binders are desperately needed. In this respect, the working under conditions which lead to the typical sigmoidal shape are always preferred with respect to reliability.

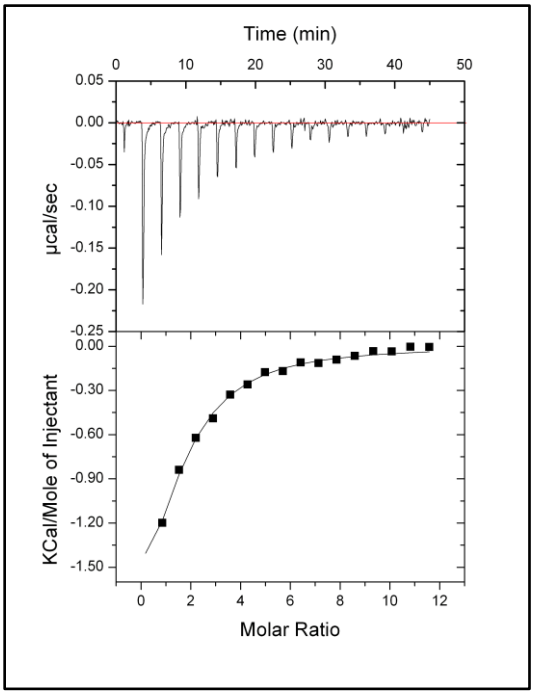
In this contribution we present a measurement strategy to reliably overcome the aformentioned problems by expanding the dynamic range of ITC towards micro- to even millimolar binding affinities by applying the displacement technique first experimentally realized by Zhang and Zhang.[11] The experiment needs two titrations following the same protocol. In a first step, a nanomolar ligand is selected and characterized directly by ITC. In a second step the nanomolar ligand is titrated into the protein solution saturated with the weak-binding ligand of interest. The apparent binding parameters of the nanomolar ligand obtained in a displacement titration can be used to calculate the entire thermodynamic profile of the weak-binding ligand by applying the displacement model of Sigurskjod, originally developed for potent binders. Although the principle of the method is known since

2. Validation of ITC displacement titration

11 years we could just find a handful of studies dealing with special problems comprising peptides binding to Src kinase SH2 domain[12] or 5'CMP binding to ribonuclease.[13] A comprehensive thermodynamic analysis of weak-binding fragments is unfortunately missing but desperately needed to clarify why fragments show activity and to finally support the FBDD with thermodynamic data comparable to the ITC support in lead optimization. Moreover, the review by Edink *et al.*[14] shows impressively that thermodynamic fragment characterization is not still applied widely and is in the found cases mainly limited to fragments showing already moderate affinities.

In our opinion, the main reasons for this reluctance is a missing experimental validation of the displacement approach and unsolved questions when working with displacement titrations. Are measured fragment affinities and enthalpies reliable, do the fragment affinities correlate well with kinetically determined inhibition constants, how much fragment is needed to saturate the protein and what is an appropriate competitive ligand are aspects responsible for the lacking acceptance of this method. In this publication, we give solutions to some of these questions in order to make this valuable ITC titration strategy better applicable to a larger community.

Figure 1: Titration curve for a weak binder from this study in a direct ITC titration. A weak binding ligand (1.5 mM) is directly measured at a thrombin concentration of 20 µM in Hepes buffer. Obviously, the sigmoidal shape is not obtained.



2.3 Methodology & Titration Protocol

2.3.1 The procedure of the displacement titration

Before the displacement titration can be performed a sufficiently potent reporter ligand must be selected. This latter ligand must be characterized in a direct ITC titration. The resulting titration curve has the typical sigmoidal shape which can be analyzed by the single-site binding isotherm[15] using Origin 7.0. Once this competitive ligand is characterized we prepare a protein solution incubated with the weak-binding ligand under investigation. Highest saturation possible of the protein with bound ligand should be attempted. This solution is injected to the sample cell. The solution of the competitive ligand is adjusted to the desired concentration and then filled in to the syringe. Usually, the concentration of this solution should be 15-20 times the applied protein concentration. Due to application of high concentrations of the weak-binding ligand in the sample cell, large signals of heat of dilution will result during the displacement titration. We, therefore highly recommend to add the weak-binding ligand at the same concentration to the solution used for the study of the competitive ligand. The subsequent experiment is performed similarly to direct ITC titrations. The strong binder will displace step-by-step the weakly bound ligand from the macromolecule in the course of the titration. The titration is finished when complete displacement of the weak binder is achieved which is indicated once only small peaks of dilution are recorded. The resulting titration curve of the displacement titration is analyzed using the Origin 7.0 software by fitting a single-site binding isotherm as there is no fitting function available for the resulting curves in case of displacement of weak-binding ligands. A displacement fitting routine for the ITC model is only available for the analysis of high-affinity ligands. The derived parameters are used as $K_{a\ obs}$ and ΔH_{obs} in equation 1 and 2[11] to calculate the thermodynamic properties $K_{a\ WL}$ and ΔH_{WL} of the weak-binding ligand. The values $K_{a\ SL}$ and ΔH_{SL} are taken from a separate experiment carried out in the beginning of the experiment.

$$K_{a\,WL} = \left(\frac{K_{a\,SL}}{K_{a\,obs}} - 1\right) \cdot \frac{1}{[L_{free}]} \qquad (1)$$

$K_{a\ obs}$ = Observed association constant of the strong binder in presence of the weak binder (displacement)

$K_{a\ WL}$ = Unknown association constant of the weak binder

$K_{a\ SL}$ = Association constant of the strong binder measured in a direct ITC titration

$[L_{free}]$ = Concentration of free weak binder in the saturated protein solution

$$\Delta \boldsymbol{H_{WL}} = (\Delta \boldsymbol{H_{SL}} - \Delta \boldsymbol{H_{obs}}) \left(\mathbf{1} + \frac{\mathbf{1}}{\boldsymbol{K_{a\,WL}} \cdot [\boldsymbol{L_{free}}]} \right) \quad (2)$$

ΔH_{WL} = Unknown enthalpy change of weak binder

ΔH_{obs} = Observed enthaly change of strong binder in presence of the weak binder (displacement)

ΔH_{SL} = Enthalpy change of strong binder measured in a direct ITC titration

$K_{a\,WL}$ = Association constant of weak binder

$[L_{free}]$ = Concentration of weak binder in the saturated protein solution

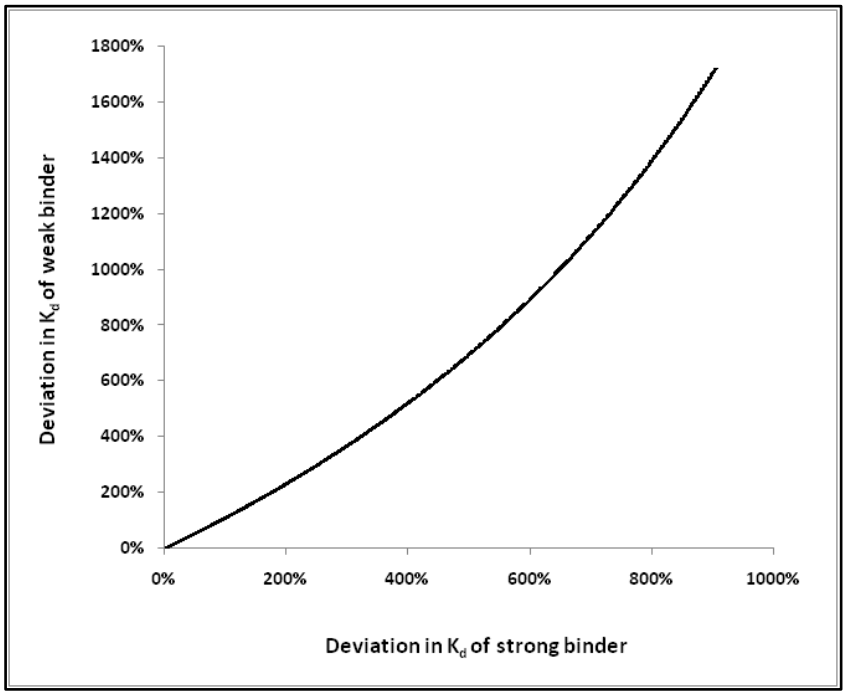
The concentration of the free ligand $[L_{free}]$ (Equ. 3) is unknown. As the concentration of the weak-binding ligand is much larger than that of the protein $[L_{tot}] >> [P_{tot}]$, the free ligand concentration $[L_{free}]$ is practically equal to $[L_{tot}]$. Thus $[L_{tot}]$ was used in the further analysis.

$$[\boldsymbol{L_{free}}] = [\boldsymbol{L_{tot}}] - [\boldsymbol{PL}] \quad (3)$$

2.3.2 The accuracy of the association constant determination of the competitive ligand

The ligand used for competitive displacement has to be selected carefully with respect to its thermodynamic profile. The accuracy of the affinity determination of the weak-binding ligand ($K_{d\,WL}$) strongly depends on the correct determination of the affinity of the strong-binding ligand ($K_{d\,SL}$) as indicated in equation 1. In order to estimate this influence of the strong-binding ligand we calculated the anticipated error for $K_{d\,WL}$ as a function of the uncertainty in measurement of $K_{d\,SL}$. The uncertainty in $K_{d\,WL}$ is in the beginning of the error estimation (cf. Figure 2) nearly directly proportional to the error in $K_{d\,SL}$ of the strong binder. The larger the measured K_d of the strong-binding ligand deviates from the actual dissociation constant the larger will be the uncertainty in the K_d determination of the weak binder. Competitive ligands with a potency of 1-50 nM should not be used as their determined association constants are mostly accompanied with large errors in ITC. If there is no choice, displacement titrations in turn must be done for these potent ligands to confirm the direct $K_{d\,SL}$ estimation. To estimate the anticipated error, an Excel file is found in the supplementary material. A strong ligand with K_d in the range of 100-1000 nM is usually associated with a negligible error and therefore perfect for a displacement experiment.

Figure 2: Error estimation. The plot illustrates the deviation of the K_d of the weak binder as a function of the deviation of K_d of the strong binder. The exact calculation of this error estimation can be found in the method section.



2.3.3 The required strength of the competitive ligand

Aside accuracy of the predetermined binding constant of the strong binder, the affinity difference with respect to the ligand to be displaced has to be considered. The competitive ligand must be strong enough to virtually displace the weak-binding ligand completely. The higher the affinity of the weak-binding ligand the higher the affinity for the competitive ligand must be. On the other hand the strong-binding ligand must not exhibit a too high affinity or, more precisely, a too strong enthalpic signal (s. below) as then the recorded signal of the strong-binding ligand alone and in the competition experiment with the weak binder will be nearly identical. It is thus recommended to apply first as competitive ligand the one with the highest affinity and then to examine whether the competition experiment shows data at least five times the standard deviation of the data of the strong-binding ligand measured in a direct ITC titration. If this is not the case, an alternative strong-binding ligand with lower affinity has to be selected in order to record a sufficiently large difference.

2.3.4 The resulting enphalpy signal

As for any ITC experiment, prerequisite to perform such measurements a sufficiently large heat signal must occur. The special situation might be given where the competition titration cannot be recorded because no heat signal is produced upon displacing of the weak binder. If both the strong and weak binder exhibit the same enthalpy value the resulting ΔH_{obs} will be close to zero according to equation 4 which results from equation 2 by rearrangement.

$$\Delta \boldsymbol{H_{obs}} = \Delta \boldsymbol{H_{SL}} - \frac{\Delta \boldsymbol{H_{WL}}}{\left(\mathbf{1} + \frac{\mathbf{1}}{\boldsymbol{K_{a\,WL}} \cdot [\boldsymbol{L_{tot}}]}\right)} \tag{4}$$

For most displacement titrations, the product of $K_{a\ WL}$ and $[L_{tot}]$ will be significantly larger than 100. The term in brackets in the denominator will thus tend to be 1 and the equation is therefore dominated by (ΔH_{SL} - ΔH_{WL}). In consequence, proper selection of the strong-binding ligand with respect to the expected enthalpy of the weak-binding ligand is crucial. Opposing thermodynamic signatures (e.g. endothermic for the strong and exothermic for the weak ligand or vice versa) will guarantee sufficiently changed heat signal in the displacement titration. On the other hand prior to actually performing the experiment no information is available about the thermodynamic signature of the ligands under investigation and the number of fully characterized strong-binding ligands with opposite thermodynamic signature is most likely limited. Fortunately, the described situation is not that problematic because mostly the difference in enthalpies with equal sign is high enough to record measureable heat signals. In our examples we used strong exothermic ligands to characterize weak exothermic ligands because we lacked an endothermic binder. Nevertheless, the differences in enthalpies were at least 5 kJ/mol.

2.3.5 Estimation of the required concentration of weak binder for sufficient saturation

The crucial factor in this section is to estimate what concentration should be applied to guarantee a sufficient saturation of the protein solution. We can precisely predict how much ligand is needed to ensure a desired degree of saturation at a given protein concentration if the binding constant of the weak-binding ligand is known from an independent experiment, such as a biochemical assay. If there is no information given about the binding affinity of the

weak binder, a very weak binding affinity must be assumed, for instance 10 mM, to calculate the required concentration of the weak binder.

The equilibrium for the binding of protein P and weak-binding ligand L is described in equation 5 and 6. [L] is the concentration of unbound ligand with respect to the total concentration of [L_t] at the beginning of the experiment and [P] the free protein with respect to the likewise total concentration [P_t]. [PL] is the concentration of the formed complex. Regarding the expressions 7 and 8 in equation 6 reveals equation 9.

$$[\boldsymbol{P}] + [\boldsymbol{L}] \rightleftharpoons [\boldsymbol{PL}] \tag{5}$$

$$\boldsymbol{K_d} = \frac{[\boldsymbol{P}] \cdot [\boldsymbol{L}]}{[\boldsymbol{PL}]} \tag{6}$$

$$[\boldsymbol{P}] = [\boldsymbol{P_t}] - [\boldsymbol{PL}] \tag{7}$$

$$[\boldsymbol{L}] = [\boldsymbol{L_t}] - [\boldsymbol{PL}] \tag{8}$$

$$\boldsymbol{K_d} = \frac{([\boldsymbol{P_t}] - [\boldsymbol{PL}])([\boldsymbol{L_t}] - [\boldsymbol{PL}])}{[\boldsymbol{PL}]} \tag{9}$$

x is assumed as the degree of saturation which is defined in equation 10. x can adopt values between 0 (no inhibition) and 1 (full inhibition). Rearranging equation 11 and substituting [PL] (Eq. 11) in equation 9 results in equation 12 which gives the quadratic equation 13.

$$\boldsymbol{0 \leq x \leq 1} \tag{10}$$

$$\boldsymbol{x} = \frac{[\boldsymbol{PL}]}{[\boldsymbol{P_t}]} \tag{11}$$

$$\boldsymbol{K_d} = \frac{([\boldsymbol{P_t}] - \boldsymbol{x}[\boldsymbol{P_t}])([\boldsymbol{L_t}] - \boldsymbol{x}[\boldsymbol{P_t}])}{\boldsymbol{x}[\boldsymbol{P_t}]} \tag{12}$$

$$\boldsymbol{K_d} = \frac{[\boldsymbol{P_t}][\boldsymbol{L_t}] - \boldsymbol{x}[\boldsymbol{P_t}]^2 - \boldsymbol{x}[\boldsymbol{P_t}][\boldsymbol{L_t}] + \boldsymbol{x}^2[\boldsymbol{P_t}]^2}{\boldsymbol{x}[\boldsymbol{P_t}]} \tag{13}$$

The quadratic equation 13 is solved for [L_t] (Eq. 17) after some rearrangements (14-16).

$$\boldsymbol{K_d} = \frac{\boldsymbol{x}[\boldsymbol{P_t}]\left(\frac{[\boldsymbol{L_t}]}{\boldsymbol{x}} - [\boldsymbol{P_t}] - [\boldsymbol{L_t}] + \boldsymbol{x}[\boldsymbol{P_t}]\right)}{\boldsymbol{x}[\boldsymbol{P_t}]} \tag{14}$$

$$\boldsymbol{K_d} = \frac{[\boldsymbol{L_t}]}{\boldsymbol{x}} - [\boldsymbol{P_t}] - [\boldsymbol{L_t}] + \boldsymbol{x}[\boldsymbol{P_t}] \tag{15}$$

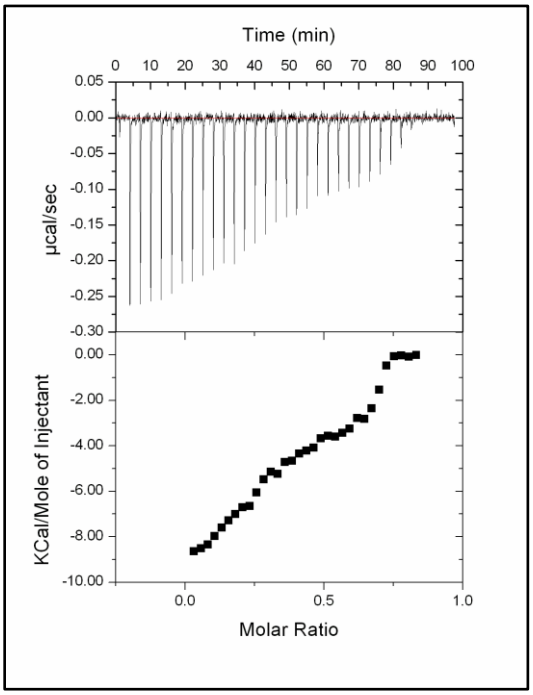
$$[\boldsymbol{L_t}] - \frac{[\boldsymbol{L_t}]}{\boldsymbol{x}} = \boldsymbol{x}[\boldsymbol{P_t}] - [\boldsymbol{P_t}] - \boldsymbol{K_d} \tag{16}$$

2. Validation of ITC displacement titration

$$[L_t] = \frac{x[P_t] - [P_t] - K_d}{1 - \frac{1}{x}} \quad (17)$$

We provide in the supporting information an Excel file integrating equation 17 in a sheet called saturation. By inserting the protein concentration [P_t] and the approximate binding constant of the weak-binding ligand from an independent experiment we can estimate the required concentration [L_t] to achieve a desired degree of saturation. In case of limited ligand availability or solubility problems it is possible to calculate for a given ligand concentration the achieved degree of saturation (x≠1). In case the titration started from incomplete saturation of the weak-binding ligand, the titration curve indicates double sigmoidal shape where two distinct processes are taking place (Figure 4). In the first step, the competitive ligand binds first to unoccupied sites of the protein. When all sites are occupied, in the second step the strong binder starts to displace the weak-binding ligand from the protein.

Figure 4: Titration curve in case of incomplete saturation in an ITC displacement titration. The experiment includes a titration of a mix of ligands into a protein solution to illustrate what happens if the protein solution is not completely saturated.



2.4 Results

2.4.1 Displacement titration for weak and strong thrombin binder with MW ≥ 250 Da

As an application of the above-described displacement titrations we selected thrombin as a model system. Fragments could not be used in this context as the assay experiment is not sensitive enough to provide accurate affinity data in this low affinity range. Thus we selected small molecules for the validation procedure. Seventeen thrombin ligands, investgated in other studies[16,17] were selected for validation of the scope and reliability of the suggested displacement protocol. We considered not only weak ligands (Series A composed of nine ligands **1**-**9**) but also high affinity ligands (Series B composed of seven ligands **10**-**16**). The displacement theory was originally developed for potent ligands and thus it was also reasonable to include a high affine series in order to cover the whole affinity range of the displacement approach. The weak binding series was analyzed by a photometric assay and the strong binding series by a fluorogenic assay prior to the performence of displacement ITC in order to compare kinetically determined inhibition constants (K_i) and ITC recorded dissociation constants (K_d). The weak-binding ligands feature low affinities between 2 and 132 µM while the high affinity series exhibits an affinity range between 12 pM and 4.8 nM in this biochemical assay.

Two competitive ligands with different thermodynamic profiles were selected to characterize the series regarding the low-affinity ligands (Table 1). Depending on the resulting titration curves either competitve ligand **17** or **18** were applied in series A. In case of the high-affinity series (Series B) also two binders (**18** and **19**) were selected with different thermodynamic profiles as competitive ligands (Table 1 and Table 2 for chemical structures). If no remarkable enthalpy signal had been observed the other competitve ligand was tested to reveal a detectable heat signal in both series. The measured affinities are plotted as ΔG values in Figure 5 and a correlation coefficient of R^2=0.98 indicates good correlation of the two binding constants K_d and K_i resulting from two completely different methods. The enthalpies (Figure 6) of weak binders are associated with a low standard deviation and spread over a range of 14 kJ/mol.

2. Validation of ITC displacement titration

Table 1: Thermodynamic data of the competitve reporter ligands used for thrombin displacement titrations.

	ΔG^0 (kJ/mol)	ΔH^0 (kJ/mol)	$-T\Delta S^0$ (kJ/mol)
17[a)]	- 47.3 ± 0.9	- 40.2 ± 0.9	- 7.1 ± 1.9
18[a)]	- 37.4 ± 0.3	- 34.5 ± 1.1	- 3.0 ± 1.4
18[b)]	- 37.9 ± 0.5	- 17.1 ± 1.1	- 20.8 ± 1.7
19[b)]	- 37.6 ± 0.2	- 37.5 ± 0.1	- 0.1 ± 0.2

[a)]The thermodynamic profiles of the competitive ligands used for testing the weak-affinity series in a displacement titration in Hepes buffer. These data were derived in a direct ITC titration.
[b)]The binding parameters for the competitive ligands applied to study high-affinity ligands by the displacement technique in Tris buffer. These data were measured in a direct ITC titration.

Figure 5: The affinities K_d (displacement ITC) and K_i (biochemical assay) are plotted as ΔG values according the equation $\Delta G = RT\ln(K_d \text{ or } K_i)$. The high correlation coefficient of R^2=0.98 indicates a good correlation of the different methods and underline the reliability of our displacement titrations.

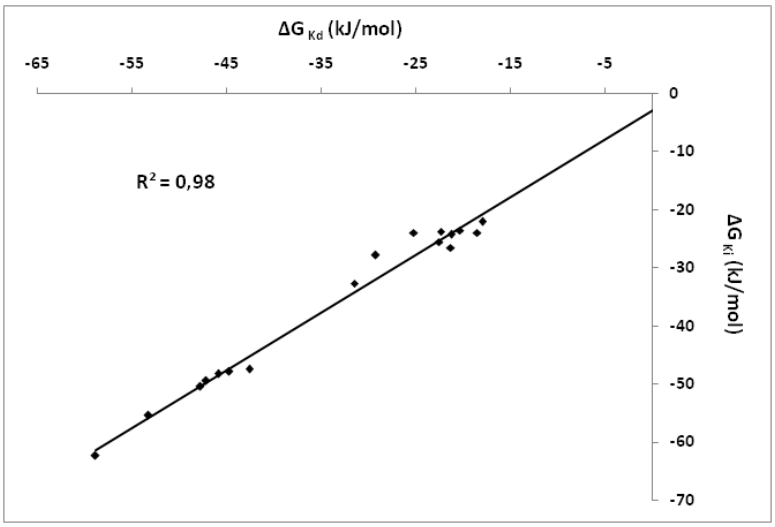
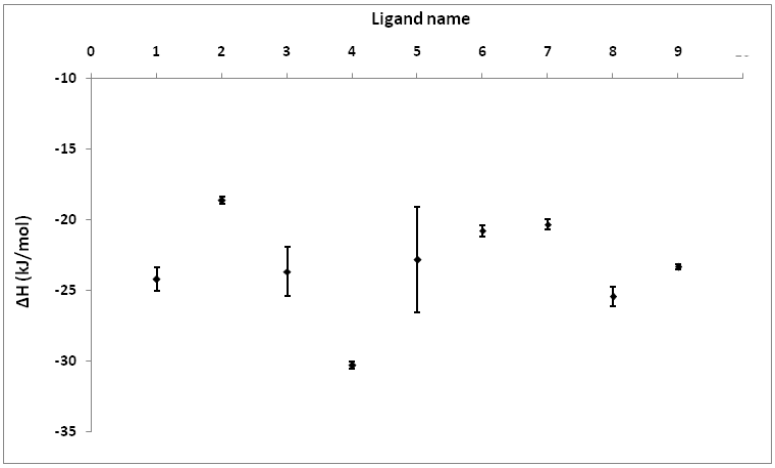


Figure 6: The estimated enthalpies of the studied weak ligands **1**-**9** (Series A) show apart of two cases very low standard deviations.



2.4.2 Fragments (MW ≤ 250 Da) measured by displacement titration

In order to demonstrate that fragments can be measured by the displacement titration we selected the aspartylprotease endothiapepsin as a model system. Four fragments (Table 2) discovered in a previous study[18] were incubated at 20 mmol/L with the protein solution (42.3 µmol/L). The HIV protease inhibitor saquinavir (named as **20**) also known to block endothiapepsin was used as competitive reporter ligand (625 µmol/L). Prior to the displacement experiment the thermodynamic profile of saquinavir was determined by direct ITC titration in acetate buffer (Figure 7a). The titration curve in displacement mode does not exhibit a double sigmoidal shape (Figure 7b) indicating sufficient saturation of the protein with the fragment. The results in Table 3 indicate a huge enthalpic contribution to binding despite the low molecular weight of the fragments which are partly compensated by unfavorable entropic contribution. The experiments were done at single buffer condition (acetate). The binding event could be superimposed by a change in protonation states which usually takes influence on the heat signal. To distinguish heat contributions caused by protonation changes from the actual binding event, the titrations would have to be conducted in three different buffers. Since, however, the ionization enthalpy of an acetate buffer amounts only to 0.49 kJ/mol[19] the obtained enthalpies can be considered as net binding enthalpies with a negligible contribution of the enthalpy of protonation from the buffer.

Considering the validation study with seventeen thrombin ligands we are keen to suggest this method as very reliable and robust also to very weak fragments despite the more elaborate experimental procedure.

2.4.3 Overcoming poor solubility of the protein

Poor solubility of proteins often limits the application of the ITC. In case of nanomolar ligands with sufficient enthalphic contributions this fact is usually not important as the required protein concentration can be obtained because only a low amount of protein is needed. In contrast, the characterization of moderate binders (1-10 µM, normally falling into the scope of direct ITC titration) with commonly small enthalpic contributions are often not accessible by direct ITC titrations simply because the high protein concentration required cannot be obtained due to solubility issues.

2. Validation of ITC displacement titration

Figure 7: *a*) The titration curve for the competitive ligand saquinavir (625 µM) in a direct ITC titration at 49 µM endothiapepsin. *b*) Titration curve for a fragment in an ITC displacement titration. The titration of saquinavir (625 µM) used as competitive ligand into a solution of endothiapepsin (49 µM) in the presence of fragment **21** (20 mM).

a) *b*)

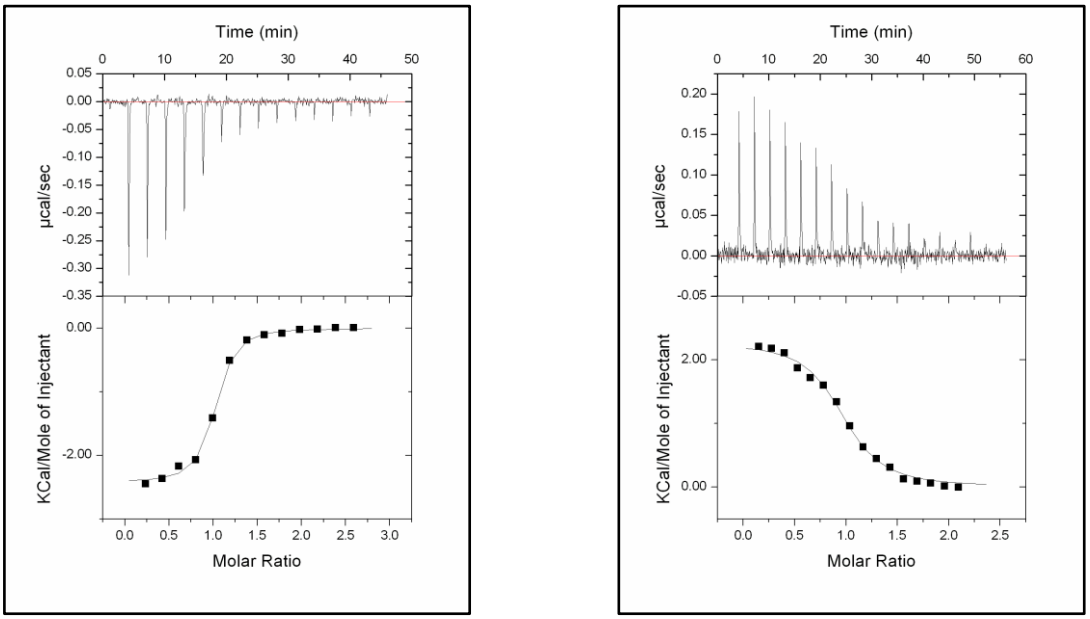
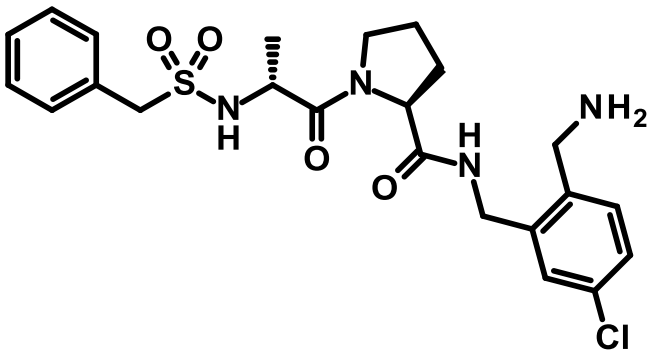
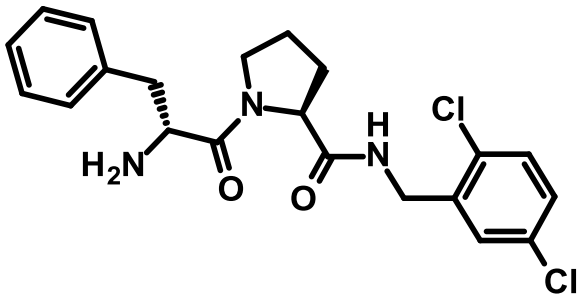
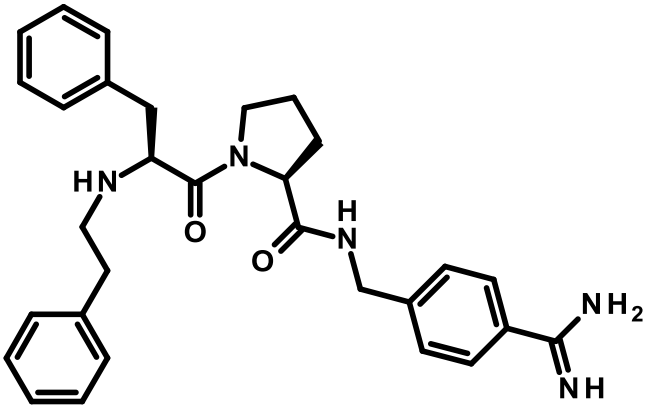
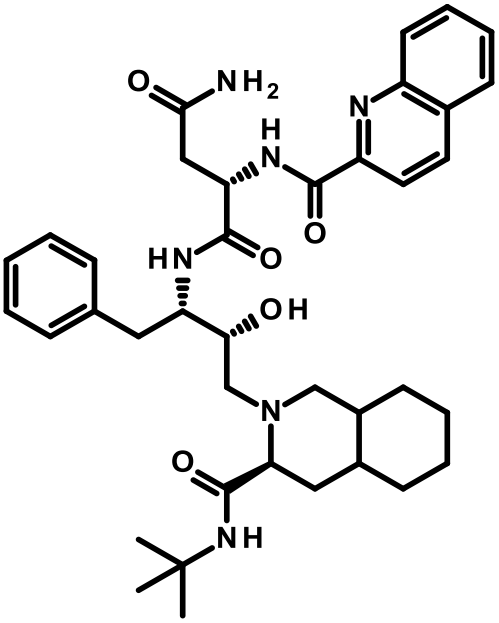
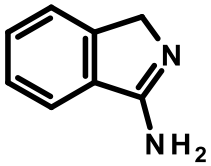
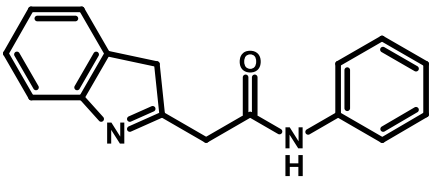
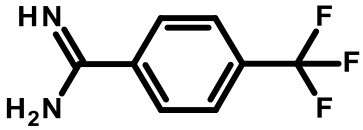


Table 2: Chemical structures of the ligands are shown with naming used in the text.

Ligand name	Chemical structure
17	
18	

19	
20 (Saquinavir)	
21	
22	
23	

2. Validation of ITC displacement titration

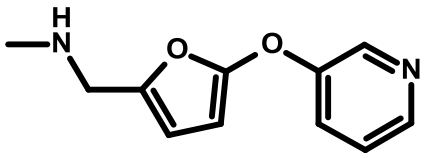
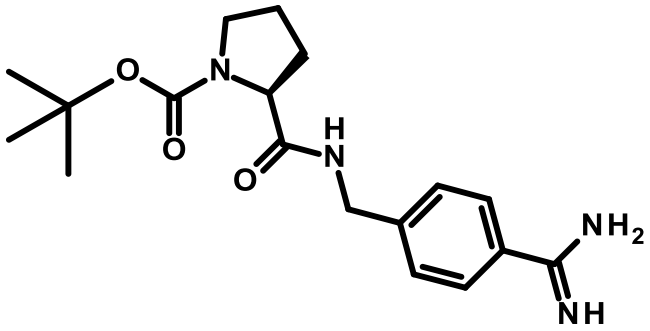
24	
25	

Table 3: Thermodynamic data of fragment binding to endothiapepsin.

	ΔG (kJ/mol)	ΔH (kJ/mol)	-TΔS (kJ/mol)
20 (Saquinavir)[a)]	- 35.6 ± 1.0	- 11.6 ± 1.3	- 24.0 ± 2.3
21[b)]	- 12.0 ± 2.2	- 28.7 ± 0.5	16.7 ± 1.8
22[b)]	- 18.1 ± 0.8	- 35.4 ± 0.8	17.3 ± 0.1
23[b)]	- 17.7 ± 0.1	- 28.9 ± 0.3	11.9 ± 0.4
24[b)]	- 19.1 ± 0.6	- 33.5 ± 1.9	14.3 ± 2.6

[a)]Direct ITC determination of the competitive ligand saquinavir to endothiapepsin in acetate buffer.
[b)]The binding profiles of the fragments studied by competitive ITC titration using saquinavir as displacing ligand.

We have chosen a thrombin ligand with a dissociation constant of K_d = 1.4 µM to demonstrate that also the low micromolar range can be addressed by the displacement titration when solubility or availability of protein material are crucial factors.

Prior to the displacement experiments the test ligand **25** was directly titrated to a 20 µmol/L thrombin solution to determine its binding parameters (K_d = 1.4 ± 0.3 µM and ΔH = - 25.3 ± 0.6 kJ/mol). In a comparable displacement titration protein concentration was lowered threefold (6 µmol/L). The competitive reporter ligand **17** at a concentration of 250 µmol/L was titrated into the protein solution saturated with **25** (1.25 mmol/L). The comparison of the affinity and enthalpy between direct and displacement approach (K_d = 7.4 ± 1.0 µM and ΔH = - 25.8 ± 0.3 kJ/mol) illustrates a good match of the binding parameters.

2.5 Discussion

Despite its potential, thermodynamic analysis of fragments is not yet applied widely in fragment-based lead discovery (FBLD). This fact does not result from limited relevance of such data but from drawbacks that are faced once working with low affinity systems. A recent review by Edink *et al.*[14] summarizes some successful examples for the application of thermodynamic data to support the optimization process in fragment-based drug discovery. Interestingly, most of the binding signatures for fragments in FBLD described in literature were collected for fragments showing already high binding affinities.[20] In contrast it is well known that fragments identified in early stages of FBLD exhibit very low affinities. Especially the use of direct ITC titrations to profile the first discovered hits is restrictive either due to huge protein consumption or insufficiently accurate binding parameter determination. Binding parameters obtained from van't Hoff analysis of, e.g. SPR biosensor data at different temperatures often show poor correlation with ITC binding signatures.[21,22,23] Clearly both techniques evaluate different aspects of binding under different conditions. However, the evaluation of thermodynamic phenomena with restricted number of data points and at a narrow temperature interval makes the data analysis and interpretation by far more complicated in case of SPR measurements. In contrast ITC directly determines enthalpy under thermodynamic equilibrium conditions at one temperature which rules out all temperature-dependent effects superimposed by changes of the complex protein-ligand-water system. Therefore, we believe van't Hoff data are less appropriate for a comprehensive thermodynamic analysis of fragments.

As presented here, ITC displacement titrations are a promising alternative in this context to characterize low-affinity systems without applying unreasonable high protein concentration or immobilizing the binding partner as it is done for SPR measurements. If the described strategy is followed highly reliable binding parameters can be obtained. Once a competitive ligand is found the whole characterization requires approximately 1.5 h including all sample preparations and the amount of protein can be lowered at least fourfold compared to a direct titration.

Especially the accurate factorization into enthalpic and entropic contributions to fragment binding by ITC provides valuable information to understand and improve binding of fragments to their target molecule. Once a series of fragments has been characterized by the ITC displacement approach, the obtained thermodynamic profiles will help to decide at very

early stages of drug discovery which fragments to carry on to further lead optimization. Choosing enthalpic binders as most promising candidates was suggested as a propitious strategy. Starting with initial enthalpic hits appears advisable as most medchem-optimization protocols more easily optimize the entropy contribution to binding. This fact is already apparent for the four fragments characterized here for endothiapepsin binding. Even though the much larger and better active-site complementary saquinavir is more potent, its enthalpic binding contribution is significantly smaller than for the four fragment hits. Saquinavir wins entropically whereas the four fragments show entropic contributions still rather detrimental to binding. Thermodynamic profiling planned in parallel to fragment-to-lead optimization of the discovered hits will help us in the future to better define the most appropriate optimization strategies.

2.6 Acknowledgments

We kindly acknowledge CSL Behring, Marburg, for supplying us with generous amounts of human thrombin from the production of Beriplast®. We thank the group of Prof. Dr. Steinmetzer and Prof. Dr. Hangauer for the synthesis of the thrombin compounds. This work was supported by the Bundesministerium für Bildung und Forschung (BMBF, Förderkennzeichen 0315161C) and the ERC Grant DrugProfilBind 268145 from the European Research Council.

2.7 Materials & Methods

2.7.1 Materials

ITC experiments were performed using an ITC200™ system from Microcal (now part of GE Healthcare), Northhampton, MA, USA. Thrombin was used from Beriplast® (CSL Behring, Marburg, Germany). Endothiapepsin was purified from Suparen® (provided by DSM food specialities) by exchanging the buffer to 0.1 M acetate buffer pH 4.6 using a Vivaspin 20 with a molecular weight cut off at 10 kDa. The protein concentration was measured by absorbance at 280 nm assuming an extinction coefficient of 1.15 for 1 mg/ml solutions.[24] Thrombin ligands from series A were synthesized in the group of Prof. Hangauer (Department of Chemistry, University of Buffalo, USA). Thrombin ligands from series B were synthesized in the group of Prof. Steinmetzer (Department of Pharmaceutical Chemistry,

Philipps-University Marburg, Germany). Saquinavir mesylate is purchased from Roche (Basel, Switzerland)

2.7.2 Bioassay

Kinetic inhibition of human thrombin (from Beriplast®, CSL Behring, Marburg, Germany) was determined photometrically for ligands with $K_i \geq 1nM$ at 405 nm using the chromogenic substrate Pefachrom tPa (LoxoGmbH, Dossenheim, Germany) as described[25] under the following conditions: 50 mM Tris–HCl, pH 7.4, 154 mM NaCl, 5 % DMSO, 0.1 % polyethylene glycol 8000 at 25°C using different concentrations of substrate and inhibitor. K_i ($n \geq 3$) were determined as described.[26]

Thrombin inhibitors with a $K_i \leq 1nM$ were characterized by a fluorogenic assay. Kinetic data were obtained using Tos-Gly-Pro-Arg-AMC[27] (tosyl-Gly-Pro-Arg-aminomethylcoumarin) as the fluorogenic substrate with a Safire II platereader (Tecan, Schweiz, *ex* = 380 nm, *em* = 460 nm). This substrate allows us to work at a protein concentration of 0.050 nM which is sufficiently lower than the lowest applied ligand concentration (0.780 nM). The experimental buffer contained 50 mM Tris-HCL, 154 mM NaCl, 0.1 % polyethylene glycol 8000 and 5 % DMSO at pH 7.4. The K_m of the substrate (1.1 ± 0.4 µM) was measured at ten different substrate concentrations and the resulting curve was analyzed using GraFit 4 software.[28] Cleavage of the substrate was measured by monitoring the change in fluorescence over a dilution series of at least ten inhibitor concentrations (500 nM – 0.780 nM) at 5 µM substrate (S) over 600 sec. The fluoroscence signal was plotted against time and after linear regression the reaction rates (v) were calculated. The latter values (v) were plotted against the respective inhibitor concentration (I) and the resulting curve was fitted with ORIGIN software using equation 18. All measurements were performed at least in triplicate.

$$\boldsymbol{v} = \frac{\boldsymbol{V_{max}} \cdot \boldsymbol{S}}{\boldsymbol{K_m} \cdot \left(\mathbf{1} + \frac{\boldsymbol{I}}{\boldsymbol{K_i}}\right) + \boldsymbol{S}} \qquad \text{Equation 18}$$

2.7.3 Experimental conditions for ITC titrations

Thrombin was freshly prepared for each experiment by dialysis of a thrombin sample in the buffer used for titration experiments (50 mM Hepes (Series A) or Tris (Series B), 100 mM NaCl, 1 % PEG8000, pH 7.8). After the dialysis the weak-binding ligand (Series A) was added

from a stock solution (50 mM, 100% DMSO) to the solution of thrombin (27 µM) in order to obtain the required ligand concentration for complete saturation (1.25 - 1.5 mM). The final DMSO concentration was subsequently adjusted to 3 %. Either **17** or **18** (500 – 625 µM) was filled in the syringe and then titrated into the stirred sample cell (200 µL) containing the thrombin solution (27 µM) saturated with the weak-binding ligand under investigation. The following analysis of the displacement titration curve is described in the results section.

The binding parameters of the strong binders (Series B) were determined performing two titrations in tris buffer (50 mM tris, 100 mM NaCl, 1% PEG8000, pH 7.8). First either **18** or **19** (500 µM) was directly titrated into a protein solution at 27 µM. As small peaks of dilution indicate complete inhibition of the protein, the titration was stopped and the resulting titration curve was analyzed by fitting a single-site-binding isotherm that yields ΔH^0 (enthalpy of binding) and K_D (dissociation constant). Afterwards the syringe was thoroughly cleaned and filled with a solution containing the strong binder (500 µM). This syringe was placed into the sample cell now filled with the protein-ligand complex from the first titration. In the following second titration the strong binder displaced step-by-step the weak-binding ligand. The analysis of the displacement titration was performed using a binding competition model. This fitting function is integrated in Origin 7.0 and is called *Competitive Binding*.

Endothiapepsin was diluted from a stock solution (147 µM) to the final concentration (49 µM) with the experimental buffer composed of 20 mM acetate, 6 % DMSO, pH 4.6 and the fragment at 6 - 20 mM to ensure sufficient saturation of the protein. Saquinavir was prepared as a stock solution in 100 % DMSO at 25 mM. Final concentration of saquinavir (625 µM) was achieved by diluting the stock in the experimental buffer containing the weak binder at 6 - 20 mM to avoid large peaks of dilutions during the titration. Prior to the displacement titration saquinavir (625 µM) was characterized by a direct titration at 49 µM Endothiapepsin.

All ITC experiments were started at 25 °C after a stable baseline had been achieved. The experimental design comprises an initial ligand injection of 0.3 µL followed by 15 injections of 1.1 - 1.4 µl with a 300 s interval between each injection.

All measurements were performed in duplicate. Raw data were collected and the area under each peak was integrated, followed by correction for heats of dilution and mixing by subtracting the final baseline consisting of small peaks of the same size to zero. The initial

data point was deleted from the integrated data because this injection usually reflects an erroneous amount of heat due to the possible exchange of liquids between syringe and cell when inserting the syringe into the calorimetric cell and the backlash error in the motorized screw mechanism in the injector.[29]

2.7.4 Excel file

The excel file contains two sheets covering the calculation of the required ligand concentration for complete inhibition of the protein and the estimated error in K_d.

The sheet for estimation of the required ligand concentration for a certain degree of inhibition is called saturation. The derivation of the applied equation can be found in the results section.

Concerning the error estimates of $K_{d\ WL}$ we calculated the observed K_d of the strong-binding ligand in a displacement mode if the titration would have been conducted with a strong binder showing 100 nM affinity to characterize a weak-binding ligand with a actual affinity of 500 µM in presence of 10 mmol/L of the weak binder. The observed K_D of the strong binder would be 2.1 µM in this displacement titration. Then we calculated the affinity for the weak binder with affinities for the strong-binding ligand which deviate systematically from the actual 100 nM affinity (101, 102, 103 nM). So we artificially introduced an error in the determination of K_d of the strong binder. Subsequently, the deviation is calculated between the actual K_d of the weak-binding ligand (500 µM) and the K_d's of the weak-binding ligand derived from the calculation of false affinities for the strong binder.

2.8 References

1 Gribbon P. & Andreas S. (2005) High-throughput drug discovery: what can we expect from HTS? *Drug Discov. Today* **10**, 17–22.

2 Hopkins A.L., Groom C.R. & Alex A. (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **9**, 430-431.

3 Abad-Zapatero C. & Metz J.T. (2005) Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* **10**, 464–469.

4 Schuffenhauer A., Ruedisser S., Marzinzik A.L., Jahnke W., Blommers M. *et al.* (2005) Library design for fragment based screening. *Curr.Top. Med. Chem.* **5**, 751–762.

5 Ladbury J.E., Klebe G. & Freire E. (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat. Rev. Drug Discov.* **9**, 23-27.

6 Velazquez-Campoy A., Todd M.J. & Freire E. (2000) HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry.* **39**, 2201-2207.

7 Freire E. (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discovery Today* **13**, 869-874.

8 Sigurskjold B.W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* **277**, 260-266.

9 Velazquez-Campoy A., Kiso Y. & Freire E. (2001) The binding energetics of first- and second-generation HIV-1 protease inhibitors: Implications for drug design. *Arch. Biochem. Biophys.* **390**, 169-175.

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

11 Zhang Y.L. & Zhang Z.Y. (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.

12 Bradshaw J.M., Mitaxov V. & Waksman G. (1999) Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. *J. Mol. Biol.* **293**, 971-985.

13 Velázquez Campoy A. & Freire E. (2005) ITC in the post-genomic era...? Priceless. *Biophys. Chem.* **115**, 115-124.

14 Edink E., Jansen C., Leurs R. & de Esch I.J.P. (2010) The heat is on: thermodynamic analysis in fragment-based drug discovery. *Drug Discovery Today: Technologies.* **7**, 189-201.

15 Wiseman T., Willistou S. & Brandts S.F. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137.

16 Biela A., Khyat M., Tan H., Kong J., Heine A. *et al.* (2012) Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin. In preparation.

17 Biela A., Sielaff F., Heine A., Steinmetzer T. & Klebe G. (2012) Enthalpic and entropic changes caused by a stepwise disruption of a water network in the S3/4 subsite of thrombin: An example of a classical hydrophobic effect. In preparation.

18 Köster H., Craan T., Brass S., Herhaus C., Zentgraf M. *et al.* (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.

19 Fukada H. & Takahashi K. (1998) Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. *Proteins* **33**, 159-166.

20 Scott A.D., Phillips C., Alex A., Flocco M., Bent A. *et al.* (2009) Thermodynamic optimisation in drug discovery: a case study using carbonic anhydrase inhibitors. *ChemMedChem* **4**, 1985-1989.

21 Chaires J.B. (1997) Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. *Biophys. Chem.* **64**, 15–23.

22 Horn J.R., Russell D., Lewis E.A. & Murphy K.P. (2001) Van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: are there significant discrepancies? *Biochemistry* **40**, 1774–1778.

23 Deinum J., Gustavsson L., Gyzander E., Kullman-Magnusson M., Edström A. *et al.* (2002) A thermodynamic characterization of the binding of thrombin inhibitors to human thrombin, combining biosensor technology, stopped-flow spectrophotometry, and microcalorimetry. *Anal Biochem.* **300**, 152-162.

24 Larson M.K. & Whitaker J.R. (1970) Endothia parasitica Protease. Parameters Affecting Stability of the Rennin-like Enzyme. *Journal of Dairy Science*, **53**, 262-269.

25 Stürzebecher J., Stürzebecher U., Vieweg H., Wagner G., Hauptmann J. *et al.* (1989) Synthetic inhibitors of bovine factor Xa and thrombin comparison of their anticoagulant efficiency. *Thromb. Res.* **54**, 245–252.

26 Dixon M. (1972) The graphical determination of K_m and K_i. *Biochem. J.* **129**, 197–202.

27 Bernett M.J., Blaber S.I., Scarisbrick I.A., Dhanarajan P., Thompson S.M. *et al.* (2002) Crystal structure and biochemical characterization of human kallikrein 6 reveals that a trypsin-like kallikrein is expressed in the central nervous system. *J. Biol. Chem.* **277**, 24562-24570.

28 Leatherbarrow R.J. (1998) GraFit Version 4 4.0 edit, Erithacus Software Limited, Staines, UK.

29 Mizoue L.S. & Tellinghuisen J. (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal. Biochem.* **326**, 125-127.

3 Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin

3.1 Introductory remarks

This study was done in cooperation with the group of Prof. Dr. Hangauer (University of Buffalo). The following text is submitted to the scientific journal *Journal of Molecular Biology*. Maan Khyat from the Hangauer group is included as second author.

3.2 Abstract

In the present study we investigate the impact of a tightly bound water molecule on ligand binding in the S1 pocket of thrombin. The S1 pocket contains a deeply buried deprotonated aspartate residue (Asp189) which is, due to its charged state, well hydrated in the uncomplexed state. We systematically studied the importance of this water molecule by evaluating a series of ligands that contains a pyridine-type P1 side chains that could potentially alter the binding properties of this water molecule. All of the pyridine derivatives retain the original hydration state albeit sometimes with a slight perturbance. In order to prevent a direct H-bond formation with Asp189, and to create a permanent positive charge on the P1 side chain that is positioned adjacent to the Asp189 carboxylate anion, the pyridine nitrogen was methylated. This methylation resulted in displacement of water but was accompanied by a loss in binding affinity. Quantum chemical calculations of the ligand solvation free energy showed that the positively charged methylpyridinium derivatives suffer a large penalty of desolvation upon binding. Consequently, they have a substantially less favorable enthalpy of binding. In addition to the ligand desolvation penalty the hydration shell around Asp189 has to be overcome which is achieved in nearly all pyridinium derivatives. Only for the ortho derivative is a partial population of a water next to Asp189 found. Possibly the gain of electrostatic interactions between the charged P1 side chain and Asp189 helps to compensate for the desolvation penalty. In all uncharged pyridine derivatives the solvation shell remains next to Asp189 partly mediating interactions between

ligand and protein. In the case of the para pyridine derivative a strongly disordered cluster of water sites is observed between ligand and Asp189.

3.3 Introduction

Water molecules play a crucial role in the mutual recognition of proteins and ligands during the binding process. For better understanding, the effects produced by water can be split into two parts: first the impact of water on the formed complex and second the influence of water on ligand and protein prior to complex formation, commonly summarized as solvation/desolvation properties. The first part is mainly characterized by the ability of water molecules to mediate interactions between target protein and ligand in the formed complex. The water incorporation is thought to be a promising concept to enhance binding affinity, as mediated through water molecules, the binding site can be extended and additional hydrogen bonds can be formed.[1] Unfortunately, up to now there are only very few cases reported in literature that clearly demonstrate under which conditions the incorporation of a water molecule into the protein-ligand interface results in a gain of Gibbs free energy of binding. In the case of binding of L-arabinose to L-arabinose binding protein (ABP) the incorporation of a water molecule is associated with an increase in binding affinity.[2,3,4] The picked-up water molecule clearly determines the specificity of the latter sugar over D-galactose wherein an additional -CH_2OH group fills the space occupied by the water molecule in the other complex and therefore explains the gain in binding affinity. Indeed, most of the reported water-mediated interactions seem to be formed to stabilize the protein-ligand complex by establishing bridges to the protein site.[5,6,7] However, usually no conclusions can be drawn about the real benefit of these water interactions as the studies lack the comparison with the corresponding water-free reference states. Moreover, it is generally believed that the displacement of the bound water molecules leads to an improved binding affinity[8], but on the other hand impressive cases have been reported where no gain in binding affinity could be observed upon water displacement.[9] This puzzling situation results from our currently rather rudimentary understanding of the influence of water molecules on ligand binding. It is widely assumed that trapping a water molecule in a binding pocket is entropically unfavorable and enthalpically favorable as water can form multiple interactions with either protein and bound ligand, thereby losing many degrees of freedom. The desired overall increase in binding affinity will only be achieved if

the entropic penalty is overcompensated by a large enthalpic benefit resulting from the formation of strong additional hydrogen bonds. Accordingly, it appears difficult to predict whether the incorporation of a water molecule at a given interface will enhance the enthalpic term enough to attain an improvement in free energy of binding. The ability of the water molecule to act simultaneously as a hydrogen-bond donor and acceptor at the interface of a protein-ligand complex, its strong tendency to form ordered structures, and its substantial contribution to heat capacity changes, are some characteristics that account for the complexity of the water problem. Aside from water's impact on the ligand bound state, one should also consider the situation of the binding partners in their solvated state prior to complex formation. Both, the protein ligand binding cavity and the ligand are solvated and thus energy is needed to remove their water shells to allow the transfer of the ligand to the protein environment.[10] The computational prediction of solvation free energies of small molecules is still an insufficiently resolved challenge in computational chemistry and modestly predictive models require computationally demanding quantumchemical or molecular dynamics calculations.[11,12] As a result they are not appropriate for the screening of large compound libraries. Moreover, the correlation between predicted and experimentally derived values is not really convincing, it is only applicable to rather small uncharged molecules and a meaningful factorization into enthalpy and entropy contributions is not possible.[13,14] In contrast the experimental approach based on vapor phase-to-water partition experiments in order to obtain data on the ligand desolvation process is promising but unfortunately limited to non-ionic and volatile molecules.[15] The enthalpic and entropic contributions for desolvation of the protein binding site are even more difficult to determine experimentally. They can be computed indirectly when all other binding contributions such as net protein-ligand interactions ($\Delta G°_{PL}$), ligand desolvation ($\Delta G°_{solvL}$) and the sum of all remaining contributions ($\Delta G°_{obs}$) are factorized into the enthalpic and entropic terms.[16] It is obvious that the overall deconvolution into the individual terms of the binding process is a complex task and can only be addressed by a combination of theoretical and experimental methods.

As the complete deconvolution is hardly tractable, we tried to reduce the complexity of the binding process. We will focus on terms which can be addressed experimentally by Isothermal Titration Calorimetry (ITC) across a congeneric series of closely related ligands. These ITC accessible binding terms include $\Delta G°_{obs}$, $\Delta H°_{obs}$ and $-T\Delta S°_{obs}$. Even though these

binding parameters describe the overall contribution to the entire binding event, the thermodynamic and crystallographic examination within a narrow series of systematically varied ligands revealed insights, ranging from enthalpy-entropy compensation[17], hydrohobic interactions[18] to cooperativity effects[19,20]. In this contribution, thrombin was chosen as a model system to study the influence of water. The S1 pocket of thrombin hosts a negatively charged Asp189 at its bottom and is useful to investigate the influence of water molecules on the binding process. Previously, we studied substituted P1 benzyl derivatives and no additional water molecule near Asp189 could be found.[21] Here, we expand our study by introducing polar nitrogen atoms into the aromatic benzyl moiety at different positions (*o-, m-* and *p*-position, see Table 1) to make the ligand competent for putative water mediated contacts with the protein, however, keeping the spacial requirements of the P1 moiety virtually unchanged. A subsequent methylation of the introduced pyridine nitrogen results in the creation of a permanent charge supposed to displace potentially picked-up water molecules as a methyl group can approximately occupy the space of a water molecule potentially interacting with the pyridine nitrogen. High resolution crystal structures tracing the displacement of water molecules from the binding pockets are investigated along with thermodynamic data to disentangle the different structural and energetic contributions to the binding event.

3.4 Results & Discussions

3.4.1 Analysis of the crystal structures and binding constants

The crystal structures of 10 tripeptide-like thrombin inhibitors with D-Phe-Pro-XXX scaffold (**2-11**, Table 1) in complex with human α-thrombin have been determined with medium to high resolution (1.90-1.27 Å). The electron density for the D-Phe-Pro portion is well defined in all examples as exemplarily shown in Figure 1 for **2**. Among the series there were no structural differences in the binding mode of the D-Phe-Pro portion. The benzyl moiety of the D-Phe portion occupies the hydrophobic S3/4 pocket and the five-membered ring of proline is well accommodated in the S2 pocket which is formed by Tyr60A and Trp60D of the 60's loop (see Figure 2 for a schematic view). Aside from hydrophobic binding features a β-sheet-like H-bond interaction is observed with Gly216 of the protein backbone. The terminal amino group and the carbonyl functionality of the inhibitor interact within a range of 2.7 - 3.1 Å with the corresponding atoms of Gly216. Superposition of all studied complexes does

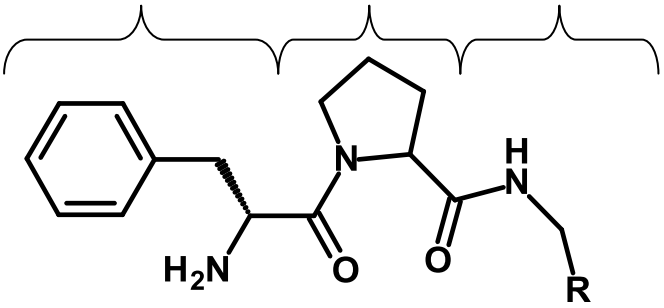
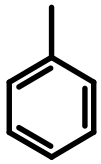
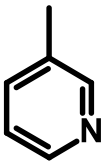
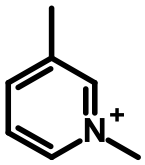
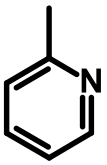
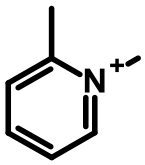
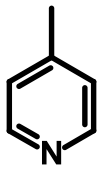
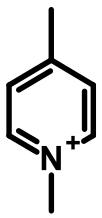
not show any significant structural differences of adjacent protein binding site residues except Glu192 at the entrance of the S1 pocket. This amino acid is observed in several orientations which are apparently not induced by the inhibitor. This is in agreement with a previously performed alignment of thrombin crystal structures from the PDB[22] (Protein Data Bank), showing the remarkable flexibility of this glutamate[23]. Despite marginal changes our structures exhibit only pronounced differences next to the modified P1 substituent.

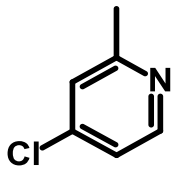
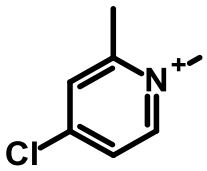
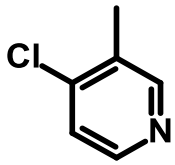
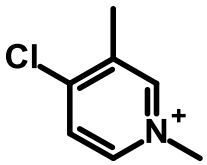
The previously studied derivative **1** with an unsubstituted benzyl group (PDB code 2ZFF[21]) will be used as reference. It shows no water molecule inbetween the terminal phenyl ring and Asp189. However, the indicated enhanced residual mobility of the aromatic portion and the entropically favored binding profile of **1** suggests some available space that could potentially accommodate an additional water molecule, provided the bound inhibitor is modified with an appropriate polar group. Accordingly, we conceived the current study of a series of pyridine derivatives.

The nitrogen atom introduced in the meta position provides the pyridine derivative **2.** It binds water-mediated to the negatively charged aspartate at the bottom of the S1 pocket (Figure 3a). This crystallographically well-defined water molecule is tetra-coordinated by four neighbouring atoms and thus can be considered as an optimal mediator of hydrogen bonds. The short distances of 2.7 Å to the pyridine nitrogen, 2.9 Å to Asp189 and 2.6 Å to a further water molecule indicate, at first glance, a favorable position of this water molecule. All interactions to this water molecule are formed in a plane. The resulting water coordination does not form a short fourth interaction to the adjacent water molecule next to Tyr228 (4.1 Å). Interestingly, the aromatic moiety of the unsubstituted benzyl group of **1** is accommodated 1.1 Å closer to Asp189 compared to **2**. Obviously, this displacement does not allow the complex with **1** to pick-up a water molecule as seen with **2**. Comparing the binding affinities of both inhibitors suggests a slightly enhanced overall binding upon release of the water molecule (K_i = 11.2 ± 6.7 µM for **1** and 33.4 ± 19.0 µM for **2**) from the complex with **1**. The corresponding N-methylated pyridinium derivative **3** displaces the water molecule picked-up by **2** (Figure 3b). The positively charged nitrogen is 4.8 Å away from the closest oxygen of Asp189. Most likely, the bulky interstitial methyl group shields the electrostatic attraction of the negatively charged Asp189 and prevents a shorter distance due to steric repulsion.

3. Impact of ligand and protein desolvation

Table 1: Chemical structures of the studied ligands together with the kinetically determined inhibition contants (K_i in µM) and the standard deviation[(a)] towards human thrombin.

<table>
<tr><td colspan="4">D-Phe for S3/4, Pro for S2, P1 variations for S1
  </td></tr>
<tr><td>1
  </td><td>11.2 ± 6.7[(21)]</td><td></td><td></td></tr>
<tr><td>2
  </td><td>33.4 ± 19.0</td><td>3
  </td><td>74.8 ± 20.5</td></tr>
<tr><td>4
  </td><td>69.3 ± 0.1</td><td>5
  </td><td>132.2 ± 35.3</td></tr>
<tr><td>6
  </td><td>64.0 ± 20.0</td><td>7
  </td><td>22.5 ± 13.2</td></tr>
</table>

8 	**1.9 ± 0.8**	**9** 	**63.5 ± 13.6**
10 	**2.6 ± 1.1**	**11** 	**55.4 ± 21.4**

[a)]The error is given as the standard deviation calculated from at least three measurements (in µM).

Figure 1: Inhibitor **2** in complex with human thrombin. View of the active site with the solvent-accessible surface of thrombin in grey.

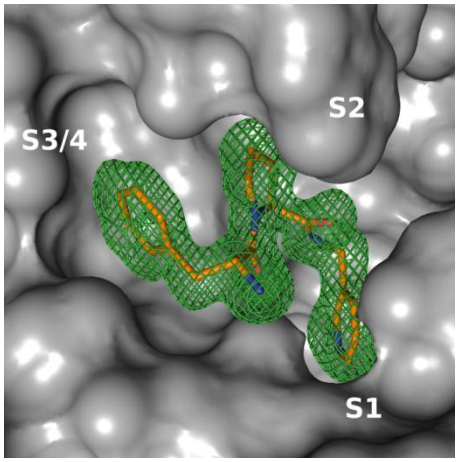
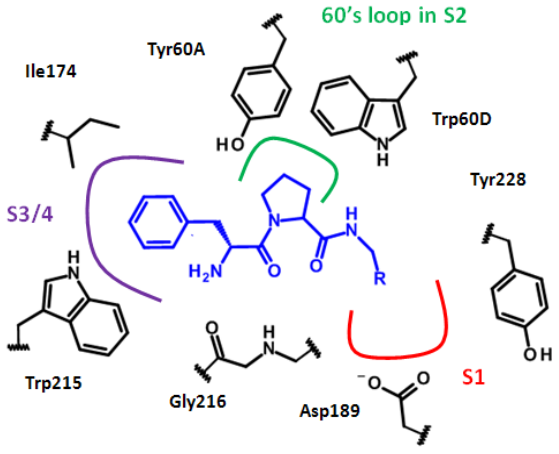


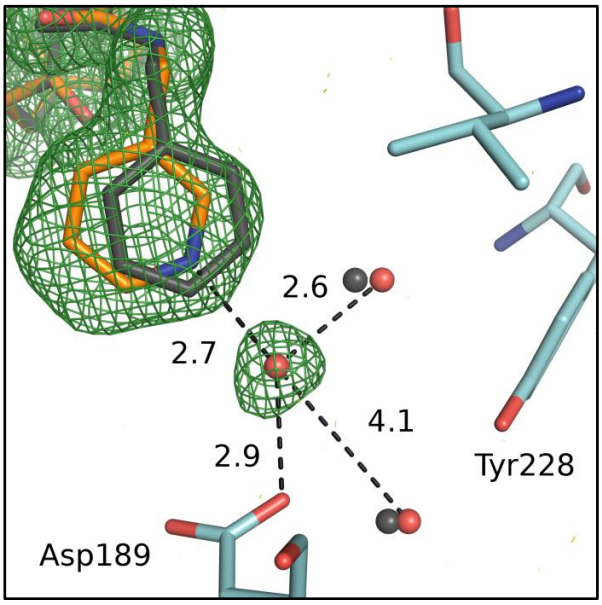
Figure 2: A schematic view of the binding pocket of thrombin in complex with the studied ligand scaffold in blue. R represents the modifications of the S1 occupant (see Table 1).



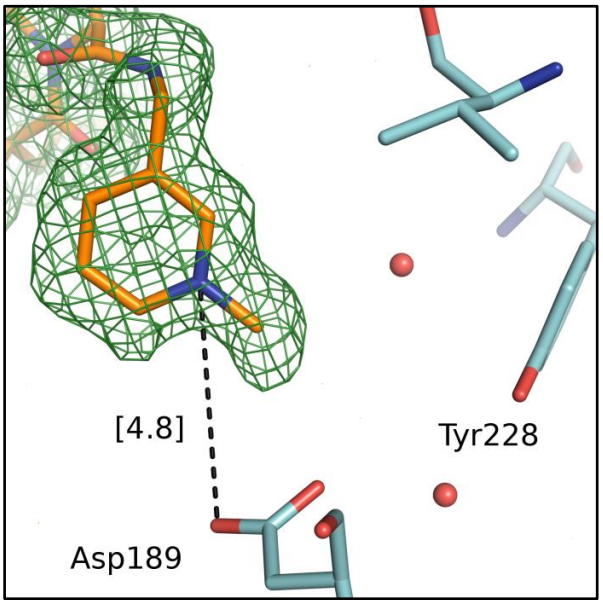
3. Impact of ligand and protein desolvation

Figure 3: Binding modes of the studied ligands in the S1 pocket. The F_o-F_c difference electron density is shown in green at 2σ for inhibitor **2** (*a*), **3** (*b*), **4** (*c*), **5** (*d*), **6** (*e*), **7** (*f*), **8** (*g*), **9** (*h*), **10** (*i*) and **11** (*j*) together with the Asp189 bound water molecule if present. The benzyl derivative **1** (black) is superimposed in *a*, *c* and *e* to emphasize the impact of the introduced nitrogen. Favourable interactions with the corresponding distances in Å are depicted with broken lines. Distances in red indicate very short contacts. Values in brackets indicate short distances found for the specific atom which are mentioned in the text. Nitrogen is shown in blue, chlorine in green, oxygen in red and carbon in orange (ligand site) and cyan (protein site).

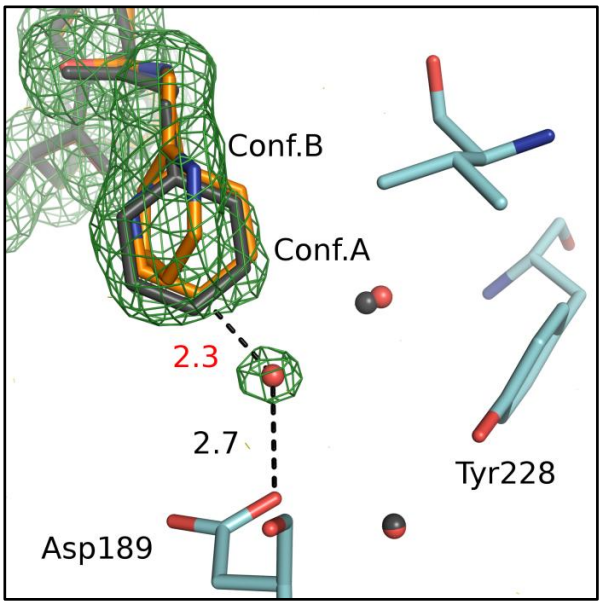
a) in complex with **2**



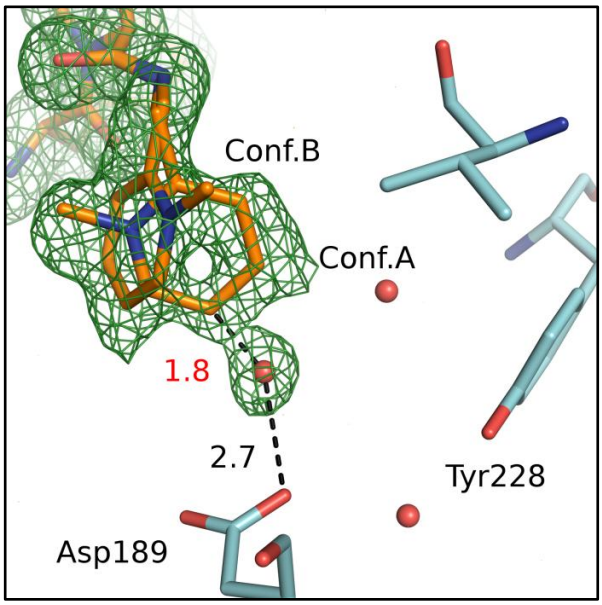
b) in complex with **3**



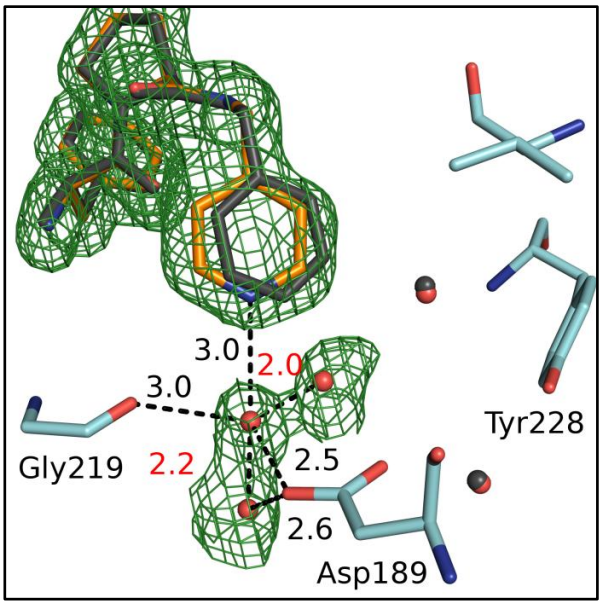
c) in complex with **4**



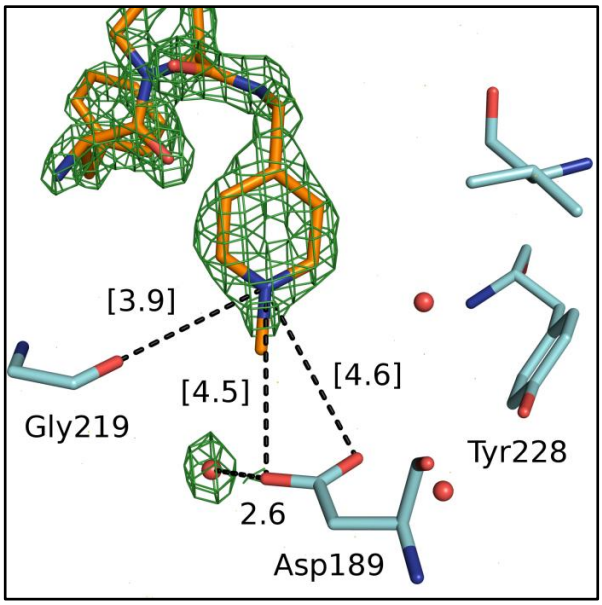
d) in complex with **5**



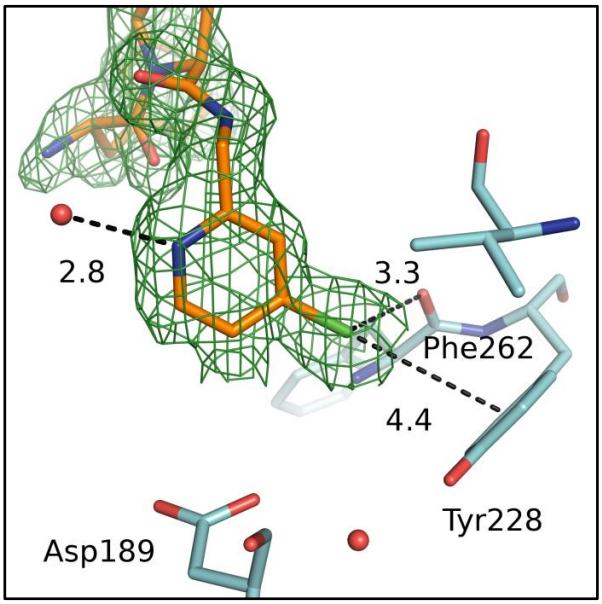
e) in complex with **6**



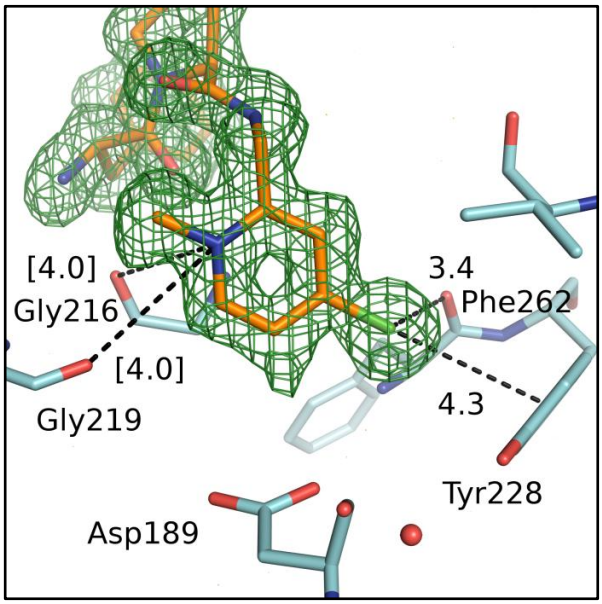
f) in complex with **7**



g) in complex with **8**

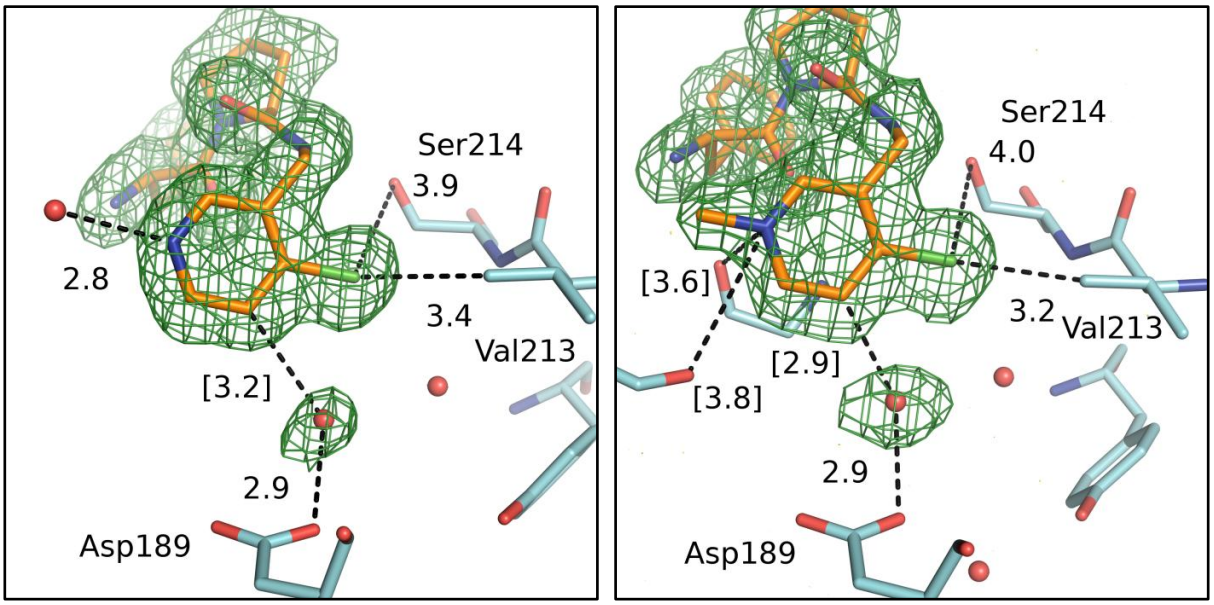


h) in complex with **9**



i) in complex with **10**

j) in complex with **11**



This suboptimal location of the positively charged nitrogen in the methylpyridinium portion results in a loss of binding affinity by a factor of 2 (K_i = 74.8 ± 20.5 µM for **3**). Supposedly, it is partly compensated by the favorable displacement of the water molecule found in the complex with **2** and the uncomplexed protein.

Interestingly, the binding mode is different when the pyridine nitrogen is placed in ortho position in **4**. The pyridine ring now adopts two alternative conformations which could be refined to 39 % occupancy for conformation A and 61 % for B (Figure 3c). Remarkably, at the above-described position in **2** again a water molecule is found close to Asp189. With respect to conformation A it would clash (2.3 Å) with a ring carbon atom of the ligand. In addition some negative electron density appeared around the water position when refined with 100 % occupancy. The twisted conformation B provides sufficient space to accommodate a water molecule. We therefore refined the water molecule with the same occupancy as the inhibitor in the twisted conformation B and a negative electron density disappeared. Obviously, the water molecule can only be present if the pyridyl moiety adopts the twisted conformation B. In conformation A, this moiety repels the water molecule. Interestingly enough, the latter conformation A corresponds to the geometry found for the benzyl portion in **1**. The pyridine nitrogen does not find a favorable interaction partner, neither in conformation A nor in B. The occurance of the second tilted conformer B cannot be explained by a favorable interaction with the protein as the closest contact to a

neighbouring protein atom is 3.7 Å (Ser195). The coplanar, stacking-type interaction observed in conformation A of aromatic moieties seems to be a general feature in the S1 pocket of trypsin-like serine proteases[24]. Most likely it renders an energetically favorable contact geometry. We assume that the pronounced desolvation costs of the highly buried aspartate are responsible for the only partially achieved displacement of the water molecule and the concurrent disorder of the P1 portion. As mentioned, the described water position is also occupied in the uncomplexed structure of thrombin (PDB code 2UUF[25]) indicating the favorable solvation of Asp189. The partial water displacement seems to correspond to the best compromise between the entropic costs associated with water trapping and the enthalpic advantage to recover the original solvation shell of Asp189. Furthermore, in conformation B the pyridine nitrogen experiences a long contact distance with unfavorable directionality to form a hydrogen bond to Ser189 and in conformation A virtually no beneficial interactions are experienced. We therefore believe that the ortho nitrogen is unfortunately placed. The low binding affinity of **4** confirms this hypothesis as it exhibits a K_i of only 69.3 ± 0.1 µM.

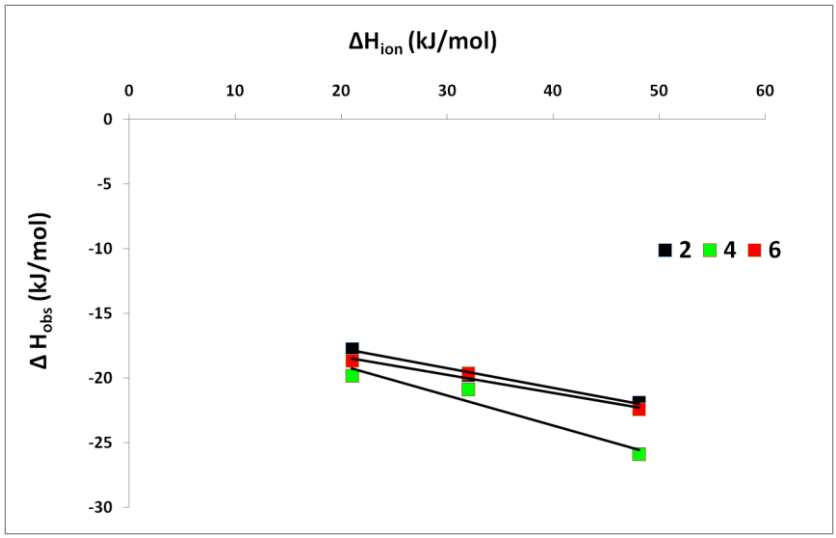
Similarly, the N-methylated ortho pyridinium derivative **5** adopts a binding mode featuring two conformations and a water molecule in very similar positions (Figure 3d) as in case of **4**. The even closer distance of 1.8 Å to one ring carbon in conformation A clearly evidences partial occupancy of the water molecule. The refinement of the occupancy revealed 59 % for conformation A which diplaces the water molecule and 41 % for B which maintains the original solvation of Asp189. Both observed binding modes show the methylated and positively charged nitrogen in an area unable to efficiently interact with Asp189 or one of the negatively polarized S1 residues. The unfavorable placement and the partial release of the water molecule account for the weakest binding affinity (K_i = 132.2 ± 35.3 µM) in this ligand series.

The crystal structure of the para pyridyl derivative **6** reveals a surprising binding feature (Figure 3e). Two well defined dump-bell shaped difference electron densities in the data set resolved to 1.30 Å arrange next to Asp189 which suggests the presence of water molecules scattered over the three sites. In a first refinement model three simultaneously present water molecules were assigned to full occupancy, however, we received very close distances of 2.0 Å and 2.2 Å between the water sites. The resulting distances do not support a model with fully occupied hydration sites. In a second model the three water sites were assigned to

50 % occupancy. A reasonable model with close distances and acceptable contacts to Asp189 (2.6 Å, 2.5 Å and 2.6 Å) could be found. The refinement of this model with a geometrical restraint to keep the close water contacts to 2.1 Å with σ = 0.3 Å resulted in B factors of about 23.0 $Å^2$ which is in agreement with those for the water position in most of our structures. In addition to the short distances to the negatively charged oxygens of Asp189 (2.5 Å) and the backbone carbonyl oxygen of Gly219 (3.0 Å), an interaction with the lone pair of the adjacent pyridine nitrogen can be found (3.0 Å). Protonation of this pyridine nitrogen can be excluded due to our ITC experiments which do not indicate superimposed protonation steps as the recorded heat signal shows no significant (≤ 0.2 mol protons) buffer dependence (Figure 4).

As the distances around the central water site suggest some kind of a distorted pentahedral coordination, occupancy by a sodium ion was also taken into consideration. Unfortunately, water and a sodium ion exhibit similar diffraction power making differentiation impossible. A CSD[26] (Cambridge Structural Database) search on sodium ion contacts performed by Harding[27] revealed a common coordination number of five or six limiting the coordination sphere to 2.7 Å. According to Harding a mean distance between a sodium ion and a neighboring oxygen of 2.42 Å is found in the CSD. A subsequent analysis of protein structures in the PDB revealed a somewhat larger distance of 2.57 Å as mean $Na^+\cdots O$ and 2.58 Å as mean $Na^+\cdots OH_2$ contact distance.[27] These values are not in aggreement with the distances found in the present complex. We therefore exclude the unlikely presence of a sodium ion and assume a binding model with a highly disordered water network near Asp189 to explain the dumb-bell shaped difference electron density. The binding affinity of **6** is rather low with a K_i of 64.0 ± 20.0 µM. Interestingly, the para pyridine derivative **6** does not experience a strong enough pKa shift to adopt a positively charged state forming a direct charge-assisted H-bond with Asp189. Instead an incomplete replacement of the favorable solvation shell of Asp189 is recognized. As a matter of fact this obstacle prevents a strong affinity contribution to complex formation and underlines the importance of the too costly desolvation of Asp189.

Figure 4: The observed enthalpy ΔH_{obs} for binding of **2**, **4** and **6** to thrombin as a function of the ionization enthalpy ΔH_{ion} of the applied three buffers tris, hepes and tricine at pH 7.8. The slope of the linear regression yields the molar ratio of protons captured during the binding reaction.



The N-methylated para-pyridinium analog **7** displaces part of the disordered water network adjacent to Asp189 found in the previous complex (Figure 3f). At one of the water sites full occupancy is observed. The orientation of the ring and the bulky methyl substituent prevent the positive charge at nitrogen to experience a short electrostatic contact to the aspartate as indicated by the long distance of 4.6 Å. We determined an inhibition constant of 22.5 ± 13.2 µM for **7** which displays the strongest binding in the methylpyridinium series. Possibly this enhanced binding reflects the closer distance of the positive charge to Asp189 compared to the complexes with the ortho or meta derivatives **3** and **5**. Nevertheless, none of the methylpyridinium derivatives exhibits a binding affinity even in the range of the highly potent benzamidine analog (K_i = 4 nM[21]) which shows impressively that sole placement of a positive charge into the S1 pocket is not sufficient for strong binding.

In the next step, we introduced a chlorine atom into ligands **2-5** in order to explore the competition between a polar, H-bond acceptor group, a positively charged substituent and a hydrophobic group with respect to binding geometry and achieved affinity. Particularly the m-chlorosubstituted benzyl derivative is known to bind with high affinity (K_i = 180 nM[21]) to the S1 pocket of thrombin experiencing strongly favorable chloro-aromatic contacts with the adjacent Tyr228. The determinant influence of this interaction is underlined by the complex geometries of **8** and **9** which bear a chlorine atom located in meta position and pyridine nitrogen or a N-methylated pyridinium nitrogen in ortho position (Figure 3g and 3h). They

orient their meta-chloro substituent towards Tyr228 to form the chlorine-aromatic interaction. The opposing positive charge in ortho position of **9** is placed similarly to **5** in conformation B. Disorder with a second conformation allowing for partial residual solvation of Asp189 is not observed. Obviously, the energetic benefit of chloro-aromatic interactions over compensates for the costs to fully desolvate Asp189.

The complexes with **8** and **9** should be compared with those of **10** and **11**, where the chloro-substituent is moved to ortho position and the pyridine nitrogen or the N-methylated pyridinium group occupy the meta position (Figure 3i and 3j). Different from the binding modes of **2** and **3** neither the pyridine nitrogen nor the positively charged N-methylated group are placed next to Asp189, instead they are pointing in opposite direction towards the carbonyl backbone groups of Gly216 and Gly219 as similiarly observed in **4** and **5** in conformation A which repels the water molecule adjacent to Asp189. The N-methyl group in **11** occupies virtually the same region as in the topological isomer **8**. Also the contacts to the neighbouring polar groups are of comparable length. The ortho chloro-substituents of **10** and **11** are oriented towards the opposite face, however for stereochemical reasons they cannot establish a short favorable contact to Tyr228. Interestingly enough, both complexes show again the crucial water molecule in the vacant space next to Asp189. The occupancy of these water molecules could be refined to 100 %. With respect to binding affinity, both ligand pairs **8/10** and **9/11** exhibit almost the same values (Table 1). This agrees with the other examples in this series, however a large drop in affinity is experienced relative to the plain meta-chloro benzyl derivative which experiences 180 nM potency[21].

3.4.2 Thermodynamic results and correlation to described structural characteristics

Due to the weak binding of the inhibitors of this series the thermodynamic characterization is challenging as it does not allow direct titrations. In a direct titration, the resulting curves show incomplete typical sigmoidal shape, thus the determination of the thermodynamic parameters will be associated with large errors and the inflection point of such titration curves, which indicates the dissociation constant, cannot be fitted properly. However, in order to obtain thermodynamic data of sufficient accuracy and to reduce consumption of protein material we profiled the ligands by displacement titrations. We followed a procedure recently suggested by us (unpublished). The solution of thrombin is saturated by the ligand

under investigation and subsequently titrated with a stronger binder to displace in a step-wise fashion the weak-binding ligand from the active site. We selected two tight binders (see Table 2 for chemical structures) with different thermodynamic profiles (Table 3) since similar enthalpies of weak and strong binders will result in a very low difference signal. If the heat signal was too low in the displacement titration with the first reference ligand, the experiment was repeated with the second tight binder in order to avoid unfortunate compensating heat effects.

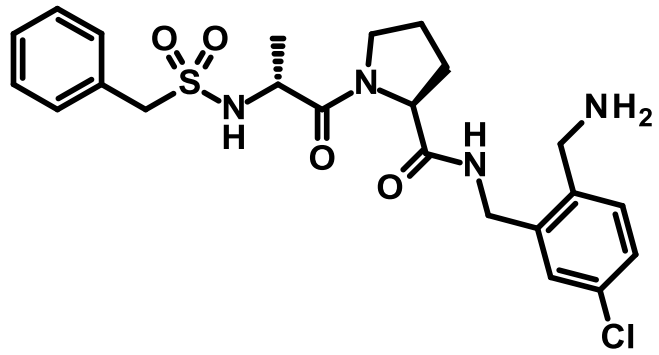
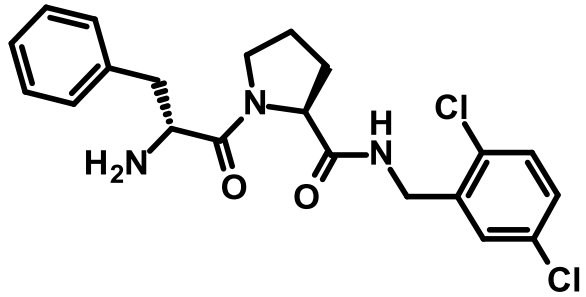
As described below, the thermodynamic results suggest that large costs in desolvation enthalpy are responsible for the weak binding of the N-methylated pyridinium derivatives (see Table 3). The closer the positive charge has been found in the crystal structures to the negatively charged Asp189, the better the price for the high desolvation enthalpy can be compensated. The largest distances are observed for the N-methylated pyridinium derivatives with the positive charge in ortho position. They place it the most remote distance from Asp189. Here we observe a loss in binding enthalpy ($\Delta\Delta H_{8/9}$) of - 7.0 kJ/mol for the chloro substituted pair **8/9** and - 4.3 kJ/mol for analogous chlorine free pair **4/5**. For the pair **2/3** bearing the nitrogen or the N-methylated group in meta position shorter distances to Asp189 are experienced which correlates with a slightly more favorable enthalpic difference by $\Delta\Delta H_{2/3}$ of 1.0 kJ/mol. Finally, for the para analogs **6/7**, the N-methylated derivative places its positive charge the closest to the negative aspartate. Here the enthalpy improvement for N-methylation amounts to $\Delta\Delta H_{6-7}$ of 4.9 kJ/mol. The series suggests that the large desolvation cost of the protein and the ligands, assumed to be of the same magnitude for all derivatives, is increasingly compensated by an enthalpic electrostatic contribution experienced in the considered derivatives along with a decreasing distance.

In addition to the desolvation effects in this series it is interesting to note, that only in some cases an entropic signal can be related to the displacement of a water molecule. Comparing the crystal structure of the benzyl derivative **1** with that of **2** shows one extra water in the latter structure. Indeed, the less solvated benzyl ligand **1** shows a favorable entropy term compared to **2** ($- T\Delta\Delta S_{2-1} = + 10.0$ kJ/mol) due to the release of an ordered water molecule.

In contrast, a clearly defined water molecule, present in the complex with **2,** is displaced from the S1 binding pocket in the complex with **3.** The expected entropic benefit cannot be detected as indicated by an unfavorable $- T\Delta\Delta S_{2-3}$ term of - 2.9 kJ/mol. We calculated the average B factors for the ligand's P1, P2 and P3 side chains and the surrounding binding site

residues to investigate whether changes in residual mobility could possibly account for the entropic effects recorded by our ITC experiments.

Table 2: The chemical structures used in the displacement titration as competitive binder together with the kinetically determined inhibition contants (K_i in nM) and the standard deviation[(a)] towards human thrombin.

Competitive Ligand 1 (**12**) 	**2.2**
Competitive Ligand 2 (**13**) 	**62.9 ± 19.8**

[a)]The error is given as the standard deviation calculated from at least three measurements (in nM).

3. Impact of ligand and protein desolvation

Table 3: Binding data ΔG^0, ΔH^0 and $-T\Delta S^0$ (kJ/mol) of ligands determined by ITC and evaluation of B-factors (Å^2) based on the crystal structures.

Ligand name	ΔG^0 (kJ/mol)	ΔH^0 (kJ/mol)	$-T\Delta S^0$ (kJ/mol)	B ligand	B binding site	B ligand/ B binding site	B D-Phe	B Pro	B D-Phe + Pro	B S1 occupant	B S1 occupant/ B D-Phe + Pro
1[21]	- 31.7	- 13.6	-18.1	26.2	18.9	**1.39**	26.2	24.9	25.5	28.2	**1.11**
2	- 25.8 ± 0.0	- 17.7 ± 0.3	- 8.1 ± 0.2	17.7	15.9	**1.11**	16.0	15.9	16.0	21.8	**1.37**
3	- 23.9 ± 0.4	- 18.7 ± 1.4	- 5.2 ± 1.0	15.2	15.1	**1.01**	15.2	14.9	15.1	15.4	**1.02**
4	- 25.0 ± 0.2	- 19.8 ± 0.4	- 5.2 ± 0.2	**Highly disordered as indicated by multiple conformations**							
5	- 21.4 ± 0.1	- 15.5 ± 0.3	- 5.9 ± 0.3	**Highly disordered as indicated by multiple conformations**							
6	- 24.2 ± 0.3	- 18.7 ± 0.1	- 5.5 ± 0.3	20.3	17.2	**1.18**	19.2	19.3	19.3	22.6	**1.17**
7	- 24.0 ± 0.1	- 23.6 ± 0.7	- 0.4 ± 0.5	16.0	16.0	**1.00**	15.3	15.8	15.6	17.0	**1.09**
8	- 31.5 ± 0.3	- 30.3 ± 0.3	- 1.2 ± 0.5	36.2	24.8	**1.46**	35.1	41.1	38.1	33.8	**0.89**
9	- 23.7 ± 0.4	- 23.3 ± 3.8	- 0.4 ± 3.4	14.1	12.0	**1.18**	14.0	14.3	14.2	14.0	**0.99**
11	- 24.3 ± 0.2	- 22.4 ± 0.8	- 1.9 ± 1.0	23.4	20.1	**1.16**	22.2	22.4	22.3	25.5	**1.14**
12	- 47.3 ± 0.9	- 40.2 ± 0.9	- 7.1 ± 1.9								
13	- 37.6 ± 0.3	-34.5 ± 1.1	- 3.1 ± 1.4								

Ligands **2**-**11** were determined in a displacement titration using ligands **12** or **13** as competitive binder. The thermodynamic data of **12** and **13** were determined in a direct ITC titration. The binding pocket includes all residues within 4 Å from the inhibitor. Errors indicate the standard deviation from at least duplicate measurements.

3. Impact of ligand and protein desolvation

We followed a protocol recently suggested in a previous contribution (Baum *et al.*[21]) by comparing ratio's of B factors (see Table 3 for definition of B factor ratio's). The ratio of B factors of the ligand's S1 occupant relative to the remaining D-Phe-Pro moiety of **2** is actually higher by a factor of 1.37 (Table 3) indicating higher residual mobility of the P1 portion in this complex compared to a value of 1.02 determined for the related N-methylated pyridinium derivative **3**. Despite similiar desolvation costs for the aspartate, **1** is entropically clearly favored compared to the charged ligand **3** ($- T\Delta\Delta S_{3/1} = - 12.9$ kJ/mol). For the complex with **1** we observe a higher B factor ratio (1.11) suggesting higher residual mobility of the benzyl substituent compared to the S1 occupant of **3** which shows almost no enhanced residual motion (1.02). It seems that the entropic benefit which is assigned to the displacement of an ordered water molecule from the binding site is compensated by an unfavorable entropic contribution arising from the reduced residual motion of the P1 portion of **3**. Furthermore, both ligands differ significantly in the free desolvation energy. Possibly, this strong effect will also influence the partitioning of enthalpy and entropy and can still be reflected in a deviating profile observed for both complexes.

The P1 side chains in **4** and **5** are scattered over at least two conformations, nevertheless, we cannot exclude even further conformers to be populated. Both derivatives exhibit similar entropic contributions ($- T\Delta\Delta S_{4/5} = + 0.7$ kJ/mol) upon complex formation and have similar crystallographic results with comparable disorder and displacement of overall 0.5 mol water molecules. The most surprising binding profile difference is indicated for the ligand pair **6**/**7** showing an unfavorable entropic signal ($- T\Delta\Delta S_{6/7} = - 5.1$ kJ/mol) upon displacement of much of the water network near Asp189 observed with **6** but lacking with **7**. In the case of **7** only one ordered water molecule adjacent to Asp189 is found whereas in **6** with experimental accuracy 1.5 water molecules can be assigned to the scattered difference electron density. This number results as an average value if all of the water positions would in fact be occupied at 50 %. Apparently, the striking difference in the thermodynamic signature of both complexes cannot be correlated with the amount of replaced water molecules since both complexes feature almost the same quantity. Nevertheless, why is **6** entropically more favored than **7**? To answer this we have to compare both structures to the apo form of thrombin. Here we find two well-ordered water molecules bound to Asp189 which occupy the most right and left sites also seen in the complex with **6** (Figure 3e). Upon binding of **6** one (or 1.5) water molecule is still present in the complex, however, highly scattered over

several positions whereas in the complex with **7** one ordered water molecule (most left site) is preserved. As it seems binding of **6** drives the water molecules present in the apo structure in a highly disordered state. As such water molecules are then quite disordered in the protein-bound state, their beneficial contribution to the ordering parameters of the complex leads to a more favored entropic signature. This entropic contribution of the disordered water molecules is missing in **7**. Beside these effects it is likely that the S1 occupant of **6** is not firmly fixed in its position because it interacts with the disordered water molecules via hydrogen bonds. Most likely the pyridine moiety will therefore experience also some influence from the "adjacent disorder". All these entropically beneficial effects on the ordering parameters of **6** in bound state are not present in **7**. For the latter we observe a more favorable enthalpic term ($\Delta\Delta H_{6/7}$ = + 4.9 kJ/mol) which most likely also reflects the strong electrostatic interaction to Asp189 across short distance.

To sum this up, the observed disorder around Asp189 corresponds to a poorly solvated hydration shell and the participating water molecules are therefore unlikely to be forming strong H-bonds requiring firmly fixed positions. We therefore assign the entropically favorable thermodynamic signature of **6** to the reorganization of the originally well ordered solvation pattern adjacent to Asp189 to a more dynamic state.

3.4.3 Calculation of free energies of solvation

Our hypothesis suggests that large costs for desolvation of the ligands and the protein binding site play a major role in the thermodynamic binding inventory and explain the observed weak binding of the positively charged inhibitors. In order to confirm the hypothesis, we calculated solvation free energies for the ligand series by a quantumchemical approach using SM8 and PCM implicit solvation models[28]. Both models show different absolute values but indicate basically the same trends that ligands carrying a positive charge are more costly to desolvate than the neutral species (Table 4). The latter ones show solvation energies in the range of - (56-66) kJ/mol whereas introducing a charge at the pyridine ring dramatically increases the solvation energies [- (218-246) kJ/mol, SM8 model]. Basically, the same trend was found in the PCM model but with lower absolute values, - (10-13) kJ/mol for the neutral and - (123-139) kJ/mol for the charged analogs.

Table 4: Solvation free energies[a] of the studied inhibitors. The calculation of these energies was done with the solvation models SM8 and PCM[28].

Ligand name	ΔG_{Solv} (kJ/mol) SM8	ΔG_{Solv} (kJ/mol) PCM
1	- 55,9	- 11,3
2	- 63,3	- 13,4
3	- 242,9	- 138,7
4	- 62,4	- 9,7
5	- 218,1	- 128,8
6	- 62.4	- 9.3
7	- 245,8	- 138,2
8	- 59,6	- 11,2
9	- 231,0	- 135,3
10	- 66,4	- 10,8
11	- 223,1	- 123,2

[a)]The values for the neutral inhibitors are in blue and the charged derivatives in red.

Table 5: Experimentally determined solvation free energies from literature of some compounds containing nitrogen either as pyridines or primary, secondary or tertiary amines[29] and the corresponding charged analogs.

Compound name	ΔG_{Solv} (kJ/mol) Exper.	ΔG_{Solv} (kJ/mol) Exper.	Compound name
aniline	- 23.0	- 301.2	anilinium
CH_3NH_2	- 19.2	- 309.6	CH_3NH_3+
$(CH_3)_2NH$	- 18.0	- 280.3	$(CH_2)_2NH_2$+
$(CH_3)_3N$	- 13.4	- 263.6	$(CH_2)_3NH$+
pyridine	- 19.6		
2-methylpyridine	- 19.3		
3-methylpyridine	- 20.0		
4-methylpyridine	- 20.6		

At first glance, the values appear rather high, however, experimental data for pyridine and other amines compared to its charged analogs support our calculated values. Pyridine and its methyl derivatives exhibits an experimental solvation energy of - (19.3-20.6) kJ/mol.[29] For our much larger inhibitors the SM8 model suggests about - 60 kJ/mol.

Introducing a charge in the molecules is associated with a large increase in solvation energy which was measured for the charged amine species to be an order of magnitude higher [ΔG_{Solv} = - (264-310) kJ/mol, Table 5]. These findings firmly support our calculations which suggest a strongly increased desolvation contribution for our N-methylated pyridinium derivatives.

3.5 Conclusion

Herein we presented a comprehensive study of 10 thrombin inhibitors demonstrating the importance of the desolvation of the ligand functional groups and deeply buried charged protein residues. These aspects have to be considered in the analysis of ligand binding to a target protein. It is generally accepted that water molecules play an important role in the ligand binding but most efforts fail to fully implement water contributions in scoring functions to predict binding affinities. Our ligand series shows impressively how decisive the cost in free energy can be if charges are present in a ligand and only weak interactions can be formed to the protein residues. Large penalties in enthalpy have to be paid for ligand desolvation especially of the charged N-methylated pyridinium derivatives. They are responsible for a large drop in the binding affinity by a factor of 20 - 30 (compared to the neutral non-methylated analogs) in case of the chloro-substituted ligands. As the ligands adopt a geometry with the chlorine substituent placed close to the neighboring tyrosine residue, the charged group is located in a region not favorable to form strong interactions to Asp189. In contrast, in the derivatives lacking the chlorine substituent, the positive charge is placed as close as possible to the negatively charged aspartate. This way, the P1 substituent compensates the price to be paid for desolvation compared to the neutral analogs (only by a factor of about 2 less in affinity). Consequently, improvement in binding affinity is obtained when the positive charge is placed adjacent to the deprotonated Asp189 in the para derivative **7**. Furthermore, the contributions to be paid to desolvate the protein binding site are important for the binding process. Especially, the desolvation of buried and

deprotonated aspartate residues requires a large free energy penalty of desolvation and has to be appropriately compensated by newly formed interactions if a high affinity complex is to be formed. In our series, we identify a strongly bound water molecule next to the deprotonated residue Asp189 which is present in the uncomplexed structure. Because of the rather costly displacement of this water molecule, even distorted and most likely less favorable ligand conformations are partially populated which still allow this water molecule to be accommodated. Its complete displacement requires high desolvation costs. Consequently, for **4** and **5** only about 50 % of the water molecules are displaced from the S1 pocket.

The 3-pyridine derivative **2** binds to the S1 pocket with the water molecule still bound to Asp189. It shows reduced binding affinity compared to the unsubstituted benzyl derivative **1**, however these complexes exhibit a reversed thermodynamic signature; **1** repels the water molecule from Asp189 and shows remarkable residual mobility in the pocket. Its binding is accompanied by a favorable entropic contribution. **2** binds with a more enthalpically favored profile. It also shows compared to **1** reduced residual mobility of its P1 portion in bound state but enhanced motion with respect to **3**. However, compared to the complex with **1**, Asp189 remains in **2** partially solvated and only the enhanced desolvation of the pyridine nitrogen has to be paid for. Overall this results in an enthalpy-driven binding of **2** and **3**.

In addition to direct and well defined water-mediated interactions, we provide evidence that also highly disordered water molecules can serve this purpose. Interestingly enough, the replacement of these disordered water molecules is determinant for an entropically unfavorable signature. Waters already disordered in the bound state, cannot make much of an impact on the entropic component once released to the bulk solvent. Consequently, the enthalpic component dominates as newly formed interactions of the repelled water molecules are experienced in the bulk phase.

3.6 Acknowledgements

We kindly acknowledge CSL Behring, Marburg, for supplying us with generous amounts of human thrombin from the production of Beriplast®. We thank the beamline support staff at SLS, BESSY and ESRF for their advice during data collection. This work was supported by the Bundesministerium für Bildung und Forschung (BMBF, Förderkennzeichen 0315161C).

3.7 Materials and Methods

3.7.1 Bioassay

Kinetic inhibition of human thrombin (from Beriplast®, CSL Behring, Marburg, Germany) was determined photometrically at 405 nm using the chromogenic substrate Pefachrom tPa (LoxoGmbH, Dossenheim, Germany) as described[30] under the following conditions: 50 mM Tris–HCl, pH 7.4, 154 mM NaCl, 5 % DMSO, 0.1 % polyethylene glycol 8000 at 25°C using different concentrations of substrate and inhibitor. K_i values ($n \geq 3$) were determined as described by Dixon[31].

3.7.2 Isothermal titration calorimetry

ITC experiments were performed using an ITC200™ system from Microcal (now part of GE Healthcare, Northhampton, MA, USA). Thrombin was freshly prepared for each experiment by dialysis of a thrombin sample in the buffer used for titration experiments (50 mM Hepes, 100 mM NaCl, 1 % PEG8000, pH 7.8). After dialysis the weak-binding ligand was added from a stock solution (50 mM, 100% DMSO) to the solution of thrombin (27 µM) in order to obtain the required ligand concentration for sufficient saturation (1.25 - 1.5 mM). The final DMSO concentration was subsequently adjusted to 3 %. Either competitive ligand **12** or **13** (500 – 625 µM) were filled in the syringe and then titrated into the stirred sample cell (200 µL) containing the thrombin solution (27 µM) saturated with the weak-binding ligand under investigation. The resulting titration curve of the displacement titration was analysed using the Origin 7.0 software by fitting a single-site binding isotherm. The derived parameters are used as $K_{a\ obs}$ and ΔH_{obs} in equation 1 and 2[32] to calculate the thermodynamic properties $K_{a\ WL}$ and ΔH_{WL} of the weak-binding ligand. The values $K_{a\ SL}$ and ΔH_{SL} for **12** or **13** were taken from a separate experiment carried out in the beginning of the experiment.

$$K_{a\,WL} = \left(\frac{K_{a\,SL}}{K_{a\,obs}} - 1\right) \cdot \frac{1}{[L_{free}]} \qquad \text{Equation 1}$$

$K_{a\ obs}$ = Observed association constant from the ITC displacement experiment

$K_{a\ WL}$ = Association constant of weak binder

$K_{a\ SL}$ = Association constant of strong binder from a direct ITC experiment

$[L_{free}]$ = Concentration of weak-binding ligand in the saturated protein solution

$$\Delta H_{WL} = (\Delta H_{SL} - \Delta H_{obs})\left(1 + \frac{1}{K_{a\,WL} \cdot [L_{free}]}\right)$$ Equation 2

ΔH_{WL} = Enthalpy change of weak binder

ΔH_{obs} = Observed enthalpy change of strong binder in the displacement experiment

ΔH_{SL} = Enthalpy change of strong binder

$K_{a\,WL}$ = Association constant of weak binder

$[L_{free}]$ = Concentration of weak binder in the saturated protein solution

The concentration of the free ligand $[L_{free}]$ (Equ. 3) is unknown. As the concentration of the weak-binding ligand is much larger than that of the protein $[L_{tot}] >> [P_{tot}]$, the free ligand concentration $[L_{free}]$ is practically equal to $[L_{tot}]$. Thus $[L_{tot}]$ was used in the further analysis.

$$[L_{free}] = [L_{tot}] - [PL]$$ Equation 3

All ITC experiments were started at 25 °C after a stable baseline had been achieved. The experimental design comprises an initial ligand injection of 0.3 µL followed by 15 injections of 1.1 - 1.4 µL with a 300 s interval between each injection.

All measurements were performed in duplicate. Raw data were collected and the area under each peak was integrated, followed by correction for heats of dilution and mixing by subtracting the final baseline consisting of small peaks of the same size to zero. The initial data point was deleted from the integrated data because this injection usually reflects an erroneous amount of heat due to the possible exchange of liquids between syringe and cell when inserting the syringe into the calorimetric cell and the backlash error in the motorized screw mechanism in the injector.[33] Representative titration curves are shown in the supplementary data.

3.7.3 Solvation energy calculation

The calculation has been completed by the program package Q-chem ver4.0. The solvation energy has been calculated by density functional theory B3LYP method with 6-31G** basis set, along with PCM or SM8 implicit solvent model respectively.[28]

3.7.4 Crystallization and soaking

Human α-thrombin (from Enzyme Research Laboratories, South Bend, USA) was dissolved in the crystallization buffer (20 mM NaH_2PO_4, 350 mM NaCl, 2 mM benzamidine, pH 7.5) at 10 mg/ml. A hirudin fragment called Acetyl-Hirudin (54-65) purchased from Bachem (Bubendorf, Switzerland) was dissolved in crystallisation buffer at 2.5 mg/ml. In the next step, 40 µL of the solution of the hirudin fragment was mixed with 160 µL of the thrombin solution. After incubation for 2 h at 4 °C, crystallization was carried out at 4 °C by the hanging-drop method. One 1 µL of the hirudin/thrombin solution was placed in the centre of a cover slip and mixed with 1 µL reservoir solution (20 mM NaH_2PO_4, 27 % polyethylene glycol 8000, pH 7.5). Immediately after the mixing of protein and reservoir buffer microseeding was done. The wells of the crystallization trays were filled with 500 µL of the reservoir buffer. Subsequently, the cover slips were placed on the wells and sealed. Crystals of good diffracting quality could be produced within 7 days. For soaking DMSO stock solutions of the inhibitors (50 mM) were diluted 1:10 with a solution containing 50 % crystallization and 50 % reservoir buffer resulting in the final soaking concentration containing 5 mM of the inhibitor and 10 % DMSO. Medium-size crystals without visible imperfections were selected and transferred into the soaking solution for 24 h.

3.7.5 Data collection and processing

Crystals were prepared for data collection at 110 K using a cryoprotectant solution of 20 % glycerol in reservoir buffer. The data sets for **2**, **3**, **4**, **5**, **7** and **10** were collected with synchrotron radiation at SLS (Villingen, Switzerland) on a Marmosaic 225 mm CCD detector. Complex structures for **8, 9** and **11** were collected at BESSY beamline 14.2 (Berlin, Germany) on a Rayonix MX 225 CCD detector. The data set for **6** was collected at ESRF ID29 beamline (Grenoble, France) on a Pilatus 6M detector. Data processing and scaling were performed using the HKL2000 package[34]. Data processing and scaling were done with XDS in case of **5**.[35]

3.7.6 Structure determination and refinement

The coordinates of human thrombin (PDB code 1H8D)[36] were used for initial rigid body refinement of the protein molecules followed by repeated cycles of maximum likelihood

energy minimization, simulated annealing and B-factor refinement using the CNS program package[37]. Refinement of structure **2** was done with SHELXL[38] and structures **3**, **4**, **5**, **6**, **7**, **8, 9, 10** and **11** were refined with PHENIX[39]. The temperature factors for structures **2**, **6** and **9** were anisotropically refined whereas for structures **3**, **4**, **5**, **7, 8, 10** and **11** TLS refinement was applied. The definition of the TLS groups were done with the TLSMD server[40,41]. A randomly chosen 5 % of all data were used for the calculation of R_{free} and were not used in the refinement. Amino acid side chains were fit into σ-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps using Coot[42]. After the first refinement cycle, water molecules and subsequently ions and ligands were located in the electron density and added to the model. Restraints were applied to bond lengths and angles, planarity of aromatic rings and van der Waals contacts. Multiple side chain conformations were built in case an appropriate electron density was observed and maintained during the refinement, and if the minor populated side chain showed at least 20 % occupancy. The final models were validated using PHENIX own validation options or MolProbity[43]. The Ramachandran plot's were calculated with PROCHECK[44]. Data collection, unit cell parameters and refinement statistics are given in Table 6. Analysis of temperature factors was done with Moleman[45] and fconv[46]. The naming of the protein amino acids was done according to Bode *et al.*[47]. The figures were prepared using Pymol 0.99. The Protein Data Bank accession codes of the coordinates and structure factors of all X-ray structures are given in Table 6.

3.7.7 Protein Data Bank and accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with the following accession codes: THR-**2** complex 3P17; THR-**3** complex 3QTO; THR-**4** complex 3SI3; THR-**5** complex 3SI4; THR-**6** complex 3SV2; THR-**7** complex 3QTV; THR-**8** complex 3SHC; THR-**9** complex 3QX5; THR-**10** complex 3SHA; THR-**11** complex 3QWC.

Table 6: Data collection and refinement statistics for the ten thrombin complex structures determined in this publication.

Structure (PDB entry)	THR-**2** complex **(3P17)**	THR-**3** complex **(3QTO)**	THR-**4** complex **(3SI3)**	THR-**5** complex **(3SI4)**	THR-**6** complex **(3SV2)**
A. Data collection and processing					
No. Crystals used	1	1	1	1	1
Wavelength (Å)	1.00	1.00	1.00	1.00	0.9686
Space group	C2	C2	C2	C2	C2
Unit cell parameters					
a, b, c (Å)	70.0, 71.4, 72.5	69.8, 71.4, 72.5	70.1, 71.4, 72.3	70.4, 71.2, 72.6	70.0, 71.4. 72.2
β (°)	100.3	100.3	100.1	100.4	100.3
Matthews coefficient ($Å^3$/Da)	2.7	2.5	2.5	2.5	2.5
Solvent content (%)	51	51	51	51	51
B. Diffraction data[a]					
Resolution range (Å)	50 – 1.43	50 – 1.52	50 – 1.55	50 – 1.27	50 – 1.30
	(1.45 – 1.43)	(1.55 – 1.52)	(1.58 – 1.55)	(1.29 – 1.27)	(1.38 – 1.30)
Unique reflections	63,603 (3,046)	53,781 (2,645)	50,679 (2,310)	91,193 (4,405)	85,743 (11,030)
R(I)sym (%)	4.6 (39.3)	6.5 (49.7)	6.6 (47.0)	3.8 (47.1)	2.7 (49.2)
Completeness (%)	98.1 (95.7)	99.8 (99.5)	99.3 (91.6)	98.4 (95.2)	93.6 (88.6)
Redundancy	2.5 (2.5)	3.1 (2.9)	3.7 (3.0)	2.7 (2.5)	3.6 (3.5)
I/σ(I)	19.5 (2.3)	17.0 (2.2)	20.0 (2.0)	26.4 (2.2)	18.2 (2.5)
C. Refinement					
Resolution range (Å)	10 – 1.43	35.7 – 1.52	35.7 – 1.55	34.6 – 1.27	35.7 – 1.30
Reflections used in refinement	56,991 / 3,021	51,053 / 2,584	47,831 / 2,396	86,393 / 4,347	80,045 / 4,025
(work/free)					
Final R values for all reflections	14.0 / 17.6	15.5 / 16.5	15.6 / 18.1	13.7 / 15.9	14.0 / 16.5
(work/free) (%)					
Protein residues (L chain/H chain)	27 / 251	28 / 251	28 / 251	28 / 251	28 / 251
Sodium ions	2	2	2	2	2
Inhibitor atoms	26	27	26[c] / 36[d]	27[c] / 38[d]	26
Water molecules	276	366	350	401	328

3. Impact of ligand and protein desolvation

RMSD from ideality					
Bond lengths (Å)	0.011	0.008	0.010	0.008	0.012
Bond angles (°)	2.2	1.1	1.1	1.1	1.3
Ramachandran plot					
Residues in most favoured regions (%)	86.1	85.8	85.8	85.8	86.2
Residues in additionally allowed	13.9	13.8	13.8	14.2	13.4
regions (%)					
Residues in generously allowed	-	0.4	0.4	-	0.4
regions (%)					
Mean B-factor ($Å^2$)					
Protein (L + H chain)	18.8	17.8	18.3	18.1	19.9
Binding site[b]	15.9	15.1	15.1	16.3	17.2
Inhibitor	17.7	15.2	14.7[e]	17.9[e]	20.3
Water molecules	31.5	32.1	31.9	35.4	33.7

3. Impact of ligand and protein desolvation

Structure (PDB entry)	THR-**7** complex **(3QTV)**	THR-**8** complex **(3SHC)**	THR-**9** complex **(3QX5)**	THR-**10** complex **(3SHA)**	THR-**11** complex **(3QWC)**
A. Data collection and processing					
No. Crystals used	1	1	1	1	1
Wavelength (Å)	1.00	0.91841	0.91841	1.00	0.91841
Space group	C2	C2	C2	C2	C2
Unit cell parameters					
a, b, c (Å)	69.9, 71.3, 72.4	70.4, 71.4, 72.8	70.2, 71.1, 72.8	69.8, 71.4, 72.2	70.6, 71.5, 72.7
β (°)	100.4	100.6	100.6	100.0	100.6
Matthews coefficient ($Å^3$/Da)	2.5	2.5	2.5	2.5	2.5
Solvent content (%)	51	51	51	51	52
B. Diffraction data[a]					
Resolution range (Å)	50 – 1.63	50 – 1.90	30 – 1.35	50 – 1.52	30 – 1.75
	(1.66 – 1.63)	(1.93 – 1.90)	(1.37 – 1.35)	(1.55 – 1.52)	(1.78 – 1.75)
Unique reflections	43,364 (2,129)	27,888 (1,411)	74,225 (2,719)	51,764 (2,454)	34,739 (1,776)
R(I)sym (%)	6.8 (45.9)	7.9 (40.0)	4.8 (29.3)	4.4 (37.4)	5.5 (40.7)
Completeness (%)	99.7 (96.1)	99.5 (100.0)	96.0 (71.2)	96.1 (92.3)	96.1 (98.4)
Redundancy	3.1 (2.4)	3.1 (2.7)	3.0 (2.0)	2.2 (2.1)	2.2 (2.1)
I/σ(I)	16.7 (1.9)	13.2 (2.7)	22.3 (2.8)	15.7 (2.2)	14.7 (2.5)
C. Refinement					
Resolution range (Å)	35.7 – 1.63	24.8 – 1.90	24.8 – 1.35	31.9 – 1.52	24.9 – 1.75
Reflections used in refinement	41,532 / 2,090	26,703 / 1,297	71,675 / 3,635	48,953 / 2,468	32,908 / 1,642
(work/free)					
Final R values for all reflections	15.5 / 17.6	15.9 / 19.7	13.2 / 15.6	16.1 / 19.0	15.4 / 17.9
(work/free) (%)					
Protein residues (L chain/H chain)	28 / 251	28 / 251	28 / 251	28 / 251	28 / 251
Sodium ions	2	2	2	2	2
Inhibitor atoms	27	27	28	27	28
Water molecules	331	223	450	328	260

RMSD from ideality					
Bond lengths (Å)	0.009	0.010	0.008	0.009	0.010
Bond angles (°)	1.075	1.062	1.127	1.091	1.070
Ramachandran plot					
Residues in most favoured regions (%)	85.4	87.4	84.6	86.2	86.2
Residues in additionally allowed	14.6	12.6	15.4	13.4	13.8
regions (%)					
Residues in generously allowed	-	-	-	0.4	-
regions (%)					
Mean B-factor ($Å^2$)					
Protein (L + H chain)	19.4	24.6	15.9	19.6	22.7
Binding site[b]	16.0	24.8	12.0	15.5	20.1
Inhibitor	16.0	36.2	14.1	15.1	23.4
Water molecules	32.8	33.3	35.0	32.6	33.7

[a)]Numbers in parenthesis are for the highest resolution shell.

[b)]Definition of the binding site: all amino acids which are 4 Å away from inhibitor.

[c)]The inhibitor atoms from the additional disordered part were not considered.

[d)]All inhibitor atoms which were included in the refinement model.

[e)]Average B value for the ordered inhibitor portion, the disordered portions were not considered.

3.8 References

1 Ladbury J.E. (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.

2 Quiocho F.A., Wilson D.K. & Vyas, N.K. (1989) Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **340**, 404-407.

3 Clarke C., Woods R.J., Gluska J., Cooper A., Nutley M.A. *et al.* (2001) Involvement of water in carbohydrate-protein binding. *J Am Chem Soc.* **123**, 12238-12247.

4 Mikol V., Papageorgiou C. & Borer X. (1995) The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J Med Chem.* **38**, 3361-3367.

5 Rejto P.A. & Verkhivker G.M. (1997) Mean Field Analysis of FKBP12 Complexes With FK506 and Rapamycin: Implications for a Role of Crystallographic Water Molecules in Molecular Recognition and Specificity. *Proteins*. **28**, 313-324.

6 Finley J.B., Atigadda V.R., Duarte F., Zhao J.J., Brouillette W.J. *et al.* (1999) Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site. *J. Mol. Biol.* **293,** 1107-1119.

7 Sleigh S.H., Seavers P.R., Wilkinson A.J., Ladbury J.E. & Tame J.R. (1999) Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J. Mol. Biol.* **291**, 393-415.

8 Lam P.Y., Jadhav P.K., Eyermann C.J., Hodge C.N., Ru Y. *et al.* (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science.* **263**, 380-384.

9 Mikol V., Papageorgiou C. & Borer X. (1995) The Role of Water Molecules in the Structure-Based Design of (5-Hydroxynorvaline)-2-cyclosporin: Synthesis, Biological Activity, and Crystallographic Analysis with Cyclophilin A. *J. Med. Chem.* **38**, 3361-3367.

10 Shimokhina N., Bronowska A. & Homans S.W. (2006) Contribution of Ligand Desolvation to Binding Thermodynamics in a Ligand-Protein Interaction. *Angew. Chem. Int. Ed.* **45**, 6374-6376.

11 Still W.C., Tempczyk A., Hawley R.C. & Hendrickson T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127-6129.

12 Tannor D. J., Marten B., Murphy R., Friesner R., Sitkoff D. *et al.* (1994) Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. *J. Am. Chem.Soc.* **116**, 11875- 11882.

13 Wang J., Wang W., Huo S., Lee M. & Kollman P.A. (2001) Solvation Model Based on Weighted Solvent Accessible Surface Area *J. Phys. Chem. B*, **105**, 5055-5067.

14 Viswanadhan V.N., Ghose A.K., Singh U.C. & Wendoloski J.J. (1999) Prediction of Solvation Free Energies of Small Organic Molecules: Additive-Constitutive Models Based on Molecular Fingerprints and Atomic Constants. *J. Chem. Inf. Comput. Sci.* **39**, 405-412.

15 Buttery R.G., Bomben J.L., Guadagni D.G. & Ling L.C. (1971) Some Considerations of the Volatilities of Organic Flavor Compounds in *Foods J. Agric. Food Chem*. **19**, 1045–1048.

16 Homans, S.W. (2007). Water, water everywhere - except where it matters? *Drug Discovery Today*. **12**, 534-539.

17 Olsson T.S., Ladbury J.E., Pitt W.R. & Williams M.A. (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci.* **20**, 1607-1618.

18 Englert L., Biela A., Zayed M., Heine A., Hangauer D. *et al.* (2010) Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: binding of phosphonamidate to the S_1'-pocket of thermolysin. *Biochim. Biophys. Acta.* **1800**, 1192-11202.

19 Baum B., Muley L., Smolinski M., Heine A., Hangauer D. *et al.* (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.

20 Muley L., Baum B., Smolinski M., Freindorf M., Heine A. *et al.* (2010) Enhancement of hydrophobic interactions and hydrogen bond strength by cooperativity: synthesis, modeling, and molecular dynamics simulations of a congeneric series of thrombin inhibitors. *J Med Chem.* **53**, 2126-2135.

21 Baum B., Mohamed M., Zayed M., Gerlach C., Heine A. *et al.* (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.

22 Berman, H., Henrick, K. & Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol.* **10**, 980.

23 Baum B., Muley L., Heine A., Smolinski M., Hangauer D. *et al.* (2009) Think twice: understanding the high potency of bis(phenyl)methane inhibitors of thrombin. *J. Mol. Biol.* **391**, 552-564.

24 Günther J., Bergner A., Hendlich M. & Klebe G. (2003) Utilising structural knowledge in drug design strategies: applications using Relibase. *J. Mol. Biol.* **326**, 621-636.

25 Ahmed H.U., Blakeley M.P., Cianci M., Cruickshank D.W.J., Hubbard J.A. *et al.* (2007) The determination of protonation states in proteins. *Acta Crystr. D* **63**, 906-922.

26 Allen F.H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B.* **58**, 380-388.

27 Harding M.M. (2002) Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Cryst. D* **58**, 872-874.

28 Cramer C.J. & Truhlar D.G. (2008) A Universal Approach to Solvation Modeling. *Acc. Chem. Res.* **41**, 760-768.

29 Wang J., Wang W., Huo S., Lee M. & Kollman P.A. (2001) Solvation Model Based on Weighted Solvent Accessible Surface Area. *J. Phys. Chem. B* **105**, 5055–5067.

30 Stürzebecher J., Stürzebecher U., Vieweg H., Wagner G., Hauptmann J. *et al.* (1989) Synthetic inhibitors of bovine factor Xa and thrombin comparison of their anticoagulant efficiency. *Thromb. Res.* **54**, 245–252.

31 Dixon M. (1972) The graphical determination of K_m and K_i. *Biochem. J.* **129**, 197–202.

32 Sigurskjold B.W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* **277**, 260-266.

33 Mizoue L.S. & Tellinghuisen J. (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal. Biochem.* **326**, 125-127.

34 Otwinowski Z. & Minor W. (1997). Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol.* **276**, 307-326.

35 Kabsch W. (2010) XDS. *Acta Cryst.* D **66**, 125-132.

36 Skordalakes E., Dodson G.G., Green D.S., Goodwin C.A., Scully M.F. *et al.* (2001). Inhibition of human alpha-thrombin by a phosphonate tripeptide proceeds via a metastable pentacoordinated phosphorus intermediate. *J. Mol. Biol.* **311**, 549–555.

37 Brunger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921.

38 Sheldrick G.M. & Schneider T.R. (1997) SHELXL: high-resolution refinement. *Macromol. Crystallogr. B* **277**, 319–343.

39 Adams P.D, Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213-221.

40 Painter J. & Merritt E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62,** 439-450.

41 Painter J. & Merritt E.A. (2006) TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* **39**, 109-111.

42 Emsley P. & Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132.

43 Chen V.B., Arendall W.B. 3rd, Headd J.J., Keedy D.A., Immormino R.M. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12-21.

44 Laskowski R.A., MacArthur M.W., Moss D.S. & Thornton J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*. **26**, 283–291.

45 Kleywegt G.J., Zou J.Y., Kjeldgaard M. & Jones T.A. (2001) Around O. *In International Tables for Crystallography* (Rossmann, M.G. & Arnold, E., eds), vol. F, pp. 353–356, Kluwer Academic Publishers, Dordrecht.

46 Neudert G. & Klebe G. (2011) fconv: format conversion, manipulation and feature computation of molecular data. *Bioinformatics.* **27**, 1021-1022.

47 Bode W., Mayr I., Baumann, U., Huber, R., Stone S.R. *et al.* (1989) The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* 8, 3467-3475.

4 Enthalpic and entropic changes caused by a stepwise disruption of a water network in the S3/4 subsite of thrombin: An example of a classical hydrophobic effect

4.1 Introductory remarks

This study was done in cooperation with the group of Prof. Dr. Steinmetzer (Philipps-University Marburg). The following text will be submitted to a scientific journal. Frank Sielaff from the Steinmetzer group will be included as second author.

4.2 Abstract

The aim of this study is to examine the influence of the displacement of firmly bound water molecules from a hydrophobic cavity on binding affinity and the thermodynamic signature. The binding of a series of ligands with gradually increasing hydrophobicity and sterical requirements to thrombin has been studied using isothermal titration calorimetry (ITC) and X-ray crystallography. The inhibitors vary systematically at the P3 position by Gly, D-Ala, D-Val, D-Leu and D-cyclohexylalanine (D-Cha) which address the well-hydrated S3/4 pocket in the apoenzyme. Two congeneric series were studied which differ only in the group addressing the S1 subpocket of thrombin. As S1 occupants 2-(aminomethyl)-5-chlorobenzylamide and 4-amidinobenzylamide were used which show different interaction pattern in the S1 pocket. Nevertheless, they turn out to be almost equally potent. Crystal structure analysis reveals that both scaffolds exhibit also different binding modes of the terminal benzylsulfonyl moiety, whereas the systematically modified P3 substituents address the hydrophobic S3/4 pocket in a comparable way. Small modifications (D-Ala) are not capable of disrupting the hydration pattern also seen in the ligand free enzyme whereas larger substitutions (up to D-Cha) gradually displace these water molecules. Accordingly for mid-size groups even a hydration pattern with partially occupied water molecules is recorded. The improvement in binding affinity from nano- to picomolar range correlates well with the observed water displacement and with the extent of buried van-der-Waals

contacts. The potency gain is mainly attributed to a favorable entropic contribution which correlates with a water release. The Gly derivatives drop out of the series since they show a deviating binding mode. Nonetheless and most remarkably, no significant changes in the thermodynamic profile are detected for the Gly vs D-Ala derivatives clearly indicating that conserved thermodynamic signature is by no means a reliable indicator for conserved binding mode. Aspects, such as differences in the accessible conformational space of Gly vs D-Ala in solution prior to protein binding and an equivalent total water release from the entire binding site might be responsible for these balanced thermodynamic profiles. An additional crystallographically remarkable finding is the binding of a second inhibitor molecule in some complexes with 4-amidinobenzylamide inhibitors. Interestingly, this second inhibitor molecule virtually maps the same region occupied by a characteristic loop formed by amino acids present in the fibrinogen α-chain, the natural thrombin substrate.

4.3 Introduction

Molecular recognition is of fundamental importance in molecular biology and drug discovery. Especially in drug design, the demand for modulating the recognition features between a small molecule ligand and a target protein is of utmost importance as it provides an opportunity to interfere with a pathogenic mechanism where the target molecule takes a crucial role. Inhibition or enhancement of the activity of this target protein is one of the most popular approaches for the treatment of the associated disease. The binding process of a small ligand to a protein, no matter whether it results in inhibition or activation, is mediated by non-covalent interactions. Hydrogen bonds and hydrophobic van-der-Waals contacts are in this context the most important interactions and they are believed to exert major part of binding affinity. Therefore, detailed understanding of the contributions of these types of interactions to the overall binding affinity is a prerequisite for the development of new and effective drugs against a target protein. In this respect, the accurate prediction of binding affinity of a drug candidate under consideration is pivotal to the idea of rational drug design. For example, how much does a newly formed hydrogen bond in a designed complex structure contribute to binding affinity or is it virtually cancelling out due to enthalpy/entropy compensation[1,2,3], what is the role of water molecules in the binding event[4,5] and how can we exploit hydrophobic interactions to achieve optimal

modulation of ligand affinity[6]. This is only a small number of questions which need to be addressed to enable medicinal chemists to optimize their lead structures towards potent drug candidates.[7] Our understanding of hydrophobic contributions to protein-ligand interactions is still rather incomplete despite enormous research efforts which went into the characterization of the hydrophobic effect. In its classical view the hydrophobic effect experienced between interacting hydrophobic species is driven by the reorganization of water molecules upon protein-ligand binding.[8] Hydrophobic cavities in proteins and hydrophobic patches on ligand surfaces are contacted by water molecules prior to complex formation. These contacting water molecules are assumed to be well ordered and thus associated with an unfavorable entropic component. According to textbook knowledge, the release of such water molecules from hydrophobic interfaces to the bulk water phase leads to a favorable entropic signal upon complex formation as the released water molecules gain additional degrees of freedom. The association of hydrophobic species is therefore assumed to be driven by an entropic enhancement. However, an impressive number of hydrophobic complex formations in biology and host-guest chemistry have been reported that follow an opposed enthalpy-driven thermodynamic profile and would thus intuitively contradict the classical view on the hydrophobic effect. Smithrud *et al.*[9], for instance, tested the binding of benzene and pyrene derivatives to cyclophanes and their results suggest a strongly exothermic signal upon complexation despite interaction of hydrophobic binding partners. The same could be observed for accommodating apolar molecules in the cavity of cyclodextrins[10,11] and in the narrow AT-rich region of the DNA minor groove[12,13]. With respect to protein-ligand complex formation examples were provided by Bingham *et al.*[14] and Englert *et al.*[15] In the first case, binding of two pyrazine derivatives with propyl or butyl substituents to the hydrophobic binding pocket of the major urinary protein (MUP-1) was studied by ITC and resulted in an enthalpy-driven process. In the example of Englert *et al.* thermolysin was selected as model protein and its large hydrophobic S_1' pocket was addressed by ligand portions of gradually increasing size (methyl, propyl, isobutyl and benzyl). The protein-ligand association was investigated by ITC and suggested also enthalpy-driven hydrophobic interactions. Notably, in both examined proteins the binding pocket appears to be suboptimally hydrated[15,16] prior to ligand binding and indeed this common feature could explain the observed paradoxical enthalpic signal as proposed by Homans[17]. In a subsequent study, a deconvolution of the binding process in the MUP-1 case was done

and it indicates that poor solvation of the uncomplexed pocket makes its desolvation even favorable. With other words, if a subsite is poorly solvated only little energy is required for its full desolvation in order to allow ligand binding. In the binding inventory the unspent enthalpy pays off in an unexpectedly high enthalpy term as the released water molecules find enthalpically favorable interactions in the bulk water phase. These examples underline the importance to study the solvation properties of hydrophobic ligands and hydrophobic protein pockets prior to complex formation to correctly consider their contributions to binding. This aspect is also decisive with respect to hydrophobic ligand optimization since addressing a poorly hydrated pocket could require a different design concept to enhance binding affinity than dealing with well-solvated pockets. To shed more light on the complex thermodynamic signature of hydrophobic binding, we investigated the occupancy of a rather well-solvated hydrophobic pocket to better characterize the driving forces for complex formation. We selected the large hydrophobic S3/4 pocket of thrombin as model system and in contrast to the above- described examples, this pocket is well hydrated in the uncomplexed state (PDB code 2UUF[18]). A feasible approach to evaluate the contributions from hydrophobic complex formation is to examine the influence of successive ligand structural variations on the binding thermodynamics. We systematically varied the P3 substituent of substrate-analogue peptidomimetics by incorporation of a Gly, D-Ala, D-Val, D-Leu and a D-Cha to become increasingly more hydrophobic and sterically demanding. To study these effects on a broader scope we synthesized two inhibitor series with two different S1 occupants (see Table 1). Either, a 2-(aminomethyl)-5-chlorobenzylamide (ACB)[19,20] and a 4-amidinobenzylamide (AMBA)[21] anchor are known to be very potent binders for the S1 pocket. An experimental approach comprising ITC measurements, kinetic characterization and high resolution crystal structure determination was applied to study the influence of a stepwise removal of water molecules from the hydrophobic S3/4 pocket of thrombin due to variations of the P3 residues.

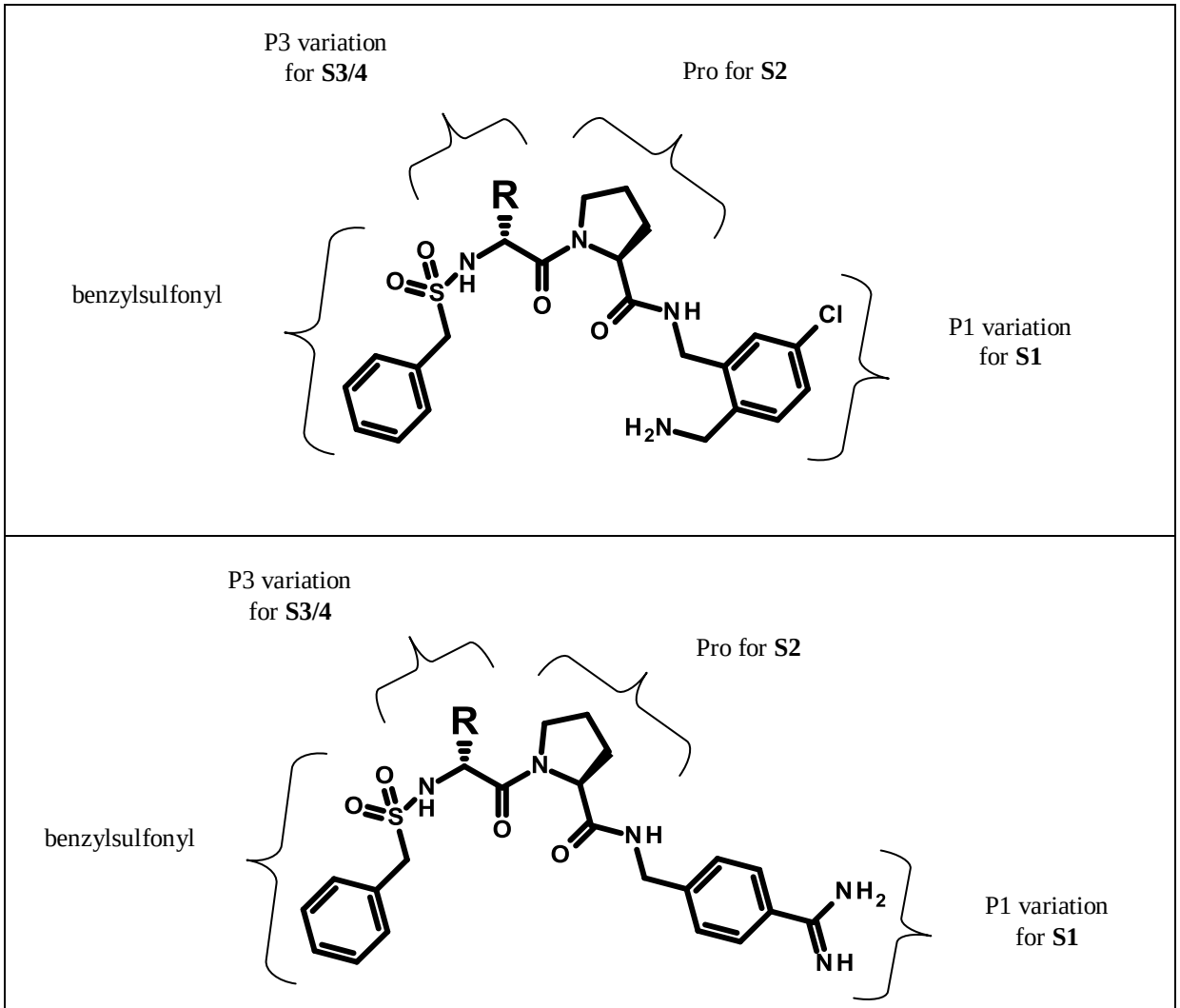
4.4 Results & Discussion

4.4.1 Introducing part to the crystal structures

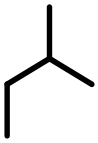
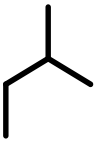
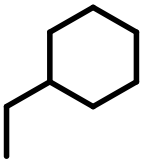
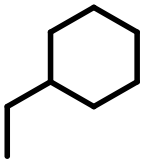
To systematically probe the replacement of water molecules from a binding pocket by crystallographic means requires high resolution crystal structures as the number of observed water molecules can suffer from the resolution of the diffraction data.[22]

4. A stepwise disruption of a water network

Table 1: Chemical structures of scaffold of the studied ligands in the ACB and AMBA series. The designed variants R are given with the kinetically determined inhibition contants (K_i in nM) towards human thrombin.



P1 substituent: ACB R:		P1 substituent: AMBA R:	
1 	1.5 ± 0.1	6 	3.7 ± 0.6
2 	2.2 ± 0.4	7 	5.2 ± 0.8

3 	**0.788 ± 0.070**	**8** 	**1.290 ± 0.347**
4 	**0.259 ± 0.024**	**9** 	**0.891 ± 0.115**
5 	**0.052 ± 0.005**	**10** 	**0.119 ± 0.010**

The error is given as the standard deviation calculated from at least three measurements (in nM).

Statistics show that, on average, at 2.0 Å resolution one water molecule per residue is included in the model, while at 1.0 Å resolution about 1.6-1.7 are crystallographically located. We therefore used synchrotron radiation to achieve the best possible resolution of our complex structures. In total, ten crystal structures of human thrombin in complex with **1-10** were determined with medium to high resolution (1.78-1.23 Å).

4.4.2 Binding mode of the scaffold of the ACB series

As the derivatives with a Gly P3 residue show completely different binding modes with respect to the remaining part of the designed series, we will present their data in a separate section. The complex structures of **2** to **5** exhibit a comparable binding mode and the series will be exemplified using ligand **2** (Figure 1a)**.** The chlorine atom of the S1 occupant points towards the center of Tyr228 with a distance of 4.3 Å which suggests a strong halogen-π interaction.[23] The aminomethylene anchor takes remarkable preorganizing influence on the overall binding mode as it is involved in strong intra- and intermolecular interactions. It hydrogen-bonds to the backbone carbonyl of Gly219 (3.1 Å) and intramolecularly to the oxygen atom of the inhibitor's own sulfonyl moiety (3.2 Å) and to the P2 carbonyl (3.1 Å). This binding motif associates a water molecule which mediates an additional contact between both moieties. As we assume the primary amino group to be largely protonated at the applied pH 7.5 this interaction pattern is supposedly charge assisted. The pyrrolidine ring

of proline is accommodated in the S2 pocket capped by Tyr60A and Trp60D of the 60's loop (Figure 1b). The methyl group of the P3-alanine is pointing into the S3/4 specificity pocket. The sulfonamide NH group and the adjacent central carbonyl functionality form an antiparallel β-sheet-like binding motif to Gly216. This binding feature has already been observed in an inhibitor series bearing a terminal amino group at this position.[24] With respect to the sulfonyl group one oxygen is engaged in a hydrogen bond to the NH of Gly219 (3.1 Å) and as mentioned with the aminomethylene anchor (3.2 Å) while the other oxygen is pointing towards the bulk solvent. The intramolecularly preorganized binding geometry within this ligand series forces the terminal benzyl group to adopt an orientation that largely exposes this group to the bulk solvent without additional van-der-Waals contacts since it is mostly surrounded by hydrophilic amino acids (Arg221, Glu217, Lys224).

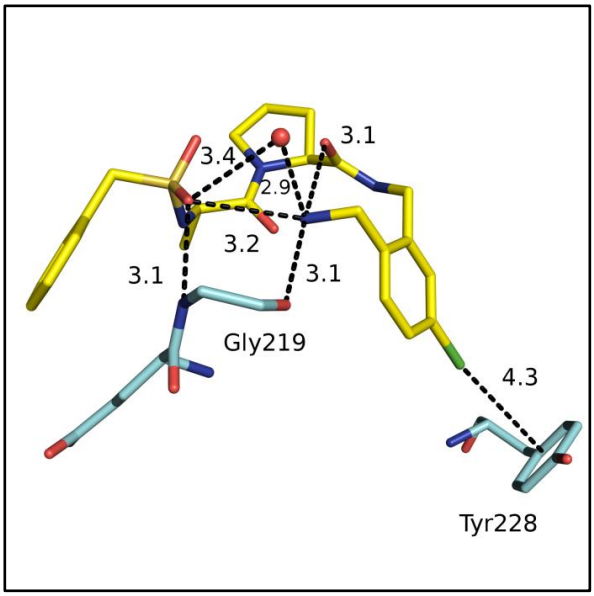
4.4.3 Binding mode of the scaffold of the AMBA series

The scaffold of ligand **7** will be used to describe the binding features within the AMBA series. The AMBA anchor is placed into the S1 pocket to form a salt bridge with short distances (2.9 and 2.8 Å) to the deprotonated Asp189 at the bottom of this pocket (Figure 1c). The proline portion fits perfectly under the 60's loop of the S2 pocket and the ligand's central amide NH and the adjacent carbonyl function form the above-mentioned β-ladder-type binding motif to the backbone of Gly216 with comparable distances (3.1 and 2.9 Å, Figure 1d). The sulfonyl takes a different role in this inhibitor series. It forms a short hydrogen bond (3.0 Å) to the backbone carbonyl of Gly219. In contrast to the ACB series, we could not observe any intramolecular hydrogen bonds in the AMBA examples. The lack of the aminomethylene anchor is most likely responsible for a different positioning of the terminal benzyl portion. It orients in opposite direction compared to the ACB series and it is now able to form a strong hydrophobic interaction to one of the sulfur atoms of a disulfide bridge (3.7 Å). A van-der-Waals contact is found to the carboxyl group of Glu192 with a distance of 3.6 Å. Additionally, strong intramolecular hydrophobic contacts are observed between the terminal benzyl moiety and the phenyl ring of the AMBA (3.6 Å) and the inhibitor's own P2 carbonyl oxygen (3.3 Å). The U-shaped inhibitor conformation most likely results from a hydrophobic collapse. It is worth mentioning that the flexibility of the Glu192 side chain is strongly reduced as the adjacent benzyl moiety restricts its accessible conformational space. In this

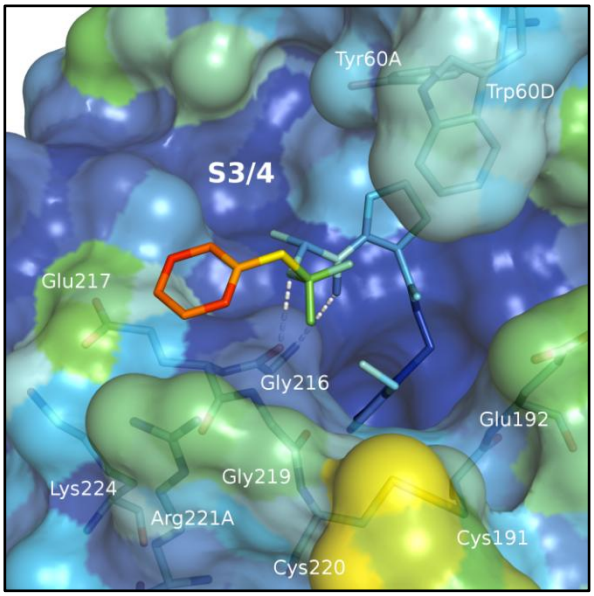
series only one single Glu192 conformer is observed, whereas it is disordered in many other thrombin complexes.[25]

Figure 1: The binding modes of ligands **2** and **7** are shown. (*a, c*) The main focus is on the preorganizing interactions between the aminomethylene and the sulfonyl group in **2** (*a*) and on the interaction pattern near the disulfide bond in **7** (*c*). (*b, d*) The binding mode of ligands **2** (*b*) and **7** (*d*) in complex with thrombin is shown with the protein's solvent-accessible surface. Colors are assigned to all atoms in (*b*) and (*d*) according to their temperature factors, from blue (low B-factor) to green and to red (high B-factor). Nitrogen is shown in blue, chlorine in green, sulfur in orange, oxygen in red and carbon in yellow (ligand site) and cyan (protein site). The distances are indicated in Å.

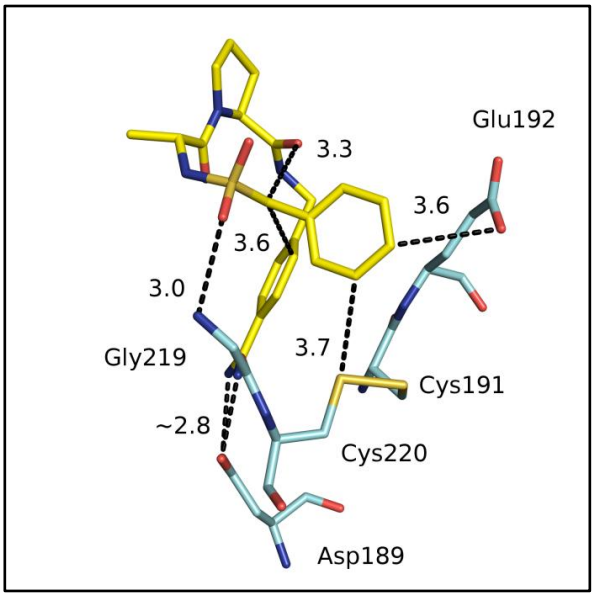
a) molecular interactions in complex with **2**



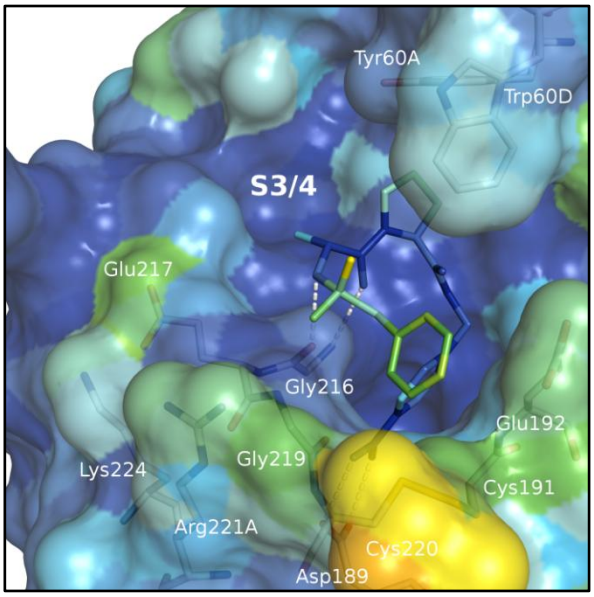
b) in complex with **2** with protein surface



c) molecular interactions in complex with **7**



d) in complex with **7** with protein surface



4.4.4 Crystallographic tracing of displacement of water molecules and implications on the binding affinity in the ACB series

In the following, focus will be attributed to the displacement of water molecules from the S3/4 pocket. Ligand **2** with P3 D-Ala leaves the hydrophobic pocket well hydrated as the methyl group penetrates only slightly and does not reach the hydration shell (Figure 2a). The distance of the methyl group to the closest water molecule is 4.0 Å. Interestingly, the water network in this pocket shows remarkable features: The most deeply buried water molecule is bound to the backbone carbonyl of Asn98 (2.9 Å) and it is oriented towards the center of the indole moiety of Trp215 (3.1 Å) to form a polar π-interaction[26]. The neighbouring water molecule is rather losely bound to the protein site (3.4 Å to Glu97A) and it is fixed in position by the two flanking water molecules (2.9 Å each). The third water is located close to the rim of the pocket and interacts with the hydroxyl group of Tyr60A (2.9 Å). Besides one van-der-Waals contact to a carbon of Trp215 (4.2 Å) the P3-methyl group is not able to form any further interactions to the hydrophobic S3/4 binding pocket. Nevertheless, **2** exhibits a binding affinity of K_i = 2.2 ± 0.4 nM although no strong interactions with the pocket residues are experienced and no water displacement could be observed.

Replacement of the P3 methyl side chain in **2** by a larger isopropyl group in **3** changes little the ligand's conformation (Figure 3). Hardly any modifications of the solvation pattern of the S3/4 pocket are apparent (Figure 2b). Obviously, the P3 D-Val seems to be too small to disrupt the S3/4 water network. The isopropyl group forms here a contact with CD1 of Ile174 (4.0 Å). Suprisingly, the isopropyl group does not point into the pocket but to the bulk solvent. This reorientation provokes a shift of the center of the terminal benzyl moiety by 1.5 Å compared to that in **2** (Figure 3). This movement is required to preserve a close contact between isopropyl group and benzyl moiety (now 3.9 Å). Furthermore, this slighly changed benzyl position induces the carboxyl group of Glu217 to adopt an orientation also found in the apo structure of thrombin whereas in the complex with **2** the carboxyl group is rotated by approximately 90° to allow binding of the benzyl portion (Figure 3). Compared to **2** no additional water molecule is displaced, the improved affinity (K_i = 0.788 ± 0.070 nM) by a factor of 3 might be related by the formation of two additional van-der-Waals contacts.

The D-Val/D-Leu exchange at P3 reveales some interesting binding features with disordered molecular portions either on the side of the protein and ligand (see Figure 2c). The P3 isobutyl side chain could be refined in two orientations with conformation A (occupancy: 58

%) inside and B (occupancy: 42 %) outside the pocket. Conformation B induces a second orientation of Ile174 preventing too close contacts with the inhibitor.

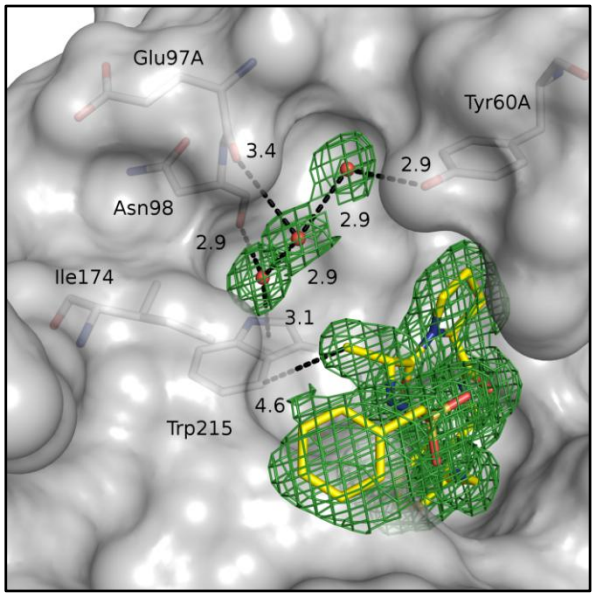
Two populated water sites can be detected in the S3/4 pocket. They would be too close to conformation A (distance of 2.3 and 2.1 Å) and refinement of these water sites with full occupancy resulted in negative density (σ = 4.5). This encouraged us to refine these sites with partial occupancy as the water molecules can only be present in unit cells hosting conformation B.

Refinement of the latter water model with an occupancy of 42 % revealed no difference electron density. The observed disorder of the P3 D-Leu portion is therefore most likely correlated to the residual hydration of the S3/4 pocket. Despite the overall hydrophobic character of the S3/4 pocket a perfect interaction network among the waters and to the protein is formed. In consequence, a significant energetic price of desolvation has to be paid for this well hydrated subsite, which has to be compensated by the binding of **4**. Apparently, as a compromise, disorder is observed where in about half of the cases the S3/4 binding pocket remains solvated. Interestingly, in a previous contribution we showed a similar example with partial replacement of water molecules upon ligand binding in case of the S1 pocket of thrombin.[5] In the present case, we assume on average replacement of one water molecule from the S3/4 pocket upon binding of **4**. The kinetic measurement revealed a binding affinity of K_i = 0.259 ± 0.024 nM which is by a factor of 3 stronger than the affinity of the D-Val derivative **3** for which no water displacement has been observed.

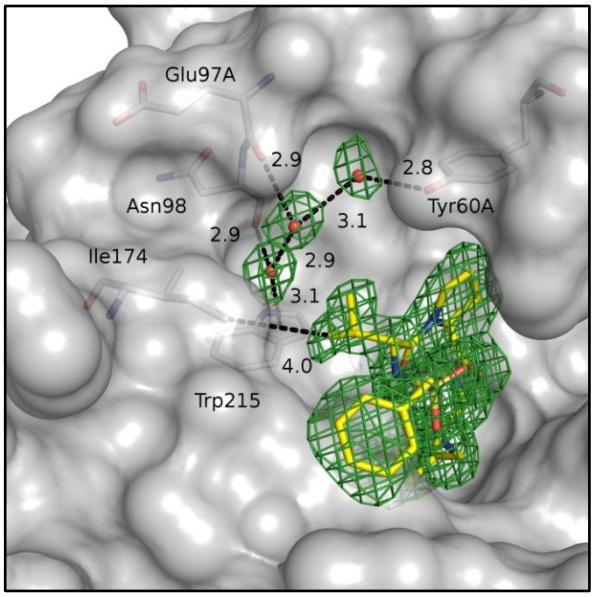
The exchange of D-Leu by D-Cha in **5** showed significant changes of the S3/4 water network with displacement of two water molecules (Figure 2d). Only one water molecule at the border of the pocket is still retained and it appears to be tightly bound to Tyr60A (2.8 Å). Detailed structural analysis shows remarkable van-der-Waals interactions between the cyclohexyl ring adopting chair conformation and the protein subsite. The contact distance to Trp215 is 3.6 Å and to Ile174 4.2 Å. Interestingly, we could not define two conformations for the sidechain of Ile174 as observed with **4**. We assume a disfavored conformation as that is not observed in the apo structure and in complexes with **2** and **3**. Some close contacts (3.4 Å) are experienced by C-H vectors at the cyclohexyl ring and the carbonyl backbone of Glu97A or the hydroxyl group of Tyr60A.

Figure 2: Binding modes of the ligands with the ACB anchor group emphasizing the structural changes in the S3/4 region and in the ligand's P3 position. The F_o-F_c difference electron density is shown in green at 2σ for inhibitor **2** (*a*), **3** (*b*), **4** (*c*) and **5** (*d*) together with the bound water molecules in the S3/4 pocket if present. Favourable interactions with the corresponding distances in Å are depicted with broken lines. Values in red indicate short distances found for the specific atom which is mentioned in the text. Nitrogen is shown in blue, chlorine in green, sulfur in orange, oxygen in red and carbon in yellow (ligand site) and white (protein site). Water molecules are red spheres.

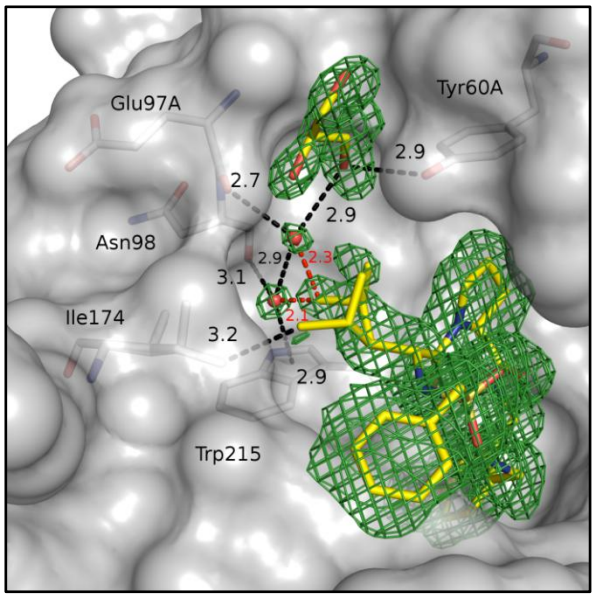
a) in complex with **2**



b) in complex with **3**



c) in complex with **4**



d) in complex with **5**

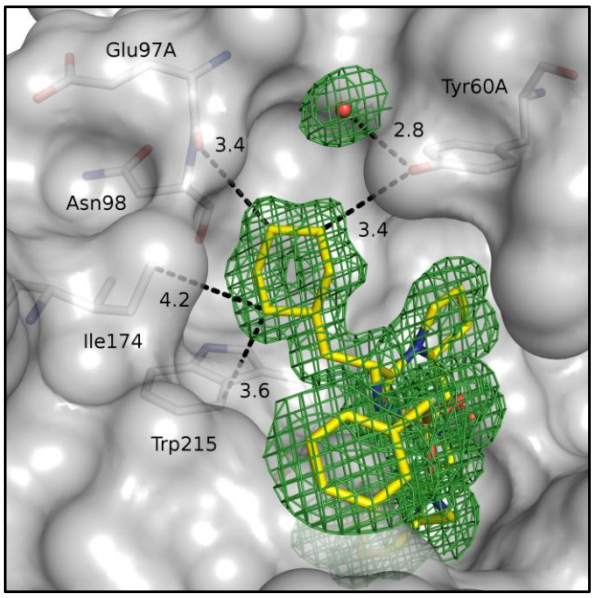
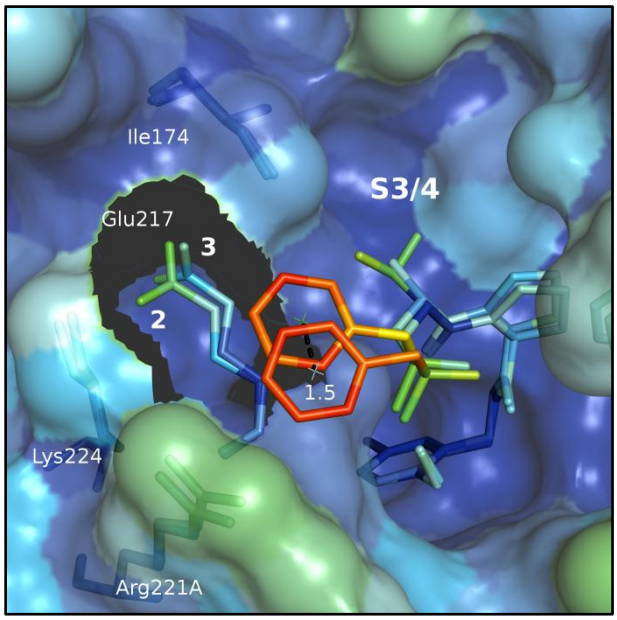


Figure 3: Superposition of ligand **2** and **3** to emphasize the shift of Glu217 and the movement of the benzyl portion of ligand **3**. Colors are assigned to all atoms according to their temperature factors, from blue (low B-factor) to green and to red (high B-factor). The solvent-accessible surface is shown. The distances with broken lines are indicated in Å.



Most likely the displacement of two firmly bound water molecules and the extensive hydrophobic contacts to the P3 cyclohexyl side chain result in a further 5-fold improved affinity (K_i = 0.052 ± 0.005 nM) compared to **4**. The overall binding affinity was improved 42-fold from 2.2 nM for **2** to 0.052 nM for **5**.

4.4.5 Crystallographic tracing of the displacement of water molecules and implications on the binding affinity in the AMBA series

Using the AMBA series, we wanted to confirm the water displacement from the S3/4 pocket and study whether the different orientation of the terminal benzyl moiety takes any impact on the interaction pattern in the hydrophobic S3/4 pocket. Unfortunately, it turned out that the displacement of the water molecules could not be analyzed directly. As expected, the D-Ala derivative **7** orients its P3 methyl group into the S3/4 pocket. However, to our surprise the remaining active site is filled by an additional inhibitor molecule showing 54 % occupancy (Figure 4a). In this case, the benzylsulfonyl portion of the second ligand fits perfectly into the S3/4 pocket resulting in partial desolvation. In agreement with the partial occupancy of this second ligand, the diffraction data indicate at a position nearly coinciding with the benzyl moiety (1.3 Å) a density peak which suggests a partially occupied water

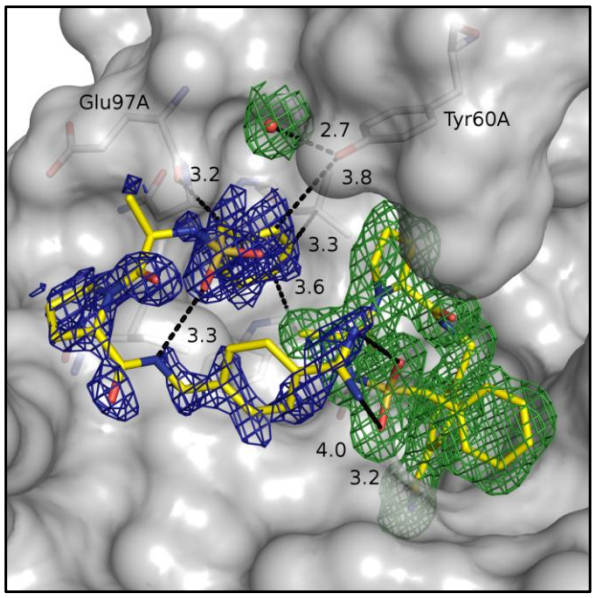
molecule as a remainder of the above described water network (Figure 5a). Occupancy of this water molecule was refined with 46 % population indicating its presence solely in absence of the second bound ligand. The hydrophobic portion of this extra ligand exhibits pronounced van-der-Waals contacts (3.3 Å to 3.6 Å) with hydrophobic residues in the active site. The sulfonyl group shows an intramolecular hydrogen bond (3.3 Å) to the NH group of its P2-P1 amide bond. The D-Ala and L-Pro portions are solvent exposed with no remarkable interactions to the protein. The positively charged amidino function is weakly hydrogen bonded (3.2 Å and 4.0 Å) to the sulfonyl part of the fully populated ligand. The overall geometry adopted by the extra ligand resembles also a U-shaped conformation. We checked whether any crystal packing effects could account for this unexpected crystallographic result. However, we could not detect any further hydrogen bonds to a symmetry-related protein molecule. The closest contact was found between the methyl group of the D-Ala portion and Glu13 of a neighbouring crystal mate exhibiting rather large distance of 5.4 Å. Ligand **7** posseses comparable potency (K_i = 5.2 ± 0.8 nM) as the corresponding D-Ala derivative of the ACB series.

The D-Val derivative **8** reveals the same behaviour also exhibiting two inhibitor molecules in the active site (Figure 4b). The isopropyl group of the first fully populated ligand points towards the S3/4 pocket. Also here the solvation pattern of this pocket is difficult to define due to the presence of a second inhibitor molecule. The binding modes of both ligands show similiarity to the previously described case**.** The extra ligand refines to an occupancy of 42 % and the adjacent water sites converge to 58 % occupancy (Figure 5b). All contacts found for **8** compare well with whose detected for **7** whereas the binding affinity of **8** is increased by a factor of 3 to K_i = 0.788 ± 0.070 nM.

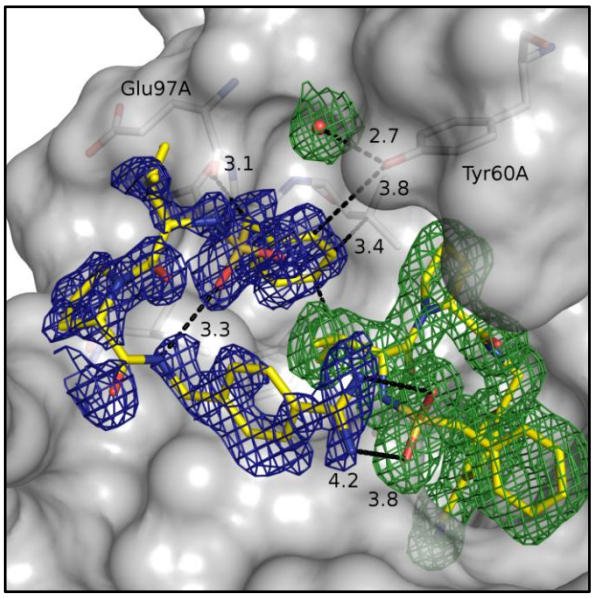
For the D-Leu derivative **9** we receive an even more complex situation in the S3/4 pocket (Figure 4c). Again, two inhibitor molecules are found in the binding site. The first ligand is fully populated and well defined. The difference electron density suggests very similar binding mode in line with the other members of the series. In contrast, the electron density of the second extra-bound ligand is poorly defined and we could only include the benzylsulfonyl part in our refinement model.

Figure 4: Binding modes of the ligands with the AMBA anchor group. Especially, the structural changes in the S3/4 region and in the ligand's P3 position are emphasized. The F_o-F_c difference electron density is shown in green for the fully populated and in blue for the partially populated ligand at 2σ for **7** (*a*), **8** (*b*), **9** (*c*) and **10** (*d*) together with the bound water molecules in the S3/4 pocket if present. Favourable interactions with the corresponding distances in Å are depicted with broken lines.

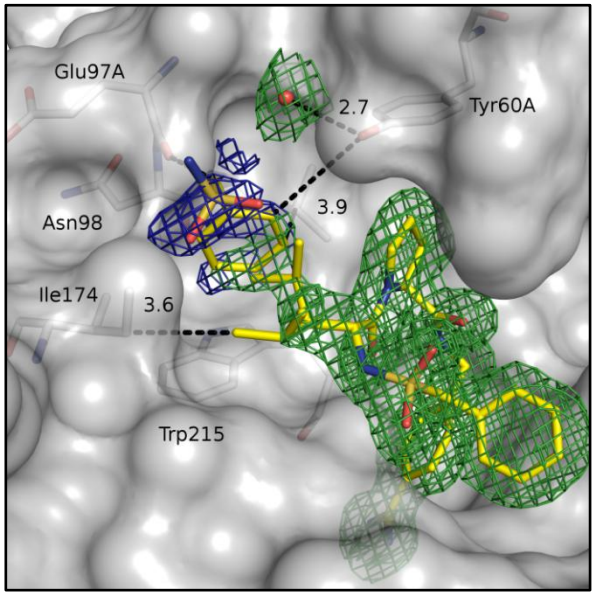
a) in complex with **7**



b) in complex with **8**



c) in complex with **9**



d) in complex with **10**

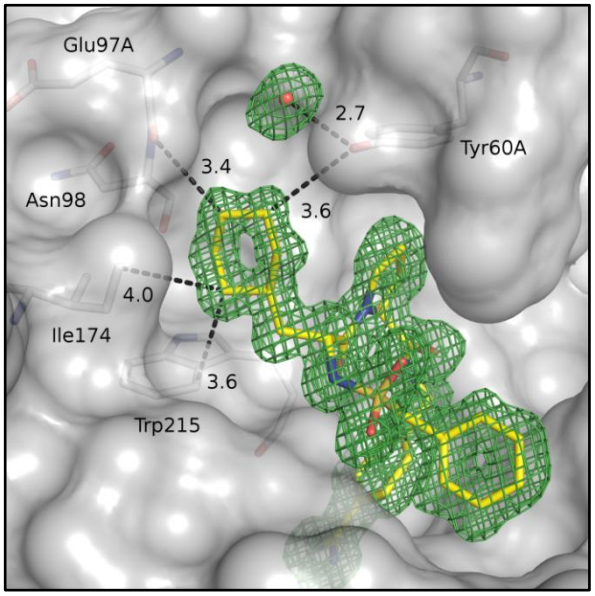
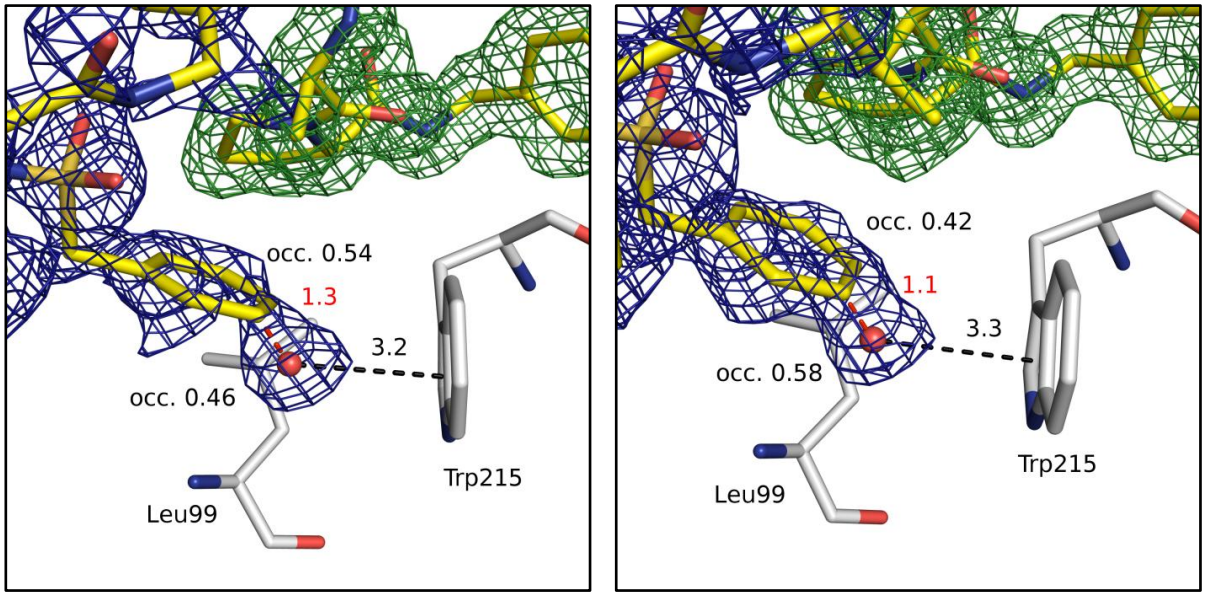


Figure 5: View from top on the S3/4 pocket to show close water contacts. The F_o-F_c difference electron density is shown in green for the fully populated and in blue for the partially populated ligand at 2σ for **7** (*a*) and **8** (*b*) together with the bound water molecule in the S3/4 pocket. Favourable interactions with the corresponding distances in Å are depicted with broken lines. Values in red indicate short distances found for the specific atoms.

a) *b*)



The D-Leu side chain of the fully populated ligand of **9** is scattered over two conformations, each populated by 50 %. The orientation outside the S3/4 pocket would allow binding of the second extra ligand forming the double hydrogen bond between the amidino group and the SO_2 group of the first molecule. Most likely, the orientation inside the pocket further reduces the occupancy of the extra ligand resulting in its minor population. In consequence, solely the benzylsulfonyl moiety could be assigned to the difference electron density. Inhibitor **9** has a slightly improved affinity likely due its larger P3 residue (K_i = 0.891 ± 0.115 nM for **9**).

The cyclohexyl portion of **10** is large enough to fully occupy the S3/4 pocket. It completely disrupts the water network concomitantly displacing two water molecules from the S3/4 pocket (Figure 4d). This ligand binds only with one copy to the active site. As no space remains in S3/4 pocket, no extra ligand binding is indicated. **10** proved to be the strongest binder in the AMBA series with an affinity of K_i = 0.119 ± 0.010 nM. The increasing hydrophobicity from methyl to cyclohexyl improved the affinity by a factor of 40. This is consistent with the first series. In general, the ACB inhibitors are always slightly more potent than the corresponding AMBA derivatives despite the lack of the salt bridge to Asp189.

4.4.6 Comparison of the binding mode of the second additionally bound ligand to the natural substrate of thrombin

Thrombin has a central role in blood coagulation. After activation from inactive prothrombin (II) to active thrombin (IIa), it converts soluble fibrinogen to insoluble fibrin which becomes crosslinked by factor XIIIa to form a stable clot.[27] Upon activation of the fibrinogen α-chain the fibrinopeptide A (FPA), which is composed of 16 amino acids (ADSGEGDFLAEGGGVR), is cleaved off. Three crystal structures of human or bovine thrombin have been determined in complex with FPA or slightly modified analogs.[28-30] Superposition of both inhibitor molecules of **8** with a FPA-analog (Figure 6) shows impressively that our bound ligands map well the binding geometry of the FPA in complex with thrombin. In early days of thrombin research it was a surprise that FPA binds with its P9 Phe into the S3/4 pocket. It then leaves the cleavage site, forms a loop of 6 residues (Leu9-Gly14) to return with its Val residue (Val15) into the S2 pocket. The second extra ligand in our study binds similarly with its benzyl portion to the S3/4 specificity pocket and the loop of FPA bulging out of the site is traced by the ligand's remaining parts.

4.4.7 Crystal structure analysis of the Gly derivatives 1 and 6

The Gly derivatives **1** and **6** of both series exhibit a different binding mode compared to the Cα substituted examples. For **1** and **6** the geometry of the respective S1 occupants and the central proline are consistent with all other members of the series. Differences occur at P3 and involve the terminal benzylsulfonyl moiety. The latter hydrophobic portion is used to occupy the S3/4 pocket which results in the displacement of two water molecules as similiarly observed for the cyclohexyl derivatives **5** and **10** (Figure 7a). The altered binding mode of **1** requires that the intramolecular water-mediated H-bond between the sulfonyl oxygen and the P1 aminomethylene anchor (present in **2**-**5**) is sacrified. The aromatic ring of the terminal benzylsulfonyl moiety in **1** orients nearly perpendicular to the indole ring of Trp215 with a distance of 3.5 Å. Ile174 adopts a single conformation and rotates out of the pocket to avoid close contact with the benzyl moiety (3.8 Å).

Figure 6: Stereo view of the active site region of thrombin in complex with the double-bounded ligand **8** and a FPA analog (PDB code 1UCY[28]) containing a stable ketomethylene group (Argψ[CO-CH_2]Gly) instead of a scissile P1-P1' peptide bond. The solvent-accessible surface of thrombin is depicted in grey. Nitrogen is shown in blue, sulfur in orange, oxygen in red and carbon in yellow for ligand **8** and in cyan for the FPA derivative.

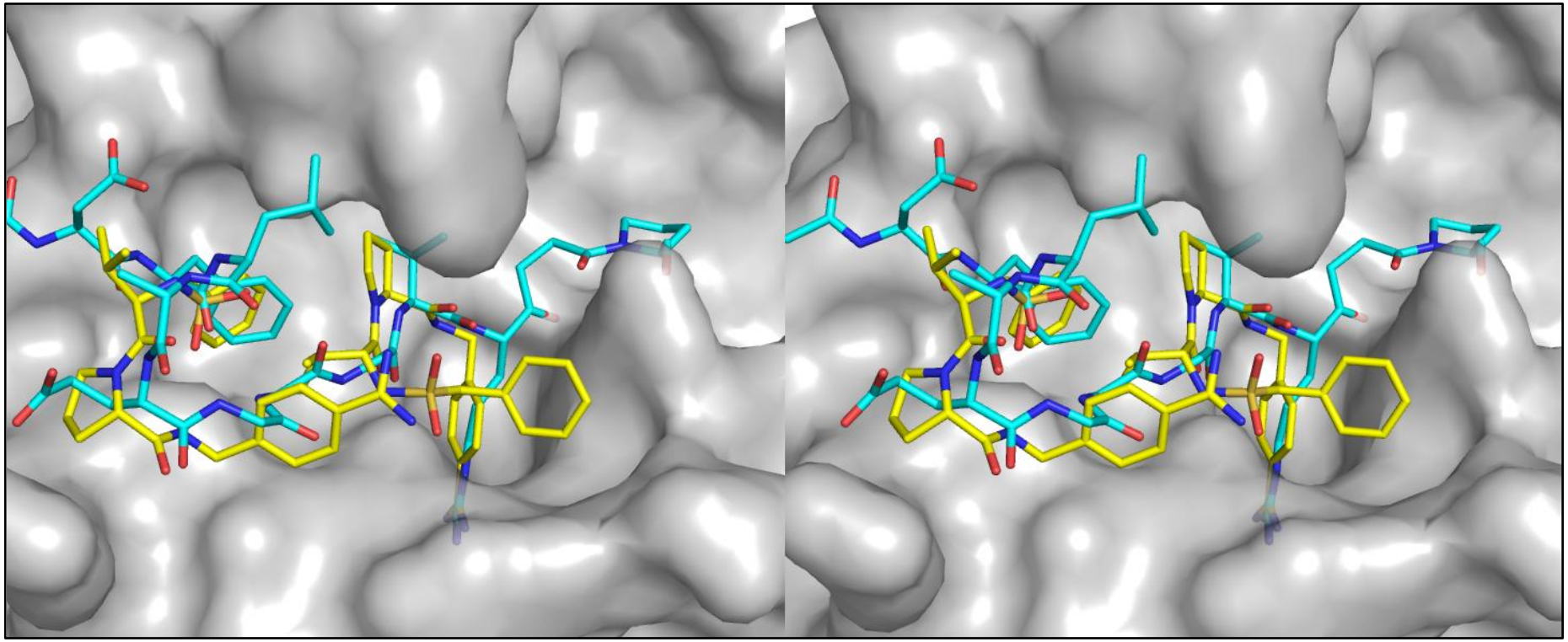
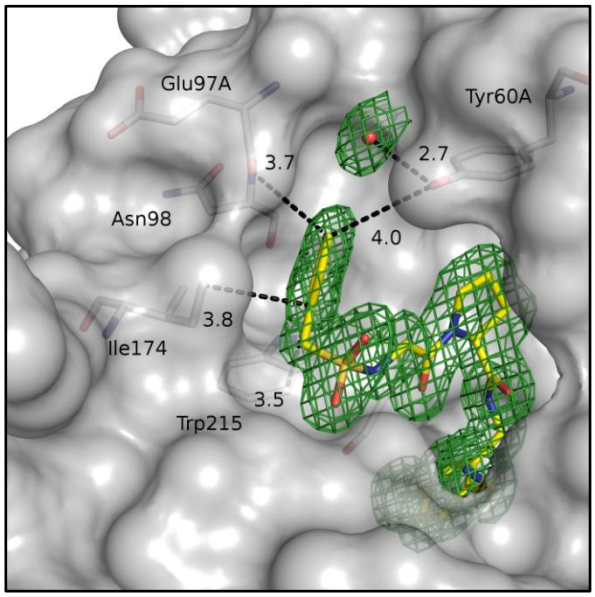
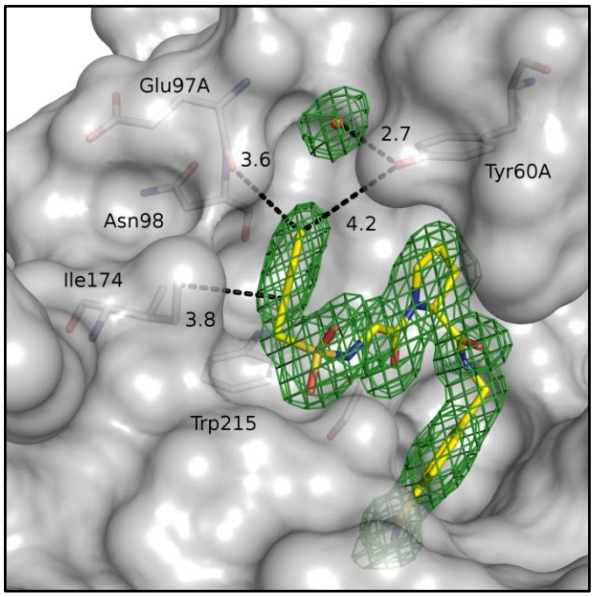


Figure 7: Binding mode of the glycine derivatives **1** (*a*) and **6** (*b*). The F_o-F_c difference electron density is shown in green at 2σ. Favourable interactions with the corresponding distances in Å are depicted with broken lines.

a) in complex with **1**



b) in complex with **6**



Both sulfonyl oxygens are exposed to the solvent. Despite of the displacement of two water molecules from the S3/4 pocket and remarkable van-der-Waals contacts, the Gly derivative **1** exhibits no significant gain in binding affinity (1.5 ± 0.1 nM) compared to the D-Ala analog **2** where the pocket is completely solvated.

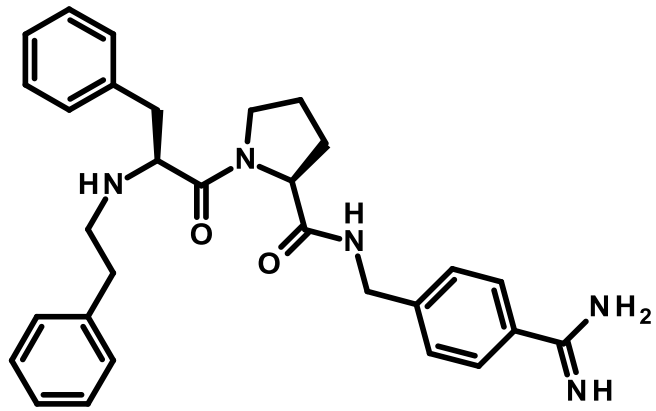
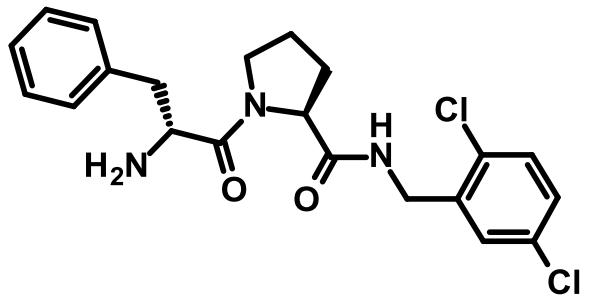
The AMBA analog **6** adopts a comparable binding mode (Figure 7b). Again its benzylsulfonyl portion is located in the S3/4 pocket and the sulfonyl oxygens are oriented towards the bulk solvent. Also this ligand exhibits nearly the same affinity as the D-Ala analog (3.7 ± 0.6 nM for **6**).

4.4.8 Thermodynamic characterization of the hydrophobic interaction in the S3/4 pocket of thrombin

We used Isothermal Titration Calorimetry (ITC) to determine the thermodynamic profiles of both series, and to factorize Gibbs free energy of binding into enthalpic and entropic components. ITC is usually restricted to a range up to 5 nM using direct titrations. However, inhibitors investigated in this study show high potency from nano to picomolar range. In such situations direct measurements of tight binders is not possible and we therefore decided to apply displacement titrations to attain an accurate thermodynamic profile.

Hence, the dissociation constant of the strong binder must be formally lowered to a range which can be recorded accurately by ITC.[31] In the displacement strategy, the protein solution is saturated by a „weak“ binder such as **11** or **12** (Table 2). In the following step the tight binder is titrated into the protein solution saturated by the first ligand (see Material & Methods for details). During this titration the „weak“ binder is displaced by the strong one and the resulting sigmoidal titration curve can be analyzed according the displacement theory developed by Sigurskjold[32].

Table 2: The chemical structures used in the displacement titration as competitive binder together with the thermodynamically determined dissociation contants (K_D in nM) towards human thrombin.

Competitive Ligand 1 (**11**) 	**232 ± 48**
Competitive Ligand 2 (**12**) 	**255 ± 16**

The error is given as the standard deviation calculated from at least three measurements (in nM).

The measured heat signals showed dependence on the three buffers applied for the titrations (Table 3). This indicated protonation reactions superimposed onto the binding event. The net binding enthalpy can be extracted from these titrations by extrapolation (Figure 8). We observed comparable dependencies for all studied ligands suggesting a release of n = 0.43 ± 0.12 mol protons upon binding (Fig. 9). For a related series of ligands

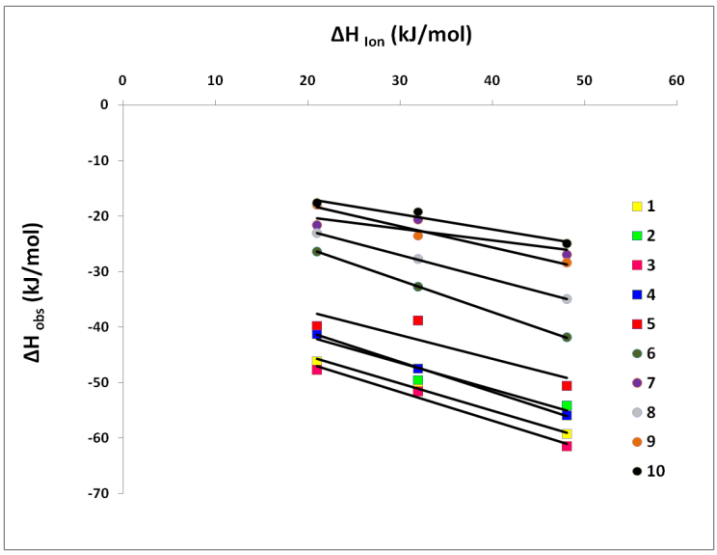
showing similar effects, Baum *et al.*[25] pointed out that His57 is responsible for the observed buffer dependency. As the functional groups in our ligands that could change protonation state (amidino group, aminomethylene group, sulfonamide group) are either too basic or very close to neutral pH, we anticipate that no ligand functional groups are involved in the protonation changes. Therefore, more likely the imidazole moiety of His57 is partially protonated in the uncomplexed state and releases protons upon ligand accommodation.[33] After buffer correction, factorization of the Gibbs free energy shows increasingly entropy-driven binding which is consistent with the classical hydrophobic effect. The D-Ala derivative **2** reveals a thermodynamic profile which is clearly dominated by its enthalpic component (ΔH = - 32.1 kJ/mol for **2**). The entropic contribution amounts to about the half of the enthalpy term (- $T\Delta S$ = - 15.1 kJ/mol for **2**). With growing size of the hydrophobic P3 substituent we observe a constant increase of the entropic component and a slight decrease of the enthalpic signal. This thermodynamic signature is reversed from an enthalpy- to a more entropy-driven binding as larger hydrophobic portions are introduced at P3. Ligand **5** with largest hydrophobic portion binds with a balanced thermodynamic profile where both, the enthalpy and entropy term, contribute almost equally to binding affinity (ΔH = - 28.7 kJ/mol and - $T\Delta S$ = - 25.8 kJ/mol for **5**). In our series we observe a slight overall decrease of enthalpy ($\Delta\Delta H_{2\text{-}5}$ = - 3.4 kJ/mol) however, fortunately the entropy term overcompensates this enthalpy loss (– $T\Delta\Delta S_{2\text{-}5}$ = 10.7 kJ/mol). Therefore, a net gain in Gibbs free energy could be observed of $\Delta\Delta G_{2\text{-}5}$ = 7.3 kJ/mol.

Table 3: Binding data ΔG^0, ΔH^0 and $-T\Delta S^0$ (kJ/mol) of ligands determined by ITC.

Ligand name	ΔG^0 (kJ/mol)	ΔH^0 (kJ/mol) in tris	ΔH^0 (kJ/mol) in tricine	ΔH^0 (kJ/mol) in hepes	ΔH^0 (kJ/mol) buffer corrected	$-T\Delta S^0$ (kJ/mol)	Number of proton release (mol/mol)
1	- 47.8 ± 0.1	- 59.3 ± 1.3	- 50.5	- 46.1	**- 35.4**	- 12.4	0.49
2	- 47.2 ± 0.9	- 54.2 ± 0.6	- 49.6	- 40.9	**- 32.1**	- 15.1	0.48
3	- 54.2 ± 0.2	- 61.5 ± 0.7	- 51.6	- 47.7	**- 36.2**	- 18.0	0.52
4	- 54.3 ± 1.4	- 55.9 ± 1.9	- 47.5	- 41.3	**- 30.1**	- 24.2	0.54
5	- 54.5 ± 0.1	- 50.6 ± 0.1	- 38.8	- 39.8	**- 28.7**	- 25.8	0.43
6	- 44.9 ± 1.1	- 41.9 ± 0.1	- 32.7	- 26.3	**- 14.3**	- 30.6	0.58
7	- 42.6 ± 0.2	- 26.9 ± 0.5	- 20.6	- 21.6	**- 15.9**	- 26.7	0.21
8	- 47.1 ± 0.1	- 35.0 ± 1.0	- 27.7	- 23.1	**- 13.8**	- 33.3	0.44
9	- 51.9 ± 1.0	- 28.4 ± 2.3	- 23.5	- 17.9	**- 10.4**	- 41.5	0.38
10	- 53.7 ± 0.8	- 24.9 ± 0.1	- 19.3	- 17.7	**- 11.4**	- 42.3	0.27
11	- 37.9 ± 0.5	- 17.1 ± 1.1				- 20.8 ± 1.7	
12	- 37.6 ± 0.2	- 37.5 ± 0.1				- 0.1 ± 0.2	

Ligands **1**-**10** were determined in a displacement titration using ligands **11** or **12** as competitive binder. The thermodynamic data of **11** and **12** were determined in a direct ITC titration in tris buffer. ΔH^0 was measured in three different buffers. Errors indicate the standard deviation from at least duplicate measurements. Since the one point measurements in hepes and tricine indicate the expected constant rate of proton release, these experiments were not performed in duplicate in order to save protein material (Figure 8).

Figure 8: The measured heat signal is plotted against the ionization enthalpy of the buffer system (Tris, hepes and tricine) in order to extract the enthalpy of binding. The circles represent the values from the ACB series (**1**-**5**) and the squares the ligands from the AMBA series (**6**-**10**).

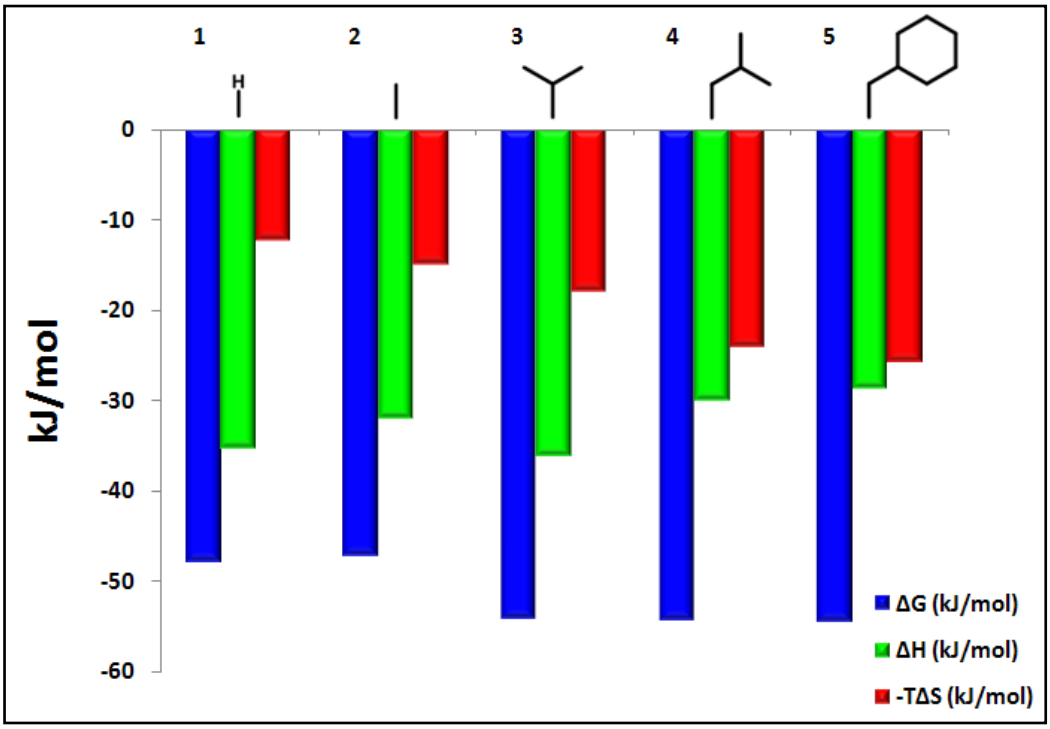


Comparable results are observed for the AMBA series, however the thermodynamic signature starts from a different partitioning (Figure 9b). Here, the entropy term clearly governs the binding across the entire series (ΔH = - 15.9 kJ/mol and - TΔS = - 26.7 kJ/mol for **7**). **6** falls out of the series, but from **7** to **10** the entropic influence gradually increases during hydrophobic optimization, consistent with the findings in the ACB series. The entropic component of **10** is about four times the enthalpic signal (ΔH = - 11.4 kJ/mol and - TΔS = - 42.3 kJ/mol for **10**) but the entropic enhancement is comparable to the ACB series. Again, we experience a slight decrease in the enthalpic term ($\Delta\Delta H_{7\text{-}10}$ = - 4.5 kJ/mol) overbalanced by a huge favorable entropic contribution (– $T\Delta\Delta S_{7\text{-}10}$ = 15.5 kJ/mol).

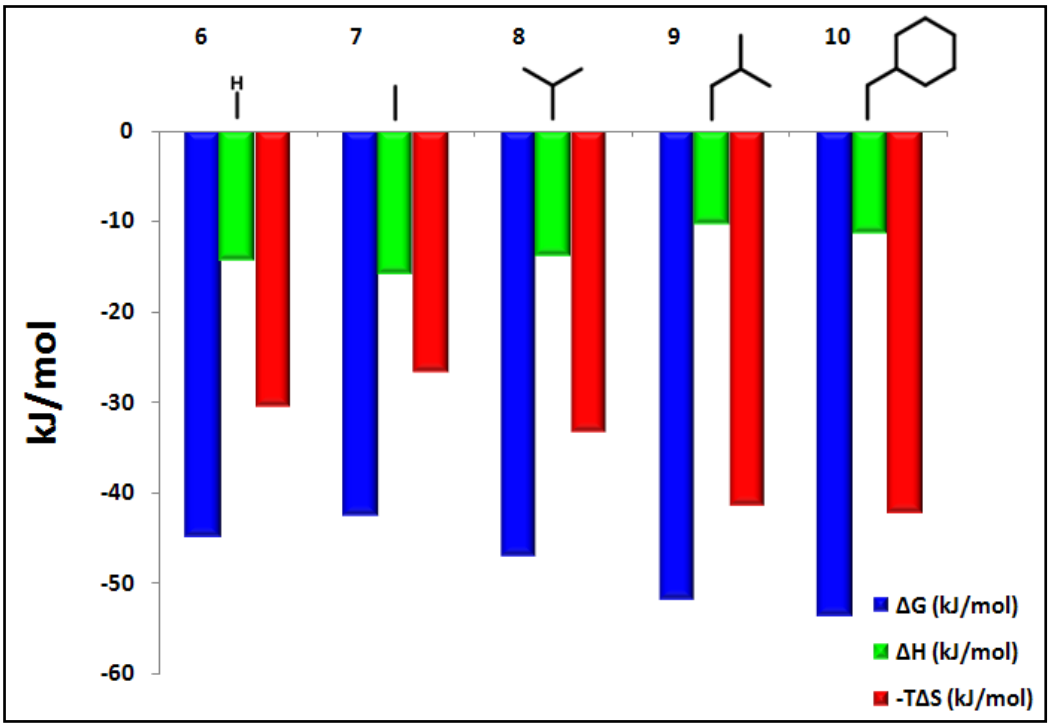
The Gly derivatives **1** and **6** adopt quite different binding modes compared to the remaining examples of the two series. Surprisingly, this is not reflected in the thermodynamic profiles, as those for **1** and **6** fit quite reasonably into the series. Therefore a closer comparison of the corresponding Gly and D-Ala derivatives should be regarded. Although **1** displaces two water molecules from the S3/4 pocket, we could only observe overall a sligthly increased entropy component (– $T\Delta\Delta S_{1\text{-}2}$ = 2.7 kJ/mol) and basically no impact on the Gibbs free energy ($\Delta\Delta G_{1\text{-}2}$ = - 0.6 kJ/mol) in comparison to **2.** As both ligands position their benzyl moieties differently, we must consider whether a displacement of water molecules might occur at the hydrophilic

Figure 9: Thermodynamic parameters determined by ITC for ligands **1**-**5** (*a*) and **6**-**10** (*b*). The buffer corrected enthalpy terms are shown.

a) ACB series



b) AMBA series



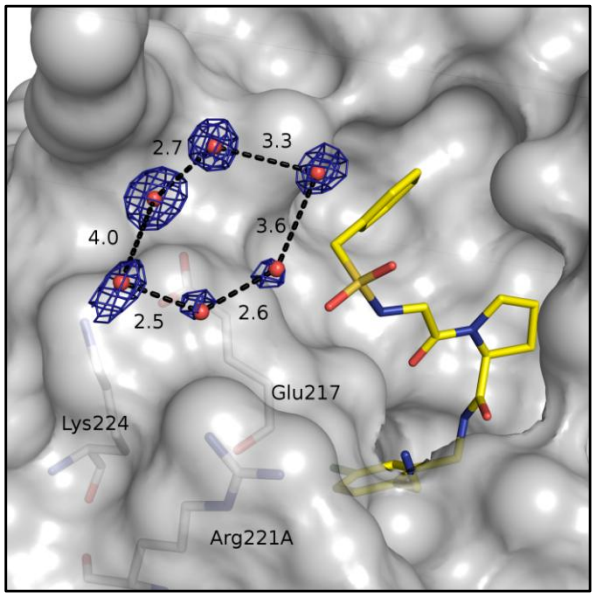
site next to Glu217, Lys224 and Arg221 which is occupied in case of **2**. **1** is able to desolvate the S3/4 pocket but leaves the adjacent Glu217/Lys224/Arg221 region solvated (Figure 10a). In contrast, ligand **2** leaves the S3/4 pocket solvated but disrupts the solvation shell in the neighboring Glu217/Lys224/Arg221 region through accommodation of the benzyl portion (Figure 10b). Due to its polar character this site is well solvated and the crystal structure of **1** leaving this area unoccupied does show here six bound water molecules whereas in **2** only three water molecules are present. Possibly the water replacement in different subpockets observed for **1** and **2** might result in very similar thermodynamic signature. Furthermore, we observe a difference in the residual mobility of the benzyl portions of **1** and **2** which should also influence the entropic term. We analyzed the B factor ratio's of the benzyl portions relative to the remaining inhibitor molecules in order to estimate the degree of mobility. The benzylsulfonyl group of **1** exhibits low mobility [B (benzyl)/B (P1+P2+P3) = 1.26] as it is tightly fixed in the S3/4 pocket whereas in **2** this portion shows substantially higher mobility [B (benzylsulfonyl)/B (P1+P2+P3) = 2.06]. This enhanced motion of the benzylsulfonyl portion is a common structural feature of **2**-**5** (1.90 for **3**, 1.69 for **4** and 1.77 for **5**). The polar region next to Glu217/Lys224/Arg221A (Figure 3) is structured as a rather shallow crevice which is not suited to perform strong interactions with benzyl portion at this protein site. As a consequence, we observe this portion with pronounced residual mobility.

A similar situation is observed for **6** and **7**. Two water molecules are released to the bulk solvent by **6** but now with a slight advantage with respect to the entropy ($-T\Delta\Delta S_{6\text{-}7}$ = - 3.9 kJ/mol). This parallels a small loss in the free energy term ($\Delta\Delta G_{6\text{-}7}$ = - 2.3 kJ/mol). Unfortunately, we can not discuss the hydration of the binding pocket as done in **1**/**2** because the second bound ligand molecule in the AMBA series prevents detailed insights. Suprisingly, the study of the B factor ratio's for the benzylsulfonyl portions reveals in this series equivalent values suggesting equal mobility [B (benzyl)/B (P1+P2+P3) = 1.58 for **6** and B (benzyl)/B (P1+P2+P3) = 1.51 for **7**]. All benzylsulfonyl moieties placed in this region next to the disulfide bridge reveal less mobility compared to the corresponding examples of the ACB series. It shows that the vicinity to the disulfide bridge fixes the benzyl portion in **7** better than the polar environment of Glu217/Lys224/Arg221A in **2** (Figure 1b and d). This positioning of the phenyl ring is also known from inhibitor complexes with the related serine proteases factor Xa[(34)] and urokinase[(35)].

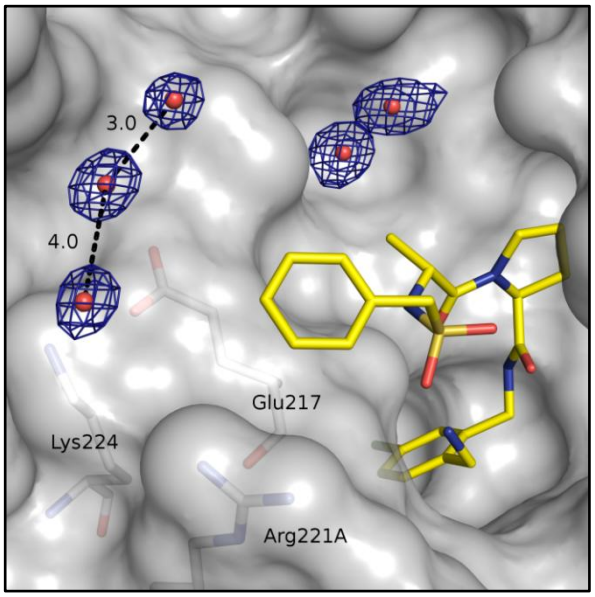
As a further aspect, that will affect the partitioning of enthalpy and entropy of the Gly and D-Ala derivatives, the conformational properties in solution has to be considered. The Gly derivative which lacks a Cα substituent will gain access to a larger conformational space in solution prior to protein binding compared to the D-Ala example. This will be related to a stronger loss in the entropic position of the Gly derivatives upon binding. This might compensate to some degree for the entropically beneficial water release and make the entropic signal less pronounced for **1** and **6**.

Figure 10: Binding mode of **1** (*a*) and **2** (*b*). The $2F_o$-F_c contour electron density is shown in blue at 1σ for water molecules. Favourable interactions with the corresponding distances in Å are depicted with broken lines.

a) in complex with **1**



b) in complex with **2**



4.5 Conclusion

We performed a comprehensive study on water displacement upon binding of hydrophobic ligand portions to thrombin's well hydrated S3/4 pocket. We wanted to collect evidence for thermodynamic signature that drives the effect. Several studies in literature showed that hydrophobic interactions are much more complex than generally believed as in some examples the binding of hydrophobic species is accompanied with a clearly favorable enthalpy component instead of the expected entropy signal. As two studies[14,15] propose that a poorly defined hydration state of the protein pocket could be an explanation for an enthalpy-driven hydrophobic effect, we decided to address a hydrophobic pocket which is

according to a high-resolution structure of uncomplexed thrombin well hydrated and accessible to water molecules. Additionally, the study was performed with pico to nanomolar ligands which is in terms of thermodynamic characterization still a rather rarely explored field. The applied displacement titration are elaborate as two titrations are necessary but the results are reliable and well reproducible.

Two series with different S1 occupants were designed and synthesized which were systematically varied by hydrophobic residues (Gly, D-Ala, D-Val, D-Leu and D-Cha) in the P3 position addressing the S3/4 pocket. In both series the affinity improvement is correlated with the size of the hydrophobic P3 side chain except the Gly inhibitors. The largest P3 substituent, D-Cha, shows highest affinity whereas D-Ala in this position is in both series the weakest binder. The ACB series was optimized by a factor of 42 which is similar to the factor obtained in the AMBA series (40).

As mentioned in the introduction, the ACB anchor is well known to show high potency towards thrombin but we found only two deposited thrombin complexes with this type of ACB anchor group (PDB codes 1ZRB[36] and 3EQ0). These sparse structural data are quite surprising since the ACB portion is high potent and less basic which may lead to even better pharmacokinetic properties.[(20)] In order to better understand the superior properties of this portion we elucidated the structural and energetic basis for the high potency of the ACB anchor group which turned out in our study to possess always slightly better binding compared to the corresponding AMBA anchor.

We observed a significant change in the binding mode of the P4-P3 segment moving from the Gly derivatives to Cα substituted residues. The Gly derivatives **1** and **6** occupy the S3/4 pocket and displace two water molecules whereas the corresponding D-Ala analogs **2** and **7** do not penetrate this pocket and leave it solvated. Suprisingly, no significant differences in the thermodynamic profile of these ligands are observed indicating that unchanged thermodynamic signature is by no means a reliable indicator for conserved binding modes. The most likely experienced beneficial entropic contributions resulting from water displacement by **1** and **6** are compensated as another site remains hydrated, differences in the residual mobility of the alternatively placed positions are experienced and differences in the accessible conformational space in solution prior to binding have to be taken into account.

The structural investigation of the AMBA series revealed some surprising results. To the best of our knowledge it is the first reported case where a small size peptidomimetic is bound twice to the active site of thrombin. Superimposing the binding modes of both inhibitor skeletons with that of the cleavage product (FPA) of the natural substrate (Fibrinogen) revealed remarkable similarities with their binding regions. It is worth mentioning that the soaking procedure was carried out at rather high inhibitor concentrations. Since we could not find evidence that the binding of the second extra ligands is only supported by crystal packing effects, possibly this additional binding mode could also be of some relevance in solution when working with high inhibitor concentration. Nevertheless, we assume the observed second bound molecule is of no relevance for enzyme inhibition under *in vivo* conditions. Interestingly, it can provide a remarkable mapping of the substrate recognition area on the surface of the protease.

The thermodynamic characterization clearly suggests that the stepwise increase of the hydrophobic P3 substituent results in a growing potency which is associated with an enhanced entropic term. Both series started out with the Gly/D-Ala derivatives having the smallest entropic contribution and end up with the cyclohexyl derivative with a thermodynamic profile, remarkably dominated by its entropic component. In this respect, our results support the current understanding of the classical hydrophobic effect being mainly of entropy driven nature and resulting from the release of firmly fixed water molecules from a well-hydrated pocket.

4.6 Acknowledgements

We kindly acknowledge CSL Behring, Marburg, for supplying us with generous amounts of human thrombin from the production of Beriplast®.

We thank the beamline support staff at SLS and BESSY for their advice during data collection. This work was supported by the Bundesministerium für Bildung und Forschung (BMBF, Förderkennzeichen 0315161C) and by the ERC grant 268145-DrugProfilBind kindly provided by the EU.

4.7 Materials and Methods

4.7.1 Bioassay

We characterized the potent inhibitors by a fluorogenic assay. Kinetic data were obtained using Tos-Gly-Pro-Arg-AMC[37] (tosyl-Gly-Pro-Arg-aminomethylcoumarin) as the fluorogenic substrate with a Safire II platereader (Tecan, Schweiz, *ex* = 380 nm, *em* = 460 nm). This substrate allows us to work at a protein concentration of 0.050 nM which is sufficiently lower than the lowest applied ligand concentration (0.780 nM). The experimental buffer contained 50 mM Tris-HCL, 154 mM NaCl, 0.1 % polyethylene glycol 8000 and 5 % DMSO at pH 7.4. The K_m of the substrate (1.1 ± 0.4 µM) was measured at ten different substrate concentrations and the resulting curve was analyzed using GraFit 4 software[38]. Cleavage of the substrate was measured by monitoring the change in fluorescence over a dilution series of at least ten inhibitor concentrations (500 nM – 0.780 nM) at 5 µM substrate (S) over 600 sec. The fluoroscence signal was plotted against time and after linear regression the reaction rates (v) were calculated. The latter values (v) were plotted against the respective inhibitor concentration (I) and the resulting curve was fitted with ORIGIN software using equation 1. All measurements were performed at least in triplicate.

$$v = \frac{V_{max} \cdot S}{K_m \cdot \left(1 + \frac{I}{K_i}\right) + S} \qquad \text{Equation 1}$$

4.7.2 Isothermal titration calorimetry

ITC experiments were performed using an ITC200™ system from Microcal (now part of GE Healthcare, Northhampton, MA, USA). Thrombin Beriplast® was freshly prepared for each experiment by dialysis of a thrombin sample in the buffer used for titration experiments (50 mM Tris, 100 mM NaCl, 0.1 % polyethylene glycol 8000, 3 % DMSO, pH 7.8). The ITC displacement experiment is composed of two titrations. The first experiment was done with one of the "weak-binding" ligands in a typical direct titration. For this purpose, after dialysis the final **11** or **12** concentration was achieved after dilution of the stock solution (40 mM, 100 % DMSO) to the required ligand concentration (0.5 mM). The final DMSO concentration was subsequently adjusted to 3 %. This titration was started after the syringe was placed in the sample cell containing the thrombin solution (27 µM). The collected data were analyzed using ORIGIN Software (Microcal Inc.), by fitting a single-site-binding isotherm that yields

ΔH^0 (enthalpy of binding) and K_D (dissociation constant). These data represent the thermodynamic profiles of the "weak-binding" ligands which are important for further analysis of the strong binders. After this first direct titration the syringe was cleaned and filled with a solution of a strong binder (0.5 mM). The syringe was then placed into the sample cell which now contains the thrombin solution saturated with the "weak-binding" ligand from the first titration. During the titration the strong ligand displaced step-by-step the "weak-binding" ligand from the binding pocket of thrombin. The experiment was stopped after small peaks of dilution indicated complete displacement of the "weak-binding" ligand from the binding site by the strong binder. The resulting curves were then analyzed using the competitive binding fitting function implemented in the ORIGIN Software which was originally developed by Sigurskjold[32]. Each strong binder was characterized by this protocol applying two titrations in tris buffer. These experiments were performed at least in duplicate.

ITC experiments in different buffers indicate a protonation reaction upon binding of the strong binders. The thermodynamic data were thus collected in addition to tris buffer in 50 mM tricine and 50 mM hepes buffer in a direct titration to achieve just the enthalpy component. Since these one point measurements indicate the expected constant rate of proton release these experiments were not performed in duplicate in order to save protein material. The observed enthalpy values were plotted against the enthalpy of ionization of the applied buffer and fitted by linear regression to determine the intersection point with the y axis which represents the enthalpy of binding corrected by the buffer contribution (Figure 8).

All ITC experiments were started at 25 °C after a stable baseline had been achieved. The experimental design comprises an initial ligand injection of 0.3 µL followed by 15 injections of 1.1 - 1.4 µL with a 180 s interval between each injection.

Raw data were collected and the area under each peak was integrated, followed by correction for heats of dilution and mixing by subtracting the final baseline consisting of small peaks of the same size to zero. The initial data point was deleted from the integrated data because this injection usually reflects an erroneous amount of heat due to the possible exchange of liquids between syringe and cell when inserting the syringe into the calorimetric cell and the backlash error in the motorized screw mechanism in the injector.[39]

4.7.3 Crystallization and soaking

Human α-thrombin (from Enzyme Research Laboratories, South Bend, USA) was dissolved in the crystallization buffer (20 mM NaH_2PO_4, 350 mM NaCl, 2 mM benzamidine, pH 7.5) at 10 mg/ml. A hirudin fragment called Acetyl-Hirudin (54-65) achieved from Bachem (Bubendorf, Switzerland) was dissolved in crystallisation buffer at 2.5 mg/mL. In the next step, 40 µL of the solution of the hirudin fragment was mixed with 160 µL of the thrombin solution. After incubation for 2 h at 4 °C, crystallization was carried out at 4 °C by the hanging-drop method. One 1 µL of the hirudin/thrombin solution was placed in the centre of a cover slip and mixed with 1 µL reservoir solution (20 mM NaH_2PO_4, 27 % polyethylene glycol 8000, pH 7.5). Immediately after the mixing of protein and reservoir buffer microseeding was done. The wells of the crystallization trays were filled with 500 µL of the reservoir buffer. Subsequently the cover slips were placed over the wells and sealed. Crystals of good diffracting quality were obtained after 7 days. For soaking, DMSO stock solutions of the inhibitors (50 mM) were diluted 1:10 with a solution containing 50 % crystallization and 50 % reservoir buffer resulting in the final soaking concentration containing 5 mM of the inhibitor and 10 % DMSO. Medium-size crystals without visible imperfections were selected and transferred into the soaking solution for 24 h.

4.7.4 Data collection and processing

Crystals were prepared for data collection at 110 K using a cryoprotectant solution of 20 % glycerol in reservoir buffer. The data sets for **1**, **2**, **5**, **6** and **7** were collected with synchrotron radiation at SLS (Villingen, Switzerland) on a Marmosaic 225 mm CCD detector. Complex structures for **3**, **4, 8, 9** and **10** were collected at BESSY beamline 14.2 (Berlin, Germany) on a Rayonix MX 225 CCD detector. Data processing and scaling were performed using the HKL2000 package[(40)].

4.7.5 Structure determination and refinement

The coordinates of human thrombin (PDB code 1H8D)[(41)] were used for initial rigid body refinement of the protein molecules followed by repeated cycles of maximum likelihood energy minimization, simulated annealing and B-factor refinement using the CNS program package[(42)]. Structures **1-10** were refined with PHENIX[(43)]. The temperature factors for

structures **5**, **8** and **10** were anisotropically refined whereas for structures **1**, **2**, **3**, **4**, **6, 7** and **9** TLS refinement was applied. The definition of the TLS groups was done with the TLSMD server[44,45]. A randomly chosen 5 % of all data were used for the calculation of R_{free} and were not used in the refinement. Amino acid side chains were fit into σ-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps using Coot[46]. After the first refinement cycle, water molecules and subsequently ions and ligands were located in the electron density and added to the model. Restraints were applied to bond lengths and angles, planarity of aromatic rings and van der Waals contacts. Multiple side chain conformations were built in case an appropriate electron density was observed and maintained during the refinement, and if the minor populated side chain showed at least 20 % occupancy. The final models were validated using PHENIX own validation options or MolProbity[47]. The Ramachandran plot's were calculated with PROCHECK[48]. Data collection, unit cell parameters and refinement statistics are given in Table 4. Analysis of temperature factors was done with Moleman[49] and fconv[50]. The naming of the protein amino acids was done according to Bode *et al.* [51]. The figures were prepared using Pymol 0.99. The Protein Data Bank accession codes of the coordinates and structure factors of all X-ray structures are given in Table 4.

4.7.6 Protein Data Bank and accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with the following accession codes: THR-**1** complex 3RML; THR-**2** complex 3RMM; THR-**3** complex 3RMN; THR-**4** complex 3T5F ; THR-**5** complex 3RMO; THR-**6** complex 3RLW; THR-7 complex 3RLY; THR-**8** complex 3RM0; THR-**9** complex 3UWJ; THR-**10** complex 3RM2.

Table 4: Data collection and refinement statistics for the thrombin complex structures **1-10**.

Structure (PDB entry)	THR-**1** complex **(3RML)**	THR-**2** complex **(3RMM)**	THR-**3** complex **(3RMN)**	THR-**4** complex **(3T5F)**	THR-**5** complex **(3RMO)**
A. Data collection and processing					
No. Crystals used	1	1	1	1	1
Wavelength (Å)	1.00	1.00	0.91841	0.91841	1.00
Space group	C2	C2	C2	C2	C2
Unit cell parameters					
a, b, c (Å)	69.3, 71.2, 72.4	70.1, 71.1, 72.7	69.9, 71.1,72.9	70.2, 71.0, 73.0	70.1, 71.0, 72.8
β (°)	99.9	100.4	100.6	100.5	100.4
Matthews coefficient ($Å^3$/Da)	2.5	2.5	2.5	2.5	2.5
Solvent content (%)	50	51	51	51	51
B. Diffraction data[a]					
Resolution range (Å)	50 – 1.53	50 - 1.58	50 – 1.78	50 – 1.45	50 – 1.40
	(1.56 – 1.53)	(1.61 – 1.58)	(1.81 – 1.78)	(1.48 – 1.45)	(1.42 – 1.40)
Unique reflections	50,478 (2,436)	48,248 (2,403)	32,972 (1,610)	60,276 (2,698)	64,957 (3,195)
R(I)sym (%)	3.5 (35.7)	5.8 (47.8)	7.8 (47.6)	3.5 (33.9)	3.4 (37.0)
Completeness (%)	96.4 (92.2)	99.8 (99.6)	97.9 (99.4)	96.5 (85.6)	94.0 (92.2)
Redundancy	2.9 (2.7)	3.8 (3.7)	3.0 (2.9)	2.0 (2.0)	2.4 (2.4)
I/σ(I)	29.2 (2.7)	22.9 (3.0)	14.2 (2.1)	20.8 (2.2)	24.2 (2.2)
C. Refinement					
Resolution range (Å)	35.7 – 1.53	49.5 – 1.58	34.4 – 1.78	22.7 – 1.45	24.7 – 1.40
Reflections used in refinement	47,916 / 2,389	45,959 / 2,302	31,209 / 1,551	57,070 – 2,875	61,099 / 3,089
(work/free)					
Final R values for all reflections	15.9 / 18.6	15.5 / 17.9	15.4 / 18.5	14.0 / 17.2	13.4 / 16.5
(work/free) (%)					
Protein residues (L chain/H chain)	28 /251	28 /251	28 /251	28 / 251	28 / 251
Sodium ions	2	2	2	2	2
Inhibitor atoms	32	33	35	36[b] / 41[c]	39
Water molecules	297	305	314	325	349

RMSD from ideality					
Bond lengths (Å)	0.009	0.009	0.008	0.009	0.008
Bond angles (°)	1.08	1.08	1.06	1.04	1.05
Ramachandran plot					
Residues in most favoured regions (%)	85.8	85.0	85.8	86.2	85.0
Residues in additionally allowed	14.2	14.6	14.2	13.4	15.0
regions (%)					
Residues in generously allowed	-	0.4	-	0.4	-
regions (%)					
Mean B-factor ($Å^2$)					
Protein (L + H chain)	22.9	21.5	20.6	19.9	18.2
Binding site[d]	18.5	16.7	15.8	16.2	14.8
Inhibitor	17.5	21.0	18.5	22.4[e]	16.5
Water molecules	33.6	33.7	32.5	32.5	36.0

Structure (PDB entry)	THR-**6** complex **(3RLW)**	THR-**7** complex **(3RLY)**	THR-**8** complex **(3RM0)**	THR-**9** complex **(3UWJ)**	THR-**10** complex **(3RM2)**
A. Data collection and processing					
No. Crystals used	1	1	1	1	1
Wavelength (Å)	1.00	1.00	0.91841	0.91841	0.91841
Space group	C2	C2	C2	C2	C2
Unit cell parameters					
a, b, c (Å)	70.5, 71.3, 72.5	70.3, 71.4, 72.4	70.5, 71.3, 72.5	70.3, 71.2, 72.4	69.8, 71.5, 71.9
β (°)	100.6	100.3	100.5	100.7	99.9
Matthews coefficient ($Å^3$/Da)	2.5	2.5	2.5	2.5	2.5
Solvent content (%)	51	51	51	51	51
B. Diffraction data[a]					
Resolution range (Å)	50 – 1.69	50 – 1.51	50 – 1.34	50 – 1.50	50 – 1.23
	(1.72 – 1.69)	(1.54 – 1.51)	(1.36 – 1.34)	(1.53 – 1.50)	(1.25 – 1.23)
Unique reflections	38,822 (1,961)	55,139 (2,742)	78,347 (3,850)	55,148 (2,816)	99,085 (4,675)
R(I)sym (%)	5.9 (49.0)	5.1 (48.7)	5.1 (51.9)	5.3 (46.8)	4.7 (38.7)
Completeness (%)	97.7 (98.8)	99.9 (99.6)	98.7 (97.3)	97.3 (98.5)	97.9 (91.7)
Redundancy	2.5 (2.4)	3.7 (3.5)	3.3 (3.3)	2.7 (2.6)	2.8 (2.1)
I/σ(I)	14.6 (2.0)	24.8 (2.7)	22.3 (2.2)	17.7 (1.9)	21.2 (2.0)
C. Refinement					
Resolution range (Å)	34.7 – 1.69	35.6 – 1.51	21.5 – 1.34	22.5 – 1.5	20.1 – 1.23
Reflections used in refinement	36,443 / 1,826	52,718 /2,659	74,108 / 3,761	52,016 / 2,635	92,538 /4,685
(work/free)					
Final R values for all reflections	16.2 / 19.2	15.5 / 17.2	14.0 / 16.6	16.0 / 18.3	14.0 / 16.3
(work/free) (%)					
Protein residues (L chain/H chain)	28 /251	28 / 251	28 / 251	28 / 251	28 /251
Sodium ions	2	2	2	2	2
Inhibitor atoms	32	33[f] / 33[g]	35[f] / 35[g]	32[f] / 11[g]	39
Water molecules	243	347	361	349	390

RMSD from ideality					
Bond lengths (Å)	0.008	0.009	0.008	0.018	0.008
Bond angles (°)	1.060	1.114	1.112	1.095	1.111
Ramachandran plot					
Residues in most favoured regions (%)	87.0	85.0	85.4	86.2	85.4
Residues in additionally allowed	13.0	15.0	14.6	13.8	14.6
regions (%)					
Residues in generously allowed	-	-	-	-	-
regions (%)					
Mean B-factor ($Å^2$)					
Protein (L + H chain)	22.9	19.1	18.0	17.7	17.4
Binding site[b]	18.5	14.9	14.4	16.5	13.9
Inhibitor	22.3	17.4[h] / 26.7[i]	15.7[h] / 16.9[i]	17.7[h] / 21.1[i]	15.0
Water molecules	31.4	32.4	31.5	33.6	33.3

[a] Numbers in parenthesis are for the highest resolution shell.

[b] The inhibitor atoms from the additional disordered part were not considered.

[c] All inhibitor atoms which were included in the refinement model.

[d] Definition of the binding site: all amino acids which are 4 Å within the inhibitor.

[e] Average B value for the ordered and fully ocupied inhibitor portion, the disordered portions were not considered.

[f] Number of inhibitor atoms for the fully occupied ligand.

[g] Number of inhibitor atoms for the partially occupied ligand.

[h] Mean B-factor for the fully occupied ligand.

[i] Mean B-factor for the partially occupied ligand.

4.8 References:

1 Krishnamurthy V.M., Bohall B.R., Semetey V. & Whitesides G.M. (2006) The Paradoxical Thermodynamic Basis for the Interaction of Ethylene Glycol, Glycine, and Sarcosine Chains with Bovine Carbonic Anhydrase II: An Unexpected Manifestation of Enthalpy/Entropy Compensation. *J. Am. Chem. Soc.* **128**, 5802-5812.

2 Williams D.H., Stephens E., O'Brien D.P. & Zhou M. (2004) Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew. Chem. Int. Ed.* **43**, 6596-6616.

3 Baum B., Muley L., Smolinski M., Heine A., Hangauer D. *et al.* (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.

4 Ladbury J.E. (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.

5 Biela A., Khyat M., Tan H., Heine A., Hangauer D. *et al.* (2011) Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin. Submitted to *J.Mol.Biol.*

6 Talhout R., Villa A., Mark A.E. & Engberts J.B. (2003) Understanding binding affinity: a combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *J. Am. Chem. Soc.* **125**, 10570-10579.

7 Whitesides G.M. & Krishnamurthy V.M. (2005) Designing ligands to bind proteins. *Q. Rev. Biophys.* **38**, 385-395.

8 Tanford C. (1978) The hydrophobic effect and the organization of living matter. *Science*. **200,** 1012-1018.

9 Smithrud D.B., Wyman T.B. & Diederich F. (1991) Enthalpically driven cyclophane-arene inclusion complexation: solvent-dependent calorimetric studies. *J. Am. Chem. Soc.* **113**, 5420-5426.

10 Eftink M.R. & Harrison J.C. (1981) Calorimetric studies of p-nitrophenol binding to α- and β-cyclodextrin. *Bioorg. Chem.* **10**, 388-398.

11 Bertrand G.L, Faulkner Jr. J.R., Han S.M. & Armstrong D.W. (1989) Substituent effects on the binding of phenols to cyclodextrins in aqueous solution. *J. Phys. Chem.* **93**, 6863-6867.

12 Breslauer K.J, Remeta D.P, Chou W.Y, Ferrante R., Curry J. *et al.* (1987) Enthalpy-entropy compensations in drug-DNA binding studies. *PNAS* **84**, 8922-8926.

13 Marky L.A & Breslauer K.J. (1987) Origins of netropsin binding affinity and specificity: correlations of thermodynamic and structural data. *PNAS* **84**, 4359-4363.

14 Bingham R.J., Findlay J.B., Hsieh S.Y., Kalverda A.P., Kjellberg A. *et al.* (2004) Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J. Am. Chem. Soc.* **126**, 1675-1681.

15 Englert L., Biela A., Zayed M., Heine A., Hangauer D. *et al.* (2010) Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: binding of phosphonamidate to the S_1'-pocket of thermolysin. *Biochim Biophys Acta.* **1800**, 1192-1202.

16 Barratt E., Bingham R.J., Warner D.J., Laughton C.A., Phillips S.E. *et al.* (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.

17 Homans S.W. (2007) Water, water everywhere - except where it matters? *Drug Discov. Today.* **12**, 534-539.

18 Ahmed H.U., Blakeley M.P., Cianci M., Cruickshank D.W.J., Hubbard J.A. *et al.* (2007) The determination of protonation states in proteins. *Acta Crystallogr. D* **63**, 906-922.

19 Rittle K.E., Barrow J.C., Cutrona K.J., Glass K.L., Krueger J.A. *et al.* (2003) Unexpected enhancement of thrombin inhibitor potency with o-aminoalkylbenzylamides in the P1 position. *Bioorg. Med. Chem. Lett.* **13**, 3477-3482.

20 Dönnecke D., Schweinitz A., Stürzebecher A., Steinmetzer P., Schuster M. *et al.* (2007) From selective substrate analogue factor Xa inhibitors to dual inhibitors of thrombin and factor Xa. Part 3. *Bioorg Med Chem Lett.* **17**, 3322-3329.

21 Dullweber F., Stubbs M.T., Musil D., Stürzebecher J. & Klebe G. (2001) Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J Mol Biol.* **313**, 593-614.

22 Carugo O. & Bordo D. (1999) How many water molecules can be detected by protein crystallography? *Acta Crys. D* **55**, 479-483.

23 Matter H., Nazaré M., Güssregen S., Will D.W., Schreuder H. *et al.* (2009) Evidence for C-Cl/C-Br···pi interactions as an important contribution to protein-ligand binding affinity. *Angew Chem Int Ed Engl.* **48**, 2911-2916.

24 Baum B., Mohamed M., Zayed M., Gerlach C., Heine A. *et al.* (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.

25 Baum B., Muley L., Heine A., Smolinski M., Hangauer D. *et al.* (2009) Think twice: understanding the high potency of bis(phenyl)methane inhibitors of thrombin. *J. Mol. Biol.* **391**, 552-564.

26 Jain A., Ramanathan V. & Sankararamakrishnan R. (2009) Lone pair···pi interactions between water oxygens and aromatic residues: quantum chemical studies based on high-resolution protein structures and model compounds. *Protein Sci.* **18**, 595-605.

27 Hogg D.H. & Blombäck B. (1978) The mechanism of the fibrinogen-thrombin reaction. *Thromb Res.* **12**, 953-964.

28 Martin P.D., Malkowski M.G., DiMaio J., Konishi Y., Ni F. *et al.* (1996) Bovine thrombin complexed with an uncleavable analog of residues 7-19 of fibrinogen A alpha: geometry of the catalytic triad and interactions of the P1', P2', and P3' substrate residues. *Biochemistry.* **35**, 13030-13039.

29 Stubbs M.T., Oschkinat H., Mayr I., Huber R., Angliker H. *et al.* (1992) The interaction of thrombin with fibrinogen. A structural basis for its specificity. *Eur J Biochem.* **206**, 187-195.

30 Martin P.D., Robertson W., Turk D., Huber R., Bode W. *et al.* (1992) The structure of residues 7-16 of the A alpha-chain of human fibrinogen bound to bovine thrombin at 2.3-Å resolution. *J Biol Chem.* **267**, 7911-7920.

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

32 Sigurskjold B.W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* **277**, 260-266.

33 Czodrowski P., Sotriffer C.A. & Klebe G. (2007) Protonation changes upon ligand binding to trypsin and thrombin: structural interpretation based on pK(a) calculations and ITC experiments. *J Mol Biol.* **367**, 1347-1356.

34 Schweinitz A., Stürzebecher A., Stürzebecher U., Schuster O., Stürzebecher, J. *et al.* (2006) New substrate analogue inhibitors of factor Xa containing 4-amidinobenzylamide as P1 residue: part 1. *Med Chem.* **2**, 349-361.

35 Schweinitz A., Steinmetzer T., Banke I.J., Arlt M.J., Stürzebecher A. *et al.* (2004) Design of novel and selective inhibitors of urokinase-type plasminogen activator with improved pharmacokinetic properties for use as antimetastatic agents. *J Biol Chem.* **279**, 33613-33622.

36 Stauffer K.J., Williams P.D., Selnick H.G., Nantermet P.G., Newton C.L. *et al.* (2005) 9-hydroxyazafluorenes and their use in thrombin inhibitors. *J. Med. Chem.* **48**, 2282-2293.

37 Bernett M.J., Blaber S.I., Scarisbrick I.A., Dhanarajan P., Thompson S.M. *et al.* (2002) Crystal structure and biochemical characterization of human kallikrein 6 reveals that a trypsin-like kallikrein is expressed in the central nervous system. *J. Biol. Chem.* **277**, 24562-24570.

38 Leatherbarrow R.J. (1998) GraFit Version 4 4.0 edit, Erithacus Software Limited, Staines, UK.

39 Mizoue L.S. & Tellinghuisen J. (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal. Biochem.* **326**, 125-127.

40 Otwinowski Z. & Minor W. (1997). Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol.* **276**, 307-326.

41 Skordalakes E., Dodson G.G., Green D.S., Goodwin C.A., Scully M.F. *et al.* (2001) Inhibition of human alpha-thrombin by a phosphonate tripeptide proceeds via a metastable pentacoordinated phosphorus intermediate. *J. Mol. Biol.* **311**, 549–555.

42 Brunger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921.

43 Adams P.D, Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213-221.

44 Painter J. & Merritt E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62,** 439-450.

45 Painter J. & Merritt E.A. (2006) TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* **39**, 109-111.

46 Emsley P. & Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132.

47 Chen V.B., Arendall W.B. 3rd, Headd J.J., Keedy D.A., Immormino R.M. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12-21.

48 Laskowski R.A., MacArthur M.W., Moss D.S. & Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*. **26**, 283–291.

49 Kleywegt G.J., Zou J.Y., Kjeldgaard M. & Jones T.A. (2001) Around O. *In International Tables for Crystallography* (Rossmann, M.G. & Arnold, E., eds), vol. F, pp. 353–356, Kluwer Academic Publishers, Dordrecht.

50 Neudert, G. & Klebe, G. (2011) fconv: format conversion, manipulation and feature computation of molecular data. *Bioinformatics.* **27**, 1021-1022.

51 Bode W., Mayr I., Baumann U., Huber R., Stone S.R. *et al.* (1989) The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467–3475.

5 Water Makes the Difference: Rearrangement of Water Solvation Layer Triggers Non-additivity of Functional Group Contributions in Protein-Ligand Binding

5.1 Introductory remarks

This study was done in cooperation with the group of Prof. Dr. Hangauer (University of Buffalo). The following text is submitted to the scientific journal *ChemMedChem*.

5.2 Abstract

The binding of four congeneric peptide-like thermolysin inhibitors has been studied by high-resolution crystal structure analysis and isothermal titration calorimetry. The ligands differ only by a terminal carboxylate and/or methyl group. A surprising non-additivity of functional group contributions for the carboxylate and/or methyl groups is detected. Adding first the methyl and then the carboxylate group results in a small Gibbs free energy increase and minor enthalpy/entropy partitioning for the first modification, whereas the second involves strong affinity increase combined with huge enthalpy/entropy changes. Adding however first the carboxylate and then the methyl group yields reverse effects: now the acidic group attachment causes minor effects whereas the added methyl group provokes huge changes. As all crystal structures show virtually identical binding modes, affinity changes are related to rearrangements of the first solvation layer next to the S_2' pocket. About 20-25 water molecules are visible next to the studied complexes. The added COO^- groups perturb the local water network in both carboxylated complexes and the attached methyl groups provide favorable interaction sites for water molecules. In all complexes, apart one example, a contiguously connected water network between protein and ligand functional groups is observed. In the complex with the carboxylated ligand, still lacking the terminal methyl group, the water network is unfavorably ruptured. This results in the surprising thermodynamic signature showing only minor affinity increase upon COO^- group attachment. Since the further added methyl group provides a favorable interaction site for water, the network can be re-established and strong affinity increase with huge enthalpy/entropy signature is then detected.

5.3 Introduction

Hit-to-lead optimization generally follows chemical modifications which involve addition and replacement of functional groups at a given parent scaffold. This idea of adding chemical groups by forming and breaking chemical bonds has strongly influenced our conceptional thinking in assigning particular functional group contributions to estimate by how much an added group will enhance binding affinity of a promising lead candidate.[1-3] Following this approach, however, completely ignores that every novel molecule is on its own a new species that will interact with the target protein in its own specific way. From a physicochemical point of view this strategy of designing molecules by adding affinity contributions of functional groups and assigning Gibbs free energy increments to particular interactions is perhaps pragmatic, but fundamentally wrong and can therefore be quite misleading. Binding affinity is a free energy, composed by enthalpy and entropy and cannot be additive in terms of functional group contributions.[4] The importance of this non-additivity (sometimes also named 'cooperativity') has been first indicated by Williams *et al.*[5-7] Kawai *et al.* reported on non-additive effects with respect to chemical ligand modifications in HCV protease binding using a chemical double-mutant cycle.[8] In a recent series of studies on ligand binding to thrombin we reported on non-additivity of hydrogen bonding related to the simultaneous formation of hydrophobic contacts.[9,10] On the molecular level the observed non-additivity is explained by differences in residual mobility of bound ligand portions: The strength of a hydrogen bond takes influence on the dynamic properties of a formed complex.

Therefore properties such as the degree of hydrophobic surface portions of a ligand in a binding pocket that becomes buried upon complex formation have to be regarded under dynamic conditions. If larger residual mobility is given, formally the burial is less efficient (or less frequently experienced on the time scale) than if the residual mobility is reduced, e.g. in consequence of an adjacent more tightly fixed hydrogen bond.

Often such effects are not highly transparent in binding affinity as, with respect to thermodynamic factorization, enthalpy and entropy mutually compensate each other and only a minor contribution to the Gibbs free energy remains.[11,12] However, by partitioning enthalpy and entropy along with the investigation of structural properties these effects become obvious.

5. Water Makes the Difference

Dynamic properties are one explanation on the molecular level for the incident of non-additivity. Likely, many more mechanisms are responsible for such phenomena. In the current study we will report on non-additivity determined by changes in the local hydration pattern that is formed next to a binding site and will be perturbed and influenced by the presence of the bound ligands.

The importance of water in the binding process receives increasing appreciation in drug design.[13-15] However, it is predominantly focused on general contributions of solvation and desolvation, the release or pick-up of water molecules in a formed complex or the role of water at interstitial sites mediating contacts between ligand and protein. The influence of water next to a binding site and mediating the transition from the bound ligand to the bulk water phase is hardly taken into consideration. However, also these immediate layers of structured (or well-ordered) water molecules arranged in contiguous, but locally perturbed networks is determining for binding and, as will be shown in this contribution, and with a an alternate approach to interpreting the data but coming to a similar general conclusion that water is involved in cooperativity by Nasief *et al.*,[16] can take substantial influence on non-additivity of functional group contributions and the partitioning of enthalpy and entropy in the binding process.[17]

As model system we investigated the binding of four congeneric peptidomimetic phosphonamidate inhibitors to thermolysin (TLN), a zinc endopeptidase from *B. thermoproteolyticus.* This protease has been broadly characterized and frequently used as representative for the entire class of zinc-dependent metalloproteinases.[18-20] A combined study of high-resolution crystallography and isothermal titration calorimetry has been performed. Lucky enough, the crystals of the four investigated complexes diffracted sufficiently well to disclose important parts of the solvation pattern of the first water layer around the S_2' pocket of the protease, the region in which the chemical structure of the studied ligand series has been modified. At their terminus either an additional methyl group or carboxylate group were attached or both modifications were considered simultaneously (Fig. 1, **1**-**4**).

5.4 Results - Crystallography

5.4.1 Resolution required for determining solvation patterns.

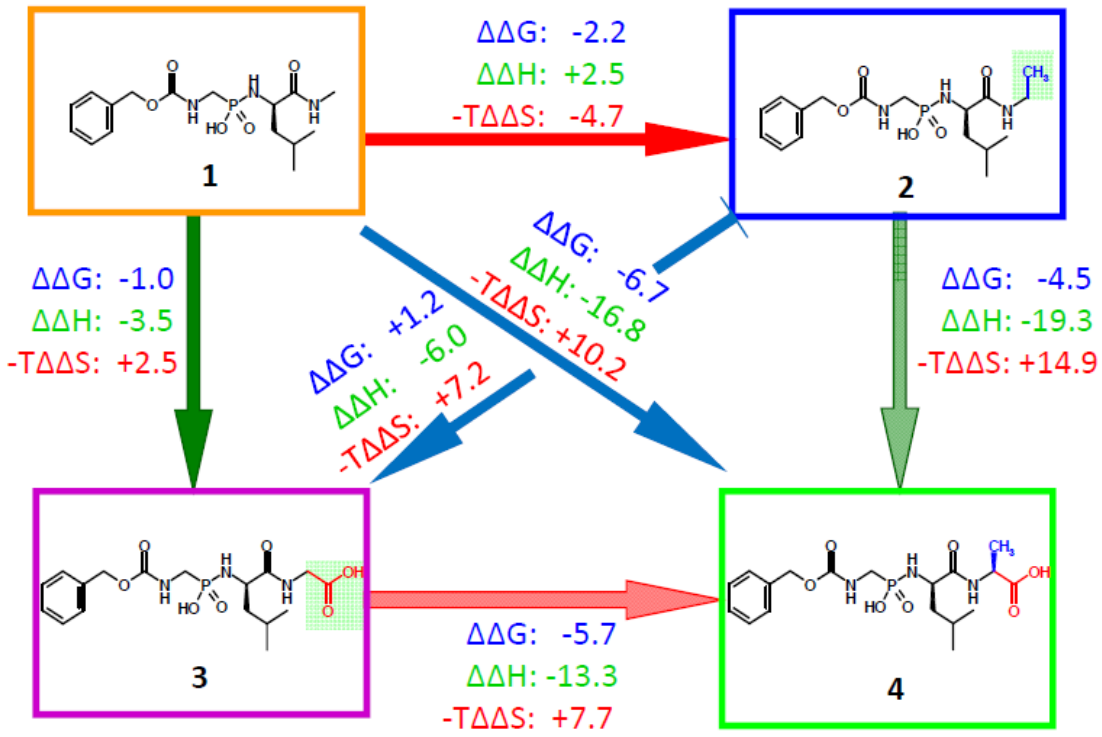
To study systematically the changes in the water arrangement next to a binding pocket by crystallography, highly resolved crystal structures are required as the number of observed water molecules can suffer from resolution of the diffraction data.[21] Statistics show that, on average, at 2.0 Å resolution one water molecule per residue is included in the model, while at 1.0 Å resolution about 1.6-1.7 are crystallographically located.[21] The data presented in this study were recorded by use of synchrotron radiation to achieve the best possible resolution of the thermolysin complexes of **1**-**4** determined with good to excellent resolution (1.60-1.28 Å). The assignment of the different water molecules was based on the interpretation of the difference electron density maps. At least a peak has to exhibit a height of three sigma in the difference electron density map to be interpreted as a water molecule.

5.4.2 Binding Mode of the Ligand Scaffolds

The crystal structures of the four inhibitors comprising a Cbz-Gly-$(PO_2)^-$-L-Leu-XXX scaffold (Cbz = Carboxybenzyl, XXX= CH_3 (**1**), CH_2CH_3 (**2**), CH_2COOH (**3**), $CH(CH_3)COOH$ (**4**), Fig. 1) have been determined in complex with thermolysin. For all, the difference electron density of the scaffold is well defined; a representative depiction is shown for **4** (Fig. 2). The Cbz-Gly-$(PO_2)^-$-L-Leu portions of all four ligands bind virtually in identical manner (Fig. 2). The terminal Cbz groups are located with highly conserved geometry next to the zinc ion in the unspecific S_1 pocket. The perpendicular orientation of the CH_2 groups to the attached phenyl rings corresponds to the preferred torsion angle for this connection. The phenyl rings of these groups interact with CD2 of Phe114 through van-der-Waals contacts (4.0 - 4.1 Å) and they find additional interactions with Phe114 and Tyr106 of a neighboring crystal mate. Furthermore, they are fixed in position by interactions of the carbonyl oxygen of the adjacent Trp115 which orients towards the center of the neighboring Cbz phenyl rings (3.8 - 4.1 Å). Below the Cbz phenyl ring in all complexes a glycerol molecule, captured from the cryo buffer, is detected and a neighboring DMSO molecule is present in the crystal packing. In all examples the carbonyl group of the Cbz carbamide group adopts two conformations. In one conformer, the carbonyl oxygen orients towards Phe114 (3.4 - 3.6 Å to CB), in the other the carbonyl oxygen is involved in an adjacent network of water molecules (s. below). The

occupancy of the two conformations refined to 50-63 % for the conformer towards Phe114 and between 37 % and 50 % towards the water network.

Figure 1: Thermodynamic cycle for the structural changes from **1** (orange) to **2** (blue), **3** (magenta) and **4** (green). Along the red pathways attachment of an additional methyl group is performed, the green pathways involve addition of an extra carboxylate group. The relative differences in the thermodynamic properties are given (ΔΔG, ΔΔH, -TΔΔS), all values in kJ/mol.

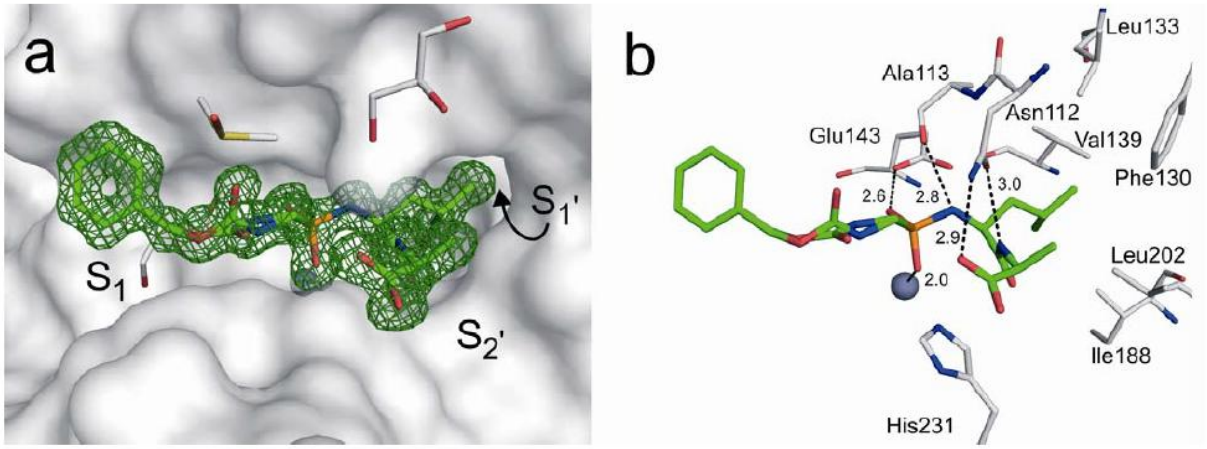


With regard to the contacts to the zinc ion, the phosphonamidate scaffold interacts in a mono-dentate fashion with one oxygen being directed towards zinc (1.9 - 2.0 Å) and the other oxygen towards OE1 of Glu143 (2.5 - 2.6 Å, Fig. 2b). The adjacent phosphonamidate NH group hydrogen-bonds to the backbone carbonyl of Ala113 (2.8 – 2.9 Å). The isobutyl group of the central leucine portion fits perfectly well into the hydrophobic S_1' pocket and forms multiple van-der-Waals contacts to the neighboring hydrophobic pocket residues (e.g. Phe130, Leu133, Val139, Ile188 and Leu202, Fig. 2b). The distances of these van-der-Waals contacts range from 3.7 Å to 4.1 Å.

The four inhibitors differ by their terminal XXX group. Nevertheless, also here remarkable conservation of the binding modes is given. Thus, the terminal CH_3 group in **2** matches well with that in **4** and the carboxylate groups of **3** and **4** superimpose perfectly. The carboxylate groups of **3** and **4** form a most likely charge-assisted stacking interaction (3.4 - 3.7 Å) with

the side chain of His231. In summary with respect to the binding modes, all four inhibitors of the series show virtually the same orientations. In all complexes above the carboxyamide terminus of Asn112 a further glycerol molecule is found.

Figure 2: Overall binding mode as observed in the representative crystal structure of TLN-**4**, water molecules omitted for clarity. (a) The protein binding pocket is indicated by the grey solvent-accessible surface and the green mesh circumvents the F_o-F_c difference electron density at 3 σ level. The Cbz group is oriented towards the S_1' pocket which shows a glycerol molecule below the phenyl ring of the Cbz group. Adjacent a DMSO molecule is found. A further glycerol molecule is detected next to Asn 112 in the S_2' pocket. (b) The ligand coordinates in mono-dentate fashion the catalytic zinc ion (blue-grey sphere). Coordinating residues are shown with prominent interactions indicated by dashed lines. Distances are given in Angstrom. Nitrogens blue, oxygens red, phosphorous orange, carbon green/grey.



5.4.3 Solvation Structure Around the Ligands

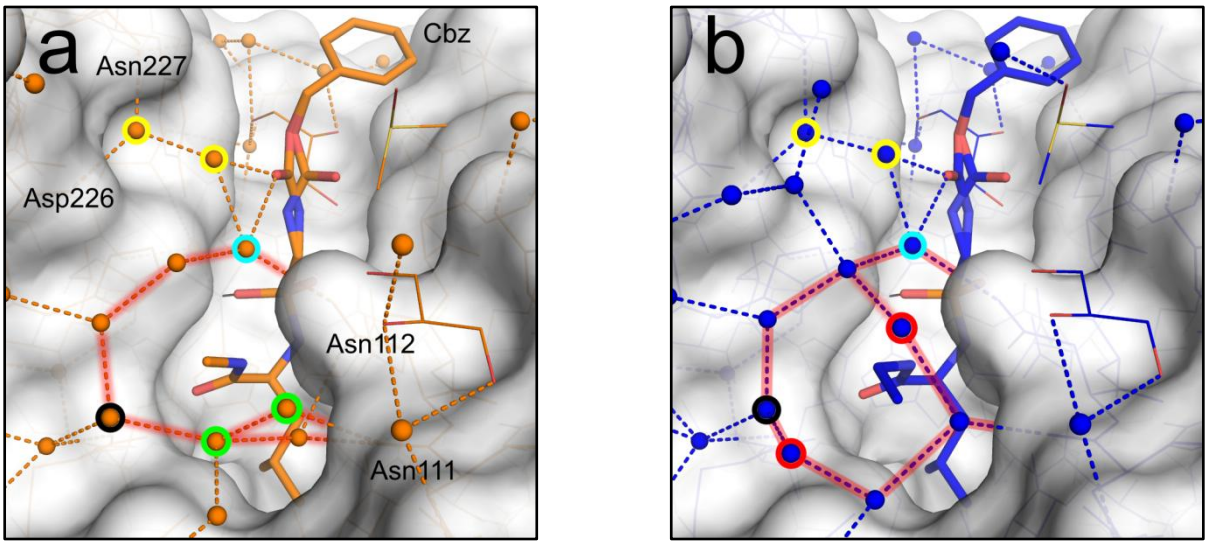
Due to the remarkable diffraction power of the studied crystals, the first layers of the solvation structure next to the binding pockets are well determined and can be compared across the four complexes. Major differences become obvious next to the S_2' pocket, in particular as here the four inhibitors differ in their substitution pattern. Obviously, the terminal methyl and negatively charged carboxylate groups induce different degrees of ordering (Fig. 3 and 4, color coding: TLN-**1** orange, TLN-**2** blue, TLN-**3** magenta, TLN-**4** green). In all complexes in the local neighborhood of the S_2' pocket (10 Å sphere around P_2' substituent) more than 20 water molecules can be assigned to the difference electron density (Fig. 3). They are found with mutually distances between 2.5 up to 3.2 Å suggesting formation of an extended H-bonding network that also involves contacts to functional groups of either the ligand or the protein. Motifs of contiguous chains and rings are

observed (Fig. 3). Cyclic patterns with five and six water molecules are generated that have frequently been detected – at least in part – next to protein surfaces.[22] The four complexes show slightly different numbers of assigned water molecules which seems to correlate with the resolution of the determined crystal structures, e.g. the complex with **4** (Fig. 3d) which diffracted to 1.28 Å resolution shows the largest number of water molecules. This fact makes it difficult to compare the absolute numbers of water molecules across the series of complexes, particular with respect to a relative inventory of released or picked-up water molecules. The distances along the water chains vary and possibly they correlate with the strength of the formed hydrogen bonds. The occupancies and B-factors of the refined water positions change with increasing distance from the protein surface and the presence of adjacent polar groups which can reinforce a more firm fixation. Accordingly, the determined spatial accuracy of individual water positions will be affected by, e.g. residual mobility. Therefore, we want to refrain from any detailed considerations of H-bond length variations.

Some confidence in the relevance of the observed water networks is suggested by the fact that recurrently the same motifs are indicated in all four complexes, but with some important local perturbations induced by the attachment of the different XXX substituents. A two-membered water chain (highlighted by yellow cycles, Fig. 3a-d) is formed between the C=O group of the Cbz portion and the carboxamide of Asn227 and the carboxylate group of Asp226. In the complexes TLN-**1** and TLN-**2** the first water molecule in this chain interacts via an additional water molecule (Fig. 3a,b, cyan cycle) with the carboxamide group of Asn112; in TLN-**3** and TLN-**4** the latter water molecule is replaced by one of the carboxylate oxygens. In the complexes with **1** and **2** a water network (highlighted by red bars) is formed that wraps around the terminal hydrophobic methyl or ethyl group. It contiguously connects the carboxyamide of Asn112 with the backbone carbonyl group of Asn111. Due to the steric demand of the larger ethyl group in TLN-**2** two water molecules hydrogen-bonded to Asn111(C=O) which are present in TLN-**1** (Fig. 3a, green cycles) are repelled from the complex. Instead, two new water molecules are detected next to the terminus of the ethyl group in TLN-**2** (Fig. 3b, red cycles). Interestingly enough in TLN-**4** the latter two water molecules are observed at very similar spatial positions (Fig. 3d, red cycles). The two derivatives **3** and **4** introduce the strongly polarizing carboxylate group. It forms along its anti direction[23] hydrogen bonds to two water molecules (Fig. 3c, d). Furthermore, it perturbs the water network along the distal syn direction and one water molecule (Fig. 3a-d, black

cycles) is dragged into an H-bond with the carboxylate group. This slight movement of the water molecule towards the carboxylate group (cf. Fig. 4 b,c) has consequences for TLN-**3**. Different from the other three cases, TLN-**3** lacks the contiguously closed water network pattern: it breaks next to the water molecule in syn direction of the COO^- group (Fig. 3c, 4d, light blue broken arrow). To trace this difference between TLN-**3** and TLN-**4** we carefully inspected the electron density map in this region (Fig. 5 a, b). Although there is little residual difference electron density at 3σ level next to the water molecule in syn direction of the carboxylate group in TLN-**3**, this density extension is too close (1.3 Å) to assign an additional water site, more likely it indicates enhanced anisotropic motion of the latter water molecule. Therefore, we can conclude that no additional water is present to mediate a fully closed water network.

Figure 3: Binding modes of **1-4** together with the local water pattern as observed in the crystal structures, the protein binding pocket is indicated by the solvent accessible surface of **4**. (a) TLN-**1**, carbon atoms and water molecules in orange, (b) TLN-**2**, carbon atoms and water molecules in blue, (c) TLN-**3**, carbon atoms and water molecules in magenta, (d) TLN-**4**, carbon atoms and water molecules in green. Important residues next to the S_2' pocket are labeled. The local water network which wraps around the terminal ligand group (**1**: CH_3, **2**: CH_2CH_3, **3**: CH2COO-, **4**: CH $(CH_3)COO^-$) is highlighted by red bars. In all complexes, a two-membered water chain (encircled in yellow) is formed between the ligands and Asn227/Asp226. In TLN-**1** and **2** the water network which wraps around the terminal groups starts with the water molecule encircled in cyan. In **1**, **2** and **4** it shows a contiguous chain connecting all the way from Asn112 to the backbone carbonyl group of Asn111. Next to the latter group two water molecules (encircled in green) are repelled from the complexes with **2** and **4**. In contrast, the latter ligands capture two water molecules (encircled in red) through van-der-Waals contacts with the terminal methyl group. The water molecule encircled in black becomes involved in a hydrogen bond with the carboxylate group in **3** and **4**. The water network breaks in TLN-3 (blue broken arrow).



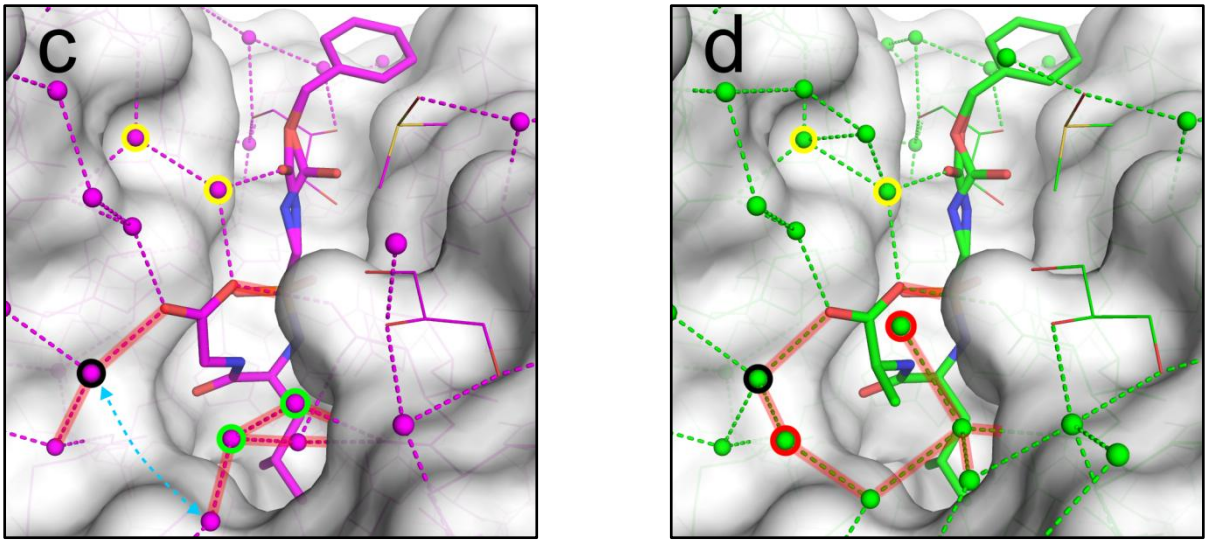
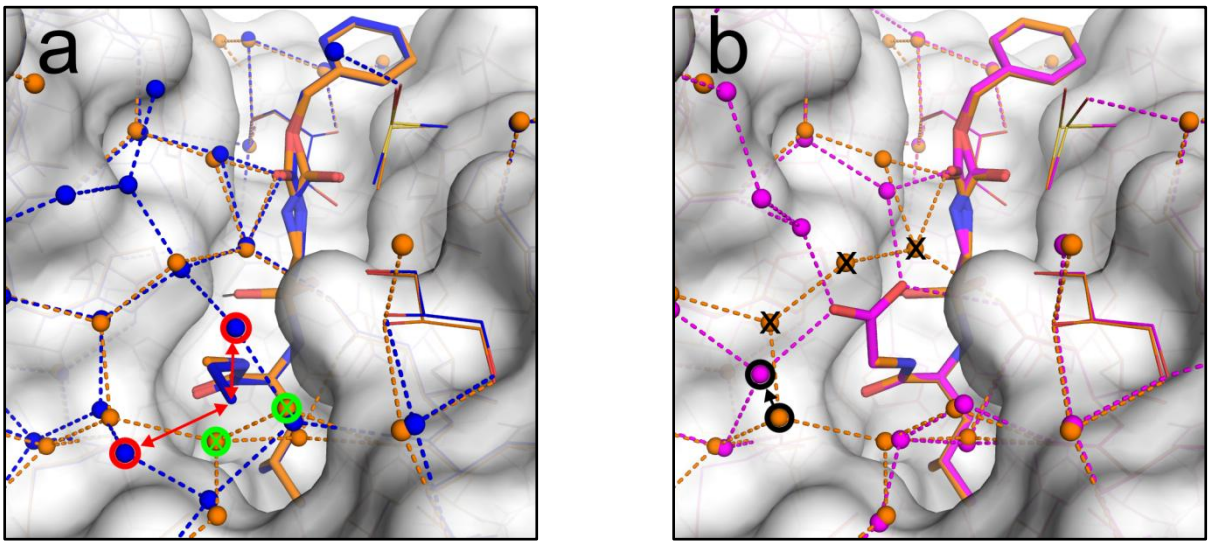


Figure 4: Mutual superposition of the crystal structures of (a) TLN-**1** (orange) and TLN-**2** (blue), (b) TLN-**1** (orange) and TLN-**3** (magenta), (c) TLN-**2** (blue) and TLN-**4** (green) and (d) TLN-**3** (magenta) and TLN-**4** (green), same view direction as in Fig. 3. (a) and (d) with the solvent accessible surface of **4**: Two water molecules (encircled and crossed in green) are repelled next to the backbone carbonyl group of Asn 111 (for labels, see Fig. 3) due to the additional steric requirements of the extra methyl group in **2**; however, two water molecules are picked-up at sites next to the added methyl group (encircled in red and arrows). (b) and (c): The introduced carboxylate group strongly perturbs the water network and replaces water molecules in close neighborhood (black crosses). Simultaneously, the water molecule next to the carboxylate's distal syn site (encircled in black) is dragged into a hydrogen bond and shifts slightly in space. In TLN-**3** this results in a rupture of the water-mediated network (d, blue broken arrow).



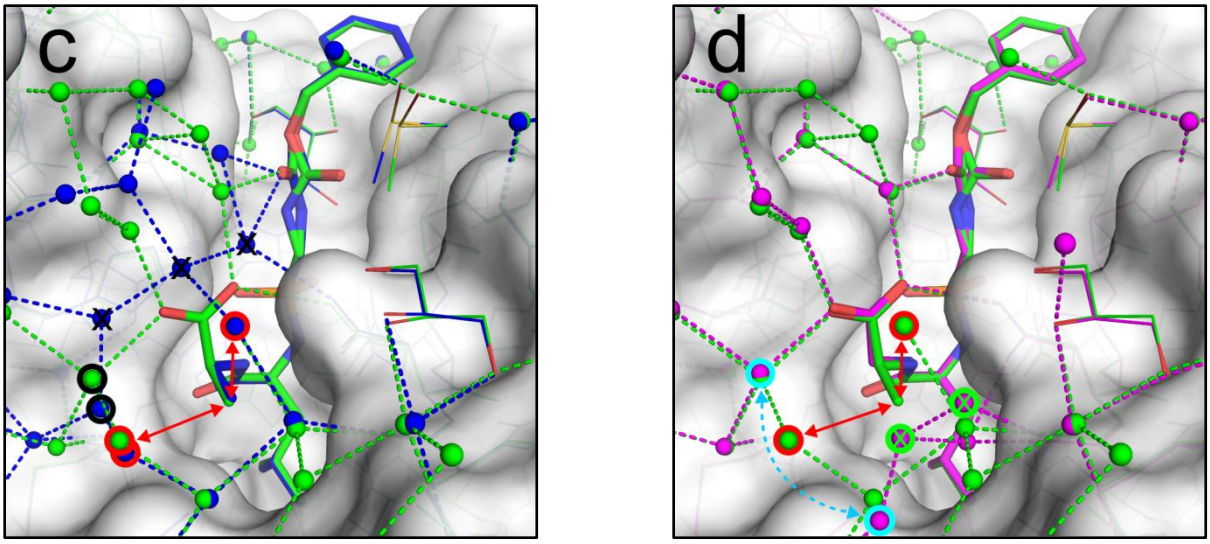
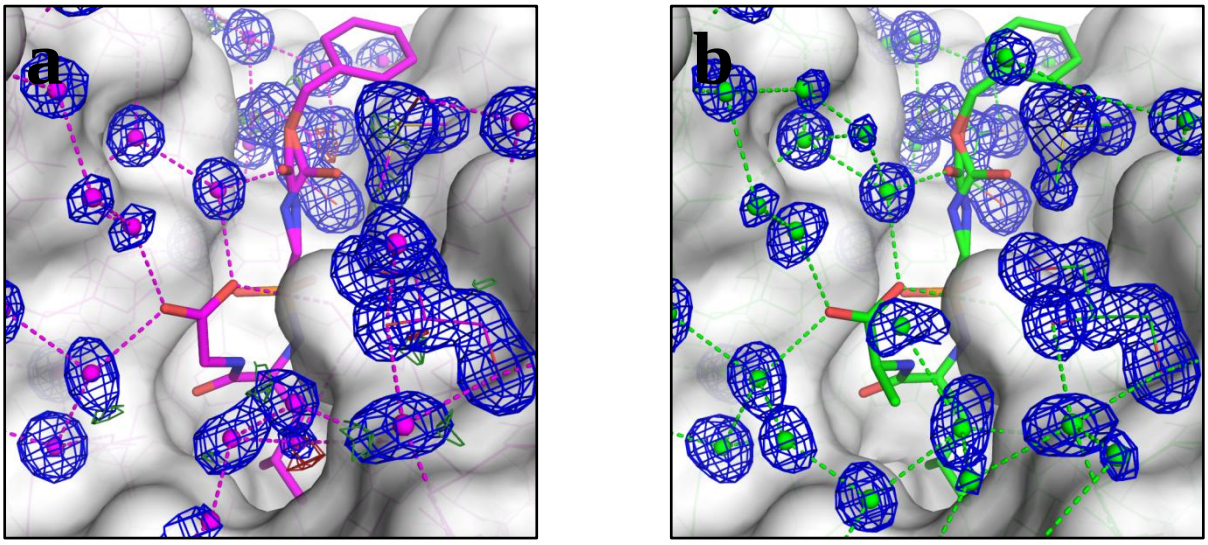


Figure 5: $2F_o–F_c$ (blue, contour level 1σ) and F_o-F_c (green, contour level 3σ) electron density in (a) TLN-**3** (ligand shown in magenta) and (b) TLN-**4** (ligand shown in green) next to the region where the water network is ruptured. The protein binding pocket is indicated by the grey solvent accessible surface of **4**. Only in TLN-**4** a contiguously connected network of water molecules is found, in TLN-**3** the network is broken.



5.5 Results - Isothermal Titration Calorimetry

5.5.1 Displacement and Protonation Steps

The binding thermodynamics have been characterized for all four inhibitors to TLN by ITC. Unfortunately, no absolute binding data could be recorded as the titrations are superimposed by the displacement of the protolysis product Val-Lys from the protein. Thermolysin tends, in particular in high concentrations, to autoprotolysis whereby the last two C-terminal residues Val-Lys are cleaved off. Therefore, all measured thermodynamic

data are overlaid by a constant contribution resulting from the displacement of Val-Lys upon inhibitor binding.[24] We can assume with high confidence that this contribution is the same in all cases; however, since we are only interested in relative differences within the ligand series, the portion of the dipeptide replacement cancels out in the inventory. In the following descriptions and discussions only reference to the relative differences will be taken (Fig. 1). Furthermore, ligand binding to TLN is superimposed by the pick-up of a proton. As described in detail elsewhere (Biela *et al.* to be published), the additional proton binds to Glu143 next to the zinc ion. The heat of ionization involving this protonation step will be constant across our narrow congeneric ligand series. Accordingly, this contribution also cancels out in the attempted relative comparison.

5.5.2 Thermodynamic Data

Ligand **1** is the weakest binder. The addition of a second methyl group to the terminal NCH_3 group of **1** to reveal **2** relates to a more exergonic binding of $\Delta\Delta G_{\mathbf{1\rightarrow 2}}$ = - 2.2 kJ/mol and partitions into an endothermic $\Delta\Delta H_{\mathbf{1\rightarrow 2}}$ contribution of 2.5 kJ/mol and an entropically favorable portion of $-T\Delta\Delta S_{\mathbf{1\rightarrow 2}}$ = - 4.7 kJ/mol (Fig. 1). These minor effects are also found if the terminus of **1** is decorated by a carboxylate function to form **3**. Gibbs free energy enhances only by $\Delta\Delta G_{\mathbf{1\rightarrow 3}}$ = - 1.0 kJ/mol, again this effect factorizes into a small exothermic $\Delta\Delta H_{\mathbf{1\rightarrow 3}}$ portion of - 3.5 kJ/mol and a nearly compensating unfavorable entropic effect of $-T\Delta\Delta S_{\mathbf{1\rightarrow 3}}$ = 2.5 kJ/mol. Remarkably, the addition of an acid group to the already methylated derivate **2** to achieve **4** improves affinity by $\Delta\Delta G_{\mathbf{2\rightarrow 4}}$ = - 4.5 kJ/mol and the addition of a methyl group to the carboxylated **3** to afford **4** reveals a $\Delta\Delta G_{\mathbf{3\rightarrow 4}}$ = - 5.7 kJ/mol. Even more surprising, these affinity enhancements factorize into a huge exothermic and an unfavorable entropy contributions ($\Delta\Delta H_{\mathbf{2\rightarrow 4}}$ = - 19.3 kJ/mol, $-T\Delta\Delta S_{\mathbf{2\rightarrow 4}}$ = 14.9 kJ/mol, $\Delta\Delta H_{\mathbf{3\rightarrow 4}}$ = - 13.3 kJ/mol, $-T\Delta\Delta S_{\mathbf{3\rightarrow 4}}$ = 7.7 kJ/mol). The thermodynamic data underscore strikingly the observation of non-additivity effects in this ligand series.

5.6 Discussion

Usually drug optimization is performed in a step-wise fashion by adding functional groups to a parent scaffold. In the present example the addition of a methyl group to **1** enhances affinity by $\Delta\Delta G$ = - 2.2 kJ/mol and the attachment of a carboxylate group only reveals an improvement of - 1 kJ/mol. In a medicinal chemistry program these rather disappointing

effects would have possibly stopped further efforts addressing the S_2' pocket. Nonetheless, the combined affinity enhancement achieved by adding both groups simultaneously to **1** featuring **4** succeeds in the remarkable effect of - 6.7 kJ/mol. As this accounts to more than one order of magnitude in the binding constant this optimization would have likely been claimed as promising.

We can describe the observations also differently (Fig. 1). Taking the pragmatic, however theoretically incorrect picture of additivities of functional group contributions into consideration, the methyl group attachment from **1** → **2** enhances ΔΔG by - 2.2 kJ/mol, whereas it adds - 5.7 kJ/mol from **3** → **4**. Similarly, the additional carboxylate group enhances affinity from **1** → **3** by only -1 kJ/mol, whereas it improves binding from **3** → **4** by - 4.5 kJ/mol. This effect of non-additivity clearly demonstrates that models of functional group additivities come to an end and can be rather misleading.

Considering the thermodynamic factorization along the four optimization pathways provides a more differentiated picture. Remarkably, the steps **1** → **2** and **1** → **3**, which improve affinity only minor, also factorize only slightly into enthalpy and entropy, even though some mutual compensation can to be recognized. Following steps **2** → **4** and **3** → **4**, along which more significant affinity enhancements are recorded, huge enthalpic and entropic effects are experienced, however, they largely compensate. Nevertheless, they suggest major changes in the binding and interaction features.

At this point crystal structure analyses should be consulted. As a matter of fact, in the present case all four inhibitors adopt virtually the same binding mode. Thus, any explanation of the observed non-additivity is hardly evident from the ligand binding geometries. Accordingly, what else can take impact on binding affinity and perturb the thermodynamic signature in this congeneric ligand series? As the addition of either a hydrophobic methyl group (**1** → **2**) or the attachment of a polar carboxylate function (**1** → **3**) result both in minor affinity enhancements, a sole explanation in terms of ligand desolvation costs cannot be made responsible for the observed trends.

For the sake of high crystal quality, our crystals diffracted to high resolution and insights into the local water structure formed next to the bound ligands becomes available. It describes the gradual transition from the ligand-accommodated binding pocket to the surrounding bulk water phase. Remarkable differences in the produced water networks are evident. In all structures several water molecules mediate in chain-type fashion interactions between the

partly exposed ligand functional groups and protein functional groups unsheltered from the interior and next to the catalytic center. Multiple polygonal and mutually fused ring patterns, composed by several water molecules are observed, but they are perturbed by the different substituents attached to **1** – **4** at their terminal end. An extensive water chain wraps around the exposed hydrophobic alkyl chain. It is even capped by an additional water molecule in the complexes with **2** and **4** as the terminal methyl group provides additional van-der-Waals contacts to host this water in favored position (Fig. 3 b, d and 4 c, d, red cycles and arrows). In contrast, the latter complexes release two water molecules originally bound to the backbone carbonyl group of Asn111 (Fig. 3 a, c and 4 a, d, green cycles). The terminal methyl groups in **2** and **4** require too much space to still host these water molecules.

Crystallography provides, with the necessary care in discussing the geometry of such surface water layers, a fair number of perturbations across the different networks. How does this structural evidence translate into changes of the thermodynamic signature? The contributions of water to ligand binding are rather tricky. In more simple cases, single water molecules that are captured on interstitial positions to mediate an interaction between ligand and protein show minor contributions to the Gibbs free energy of binding.[25,26] Nevertheless, much larger changes in enthalpy and entropy are experienced. An enthalpic gain is found due to the formation of new water-mediated hydrogen bonds between protein and ligand. This effect is virtually compensated by entropy as capturing the water molecule is entropically unfavorable. Possibly these effects are less outweighed if charges are involved.[27] Thus, on first glance, it appears rather difficult to relate the observed network changes with the thermodynamic signature discovered for the different complexes.

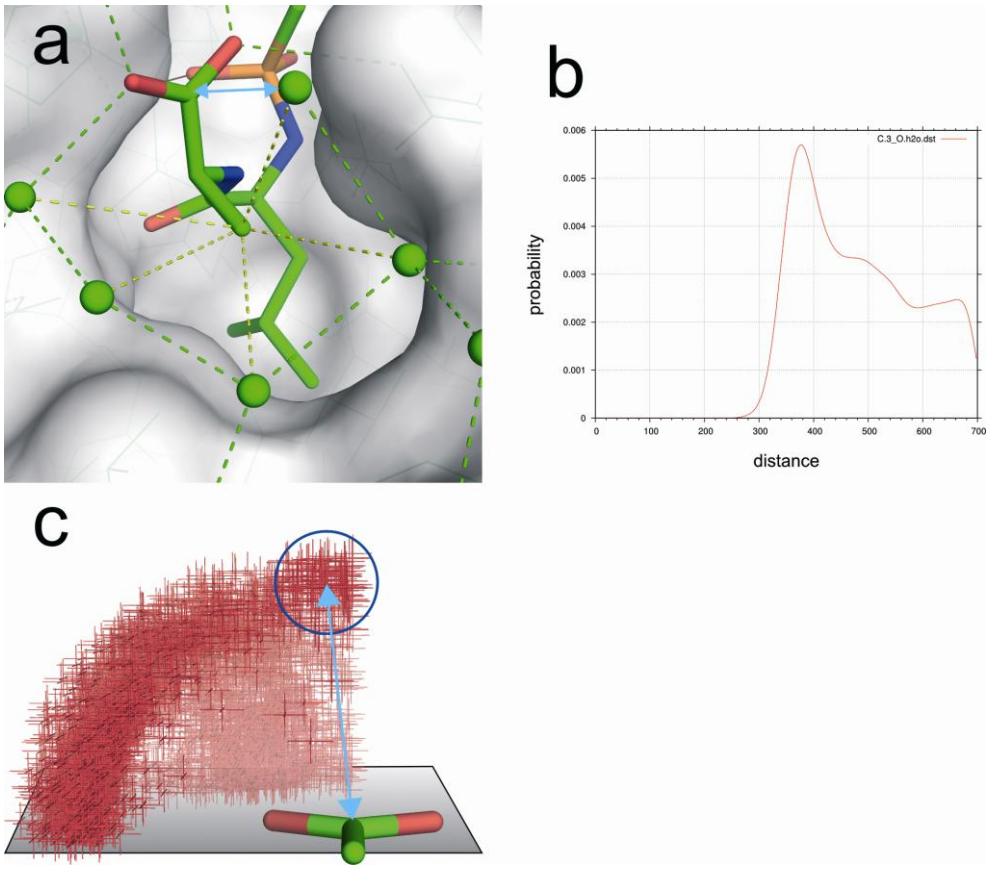
In three of the four complexes (**1**, **2**, **4**) a contiguous water chain spans from the carbonyl oxygen of the Cbz group all the way through to the backbone carbonyl oxygen of Asn111 and wraps around the terminal ends of the bound ligands. The attached COO^- groups in **3** and **4** replace some (most likely three, Fig. 4 b, c) of the water molecules present in TLN-**1** and TLN-**2** and drag the water network towards their polar oxygens (Fig. 3 a, c and Fig. 4 b, c, black cycles), particularly to form the syn hydrogen bond. In the complex with **3**, however, the generated water network remains ruptured and leaves a supposedly unfavorable break. In **4** the water chain shows up again with continuous geometry, most likely supported by the stabilizing van-der-Waals contacts to the now adjacent methyl group. This observation is underscored by the fact that in TLN-**2** at very similar site, an additional water molecule is

captured (Fig. 4 c, red cycles and arrows). The stabilizing effect of the methyl group is also indicated by a second water binding site which only exists in TLN-**2** and TLN-**4** (Fig. 3 b, d, 4 a, c, d, red cycles and arrows). To support this stabilizing contribution of this water-to-methyl contact we inspected the occurrence frequency of such contacts in crystal structures (Fig. 6).[28]

Compared to **1**, **3** shows only minor affinity enhancements with little enthalpy and unfavorable entropy even though a polar carboxylate group is introduced. Supposedly the costly water network rupture prevents any affinity enhancement resulting from the added polar group.

The addition of a methyl group from **3** → **4** results in a huge affinity increase, which is enthalpically driven and partly compensated by an entropic loss. The wrapping water chain is closed again and further stabilized by the water molecule picked up at the capping position (Fig. 3 b, d, 4 a, c, d, red cycle). This latter site likely experiences electrostatic interactions in TLN-**4** with the adjacent carboxylate group (3.6 Å); in TLN-**3** no water molecule is found at this position. Following the path from **1** → **2** shows minor exergonic effects but falls into a range also found in other examples for methyl group attachments.[29,30] The observed small enthalpy/entropy signature most likely results from the higher desolvation price to be paid for the added hydrophobic ligand portion. Finally, the step **2** → **4** which adds the COO^- group to the ligand returns a fairly strong affinity enhancement. It factorizes into a huge enthalpic gain and a strong entropic loss. This agrees with a signature expected for the introduction of a polar functional group which creates significant electrostatic effects and restructures the local water network (Fig. 4 b, c, black crosses and arrows).

Figure 6: (a) Five water molecules in TLN-4 with close contacts (yellow dashed line) to the terminal methyl group of the ligand (3.8 – 4.0 Å). One water molecule occupies a site perpendicular to the carboxylate group (cyan arrow). (b) Frequency distribution of intermolecular contacts between a water molecule and a terminal methyl group in crystals structures deposited in the PDB. The histographic distribution has been transformed according to the formalism in Drugscore (DSX)[28] into the depicted smoothed distribution. The maximum indicates high occurrence frequency at 380 pm (3.8 Å), correlating with an energetically favorable arrangement. (c) Distribution of intermolecular contacts as compiled in IsoStar between a carboxylate group and water molecules. Apart from formed hydrogen bonds in the plane of the carboxylate group also at a position above the best plane contacting water molecules are observed. The frequent population at this site indicates favorable electrostatic interactions between a water molecule and a carboxylate group.



5.7 Conclusions

The present structural and thermodynamic study of binding properties of four closely related, highly congeneric ligands demonstrates the effects of non-additivity of functional group contributions. As the binding mode of all four inhibitors remains virtually unchanged, the observed affinity differences must result from other contributions. Attachment of a methyl group to the parent scaffold increases affinity by about - 2 kJ/mol, a value expected for a favorable methyl group placement. It also factorizes only slightly into enthalpy and entropy. Adding a further carboxylate group results in a pronounced affinity gain combined

with huge, however, largely compensating exothermic and entropically unfavorable effects. Also such a profile is perhaps expected for the attachment of a polar group, well suited to establish strong interactions with its environment. It introduces charge-assisted hydrogen bonds assumed to be enthalpically favorable. At the same time the intermolecular contacts become more stringent which provokes a loss in entropy. Performing the attachments with inverted sequence results in effects with reverse order. Why is the carboxylate attachment now only yielding about - 1 kJ/mol in Gibbs free energy and why is hardly any enthalpy/entropy compensation detected? Instead the subsequently added methyl group provokes the strong affinity increase with huge enthalpy/entropy compensation. Three of the complexes show a contiguous and complete water network which wraps around the terminal hydrophobic substituent and involves the carboxylate group. In the complex with the carboxylated ligand **3**, which still lacks the additional methyl group and does not achieve the expected affinity boost, the water network is broken and appears incomplete. Supposedly this rupture is enthalpically unfavorable and entropically beneficial as it captures less water molecules. Overall this counterbalances any favorable contributions resulting from the introduction of the polar functional group. This is, however curated with the introduction of an additional methyl group as it allows the network to be re-established via the supply of a favorable interaction site to host a water molecule (Fig. 6 a).

Two aspects can be learned from this study. First, non-additivity of functional group contributions can have different molecular origins; apart from dynamics also local solvation patters can be responsible. Second, binding affinity and the enthalpy/entropy signature can be strongly influenced by the water network formed as solvation layer next to a binding pocket, as similarly concluded by Nasief et al.[16] even though using different arguments based on different considerations. Regarding our currently used routine computer tools for modeling[31] none of these effects will be reflected by any means. This demands for future developments to improve such tools.

5.8 Materials and Methods

5.8.1 Synthesis

The compounds were kindly provided by Nasief and Hangauer (SUNY, Buffalo, USA) and the synthesis is described and discussed elsewhere.[16]

5.8.2 Isothermal titration calorimetry (ITC)

ITC experiments were performed using an ITC200™ system from MicroCal (now part of GE Healthcare, Northhampton, MA, USA). Thermolysin from Calbiochem was freshly prepared for each experiment in the buffer applied in the measurements. The comparison of the thermodynamic data of the dialysed and undialysed protein solution revealed no significant difference. Therefore the undialysed protein solution was used in this study. The inhibitor solution was freshly prepared in the experimental buffer to the required concentration. The ITC experiments were done at 25 °C in a buffer composed of 20 mM Hepes, 500 mM NaSCN, 2 mM $CaCl_2$ at pH 7.5. Inhibitor solutions (1-1.5 mM) were titrated into the sample cell containing the TLN solution (45.0 µM) after a stable baseline was achieved. ITC experiments comprised of initial ligand injection of 0.3 µl followed by 15 injections of 1-2 µl with a 300 s interval between each injection. The ITC cell volume was 200 µl. Raw data were collected and the area under each peak was integrated, followed by correction for heats of dilution and mixing by subtracting the final baseline consisting of small peaks of the same size to zero. The initial data point was deleted from the integrated data because this injection usually reflects an erroneous amount of heat due to the possible exchange of liquids between syringe and cell when inserting the syringe into the calorimetric cell and the backlash error in the motorized screw mechanism in the injector.[32] Data were analysed using ORIGIN Software (Microcal Inc.), by fitting a single-site-binding isotherm [33] that yields ΔH_0 (enthalpy of binding) and K_d (dissociation constant). Measurements were performed at least in duplicate and show on the mean esds of 0.4 kJ/mol.

5.8.3 Crystallization and soaking

Native thermolysin (purchased from Calbiochem) was crystallized as described by Holmes and Matthews [34] applying slight modifications. Thermolysin was dissolved in 100 % DMSO at 8 mM. Subsequently, the DMSO solution was diluted with the buffer 1:1 (100 mM Tris/HCl, 5 M CsCl, pH 7.5). The suspension was then vortexed and centrifuged for two minutes. The so achieved supernatant was used in the crystallization trial. This final solution contained 50 % DMSO and 2.5 M CsCl. The final protein concentration in the crystallization drop was 4.0 mM. Crystals were grown at 18 °C by the sitting drop vapor diffusion method using water as reservoir solution. Protein-ligand complex crystals were obtained via soaking crystals in a solution of 100 mM Tris/HCl, 2 mM CaCl and 10 % DMSO. For soaking ligand

concentration varied between 1 and 5 mM. When low solubility was given, saturated solutions were used. Crystals were soaked between one and two days before freezing.

5.8.4 Data collection and processing

Crystals were prepared for data collection at 110 K using a cryoprotectant solution of 20 % glycerol in a buffer containing 10 mM Tris/HCl, 10 mM $Ca(CH_3COO)_2$, 5 % DMSO at pH 7.3. The data sets for **1** and **3** were collected with synchrotron radiation at SLS (Villingen, Switzerland) on a Marmosaic 225 mm CCD detector. Complex structures for **2** and **4** were collected at BESSY beamline 14.2 (Berlin, Germany) on a Rayonix MX 225 CCD detector. Data processing and scaling were performed using the HKL2000 package.[35]

5.8.5 Structure determination and refinement

The coordinates of thermolysin (PDB code 8TLN)[36] were used for initial rigid body refinement of the protein molecules followed by repeated cycles of maximum likelihood energy minimization, simulated annealing and B-factor refinement using the CNS program package.[37] Subsequently, all structures were refined with PHENIX.[38] The temperature factors for structure **4** were anisotropically refined whereas for structures **1** – **3** TLS refinement was applied. The definition of the TLS groups was done with the TLSMD server.[39,40] A randomly chosen 5 % of all data were used for the calculation of R_{free} and were not used in the refinement. Amino acid side chains were fit into σ-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps using Coot.[41] After the first refinement cycle, water molecules and subsequently ions and ligands were located in the electron density and added to the model. Restraints were applied to bond lengths and angles, planarity of aromatic rings and van der Waals contacts. Multiple side chain conformations were built in case an appropriate electron density was observed and maintained during the refinement, and if the minor populated side chain showed at least 20 % occupancy. The final models were validated using PHENIX own validation options or MolProbity.[42] The Ramachandran plot's were calculated with PROCHECK.[43] Data collection, unit cell parameters and refinement statistics are given in Table 1. Analysis of temperature factors was done with Moleman[44] and fconv.[45] The figures were prepared using Pymol 1.2r3pre.[46] The Protein Data Bank accession codes of the coordinates and structure factors of all X-ray structures are given below.

5.8.6 Protein Data Bank and accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with the following accession codes: TLN-**1** complex 3T73; TLN-**2** complex 3T8F; TLN-**3** complex 3T8G; TLN-**4** complex 3T74.

Table 1: Data collection and refinement statistics for the four thermolysin complex structures determined in this publication.

Structure (PDB entry)	TLN-**3** complex **(3T8G)**	TLN-**4** complex **(3T74)**	TLN-**1** complex **(3T73)**	TLN-**2** complex **(3T8F)**
A. Data collection and processing				
No. Crystals used	1	1	1	1
Wavelength (Å)	1.00	0.91841	1.00	0.91841
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters				
a, b (Å)	92.7, 92.7	92.6, 92.6	92.7, 92.7	92.4, 92.4
c (Å)	130.8	131.0	130.1	131.2
Matthews coefficient ($Å^3$/Da)	2.4	2.4	2.4	2.4
Solvent content (%)	48	48	48	48
B. Diffraction data[a]				
Resolution range (Å)	50 – 1.50	50 – 1.28	50 – 1.60	50 – 1.44
	(1.53 – 1.50)	(1.30 – 1.28)	(1.63 – 1.60)	(1.46 – 1.44)
Unique reflections	53,702	83,375	44,226	60,157
R(I)sym (%)	7.9 (49.9)	5.2 (32.6)	8.0 (49.5)	6.2 (27.2)
Completeness (%)	99.9 (99.5)	97.3 (95.4)	100.0 (100.0)	99.7 (96.2)
Redundancy	20.7 (18.5)	5.7 (5.7)	16.3 (15.1)	5.0 (4.3)
I/σ(I)	41.1 (6.6)	30.4 (5.3)	35.9 (6.5)	22.0 (5.4)
C. Refinement				
Resolution range (Å)	40.1 – 1.50	22.8 – 1.28	43.7 – 1.60	30.4 – 1.44
Reflections used in refinement (work/free) (%)	52,333 / 2,649	80,965 / 4,095	42,664 / 2,135	58,574 / 2,951
Final R values for all reflections (work/free) (%)	14.2 / 15.6	10.5 /12.7	14.3 / 16.6	14.2 / 15.9
Protein residues	316	316	316	316
Calcium /zinc ions	4 / 1	4 / 1	4 / 1	4 / 1
Inhibitor atoms	28[b] / 32[c]	29[b] / 33[c]	25[b] / 29[c]	26[b] / 30[c]
Water molecules	442	499	400	477
RMSD from ideality				
Bond lengths (Å)	0.012	0.013	0.012	0.012

Bond angles (°)	1.054	1.318	1.055	1.048
Ramachandran plot				
Residues in most favoured regions (%)	88.1	88.1	88.5	88.9
Residues in additionally allowed regions (%)	10.7	10.7	10.7	10.4
Residues in generously allowed regions (%)	0.7	0.7	0.4	0.4
Residues in disallowed regions (%)	0.4	0.4	0.4	0.4
Mean B-factor (Å^2)				
Protein	11.9	9.4	14.2	10.3
Binding site[d]	9.1	6.7	11.2	7.6
Inhibitor	12.0[e]	8.4[e]	14.4[e]	9.2[e]
Water molecules	28.0	27.1	30.5	26.9

[a] Numbers in parenthesis are for the highest resolution shell.
[b] The inhibitor atoms from the additional disordered part were not considered
[c] All inhibitor atoms which were included in the refinement model
[d] Definition of the binding site: all amino acids which are 4 Å away from inhibitor.
[e] Average B value for the ordered and fully occupied inhibitor portion, the disordered portions were not considered

5.9 References

1 Andrews P.R., Craik D.J. & Martin J.L. (1984) Functional group contributions to drug-receptor interactions. *J Med Chem* **27**, 1648-1657.

2 Kuntz I.D., Chen K., Sharp K.A. & Kollman P.A. (1999) The maximal affinity of ligands. *Proc Natl Acad Sci U S A* **96**, 9997-10002.

3 Dill K.A. (1997) Additivity principles in biochemistry. *J Biol Chem* **272**, 701-704.

4 Mark A.E. & van Gunsteren W.F. (1994) Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J Mol Biol* **240**, 167-176.

5 Williams D.H., Searle M.S., Mackay J.P., Gerhard U. & Maplestone R.A. (1993) Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. *Proc Natl Acad Sci U S A* **90**, 1172-1178.

6 Williams D.H., Stephens E., O'Brien D.P. & Zhou M. (2004) Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew Chem Int Ed Engl* **43**, 6596-6616.

7 Williams D.H., Zhou M. & Stephens E. (2006) Ligand binding energy and enzyme efficiency from reductions in protein dynamics. *J Mol Biol* **355**, 760-767.

8 Kawai S.H., Bailey M.D, Halmos T., Forgione P., Laplante S.R. *et al.* (2008) The use of chemical double-mutant cycles in biomolecular recognition studies: application to HCV NS3 protease inhibitors. *ChemMedChem* **3**, 1654-1657.

9 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.

10 Muley L., Baum B., Smolinski M., Freindorf M., Heine A., Klebe G. & Hangauer D.G. (2010) Enhancement of hydrophobic interactions and hydrogen bond strength by cooperativity: synthesis, modeling, and molecular dynamics simulations of a congeneric series of thrombin inhibitors. J Med Chem **53**, 2126-2135.

11 Dunitz J.D. (1995) Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol* **2**, 709-712.

12 Whitesides G.M. & Krishnamurthy V.M. (2005) Designing ligands to bind proteins. *Q Rev Biophys* **38**, 385-395.

13 Ladbury J.E. (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.

14 Homans S.W. (2007) Water, water everywhere--except where it matters? *Drug Discov Today* **12**, 534-539.

15 Wang L., Berne B.J. & Friesner R.A. (2010) Ligand binding to protein-binding pockets with wet and dry regions. *Proc Natl Acad Sci U S A* **108**, 1326-1330.

16 Nasief N.N., Tan H., Kong J. & Hangauer D. submitted.

17 Dunitz J.D. (1994) The entropic cost of bound water in crystals and biomolecules. Science **264**, 670.

18 Morgan B.P., Scholtz J.M., Ballinger M.D., Zipkin I.D. & Bartlett P.A. (1991) Differential binding energy: A detailed evaluation of the influence of hydrogen-bonding and hydrophobic groups on the inhibition of Thermolysin by phosphorus-containing inhibitors. *J Am Chem Soc* **113**, 297-307.

19 Hangauer D.G., Monzingo A.F., Matthews B.W. (1984) An interactive computer graphics study of thermolysin-catalyzed peptide cleavage and inhibition by N-carboxymethyl dipeptides. *Biochemistry* **23**, 5730-5741.

20 Holden H.M., Tronrud D.E., Monzingo A.F., Weaver L.H. & Matthews B.W. (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.

21 Carugo O. & Bordo D. (1999) How many water molecules can be detected by protein crystallography? *Acta Crystallogr D Biol Crystallogr* **55**, 479-483.

22 Lee J., Kim S.H. (2009) Water polygons in high-resolution protein crystal structures. *Protein Sci* **18**, 1370-1376.

23 Görbitz C.H. & Etter M.C. (1992) Hydrogen bonds to carboxylate groups. Syn/anti distributions and steric effects. *J Am Chem Soc* **114**, 627-631.

24 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 S_1'-pocket of thermolysin. *Biochim Biophys Acta* **1800**, 1192-1202.

25 Petrova T., Steuber H., Hazemann I., Cousido-Siah A., Mitschler A., Chung R., Oka M., Klebe G., El-Kabbani O., Joachimiak A., Podjarny A. (2005) Factorizing selectivity determinants of inhibitor binding toward aldose and aldehyde reductases: structural and thermodynamic properties of the aldose reductase mutant Leu300Pro-fidarestat complex. *J Med Chem* **48**, 5659-5665.

26 Biela A., Sielaff F., Terwesten F., Heine A. & Klebe G. Ligand binding gradually disrupts water network in thrombin: Enthalpic and entropic changes reveal classical hydrophobic effect. *J Med Chem*:submitted.

27 Steuber H., Heine A. & Klebe G. (2007) Structural and thermodynamic study on aldose reductase: nitro-substituted inhibitors with strong enthalpic binding contribution. *J Mol Biol* **368**, 618-638.

28 Neudert G. & Klebe G. (2011) DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *J Chem Inf Model* **51**, 2731-2745.

29 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.

30 Barreiro E.J., Kummerle A.E. & Fraga C.A. (2011) The methylation effect in medicinal chemistry. *Chem Rev* **111**, 5215-5246.

31 Warren G.L., Andrews C.W., Capelli A.M., Clarke B., LaLonde J., Lambert M.H., Lindvall M., Nevins N., Semus S.F., Senger S., Tedesco G., Wall I.D., Woolven J.M., Peishoff C.E., Head M.S. (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* **49**, 5912-5931.

32 Mizoue L.S. & Tellinghuisen J. (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal Biochem* **326**, 125-127.

33 Wiseman T., Williston S., Brandts J.F. & Lin L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**, 131-137.

34 Holmes M.A. & Matthews B.W. (1982) Structure of thermolysin refined at 1.6 A resolution. *J Mol Biol* **160**, 623-639.

35 Otwinowski Z. & Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzymol* **276**, 307-326.

36 Holland D.R., Tronrud D.E., Pley H.W., Flaherty K.M., Stark W., Jansonius J.N., McKay D.B. & Matthews B.W. (1992) Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* **31**, 11310-11316.

37 Brunger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.S., Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T. & Warren G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921.

38 Adams P.D., Afonine P.V., Bunkoczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C. & Zwart P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.

39 Painter J. & Merritt E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-450.

40 Painter J. & Merritt E.A. (2006) TLSMD web server for the generation of multi-group TLS models. *J Appl Cryst* **39**, 109-111.

41 Emsley P. & Cowtan K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

42 Chen V.B., Arendall W.Br., Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S. & Richardson D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

43 Laskowski R.A., MacArthur M.W., Moss D.S. & Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallog* **26**, 283-291.

44 Kleywegt G.J, Zou J.Y., Kjeldgaard M. & Jones T.A. (2001) Around O. Rossmann M.G., Arnold E., editors. Dordrecht: Kluwer Academic Publishers; 2001. 353-356 p.

45 Neudert G. & Klebe G. (2011) fconv: Format conversion, manipulation and feature computation of molecular data. *Bioinformatics* **27**, 1021-1022.

46 The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

46 Bruno I.J., Cole J.C., Lommerse J.P., Rowland R.S., Taylor R. & Verdonk M.L. (1997) IsoStar: a library of information about nonbonded interactions. *J Comput Aided Mol* **11**, 525-537.

6 Dissecting the Hydrophobic Effect on Molecular Level: The Role of Water, Enthalpy and Entropy in Ligand Binding to Thermolysin

6.1 Introductory remarks

This study was done in cooperation with the group of Prof. Dr. Hangauer (University of Buffalo). The following text will be submitted to a scientific journal. Nader Nasief from the Hangauer group is included as second author.

6.2 Abstract

The hydrophobic effect in protein-ligand complex formation is related to the successive replacement of binding-site water molecules by hydrophobic ligand portions. Even though it is assumed that the hydrophobic effect is primarily entropy-driven, detailed studies have revealed a much more complex picture involving either enthalpic or entropic effects. In thermolysin, adjacent to the ligand's P_2'substituents of increasing size and hydrophobicity, we observed in eight congeneric high resolution crystal structures systematic changes and perturbations of the local water network in the first solvation layer around the protein. These changes correlate remarkably well with differences in the thermodynamic signature of protein-ligand binding. The substituents of increasing size and hydrophobicity replace a growing number of water molecules, however, depending on the local binding features of the perturbed water molecules and their embedding in the local water network a more enthalpic or entropic signature is created. With respect to the Gibbs free energy of binding pronounced enthalpy/entropy compensation is in operation. However, there are no arguments why the hydrophobic effect should be predominantly enthalpic or entropic, small structural changes in the binding features of the water molecules finally determine whether hydrophobic binding to the hydrophobic pocket results in a more enthalpy or entropy-driven signature.

6. Dissecting the Hydrophobic Effect on Molecular Level

6.3 Introduction

The hydrophobic effect is attributed to the general driving force of nonpolar substances with extented lipophilic molecular surfaces to aggregate in aqueous solution excluding water molecules from the formed interfaces.[1,2] It is usually quoted to explain why a mixture of oil and water spontaneously separates, why soluble proteins fold with architecture showing hydrophobic core and hydrophilic outer surface[3,4], why membrane components assemble in lipid bilayers and micelles, why membrane proteins accommodate in membrane segments or why small molecules associate in protein binding pockets with mutual burial of hydrophobic surface portions.[5] In the later instance, it is a general strategy in medicinal chemistry to enhance protein-ligand binding through tailored increase of the ligand's hydrophobic surface which becomes buried in hydrophobic pockets of the target protein. In all cases, the hydrophobic effect serves as an explanation for the association. On molecular level, the observed phenomena are related to the displacement of water molecules and entropic effects are made responsible to drive this association. Arguments are related to a change in the degrees of ordering and the extent of dynamic properties of the water molecules which are assumed to be more ordered and fixed next to surfaces than in the bulk water phase to where they are released upon hydrophobic association. Recent studies have demonstrated, however, that hydrophobic binding phenomena can originate either from enthalpy-driven or entropy-driven binding, which makes the assumed clear-cut picture more blurred[6-12].

To create a better understanding on the molecular level for the hydrophobic effect in case of protein-ligand binding we embarked on the systematic study of thermolysin (TLN)[6], a thermostable bacterial zinc metalloprotease from *Bacillus thermoproteolyticus* which exhibits three specificity pockets of predominantly hydrophobic nature (Figure 1). Despite of remarkable sequence differences to other zinc metalloproteases, this protein has been considered prototype for the entire class of enzymes[13] due to a highly conserved active-site architecture. It has frequently been used as surrogate for the development of new drugs against other metalloenzymes or served as a model system to test ideas and methodological concepts, e.g. in identifying potential drug candidates[14-17] or to investigate probe molecules[18,19] and fragments.[20] TLN was one of the first metalloproteases for which a crystal structure was determined.[21,22] Its easily accessible catalytic center consists of the central zinc ion coordinated by His142, His146 and Glu166. The adjacent S_1 subsite is rather unspecific and

accommodates hydrophobic ligand portions.[23] In contrast, the S_1' pocket is a deep and well-defined cavity and it also accommodates preferentially hydrophobic residues such as the side chains of valine, leucine, or phenylalanine and determines substrate specificity of TLN.[24] The neighboring S_2' pocket is shallow and flat. Even though it has clearly hydrophobic characteristics, it is well accessible to the bulk water phase. Nevertheless, with respect to the binding of ligand side chains, it hosts similar groups as the S_1' pocket. Based on recently determined high-resolution crystal structures, a sophisticated arrangement of a complex water network has been observed next to the S_2' pocket that takes dominant influence on binding properties of accommodated ligands.[25,26]

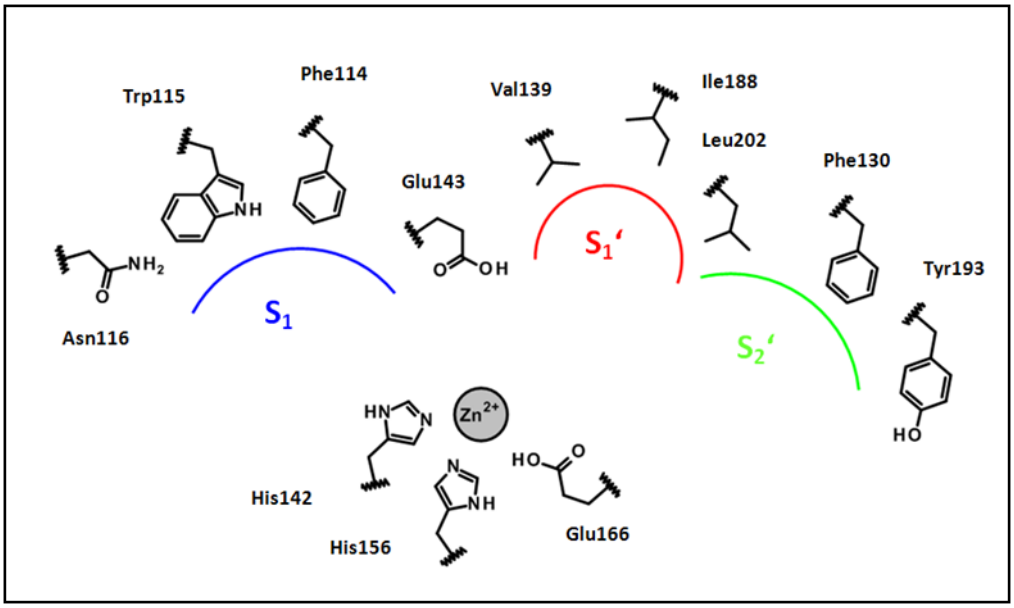
Potent TLN inhibitors are often designed as transition state analogs addressing both hydrophobic S_1' and S_2' pockets, e.g., the highly potent phosphoramidon comprises a phosphinate group to coordinate the catalytic zinc ion in bidentate fashion.[23,27,28] Besides zinc binding, the occupancy of both hydrophobic cavities is of utmost importance. Inhibitors occupying the S_1' pocket with an isobutyl instead of a methyl side chain exhibit a dramatic 800-fold enhanced potency whereas for S_2' a 50-fold increase is experienced.[6] Furthermore, increasing the hydrophobic interactions in the S_1' pocket is strikingly enthalpy-driven and not, as assumed for the classical hydrophobic effect, entropically beneficial.[6] The favorable enthalpic signal, observed for growing hydrophobicity of the P_1' substituent in the congeneric series, was attributed to "poor solvation" of the S_1' pocket. The latter property most likely does not relate to a water-free enzyme pocket in unbound state but to the accommodation of several highly mobile water molecules which are scattered over multiple positions and thus are hardly detectable by crystal structure analysis.

Homans *et al.*[7,8] reported on a similar enthalpy-driven hydrophobic interaction which has also been attributed to suboptimal solvation of a protein pocket. The desolvation enthalpy of the protein binding pocket is strongly reduced that overall the thermodynamic signature shifts to an overwhelmingly enthalpy-driven Gibbs free energy of the binding process. Synder *et al.*[9] also received an enthalpy-driven thermodynamic profile for heterocyclic aromatic sulfonamides with increasing hydrophobic properties against carbonic anhydrase. They explain the observed differences with changes in the number and organization of well-ordered water molecules in the binding site. In a study regarding the displacement of several well-ordered water molecules from the $S_{3/4}$ pocket of thrombin by increasingly hydrophobic P_3 substituents of peptidomimetic inhibitors we observed an entropy-driven signal.[10] In the field of host-guest chemistry several

examples of complex formation have been reported that either relate to enthalpy or entropy increase.[11,12] These studies indicate that the thermodynamic signature of hydrophobic binding is determined by changes in the water structure and the observed thermodynamic signature is obviously dependent on the status of the water molecules being reorganized during the binding process.

Since the S_1' and S_2' pockets of TLN exhibit opposing features with respect to the observed solvation pattern and shape but can host chemically similar molecular ligand portions, we embarked on the study of a congeneric series of peptidomimetics with step-by-step modifications in the P_2' side chain addressing the S_2' pocket. The complex structures are analyzed crystallographically to investigate changes in the water structure, modulated by interactions of the P_2' side chains. Isothermal Titration Calorimetry (ITC) data were recorded to complement crystallographic findings and to reveal the driving forces associated with hydrophobic binding to the S_2' pocket of TLN.

Figure 1: A schematic view of the binding pocket of uncomplexed thermolysin. Asn111 and Asn112 are not shown for clarity.



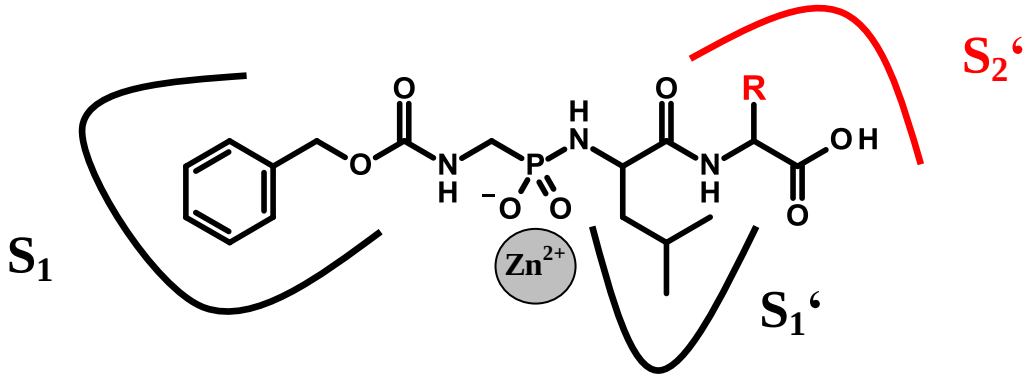
6.4 Results

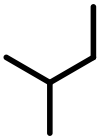
6.4.1 Analysis of the binding mode of the parent scaffold

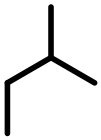
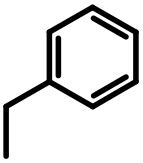
The crystal structures of 8 TLN inhibitors containing the Cbz-Gly-$(PO_2)^-$-L-Leu-L-X scaffold (Cbz = carboxybenzyl, X = Gly **1**, Ala **2**, Et-Gly **3**, Val **4**, nPr-Gly **5**, Ile **6**, Leu **7**, Phe **8**; s. Table 1) in complex with thermolysin have been determined with high resolution (1.28 - 1.66 Å). Crystal

structures with ligands **1**, **2** and **7** have been studied previously by us,[6,25] for TLN-**7** the original data have been newly refined to apply the same protocol to all complexes. As the binding mode of the parent scaffold has already been described,[25] we will only briefly depict the predominant interactions of the ligand to TLN's active site and focus on novel structural features in the series.

Table 1: The designed ligands are given along with the kinetically determined inhibition contants and the binding free energies.



Ligand name	Modification R	K_i (nM)	ΔG (kJ/mol) from K_i
1		472.3 ± 36.0	- 36.1 ± 0.2
2		19.1 ± 0.6	- 44.0 ± 0.1
3		4.8 ± 0.2	- 47.4 ± 0.1
4		5.7 ± 1.2	- 47.0 ± 0.5
5		5.5 ± 0.9	- 47.1 ± 0.4
6		12.2 ± 1.3	- 45.1 ± 0.3

7		10.3 ± 1.3	- 45.6 ± 0.3
8		54.0 ± 13.6	- 41.5 ± 0.6

The electron density of the scaffold is well defined for all the studied inhibitors (representatively shown for **3** in Figure 2), and no significant changes in the binding mode are observed among all ligands (Figure 3). Their Cbz moiety binds to the unspecific S_1 pocket, the central phosphonamidate group coordinates the zinc ion, and the leucyl P_1' and structurally varied P_2' substituents interact with the hydrophobic environment of the S_1' and S_2' pockets. The phenyl ring of the Cbz moiety is able to interact with CD2 of Phe114 through van-der-Waals contacts (4.0 Å, Figure 4 a). Furthermore, this moiety is fixed in position by the interaction with the carbonyl oxygen of the adjacent Trp115 (3.5-3.7 Å) and a glycerol molecule immobilized below the phenyl ring. The binding position of the Cbz moiety is clamped by intermolecular contacts to a neighboring mate in crystal packing. In all studied inhibitors, apart **5** and **7**, we observed a second conformer for the Cbz carbonyl group. One conformer orients the carbonyl oxygen towards Phe114 (3.5 Å, Figure 4 a), in the second the carbonyl oxygen interacts water-mediated through short H-bonds with unperturbed geometry (distances 2.4 and 2.7 Å, angle: 104.6°) with the inhibitor's terminal carboxylate group. The occupancy of the second conformer refines to 37 % for **3**, across the whole series it ranges between 37 % and 48 %. Likely, resolving this binding feature strongly depends on the accuracy of the determined crystal structure. For TLN-**5** and TLN-**7** where no additional conformer for the Cbz carbonyl group of the ligand is evidenced, the resolution amounts to 1.66 Å and 1.56 Å respectively, in all the other structures with the additional conformer present, a resolution < 1.60 Å is experienced.

The inhibitor's phosphonamidate group interacts with the zinc ion in monodentate fashion positioning one oxygen towards the zinc ion (2.0 Å) and the other towards OE1 of Glu143 (2.6 Å, Figure 4 b). The NH group of the phosphonamidate is strongly hydrogen-bonded to Ala113 (2.8 Å). The isobutyl group of the leucine fragment fits perfectly well into the S_1' pocket forming

multiple van-der-Waals contacts (3.7 - 4.1 Å) to hydrophobic residues in this pocket (e.g. Leu133, Val139, Ile188 and Leu202, Figure 4 c).

Figure 2: Inhibitor **3** in complex with TLN. View of the active site with the solvent-accessible surface of the protein in grey. Water molecules are not shown for clarity, the ligand is visualized by F_o-F_c difference electron density (green) at 2σ. The blue sphere represents the zinc ion.

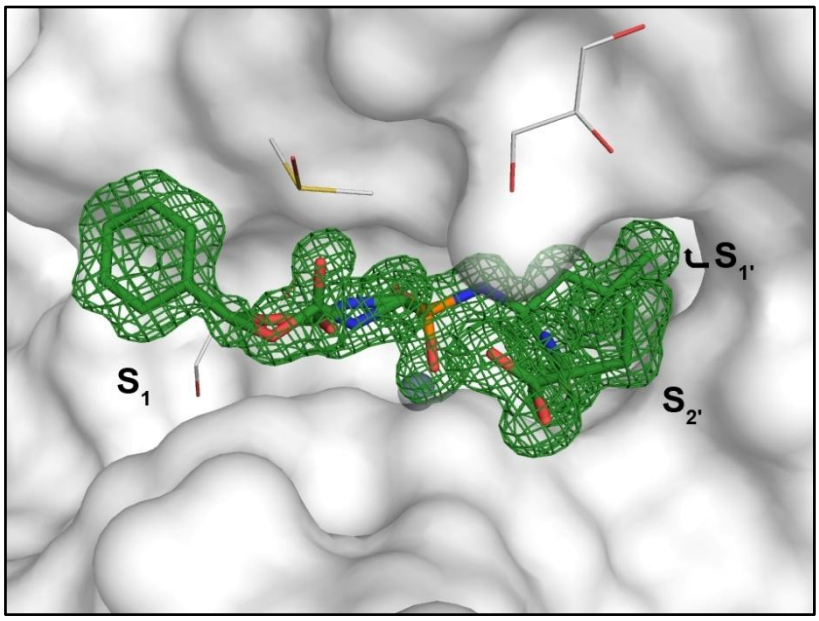


Figure 3: Superposition of TLN-**1**, TLN-**2**, TLN-**3**, TLN-**4**, TLN-**5**, TLN-**6**, TLN-**7** and TLN-**8**. Water molecules are omitted for clarity.

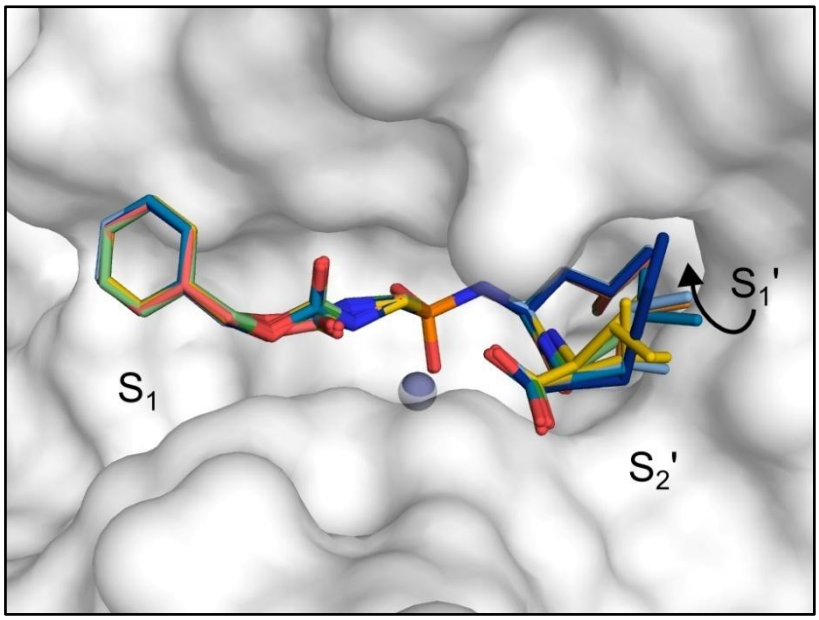
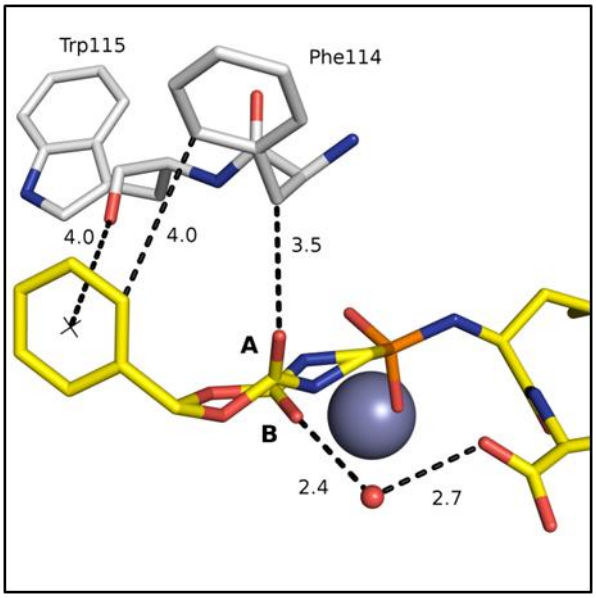
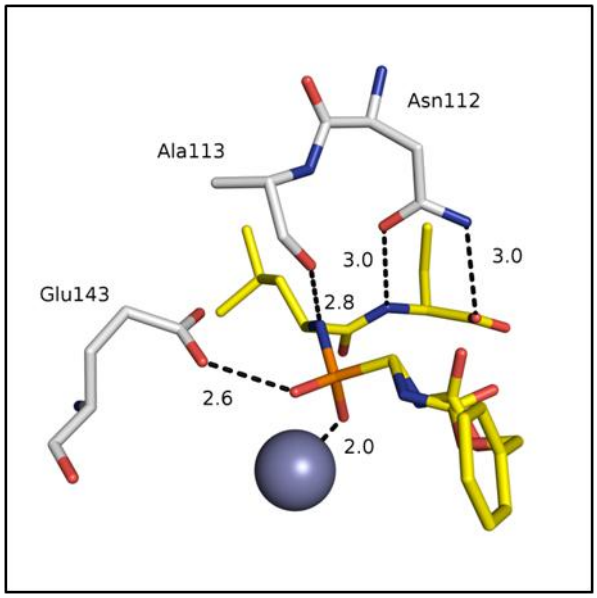


Figure 4: Crystallographically determined binding mode of **3**. The detailed interactions are illustrated next to the Cbz group (*a*), the phosphonamidate (*b*) and the P_1' side chain in the S_1'pocket (*c*). Oxygen is coloured in red, nitrogen in blue, phosphorous in orange, carbon in white (protein) and in yellow (inhibitor). Distances in Å are depicted as dotted lines.

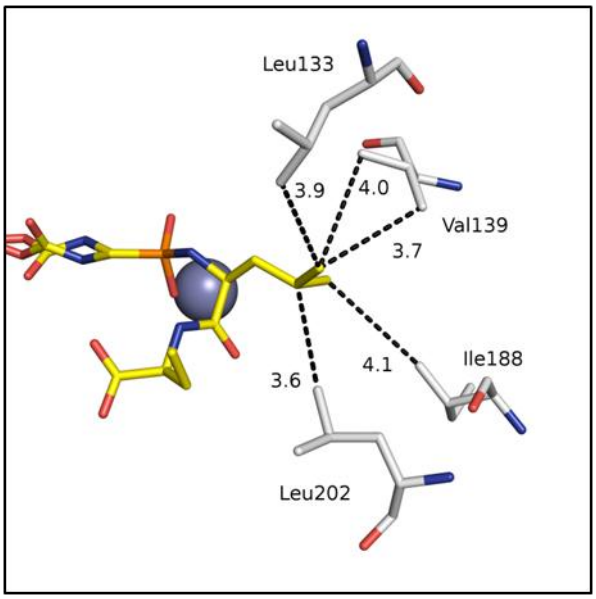
a)



b)



c)



6.4.2 Analysis of the binding mode of the hydrophobic P_2' substituent in the S_2' pocket

Major differences between the complexes are indicated with respect to the water network adjacent to the S_2' pocket. This network is perturbed and modulated by the size of the P_2' substituent. Unfortunately, not all complexes could be determined to the same resolution, the valyl **4** derivative even shows some disorder in this crucial region. For the related n-propylglycyl

5 residual difference density indicates that some disorder of the side chain might be given, however we decided to describe the final density by one model. Local disorder makes the reliable detection of water molecules with increasing distance from the protein surface or from polar ligand functional groups difficult; particularly if chains of contiguously connected water molecules are analyzed. Therefore, the diffraction properties of the crucial water molecules have been thoroughly inspected by difference electron density maps (F_o-F_c) to examine the accuracy and reliability of the hydration properties of the S_2' pocket. Particularly, the B-factors and occupancies which are highly correlated have been regarded with care. These limitations complicate a straight-forward comparison of absolute numbers of water molecules across the ligand series, especially considering the relative inventory of released or picked-up water molecules. The distances along the water network vary and may even correlate with the strength of formed hydrogen bonds. However, the determined spatial accuracy of individual water positions can be affected by, e.g. residual mobility, disorder or partial occupancy which limits positional accuracy. Therefore, we refrained from any detailed analysis of H-bond length variations.

6.4.3 Water network patterns

Virtually the same solvation pattern is observed for all complexes next to the Cbz carbonyl and negatively charged terminal carboxylate group (Figure 5, upper part). A network of at least seven mutually connected water molecules mediates interactions between the latter two ligand functional groups and Asp226, Asn227 and Asn112. In two complexes water molecules corresponding to the second solvation shell are indicated (TLN-**4**, TLN-**5**).

A more complex pattern is observed next to the area of the S_2' pocket where the growing P_2' side chain extends and perturbs the water network (Figure 5, lower part). In a previous study the crystal structures of TLN-**1** and TLN-**2** have been compared.[25] The glycine derivative **1** shows two water molecules hydrogen-bonded to the backbone carbonyl group of Asn111 which are repelled from the TLN-**2** complex (Figure 6 a, encircled in cyan) due to the steric conflicts of the attached methyl group in the alanyl derivative **2**. In contrast, the latter recruits two additional water molecules (Figure 6 b, encircled in yellow and green) which are picked-up and find favorable van-der-Waals contacts with the terminal methyl group. In TLN-**2** the water

network establishes a contiguously connected water chain from the ligand's carboxylate group to Asn111(C=O), whereas in TLN-**1** the water network is ruptured.

Figure 5: Binding modes of the ligands **1**-**8** next to the S_2' pocket. All eight ligands (color coding, cf. Fig. 6) are superimposed and the corresponding solvation shells are shown as colored spheres. In the upper part a virtually identical water network pattern is indicated, differences in the number of water molecules assigned to the density are most likely attributed to differences in the resolution of the individual structure determinations (1.28 – 1.66 Å). In the lower part next to the varied P_2' substituent of the ligand pronounced differences in the water network pattern can be detected.

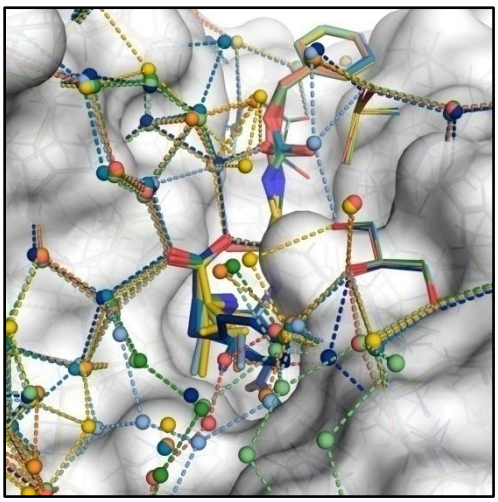
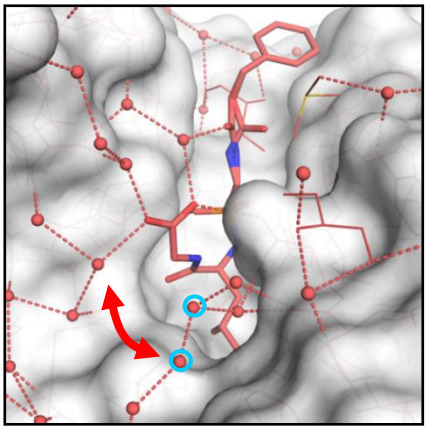
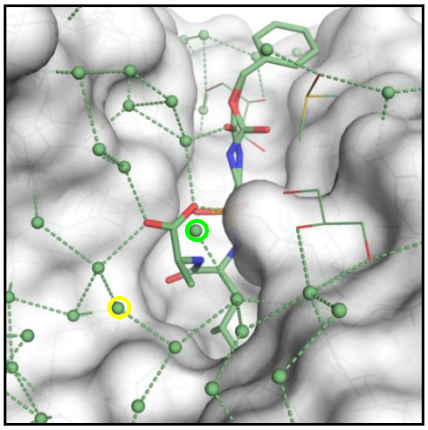


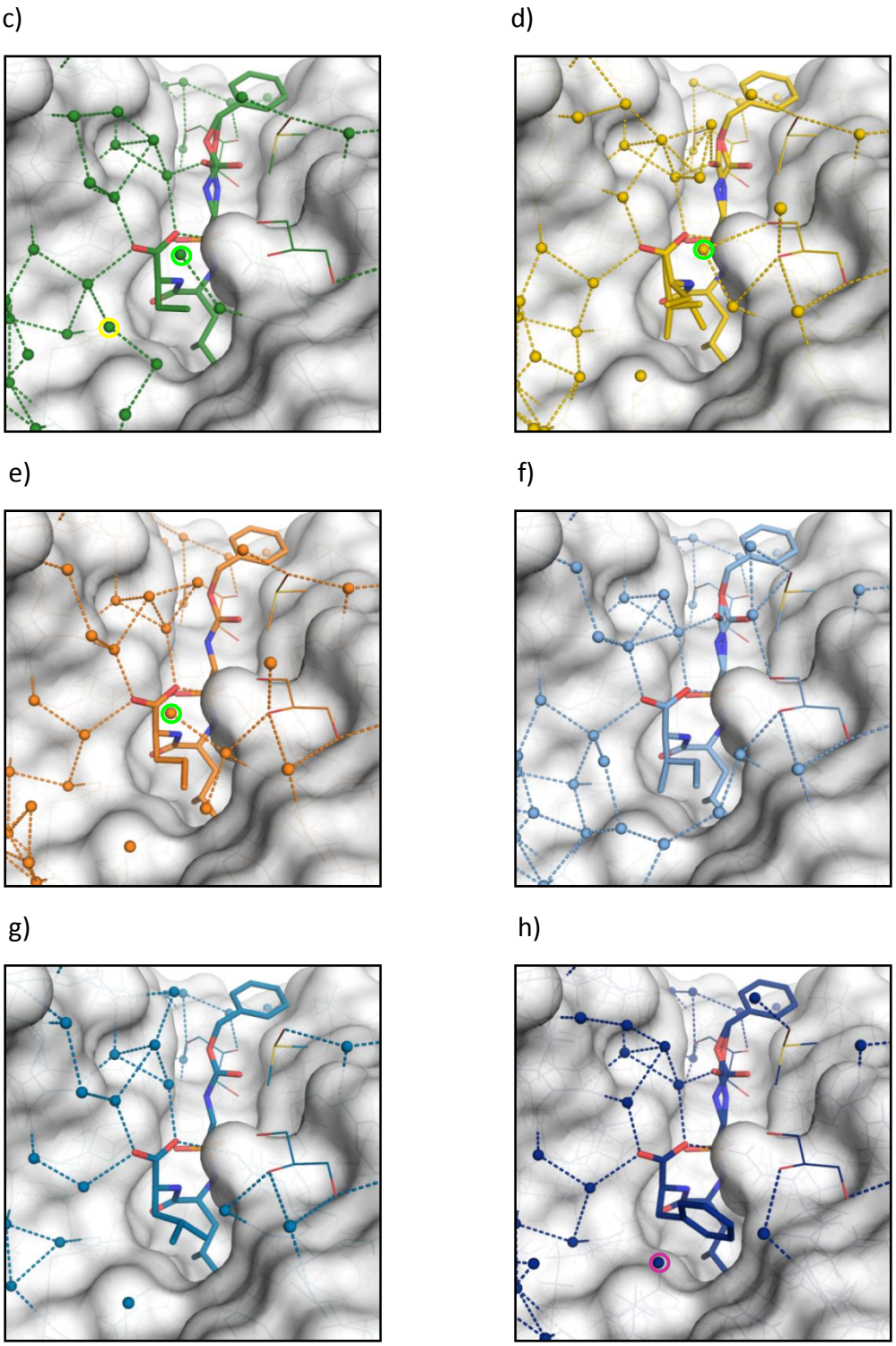
Figure 6: Binding modes of the ligands **1**-**8** (a-h) next to the S_2' pocket. Each complex is shown with a different color, heteroatoms in atom-type color coding, water molecules as spheres with the same color as the parent structure. The same colors are used in Fig. 5. In TLN-**1** two water molecules (encircled in cyan) are present that are replaced in the other complexes due to steric requirement of the growing P_2' substituent. TLN-**1** shows a break in a contiguously connected water network (red arrow) which is closed in TLN-**2** and TLN-**3** by the pick-up of an additional water molecule (encircled in yellow) which is stabilized by favorable van der Waals contacts with the P_2'-methyl or ethyl group in **2** or **3**. Similar favorable van der Waals interactions help to accommodate a water molecule at a position capping the ligand's carboxylate group (encircled in green) in **2**, **3**, **4**, and **5**. In the complexes with **6**, **7** and **8** this water molecule is repelled, whereas **8** picks up a water molecule (encircled in magenta) next to the benzyl moiety of the ligand's P_2' substituent.

a)



b)



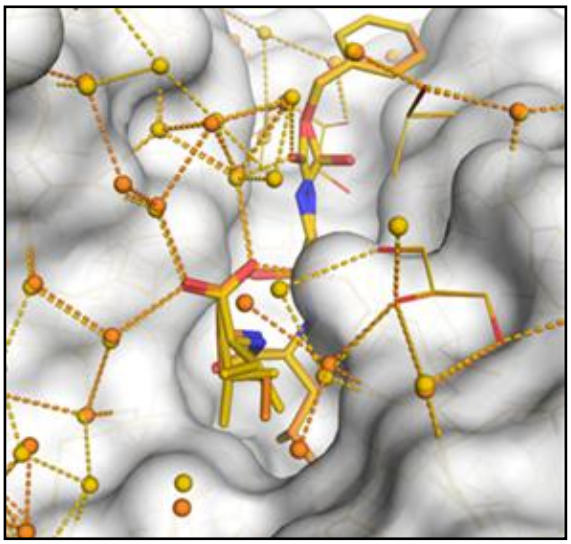


A comparison of the water networks in TLN-**2** and TLN-**3** (Figure 6 b, c) suggests nearly identical patterns, whereas those in TLN-**4** and TLN-**5** (Figure 6 d, e) seem to deviate and they are disconnected at the lower left rim of the pocket, nonetheless the valyl **4** and n-propyl **5** derivatives display very similar water network patterns. In TLN-**4**, the propyl side chain is scattered over at least two conformations. This partial disorder is translated to the neighboring

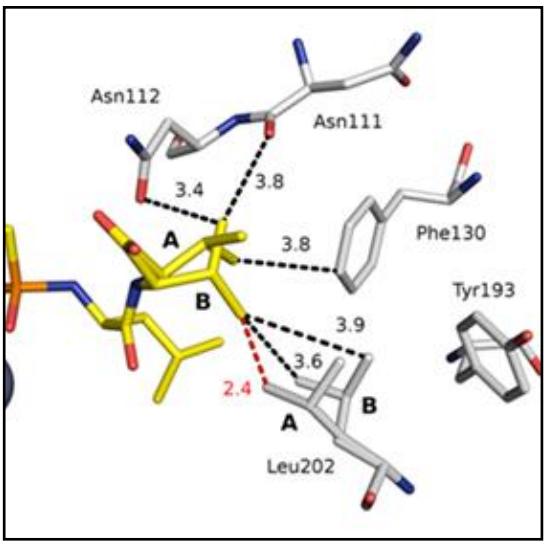
Leu202 residue as its iso-butyl side chain adopts two conformations which refine to 55 % and 45 % occupancy, respectively (Fig. 7). Both conformations occur in correlated manner due to mutual steric interference. Considering the water molecules picked up by TLN-**2** and **3** compared to TLN-**1**, the complexes with **4** and **5** show the water molecule capping the position of the carboxylate group (Figure 6 d, e, encircled in green).

Figure 7: (a) Crystallographically determined binding modes of TNL-4 (yellow) and TLN-5 (orange), heteroatoms type-coded, water molecules indicated as spheres with the same color as used for the respective crystal structure. The side chain of the P_2' substituent is observed with disorder over at least two conformations. (b) In case of the valyl derivative 4 the difference density suggests that also the side chain of the adjacent Leu202 residue adopts two conformations. Most likely conformation A corresponds for steric reason to the conformer A of the ligand, whereas conformation B aligns with ligand conformer B of the P_2' substituent. Selected distances in Å are depicted as dotted lines.

a)



b)

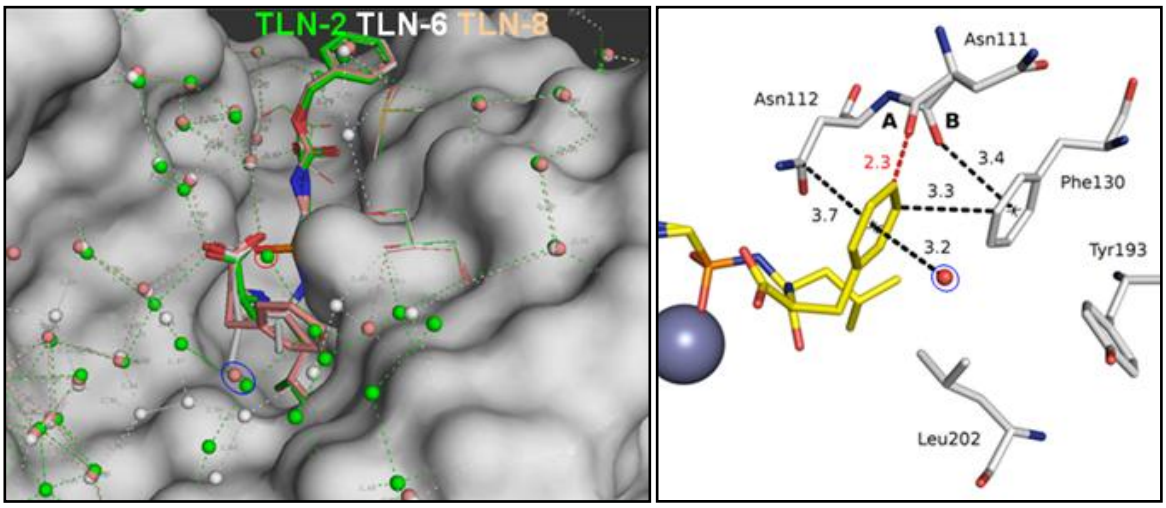


The complex of the Ile derivative **6** (Figure 6 f) makes again a contiguously connected water network evident which wraps around the terminal hydrophobic group as also seen in TLN-**2** and TLN-**3**. The network takes a more extended detour around the butyl group compared to TLN-**2**. TLN-**7**, reported in a previous study (PDB code 3FWD[6]), was re-refined in the present study to apply exactly the same refinement protocol and program suite. Even though the indicated network is not as complete as for TLN-**6**, a related pattern is indicated for TLN-**7** (Figure 6 g). Remarkably, for both complexes the water molecule at the position capping the carboxylate group is no longer present for steric reasons, the site occupied in all complexes with **2**, **3**, **4**, and **5** (Figure 6 b - e, encircled in green) would sterically interfere with the more bulky butyl side

chains in TLN-**6** and TLN-**7**. Finally, TLN-**8** with a terminal benzyl moiety shows the least amount of ordered water molecules next to the S_2' pocket. Is seems that nearly all water molecules observed in other complexes close to the lower rim of the S_2' pocket are either repelled or not sufficiently well ordered. Most likely this correlates with increasing steric requirements of the benzyl group which fills the S_2' pocket quite significantly. Remarkably, however, one water molecule returns back into the complex next to the ligand binding site that is also found in TLN-**2** and TLN-**3** (Figure 6 h, encircled in magenta) and occupies a site very close to the ligand. This site is clearly not accommodated in TLN-**6** and TLN-**7**. At this site in TLN-**8** a water molecule finds stabilizing interactions (3.2 Å) with the π-system of the ligand's neighboring phenyl ring (Fig. 8). Interestingly enough, the water molecule at the position capping the ligand's carboxylate group is missing in TLN-**8**, apparently again due to steric interference with the ligand's terminal P_2' substituent.

Figure 8: (a) Crystallographically determined binding modes of TLN-**2** (light green) and TLN-**6** (white), and TLN-**8** (beige), heteroatoms type-coded, water molecules indicated as spheres with the same color as used for the respective crystal structure. In TLN-**2** and TLN-**6**, a contiguously connected chain of water molecules is indicated that wraps around the protruding P_2' substituent. With the larger Ile residue the network takes a larger detour. Only TLN-**2** with the small P_2' substituent allows accommodation of a water molecule capping the position of the ligand's carboxylate group (encircled in red). In the Phe complex TLN-**8** a similarly connected network is not indicated, however this complex picks up a water molecule in close neighborhood of the ligand (encircled in blue). It almost matches with the position of a water molecule found in TLN-**2**. (b) This water molecule in TLN-**8** is stabilized through interactions with the aromatic system of the P_2'-Phe substituent in **8** (encircled in blue). Furthermore the Phe substituent penetrates deeply into the S_2' pocket and pushes the backbone carbonyl group of Asn111 out of space. This adopts two alternative conformations A and B, both are most likely energetically destabilized as indicated by unfavorable geometric distortions.

a) b)



6. Dissecting the Hydrophobic Effect on Molecular Level

The bulky benzyl group in TLN-**8** takes impact on the neighboring protein molecule. It interfers with the carbonyl group of Asn111 which is pushed to a different position giving rise to two split conformations (Fig. 8). This perturbance seems to parallel a partial loss of planarity of the peptide bond between Asn111 and Asn112 which is not observed in the other complexes of the series (ω angle deviates from planarity by 10.4° and -17.3°, more details Table 2). Apparently, the carbonyl oxygen evades in two directions to create enough space for the large benzyl side chain of **8**. The ϕ angle in conformation B is - 3.7° indicating the amide nitrogen to adopt nearly planar geometry. The ϕ values in the other complexes range between 51-55°. The ψ angles of conformation A (- 158.9°) and B (- 125.0°) also show remarkable differences from the other complexes where ψ ranges between 166-168°. Despite the efforts on the protein side to generate the required space for **8**, conformer A is still in very close distance to one of the ring carbons of **8** (2.3 Å). The pronounced deviation from planarity of the peptide bond in B, even avoiding a steric clash with the ligand's phenyl moiety, is supposed to be quite unfavorable because it disturbs the electronic conjugation across the amide bond. A closer inspection of the B-factors in the terminal part of the phenyl ring indicates that this part of the ring is scattered over more than one orientation which corresponds to the evasion of the carbonyl group in two directions. Despite these slight perturbances, the phenyl ring of **8** achieves pronounced π-stacking with the carbamide group of Asn112 (3.7 Å) and extended van-der-Waals contacts to the phenyl side chain of Phe130 (3.3 Å) through mutual parallel packing.

Table 2: Torsion angles of the peptide bond Asn111-Asn112 are listed in the respective structure.

	Ω (°)	**Ψ (°)**	**Φ (°)**
TLN-**1**	179.9	51.4	- 166.3
TLN-**2**	178.7	53.1	- 167.8
TLN-**3**	179.2	54.0	- 167.8
TLN-**4**	178.6	54.3	- 167.2
TLN-**5**	177.5	54.9	- 166.8
TLN-**6**	177.3	55.3	- 167.8
TLN-**7**	176.7	55.5	- 167.4

TLN-**8** (Conf. A)	169.6	53.7	- 158.9
TLN-**8** (Conf. B)	- 162.7	- 3.7	- 125.0

6.4.4 Thermodynamic binding data

The thermodynamic binding properties of each inhibitor have been determined by ITC. Unfortunately, no absolute values could be recorded as the experiments are superimposed by the displacement of the cleavage product Val-Lys produced by autoprotolyses at high TLN concentrations. In this respect, all measured thermodynamic values are associated with a constant contribution resulting from the displacement of Val-Lys upon inhibitor binding. Since we focus on relative differences within the ligand series, replacement of the dipeptides cancels out. To obtain an independent confirmation of the binding data we recorded the enzyme kinetic inhibition data of all inhibitors **1**-**8**, despite a constant offset they show very similar increments among the compounds (Table 1)**.**

Furthermore, a buffer dependence of our recorded ITC data is given which shows that all complexes pick-up one proton per mol formed ligand-protein complex. It could be shown that Glu143 next to the catalytic zinc ion changes its protonation state upon inhibitor binding. Englert *et al.*[29] reported on a buffer dependence of the enthalpy term in a series of phosponamidates. Similarly also in our phosphonamidate series a protonation step is superimposed to the binding process. The origin, however of the protonation change remained unclear. In 1987 Holden *et al.*[30] stated that the decrease of bond angles at the phosphonamidate nitrogen and the concurrent increase in P-N bond length in the complex is an indication for a cationic state of nitrogen. Later Copié *et al.*[31] could not confirm this hypothesis by NMR measurements. In order to provide additional evidence for the assumed protonation step, we studied two ligands supposedly capable to trace the ligand functional group responsible for proton uptake. In the phosphonester **9** the amidate nitrogen is replaced by an oxygen and the phosphonamidate analog **10** which lacks the terminal carboxylate group (Figure 9). **9** records possible protonation of the amidate nitrogen (i.e. -NH_2^+-) and **10** elucidates whether the carboxylate group is involved in the protonation reaction (i.e. -COOH). ITC measurements of **9** and **10** in three different buffers revealed similar buffer dependencies in all cases (Figure 10). Thus for all ligands (**1**, **2**, **9**, **10** displayed), in molar ratio one proton is taken

up per mol formed ligand-protein complex, suggesting that the protein and not the ligand changes protonation state.

Figure 9: Depiction of the strategy to elucidate whether the ligand is involved in the protonation reaction.

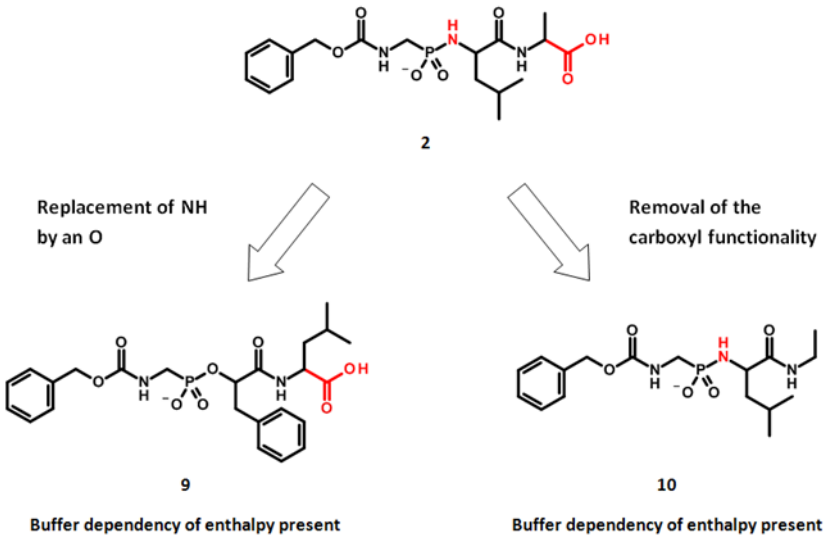
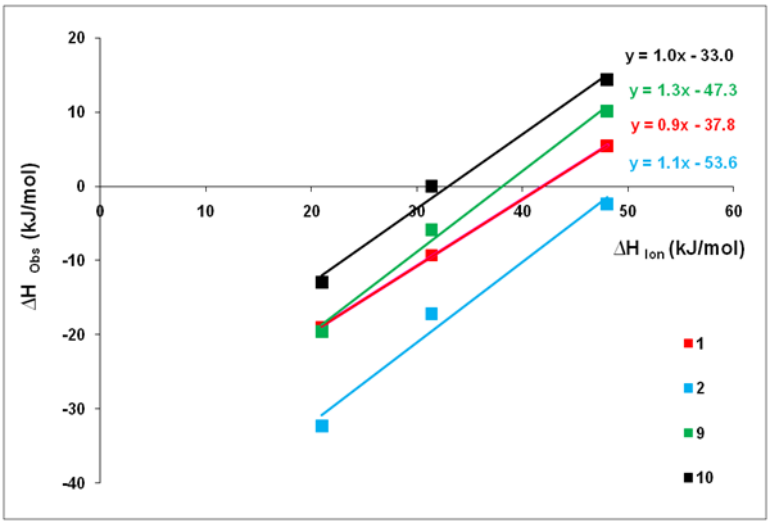


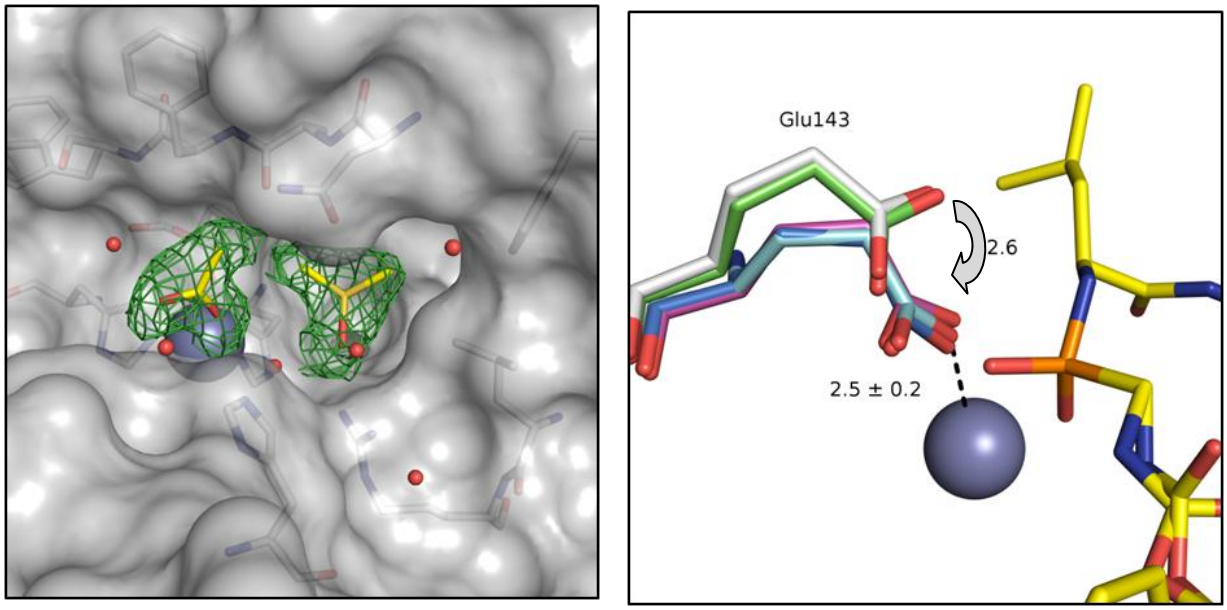
Figure 10: The observed enthalpy ΔH_{obs} for binding of **1**, **2**, **9** and **10** to thermolysin as a function of the deprotonation enthalpy ΔH_{ion} of the three applied buffers Tris, Hepes and Aces at pH 7.5. The slope of the linear regression yields the molar ratio of protons captured during the binding reaction. For all investigated ligands the same amount of protons is picked-up by the complex (0.9-1.3 mol).



To further trace where protonation might occur, a reliable structure without the Val-Lys dipeptide would be required. As all trials failed to remove the dipeptide, e.g. by dialysis, we attempted to stop TLN self-digestion by removal of zinc, soaking crystals with the chelator 1,10-phenanthroline since the zinc-free protein has been reported as inactive.[32] Refinement of

diffraction data collected with thus treated crystals indicates that 50 % of the zinc ions were still present in the active site. Fortunately, the partial zinc removal was sufficient to obtain crystals free of the protolysis product Val-Lys. Instead, a DMSO molecule accommodates the S_1' pocket picked up from the crystallization buffer and an acetate ion coordinates the zinc ion (Figure 11 a). This structure shows Glu143 in proximal geometry to the zinc ion (Figure 11 b). This suggests the glutamate to be deprotonated in the apo structure as it allows forming a charge-assisted interaction of its carboxylate with the positively charged zinc ion. In contrast, in the complexes with bound phosphonamidates Glu143 is pushed away by 2.6 Å from the zinc ion to interact with the oxygen of the phosphonamidate moiety. This geometry is also observed in the zinc free protein (PDB code 3FB0).[33] Considering the uncomplexed and complexed TLN, Glu143 switches its interaction partner from a positively charged to partially negatively charged species, which most likely requires a change in protonation state of the latter residue, otherwise formation of a hydrogen bond between its carboxylate group and the phosphonamidate oxygen appears impossible.

Figure 11: a) Crystal structure of dipeptide free TLN but in complex with DMSO and an acetate ion in the active site is shown. The F_o-F_c difference electron density is shown in green at 2σ. The solvent accessible surface of TLN is depicted in grey. Water molecules are displayed as red spheres. b) Superposition of different TLN structures: in white complexed with **3**, in green a zinc free TLN (3FB0), in cyan complexed with ruthenium-III (3FXS), in magenta complexed with copper-II (3FBO) and in blue complexed with DMSO (our apo structure).



As our apo structure exhibits only partial population of the zinc ion, we consulted the catalytically inactive TLN iso forms where the zinc ion has been exchanged by either copper-II

(PDB code 3FBO[33]) or ruthenium-III (PDB code 3FXS[33]) (Figure 11 b). Also in these structures, Glu143 is in a geometry directly coordinating to the central metal ion which also suggests deprotonated state. Interestingly, these observations are in agreement with the proposed catalytic mechanism of TLN which involves displacement of a zinc-bound water molecule towards the deprotonated Glu143 by the incoming substrate.[34,35] A negatively charged glutamate enhances in this step the nucleophilicity of the water molecule and transfers a proton to the leaving amine. As this residue is not directly involved in ligand binding of the P_2' portion, the thermodynamic data will be affected for all complexes in the same way, thus in a relative comparison also this contribution cancels out.

Figure 12 illustrates that, apart from TLN-**1**, binding becomes increasingly more entropic with growing size of the attached hydrophobic P_2' substituent, nevertheless, also significant enthalpy-entropy compensation is observed. Across the series, the absolute changes in enthalpy and entropy are much larger than those of the free energy due to mutual compensating effects.

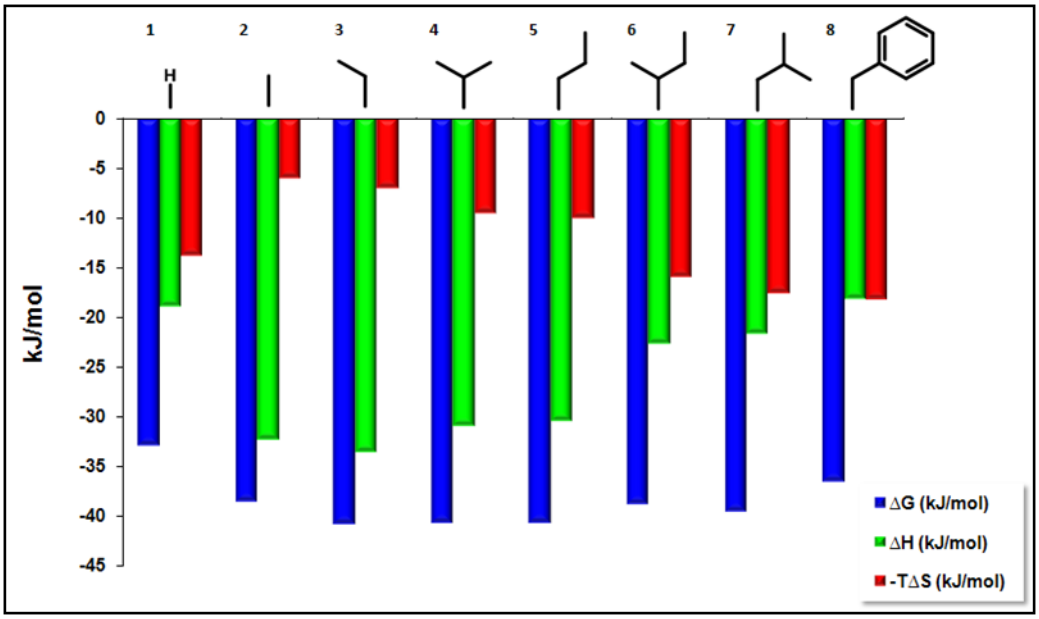
A remarkable gain in potency ($\Delta\Delta G_{1/2}$ = - 5.7 kJ/mol) is obtained for **2**, compared to **1** which is mainly due to an increase in enthalpy ($\Delta\Delta H_{1/2}$ = - 13.4 kJ/mol). This effect is partly compensated by the lower entropic signal of **2** relative to **1** (- $T\Delta\Delta S_{1/2}$ = 7.7 kJ/mol) (Table 3). Across the remaining series the Gibbs free energy seem to improve continuously from **1** to **5**, the remaining ligands **6**, **7** and **8** loose slightly in affinity (Figure 12).

Table 3: Binding data ΔG^0, ΔH^0 and - $T\Delta S^0$ (kJ/mol) of ligands determined by ITC.

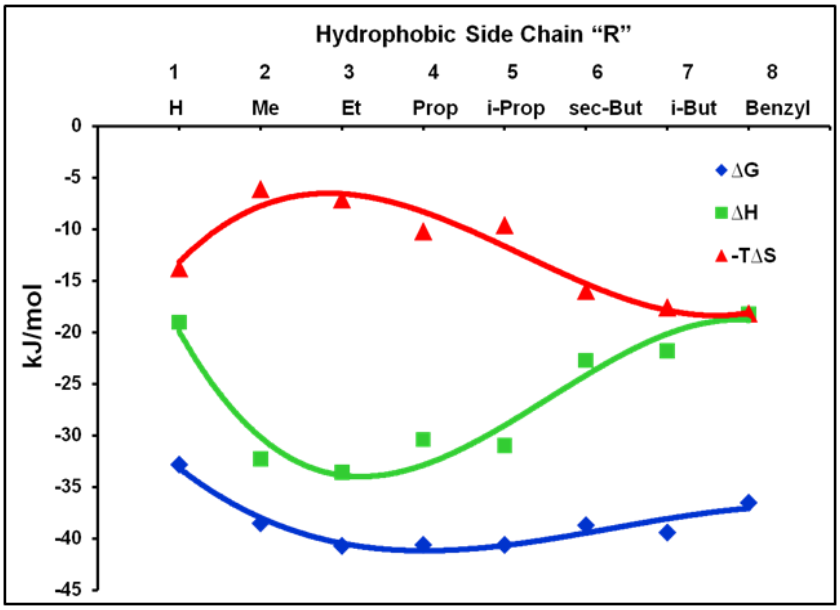
Ligand name	ΔG^0 (kJ/mol)	ΔH^0 (kJ/mol)	$-T\Delta S^0$ (kJ/mol)
1	- 32.8 ± 0.1	- 19.0 ± 0.5	- 13.8 ± 0.4
2	- 38.5 ± 0.3	- 32.4 ± 0.3	- 6.1 ± 0.1
3	- 40.7 ± 0.2	- 33.6 ± 0.7	- 7.1 ± 0.9
4	- 40.6 ± 0.1	- 31.0 ± 0.2	- 9.6 ± 0.1
5	- 40.6 ± 0.4	- 30.4 ± 0.2	- 10.2 ± 0.3
6	- 38.7 ± 0.2	- 22.7 ± 0.9	- 16.0 ± 0.7
7	- 39.4 ± 0.8	- 21.8 ± 0.1	- 17.6 ± 0.7
8	- 36.5 ± 0.1	- 18.3 ± 0.4	- 18.2 ± 0.5

Figure 12: a) The diagram shows the thermodynamic results for ΔG (blue), ΔH (green) and -TΔS as obtained by ITC. The experiments were performed in Hepes buffer and not corrected for superimposed protonation steps and replacement of the autocleavage product Val-Lys. b) To show the relative differences, mutual enthalpy-entropy compensation leading to minor free energy changes and stepwise changes in terms of related pairs the thermodynamic data is depicted in an alternative way.

a)



b)



Considering closely related ligand pairs some systematics seem to be given in the series. The pairs **2/3** ($\Delta\Delta H_{2/3}$ = - 1.2 kJ/mol, - $T\Delta\Delta S_{2/3}$ = - 1.0 kJ/mol), **4/5** ($\Delta\Delta H_{4/5}$ = - 0.6 kJ/mol, - $T\Delta\Delta S_{4/5}$ = 0.6 kJ/mol) and **6/7** ($\Delta\Delta H_{6/7}$ = 0.9 kJ/mol, - $T\Delta\Delta S_{6/7}$ = - 1.6 kJ/mol) exhibit very similar properties

to each other, whereas the steps from **3** to **4** ($\Delta\Delta H_{3/4}$ = 2.6 kJ/mol, - $T\Delta\Delta S_{3/4}$ = - 2.5 kJ/mol) and **5** to **6** ($\Delta\Delta H_{5/6}$ = 7.7 kJ/mol and - $T\Delta\Delta S_{5/6}$ = - 5.8 kJ/mol) involve larger changes. The phenylalanine derivative shows again an altered thermodynamic signature with balanced enthalpic and entropic portions.

6.5 Discussion

In the studied series of peptide-like transition state analog inhibitors the terminal hydrophobic substituent is gradually grown into the S_2' pocket of thermolysin. This pocket opens as a flat and widely opened crevice towards the bulk solvent. Apparently, it can host substituents up to the size of a benzyl moiety. This group fills the pocket quite substantially. It pushes to the limits: the carbonyl group of an adjacent backbone peptide group has to move out of space for steric reasons. It evades in two directions, so that two alternative geometries are produced, both likely to be strained and energetically disfavored.

Even though, the hydrophobic surface grows from hydrogen in the Gly to phenyl in the Phe derivative (**1** - **8**) by approximately 160 $Å^2$, overall the Gibbs free energy improves only by -3.7 kJ/mol. This is a minor contribution considering as a rough estimate about - 2 to - 3 kJ/mol per methyl group that becomes buried upon protein binding.[36] In total we would expect a much larger value as hydrophobic effect for this change. Interestingly enough, the affinity shows an optimum with an ethyl (**3**), iso-propyl (**4**) or n-propyl (**5**) substituent, even though in terms of size these groups do not yet fill the S_2' pocket optimally. More remarkable is the trend in enthalpy/entropy partitioning (Fig. 12). The absolute difference is largest with the most potent inhibitors and changes are observed not constantly across the series but systematically distinct pairs are indicated: **2**/**3**, **4**/**5** and **6**/**7** exhibit closely similar thermodynamic profiles (Figure 12). This suggests similarities in the structural properties.

Analyzing the binding modes of **1** - **8**, the parent scaffold remains virtually unchanged across the entire series. Also the interaction pattern next to the ligand's Cbz group and terminal carboxylate groups which involves at least seven conserved water molecules and mediates a complex network between ligand and protein functional groups is highly conserved (Fig. 5 upper part) and no changes are visible for the contact to the glycerol and DMSO molecules picked up from the cryo buffer. Thus, the only differences occur next to the hydrophobic P_2'

substituent of increasing size (Figure 5 lower part). Here, the network of adjacent water molecules is highly perturbed.

A huge change in the thermodynamic profile is experienced by adding a methyl group to the glycine derivative **1** to feature an alanine substituent **2**. One difference is clearly given by the fact that the Gly derivative has in solution, prior to protein binding, access to a larger conformational space than the Ala analog. In consequence, TLN-**2** will experience a smaller loss in entropy than TLN-**1**. However, additional effects are in operation. As already shown in previous studies[25] also involving the non-carboxylated analogs of **1** and **2**, this effect can be explained by the rupture of the contiguously connected water network which wraps around the terminal methyl group in **2** (Figure 6 a, b). Obviously, this methyl group provides favorable interaction sites for two additional water molecules which stabilize their binding positions through van-der-Waals contacts. Vice versa, two water molecules H-bonded in TLN-**1** to Asn111(C=O) are repelled from the complex due to steric conflicts with the attached methyl group in **2**. The rupture of the contiguously connected H-bonding network disfavors exothermic binding of **1** whereas binding of **2** is entropically less favorable due to a stronger fixation of the water network. Accordingly, the step from **1** to the more hydrophobic **2** is enthalpy-driven and we would have to classify this change as “non-classical hydrophobic effect”.

The ethyl derivative **3** shows nearly an identical thermodynamic signature as **2**. As either enthalpy and entropy increases, ΔΔG improves by - 2.2 kJ/mol; a value found in the typical range for favorably placed methyl groups. Structurally, TLN-**3** is nearly identical with TLN-**2**. The iso-propyl and n-propyl derivatives **4** and **5** share again very similar thermodynamic properties, but they are found with different enthalpy/entropy values compared to the former pair **2**/**3**. In TLN-**4** the substituent is scattered over two conformations and also a slightly disordered side-chain of Leu202 is detected. In TLN-**5** some disorder of the P_2’ substituent might be given, but it is less evident. Side-chain disorder translates into an entropic advantage of a formed complex. Next to the far end of the pocket two water molecules mediating the water network in TLN-**2** and TLN-**3** cannot be detected and the network appears incomplete in TLN-**4** and TLN-**5**. This should result, as seen for TLN-**1**, in an enthalpic loss and an entropic advantage. Both aspects, disorder and water release, will support the entropic advantage of TLN-**4** and -**5**, however with respect to the free energy, the latter advantage is nearly canceled out. Nonetheless it is remarkable, that **4** and **5** share not only similar thermodynamic signature but their detailed water networks resembles significantly and some similarity with respect a side-chain disorder

seems to be given. The hydrophobic effect related to the change from **2**/**3** to **4**/**5** must be classified as "classical entropy-driven".

All four complexes TLN-**2**, TLN-**3**, TLN-**4** and TLN-**5** host one water molecule at a position capping the ligand's carboxylate group. This position should be energetically favored for electrostatic reasons. Frequent occupancy of such carboxylate-water contacts has been highlighted by Pauline *et al.*[37] in protein structures and a clear preference for this geometry can be found in the compilation of water-carboxylate contacts as compiled in IsoStar.[38]

TLN-**6** and TLN-**7** place in ordered fashion a sec-butyl or isobutyl group into the S_2' pocket. Again both complexes experience a very similar enthalpy/entropy profile with strong enthalpic loss and entropic gain over the pair **4**/**5**. With respect to their water solvation pattern, quite similar networks are observed that wrap around the terminal hydrophobic substituent and orient along the rim of the S_2' pocket. Compared to TLN-**2** and TLN-**3**, in TLN-**6** two water molecules are shifted in space to more remote positions to create an expanded network. Remarkably, the capping water above the carboxylate group is no longer seen due to steric conflicts with the size-increased butyl substituent. Obviously, the release of this special water is one determinant for the observed enthalpy loss and entropy gain which we would relate to the "classical hydrophobic effect".

Finally, the benzyl derivative **8** looses 2.9 kJ/mol in ΔΔG with respect to **7**. This price is paid by enthalpy in agreement with the steric clash provoked with the backbone carbonyl group of Asn111. Furthermore, also in this complex the capping water is repelled from the site above the carboxylate group. The solvation structure at the far end of the S_2' pocket appears to be thinned out, however, one water site close to the ligand is newly populated that has already been accommodated in TLN-**2** and -**3**. A water molecule is most likely stabilized at this pivotal position perpendicular to the terminal aromatic benzyl moiety.

All described structural features next to the S_2' pocket are additionally modulated by further contributions arising from desolvation differences of the ligands of increasing size.[39,40] Furthermore, slight differences will also relate to the deviating conformational properties of the P_2' substituents which exhibit different numbers of rotatable bonds. These contributions are overlaid to the perturbances of the water network and as a sum they describe quantitatively the observed thermodynamic profiles. As thermolysin is a very rigid protein influences resulting from changes of the residual mobility of protein residues or induced-fit adaptations, apart of TLN-**8**, will be of minor importance. Even so our discussion only intends a qualitative

correlation, it is remarkable that complexes with side chains of comparable size and number of rotatable bonds show very similar thermodynamic signature.

6.6 Conclusion

The present series allows tracing details how the first solvation layer around a binding pocket takes impact on binding affinity of ligands. It also shows that water networks can take significant influence modulating structure-activity relationships. Growing hydrophobicity of ligand functional groups placed into hydrophobic binding pockets is usually discussed in terms of the hydrophobic effect. This effect has been related to enthalpic or entropic signature[6,7,11,12], however, most important it relates to changes in the local water structure.[6,8,9] The present series of closely related complexes shows that both, enthalpy and entropy are involved and many detailed structural phenomena determine the final signature. If a contiguously connected water network ruptures, an enthalpic loss and entropic gain are experienced.[25] This results from a loss of hydrogen bonds which allows the system to activate and distribute energy over more degrees of freedom. The displacement of ordered water molecules can result in an overall entropic signal, but the release of largely disordered waters can also reveal a predominately enthalpic signal.[6-9] However, particularly if ligand portions are accommodated in pockets opening to the bulk solvent and parts of the placed ligand are exposed to the water phase, also new binding sites for water molecules can be generated, e.g. in our study the position capping the carboxylate group or the site found on top of the benzyl ring. The capping water seems to provide a significant contribution. The three complexes TLN-**6**, -**7** and -**8** lack this water molecule due to steric conflicts and they lose in Gibbs free energy predominantly for enthalpic reasons. This is partly compensated by entropy as these complexes do not capture the water molecule, a process which is entropically unfavorable. All these phenomena contribute on the molecular level to the finally determined hydrophobic effect. In summary, there are no arguments why the hydrophobic effect should be predominantly “entropic” or “enthalpic”; small structural changes in the binding features of water molecules on the molecular level determine whether hydrophobic binding to hydrophobic pockets results in a more enthalpy or entropy-driven signature.

Admittedly, the present study goes towards the limits of experimental accuracy, accomplishable in contemporary protein-ligand structural work. Assignment of water molecules

to the difference electron density is crucial and depends on resolution and local ordering phenomena. However, in the present series eight crystal structures indicate similar and internally consistent solvation pattern which gives confidence about the necessary reliability to underscore its relevance. Our structural data originate from crystal environment taken from flash-cooled crystals which should capture a frozen-in image of the properties at ambient.[41] The thermodynamic data were recorded in solution. Nevertheless, a very consistent picture is received. Surprising pairwise systematics in the thermodynamic data are experienced for complexes of related ligands and they are convincingly well reflected by the structural properties. The present study unravels small but important details. Computational methods simulate molecular properties at atomic level and usually they are determined by the summation of many small details.[42] However, details as the ones observed in the present study are at present hardly regarded as relevant, simply because we are not really aware about their importance for protein-ligand binding, structure-activity relationships or rational drug design in general. Taking them into account will make our simulations not necessarily simpler, however, hopefully the present examples give the necessary evidence what is required to improve simulation models.

6.7 Materials & Methods

6.7.1 Crystallization and soaking

Native thermolysin (purchased from Calbiochem) was crystallized as described by Holmes and Matthews[43] applying slight modifications. Thermolysin was dissolved in 100 % DMSO at 8 mM. Subsequently, the DMSO solution was diluted with the buffer 1:1 (100 mM Tris/HCl, 5 M CsCl, pH 7.5). The suspension was then vortexed and centrifuged for two minutes. The so achieved supernatant was used in the crystallization trial. This final solution contained 50 % DMSO and 2.5 M CsCl. The final protein concentration in the crystallization drop was 4.0 mM. Crystals were grown at 18 °C by the sitting drop vapor diffusion method using water as reservoir solution. Protein-ligand complex crystals were obtained by soaking crystals in a solution of 100 mM Tris/HCl, 2 mM CaCl and 10 % DMSO. For soaking, ligand concentration varied between 1 and 5 mM. When low solubility was given, saturated solutions were used. Crystals were soaked between one and two days before freezing.

6.7.2 Data collection and processing

Crystals were prepared for data collection at 110 K using a cryoprotectant solution of 20 % glycerol in a buffer containing 10 mM Tris/HCl, 10 mM $Ca(CH_3COO)_2$, 5 % DMSO at pH 7.3. The data sets for **3-6**, **8** and the apo structure were collected with synchrotron radiation at BESSY beamline 14.2 (Berlin, Germany) on a Rayonix MX 225 CCD detector. Data processing and scaling were performed using the HKL2000 package[44].

6.7.3 Structure determination of the complex structures and refinement

The complex structures were determined by molecular replacement. The coordinates of TLN (PDB code 8TLN)[45] were used for initial rigid body refinement of the protein molecules followed by repeated cycles of maximum likelihood energy minimization, simulated annealing and B-factor refinement using the CNS program package[46]. All structures were then refined with PHENIX.[47] The temperature factors for structures **3**, **4**, **6** and **8** were anisotropically refined whereas for structures **5** TLS refinement was applied. The TLN-**7** complex (3FWD) was re-refined with PHENIX to ensure a comparable refinement protocol and to reduce potential bias caused by the refinement program. The resulting structure remained virtually unchanged, however, now a total of 331 water molecules were refined. Also TLS refinement was applied. For the newly refined structure, the R-value could be improved to 15.2% with R_{free} being 17.0%. The definition of the TLS groups was done with the TLSMD server[48,49]. A randomly chosen 5 % of all data were used for the calculation of R_{free} and were not used in the refinement. Amino acid side chains were fitted into σ-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps using Coot[50]. After the first refinement cycle, water molecules and subsequently ions and ligands were located in the electron density and added to the model. Restraints were applied to bond lengths and angles, planarity of aromatic rings and van-der-Waals contacts. Multiple side chain conformations were built in case an appropriate electron density was observed and maintained during the refinement, and if the minor populated side chain showed at least 20 % occupancy. The final models were validated using PHENIX own validation options or MolProbity[51]. The Ramachandran plots were calculated with PROCHECK[52]. Thr26 of TLN lies in a forbidden region of the Ramachandran plot, but this is frequently observed for residues located at the tip of a γ-like turn.[53] Data collection, unit cell parameters and refinement statistics are given in Table 4. Analysis of temperature factors was done with Moleman[54] and fconv[55].

6.7.4 Isothermal titration calorimetry (ITC)

ITC experiments were performed using an ITC200™ system from MicroCal (now part of GE Healthcare, Northhampton, MA, USA). Thermolysin from Calbiochem was freshly prepared for each experiment in the buffer applied in the measurements. The comparison of the thermodynamic data of the dialysed and undialysed protein solution revealed no significant difference. Therefore, the undialysed protein solution was used in this study. The inhibitor solution was freshly prepared in the experimental buffer to the required concentration. The ITC experiments were done at 25 °C in a buffer composed of 20 mM Hepes, 500 mM NaSCN, 2 mM $CaCl_2$ at pH 7.5. Inhibitor solutions (1-1.5 mM) were titrated into the sample cell containing the TLN solution (45.0 µM) after a stable baseline was achieved. ITC experiments comprised an initial ligand injection of 0.3 µl followed by 15 injections of 1-2 µl with a 300 s interval between each injection. The ITC cell volume was 200 µl. Raw data were collected and the area under each peak was integrated, followed by correction for heats of dilution and mixing by subtracting the final baseline consisting of small peaks of the same size to zero. The initial data point was deleted from the integrated data because this injection usually reflects an erroneous amount of heat due to the possible exchange of liquids between syringe and cell when inserting the syringe into the calorimetric cell and the backlash error in the motorized screw mechanism in the injector.[56] Data were analysed using ORIGIN Software (Microcal Inc.), by fitting a single-site-binding isotherm[57] that yields ΔH^0 (enthalpy of binding) and K_D (dissociation constant). Measurements were performed at least in duplicate. The absolute values for ΔG, ΔH and –TΔS together with the estimated errors are listed in Table 3, the relative differences ΔΔG, ΔΔH and –TΔΔs are given in Figure 12.

6.7.5 Data deposition

Coordinates and structure factors have been deposited in the Protein Data Bank with the following accession codes: TLN-**3** complex 3T87; TLN-**4** complex 3T8H; TLN-**5** complex 3T8C; TLN-**6** complex 3T8D; TLN-**8** complex 4D9W, TLN-**1** and TLN-**2** were previously reported as 3T8G and 3T74, respectively. TLN-**7** (3FWD) has been newly refined and deposited replacing the former refinement results.

Table 4: Data collection and refinement statistics for the six TLN complex structures.

Structure (PDB entry)	TLN-**3** complex **(3T87)**	TLN-**4** complex **(3T8H)**	TLN-**5** complex **(3T8C)**	TLN-**6** complex **(3T8D)**	TLN-**8** complex **(4D9W)**	Apo TLN **(4D91)**
A. Data collection and processing						
No. Crystals used	1	1	1	1	1	1
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841	0.91841	1.9
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters						
a, b (Å)	92.7	92.8	92.6	92.8	93.1	93.4
c (Å)	131.1	130.1	130.4	130.1	130.9	129.9
Matthews coefficient ($Å^3$/Da)	2.4	2.4	2.4	2.4	2.4	2.4
Solvent content (%)	48	48	48	48	48	48
B. Diffraction data[a]						
Resolution range (Å)	50 – 1.28	50 – 1.45	50 – 1.66	50 – 1.41	50 – 1.38	50 – 1.90
	(1.30 – 1.28)	(1.48 – 1.45)	(1.69 – 1.66)	(1.43 – 1.41)	(1.40 – 1.38)	(1.93 – 1.90)
Unique reflections	83,248	58,688	39,580	64,097	68,072	49,334
R(I)sym (%)	6.5 (37.8)	5.3 (36.3)	5.4 (22.0)	6.2 (49.4)	6.7 (50.0)	5.6 (33.3)
Completeness (%)	97.0 (99.1)	99.5 (99.1)	99.7 (99.7)	99.8 (100.0)	98.6 (99.9)	98.3 (95.9)
Redundancy	10.7 (9.0)	5.8 (5.7)	5.5 (5.4)	4.7 (4.6)	5.8 (5.9)	30.6 (27.5)
I/σ(I)	30.6 (5.2)	30.0 (4.6)	28.8 (7.6)	23.3 (3.0)	23.4 (3.4)	82.5 (11.3)
C. Refinement						
Resolution range (Å)	34.2 – 1.28	40.2 – 1.45	43.6 – 1.45	31.7 – 1.41	22.3 – 1.38	34.3 – 1.90
Reflections used in refinement	80,710 / 4,106	56,621 / 2,860	38,705 / 1,949	60,892 / 3,082	64,916 / 3,315	48,848 / 2,393
(work/free)						
Final R values for all reflections	10.5 /12.7	12.3 /15.3	13.9 /16.6	12.6 / 15.3	10.9 / 14.0	15.0 / 18.4
(work/free) (%)						
Protein residues	316	316	316	316	316	316

Calcium /zinc ions	4 / 1	4 / 1	4 / 1	4 / 1	4 / 1	4 / 1
Inhibitor atoms	30[b] / 34[c]	31[b] / 41[c]	31	32[b] / 36[c]	35[b] / 39[c]	-
Water molecules	481	446	432	450	474	368
RMSD from ideality						
Bond lengths (Å)	0.014	0.011	0.011	0.012	0.014	0.017
Bond angles (°)	1.353	1.047	1.030	1.050	1.275	0.979
Ramachandran plot						
Residues in most favoured regions (%)	87.8	88.1	87.8	88.5	87.8	88.5
Residues in additionally allowed	11.1	11.1	11.5	10.7	11.5	10.4
regions (%)						
Residues in generously allowed	0.7	0.4	0.4	0.4	0.4	0.7
regions (%)						
Residues in disallowed regions (%)	0.4	0.4	0.4	0.4	0.4	0.4
Mean B-factor (Å^2)						
Protein	9.9	11.8	11.0	11.4	11.2	17.1
Binding site[d]	6.9	9.1	8.2	8.9	9.6	-
Inhibitor	9.8[e]	11.7[e]	11.2	11.4[e]	14.7[e]	-
Water molecules	26.5	30.6	28.0	29.6	29.6	32.5

[a] Numbers in parenthesis are for the highest resolution shell.

[b] The inhibitor atoms from the additional disordered part were not considered

[c] All inhibitor atoms which were included in the refinement model

[d] Definition of the binding site: all amino acids which are within 4 Å from inhibitor.

[e] Average B value for the ordered and fully occupied inhibitor portion, the disordered portions were not considered

6.7.6 Phase determination and model building of apo structure of TLN

Structure determination utilized the graphical interface of hkl2map (version 0.2)[58], which connects several programs from the SHELX-suite[59] to guide the user from analysis of scaled diffraction data (SHELXC), via substructure determination (SHELXD), to phasing (SHELXE). In addition to conventional SHELXE, we were able to use the new beta-test version (kindly provided by Prof. George Sheldrick), which includes auto-tracing of the protein backbone via a polyalanine model. Six sites were found which correspond to four calcium, one zinc and one sulfur atom from a DMSO molecule. 20 cycles of density modification were done resulting in a pseudo-free correlation coefficient of 54.5 % (see Table 5 and 6 for further statistics). The initial phases achieved with hkl2map were submitted to the automatic model-building program ARP/wARP (version 7.0.1) included the program suite of CCP4. The TLN structure was built on the basis of the poly-Ala trace obtained from SHELXE_beta. After 200 building cycles, 311 residues were found in four chains with a connectivity index of 0.965. After loop building the structure contained 314 residues out of 316.

Table 5: Phasing statistics.

	TLN apo structure
Resolution range in Å	50 – 1.9
D''/sig(d'') (SHELXC) max/min	3.72 – 0.96
Resolution cut off (SHELXD) in Å	2.5
CCmax (SHELXD) in %	52.6
Number of sites	6
Mean FOM (SHELXE_beta)	0.561
CC in % (SHELXE_beta)	59.3
CC in % (SHELXE_beta) for best cycle 7	44.2
Connecitivity index (SHELXE_beta) for best cycle 7	0.764
Contrast (SHELX_beta) for best cycle 7	0.438
Residues (SHELX_beta) for best cycle 7	310 in 11 chains

Table 6: Data collection and refinement statistics for apo structure.

Structure (PDB entry)	Apo TLN **(4D91)**
A. Data collection and processing	
No. Crystals used	1
Wavelength (Å)	1.9
Space group	$P6_122$
Unit cell parameters	
a, b (Å)	93.4
c (Å)	129.9
Matthews coefficient ($Å^3$/Da)	2.4
Solvent content (%)	48
B. Diffraction data	
Resolution range (Å)	50 – 1.90
	(1.93 – 1.90)
Unique reflections	49,334
R(I)sym (%)	5.6 (33.3)
Completeness (%)	98.3 (95.9)
Redundancy	30.6 (27.5)
I/σ(I)	82.5 (11.3)
C. Refinement	
Resolution range (Å)	34.3 – 1.90
Reflections used in refinement	48,848 / 2,393
(work/free)	
Final R values for all reflections	15.0 / 18.4
(work/free) (%)	
Protein residues	316
Calcium /zinc ions	4 / 1
Inhibitor atoms	-
Water molecules	368
RMSD from ideality	
Bond lengths (Å)	0.017
Bond angles (°)	0.979
Ramachandran plot	
Residues in most favoured regions (%)	88.5
Residues in additionally allowed	10.4
regions (%)	
Residues in generously allowed	0.7
regions (%)	
Residues in disallowed regions (%)	0.4
Mean B-factor ($Å^2$)	
Protein	17.1
Binding site	-
Inhibitor	-
Water molecules	32.5

6.8 References

1 Nic M., Jirat J. & Kosata B. (2006) *"hydrophobic interaction"*. *IUPAC Compendium of Chemical Terminology* (Online ed.). doi:10.1351/goldbook.H02907. ISBN 0-9678550-9-8.

2 Chandler D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640-647.

3 Kauzmann W. (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14,** 1-63.

4 Charton M. & Charton B.I. (1982) The structural dependence of amino acid hydrophobicity parameters. *J Theor Biol* **99,** 629-644.

5 Whitesides G.M. & Krishnamurthy V.M. (2005) Designing ligands to bind proteins. *Q Rev Biophys* **38,** 385-395.

6 Englert L. *et al.* (2010) Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: binding of phosphonamidate to the S_1'-pocket of thermolysin. *Biochim Biophys Acta* **1800,** 1192-1202.

7 Bingham R.J. *et al.* (2004) Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J Am Chem Soc* **126,** 1675-1681.

8 Homans S.W. (2007) Water, water everywhere--except where it matters? *Drug Discov Today* **12,** 534-539.

9 Snyder P.W. *et al.* (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci U S A* **108,** 17889-17894.

10 Biela A. *et al.* (Ligand binding gradually disrupts water network in thrombin: Enthalpic and entropic changes reveal classical hydrophobic effect. *J. Med. Chem.*:submitted.

11 Smithrud D.B., Wyman T.B. & Diederich F. (1991) Enthalpically driven cyclophane-arene inclusion complexation: solvent-dependent calorimetric studies. *J. Am. Chem. Soc.* **113,** 5420-5426.

12 Breslauer K.J. *et al.* (1987) Enthalpy-entropy compensations in drug-DNA binding studies. *Proc Natl Acad Sci U S A* **84,** 8922-8926.

13 Bukrinsky J.T., Bjerrum M.J. & Kadziola A. (1998) Native carboxypeptidase A in a new crystal environment reveals a different conformation of the important tyrosine 248. *Biochemistry* **37,** 16555-16564.

14 Selkti M*. et al.* (2003) Interactions of a new alpha-aminophosphinic derivative inside the active site of TLN (thermolysin): a model for zinc-metalloendopeptidase inhibition. *Acta Crystallogr D Biol Crystallogr* **59,** 1200-1205.

15 Gaucher J.F. *et al.* (1999) Crystal structures of alpha-mercaptoacyldipeptides in the thermolysin active site: structural parameters for a Zn monodentation or bidentation in metalloendopeptidases. *Biochemistry* **38,** 12569-12576.

16 Takeuchi T., Bottcher A., Quezada C.M., Meade T.J., & Gray H.B. (1999) Inhibition of thermolysin and human alpha-thrombin by cobalt(III) Schiff base complexes. *Bioorg Med Chem* **7,** 815-819.

17 Lee K.J. & Kim D.H. (1998) Inhibition of thermolysin with nitrone-bearing substrate analogs: a new type of thermolysin inhibitors. *Bioorg Med Chem Lett* **8,** 323-326.

18 English A.C., Groom C.R., & Hubbard R.E. (2001) Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Engineering.* **14,** 47-59.

19 Behnen J*. et al.* Experimental and computational active site mapping as a starting point to fragment-based lead discovery. *ChemMedChem* **7,** 248-261.

20 Englert L*. et al.* (2010) Fragment-based lead discovery: screening and optimizing fragments for thermolysin inhibition. *ChemMedChem* **5,** 930-940.

21 Matthews B.W., Weaver L.H. & Kester W.R. (1974) The conformation of thermolysin. *J Biol Chem* **249**, 8030-8044.

22 Holmes M.A. & Matthews B.W. (1982) Structure of thermolysin refined at 1.6 A resolution. *J Mol Biol* **160**, 623-639.

23 Holden H.M., Tronrud D.E., Monzingo A.F., Weaver L.H. & Matthews B.W. (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.

24 Morihara K. (1967) The specificities of various neutral and alkaline proteinases from microorganisms. *Biochem Biophys Res Commun* **26**, 656-661.

25 Biela A., Betz M., Heine A., & Klebe G. (submitted) Water Makes the Difference: Rearrangement of Water Solvation Layer Triggers Non-additivity of Functional Group Contributions in Protein-Ligand Binding.

26 Nasief N.N., Tan H., Kong J., & Hangauer D. Side Chain Cooperativity Enhances Binding Affinity of Thermolysin Phosphonamidate Inhibitors: Polarization of the Water Matrix Mediates Positive Cooperativity in the S_2' Pocket.:submitted.

27 Tronrud D.E., Monzingo A.F. & Matthews B.W. (1986) Crystallographic structural analysis of phosphoramidates as inhibitors and transition-state analogs of thermolysin. *Eur J Biochem* **157,** 261-268.

28 Bartlett P.A. & Marlowe C.K. (1983) Phosphonamidates as transition-state analogue inhibitors of thermolysin. *Biochemistry* **22,** 4618-4624.

29 Englert L*. et al.* (2010) Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: binding of phosphonamidate to the S_1'-pocket of thermolysin. *Biochim Biophys Acta* **1800,** 1192-1202.

30 Holden H.M., Tronrud D.E., Monzingo A.F., Weaver L.H., & Matthews B.W. (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.

31 Copie V*. et al.* (1990) Inhibition of thermolysin by phosphonamidate transition-state analogues: measurement of 31P-15N bond lengths and chemical shifts in two enzyme-inhibitor complexes by solid-state nuclear magnetic resonance. *Biochemistry* **29,** 9176-9184.

32 Holmquist B. & Vallee B.L. (1974) Metal substitutions and inhibition of thermolysin: spectra of the cobalt enzyme. *J Biol Chem* **249,** 4601-4607.

33 Englert L. (2009) Probing Metalloenzymes: Virtual Screening for Inhibitors of Human Tyrosinase and Studies on the Driving Forces of Zinc-Protease Inhibition. PhD thesis

34 Matthews B.W. (1988) Structural basis of the action of thermolysin and related zinc peptidases. *Acc. Chem. Res.* **21,** 333-340.

35 Lipscomb W.N. & Strater N. (1996) Recent Advances in Zinc Enzymology. *Chem Rev* **96,** 2375-2434.

36 Leung C.S., Leung S.S., Tirado-Rives J. & Jorgensen W.L. (2012) Methyl effects on protein-ligand binding. *J Med Chem* **55,** 4489-4500.

37 Paulini R., Muller K. & Diederich F. (2005) Orthogonal multipolar interactions in structural chemistry and biology. *Angew Chem Int Ed Engl* **44,** 1788-1805.

38 Bruno I.J. *et al.* (1997) IsoStar: a library of information about nonbonded interactions. *J Comput Aided Mol Des* **11,** 525-537.

39 Abraham M.H., Whiting G.S., Fuchs R. & Chambers E.J. (1990) Thermodynamics of Solute Transfer from Water to Hexadecane. *J. CHEM. SOC. PERKIN TRANS.* **2,** 291-300.

40 Jorgensen W.L., Ulmschneider J.P. & Tirado-Rives J. (2004) Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J. Phys. Chem. B* **108,** 16264-16270.

41 Nakasako M. (2004) Water-protein interactions from high-resolution protein crystallography. *Philos Trans R Soc Lond B Biol Sci* **359,** 1191-1204.

42 Huggins D.J., Sherman W., & Tidor B. (2012) Rational approaches to improving selectivity in drug design. *J Med Chem* **55,** 1424-1444.

43 Holmes M.A. & Matthews B.W. (1982) Structure of thermolysin refined at 1.6 A resolution. *J Mol Biol* **160,** 623-639.

44 Otwinowski Z. & Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276,** 307-326.

45 Holland D.R. *et al.* (1992) Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* **31,** 11310-11316.

46 Brunger A.T. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54,** 905-921.

47 Adams P.D. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66,** 213-221.

48 Painter J. & Merritt E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62,** 439-450.

49 Painter J. & Merritt E.A. (2006) TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* **39**, 109-111.

50 Emsley P. & Cowtan K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60,** 2126-2132.

51 Chen V.B. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66,** 12-21.

52 Laskowski R.A., MacArthur M.W., Moss D.S. & Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26,** 283-291.

53 Matthews B.W. (1972) The γ Turn. Evidence for a New Folded Conformation in Proteins. *Macromolecules* **5,** 818-819.

54 Kleywegt G.J., Zou J.Y., Kjeldgaard M. & Jones T.A. (2001) *Around O.* (Kluwer Academic Publishers, Dordrecht) pp 353-356.

55 Neudert G. & Klebe G. (2011) fconv: Format conversion, manipulation and feature computation of molecular data. *Bioinformatics* **27,** 1021-1022.

56 Mizoue L.S. & Tellinghuisen J. (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal Biochem* **326,** 125-127.

57 Wiseman T., Williston S., Brandts J.F. & Lin L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179,** 131-137.

58 Pape T., Schneider T.R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Cryst* **37,** 843-844.

59 Sheldrick G. (2008) A short history of SHELX. *Acta Crystallographica Section A* **64,** 112-122.

7 Untersuchung der Kooperativität zwischen einer Wasserstoffbrücke und hydrophoben Kontakten in der S3/4 Tasche von Thrombin

7.1 Einleitende Bemerkungen/Introductory Remarks

Diese Arbeit wurde in Kooperation mit der Arbeitsgruppe von Prof. Dr. Hangauer (University of Buffalo, USA) durchgeführt. Das Manuskript für die Veröffentlichung wird wie mit unserem Kooperationspartner vereinbart von Nader Nasief als Erstautor in englischer Sprache angefertigt. Der vorliegende Text wird daher in keiner Zeitschrift sondern nur in dieser Dissertation erscheinen. Ein herzlicher Dank geht an Nader Nasief, der an der Synthese der Verbindungen und der Interpretation der Daten beteiligt war, und an Prof. Dr. Engels, der die quantenmechanischen Rechnungen durchgeführt hat.

7.2 Einleitung/Introduction

Im Bereich des computergestützten Designs (Docking) ist die größte Herausforderung die richtige Vorhersage der Bindungsaffinität für einen gegebenen Liganden. Nur wenn dies gelingt können aufwendige und kostspielige experimentelle Suchverfahren (HTS, FBDD) durch die materialschonenden Computermethoden in Teilen ersetzt werden. Für den Erfolg der richtigen Beurteilung der Bindungsaffinität im Docking spielen Scoring-Funktionen (z.B. DrugScore[(1)], GoldScore[(2)]) die entscheidende Rolle. Wenn die Scoring-Funktion alle Facetten des Bindungsprozesses berücksichtigen würde, wäre eine optimale Vorhersage möglich. Leider existieren innerhalb des Bindungsvorgangs zwischen Protein und Ligand etliche Phänomene (Wassereffekte[(3)], Enthalpie-Entropie-Kompensation[(4)], Kooperativität[(5)]) für die es zur Zeit keine endgültige Erklärung gibt. Eine Einbeziehung dieser Effekte in Scoring-Funktionen ist daher nicht zu rechtfertigen, da die Datenlage an experimentell im Detail untersuchten Protein-Ligand Komplexen einfach zu gering ist, um diese Prozesse ausreichend zu beschreiben. Es ist daher nicht verwunderlich, dass die richtige Abschätzung

der Bindungsaffinität im Docking solange noch Utopie bleibt bis der Bindungsvorgang in allen seinen kleinen Details verstanden ist.

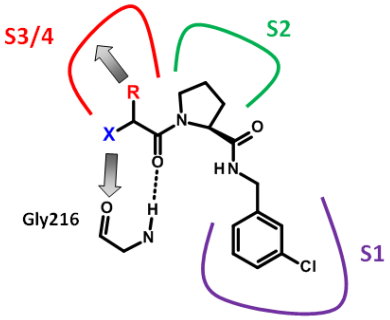
Gerade kooperative Effekte sind zwar lange bekannt, aber weit davon entfernt theoretisch quantifiziert zu werden. Im Grunde besagt die Kooperativität, dass zwei miteinander verbundene Liganden eine andere Freie Energie (ΔG_{L1-L2}) liefern als die Summe der zwei einzelnen Bestandteile ($\Delta G_{L1} + \Delta G_{L2}$). Vor der Entdeckung der Kooperativität ist man von einem rein additiven Konzept ausgegangen in dem alle positiven und negativen Bindungsbeiträge im Komplex aufsummiert wurden ohne zu berücksichtigen, dass die Größe der Beiträge teils stark von einander abhängen. Aus theoretischen Gründen ist dies nicht weiter verwunderlich, es wurde bereits mehrfach darauf hingewiesen, dass vor allem bei entropischen Bindungsbeiträgen eine Additivität nicht erfüllt sein kann. Baum *et al.*[6] konnten z.B. zeigen, dass die Stärke der hydrophoben Wechselwirkung zur S3/4 Tasche von Thrombin stark davon abhängt, ob eine Wasserstoffbrücke zu Gly216 vorhanden war oder nicht. Bei Vorhandensein der Wasserstoffbrücke zu Gly216 war die hydrophobe Optimierung zur S3/4 Tasche signifikant besser (- 0,127 $kJmol^{-1}Å^{-2}$ Affinitätsbeitrag bezogen auf Oberflächen-beiträge) als bei Nichtvorhandensein dieser Wasserstoffbrücke (- 0,077 $kJmol^{-1}Å^{-2}$). Das heißt mit anderen Worten, dass die Stärke der van-der-Waals Wechselwirkung von der Ausbildung einer anderen Wechselwirkung im Komplex abhängt. Würde man von einem additiven Konzept ausgehen, müsste der Gewinn an Bindungsaffinität pro $Å^2$ vergrabener hydrophober Oberfläche vergleichbar sein. Die zusätzliche Wasserstoffbrücke in Thrombin sorgt aber für einen um 65 % effizienteren Weg einen Inhibitor hydrophob zu optimieren.

Wenn man sich diesen Unterschied vor Augen führt, wirkt es nahezu naiv zu versuchen, den Beitrag zu bestimmen, den eine Wasserstoffbrücke[7] oder ein van-der-Waals Kontakt[8] zur Bindung beiträgt. Dieses Unterfangen muss im Detail als unmöglich gelten, da alle Kontakte miteinander zusammenhängen und jede Interaktion praktisch nicht isoliert betrachtet werden kann.

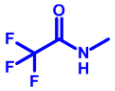
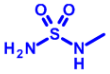
Aufbauend auf der Hypothese, dass die hydrophobe Optimierung zur S3/4 Tasche von Thrombin vom Vorhandensein einer Wasserstoffbrücke abhängt, haben wir fünf neue Serien entworfen und synthetisiert, in denen versucht wird, die Stärke dieser Wasserstoffbrücke zu modifizieren (Abbildung 1). Unmittelbar neben der NH-Funktionalität des Liganden sind verschiedene Reste eingefügt, die alle in unterschiedlichem Ausmaß einen Einfluss auf die Polarität der NH-Bindung ausüben. Der hydrophobe Rest in P3 wird in jeder Serie

systematisch variiert, um schrittweise eine immer größere hydrophobe Kontaktoberfläche zur S3/4 Tasche zu generieren. Kinetische Untersuchung der Ligandserien soll nun zeigen, ob die hydrophobe Interaktion nicht nur vom Vorhandensein einer Wasserstoffbrücke abhängt, sondern ob das Ausmaß der hydrophoben Wechselwirkung auch mit der Stärke der Wasserstoffbrücke korreliert.

Abbildung 1: Schematische Darstellung der Bindetasche von Thrombin im Komplex mit dem Grundgerüst der untersuchten Serien. Serie 1 und 2 wurden in der Doktorarbeit von Bernhard Baum[9] untersucht.



X =

1	2	3	4	5	6	7
						

7.3 Ergebnisse & Diskussion/Results & Discussion

7.3.1 Inhibitionskonstanten und die Korrelation mit den hydrophoben Kontaktoberflächen

Tabelle 1 zeigt die untersuchten Serien mit den jeweiligen hydrophoben Modifizierungen und den Inhibitionskonstanten. In jeder Serie konnte die Bindungsaffinität schrittweise durch immer größer werdende hydrophobe Reste verbessert werden. Der Gewinn an Bindungsaffinität pro $Å^2$ vergrabener hydrophober Oberfläche (Berechnung siehe Material & Methoden, Abschnitt 6.6.2) ist jedoch eindeutig abhängig von der Substitution an der terminalen Aminogruppe. Abbildung 2 zeigt die Auftragung der Freien Energie (ΔG) gegen die jeweilige berechnete hydrophobe Kontaktoberfläche ($Å^2$). Die Korrelationskoeffizienten (R^2) variieren zwischen 0,8179 - 0,950, was auf einen zuverlässigen linearen Zusammenhang hindeutet. Die Steigungen sind der Größe nach in Tabelle 2 zusammengefasst. In der Serie 1 ist, wie erwartet, die schwächste Kooperativität erkennbar (- 0,077 $kJmol^{-1}Å^{-2}$). Die Serie 3 mit der acetylierten NH-Gruppe zeigt eine kleine Verbesserung um 7,8 % (- 0,083 $kJmol^{-1}Å^{-2}$)

während die Serie 4 mit der Formyl-Gruppe eine signifikante Verbesserung um 29,9 % (- 0,100 $kJmol^{-1}Å^{-2}$) im Vergleich zu Serie 1 erzielt. Die Trifluoracetylgruppe in Serie 5 hat sogar einen um 48,1 % verbesserten Gewinn pro vergrabenener hydrophober Oberfläche (- 0,114 $kJmol^{-1}Å^{-2}$). Die Methylsulfonierung der Aminogruppe führt zu einer weiteren Steigerung des kooperativen Effekts. In der methylsulfonierten Serie 6 ($-NHSO_2CH_3$) konnte eine Steigerung um 61,0 % auf - 0,124 $kJmol^{-1}Å^{-2}$ und in Serie 7 ($-NHSO_2NH_2$) um 76,6 % auf - 0,136 $kJmol^{-1}Å^{-2}$ erzielt werden. Damit war die Serie 7 diejenige mit dem größten kooperativen Effekt. Die Serie 2 mit der freien Aminogruppe liegt dagegen zwischen den sulfonierten Liganden mit einer Verbesserung von 64,9 %. Es ist verwunderlich, dass die sulfonierten Serien eine ähnliche Kooperativität zeigen wie die Serie mit der freien Aminogruppe, die höchstwahrscheinlich protoniert vorliegt. Somit ergibt sich eine starke ladungsunterstützte Wasserstoffbrücke zu Gly216. Obwohl deutliche Unterschiede in den kooperativen Effekten erkennbar sind, steht die Frage noch aus, ob die Substitutionen an der freien Aminogruppe auch wirklich die Stärke der Wasserstoffbrücke modifizieren. Um dies zu überprüfen, haben wir pK_a Werte aus der Literatur für die jeweiligen Substitutionen herausgesucht (Abbildung 3). Die gefundenen pK_a Werte korrelieren dabei eindeutig mit den Werten für die Stärke der hydrophoben Wechselwirkung aus Tabelle 2.

7. Kooperativität

Tabelle 1: Chemische Strukturen der untersuchten Liganden mit den Inhibitionskonstanten in µM in schwarz, UBTHR-ID in grün, Freie Energien ΔG (kJ/mol) aus K_i in lila und die hydrophoben Kontaktoberflächen (Å^2) in orange nach einer Modellierung mit dem Programm Sybyl.

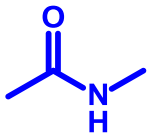
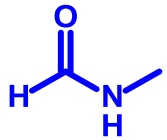
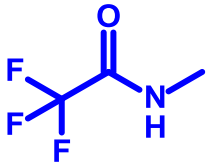
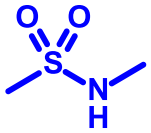
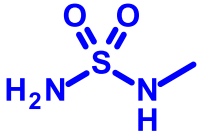
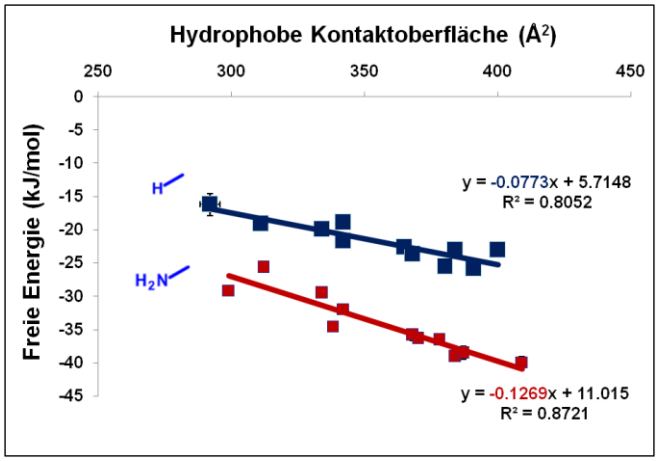
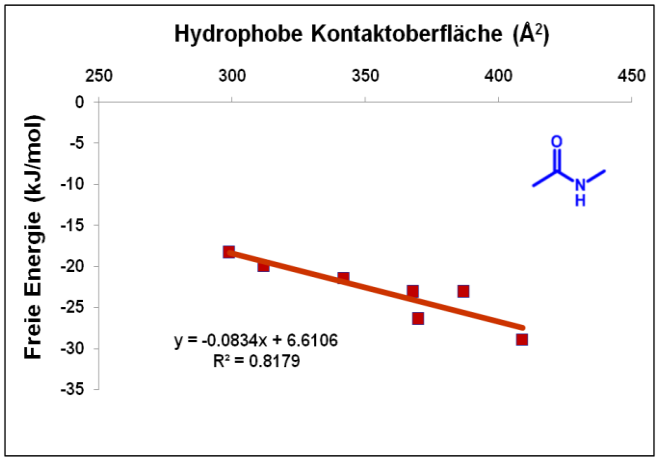
R → / X ↓	H								
3 	634,7 ± 138,9 87 - 18,3 / 299	330,9 88 - 19,9 / 312		180,6 ± 53,4 195 - 21,4 / 342	88,5 ± 36,0 89 - 23,1 / 368		24,2 ± 5,3 90 - 26,4 / 370	88,8 ± 7,2 156 - 23,1 / 387	8,7 ± 1.0 91 - 28,9 / 409
4 		308,7 ± 59,9 200 - 20,1 / 313	248,9 ± 46.0 201 - 20,6 / 337	74,7 ± 18,5 202 - 23,6 / 343	25,1 ± 6,2 203 - 26,3 / 366			32,6 ± 10,1 169 - 25,6 / 388	5,7 ± 2,7 204 - 30,1 / 409
5 		636,8 ± 86,4 144 - 18,2 / 312	349,3 ± 36,7 198 - 19,7 / 334	105,1 ± 20,4 199 - 22,7 / 342	58,6 ± 32,0 145 - 24,2 / 369	12,1 ± 1,4 147 - 28,1 / 378		24,2 ± 14,0 155 - 26,4 / 386	9,3 ± 2,3 148 -28,7 / 409
6 	133,2 ± 10,0 117 - 22,1 / 299	65,6 ± 4,4 149 - 23,9 / 312	34,4 ± 8,2 196 - 25,5 / 333	10,7 ± 1,2 197 - 28,4 / 343	2,5 ± 1,3 150 - 31,9 / 368	0,624 ± 0,101 151 - 35,4 / 376		6,0 ± 2,9 154 - 29,8 / 387	0,524 ± 0,064 152 - 35,8 / 409
7 	154,4 ± 18,7 162 - 21,8 / 299	50,5 ± 17,9 205 - 24,6 / 314	17,6 ± 5,4 206 - 27,2 / 334	4,9 ± 0,5 207 - 30,3 / 343	2,6 ± 1,1 208 - 32,1 / 370	0,524 ± 0,016 209 - 35,9 / 378		1,3 ± 0,5 160 - 33,7 / 387	0,386 ± 0,028 210 - 36,6 / 410

Abbildung 2: Graphische Auftragung der Freien Energien gegen die hydrophoben Kontaktoberflächen. Serie 1 und 2 in *a*, 3 in *b*, 4 in *c*, 5 in *d*, 6 in *e* und 7 in *f*.

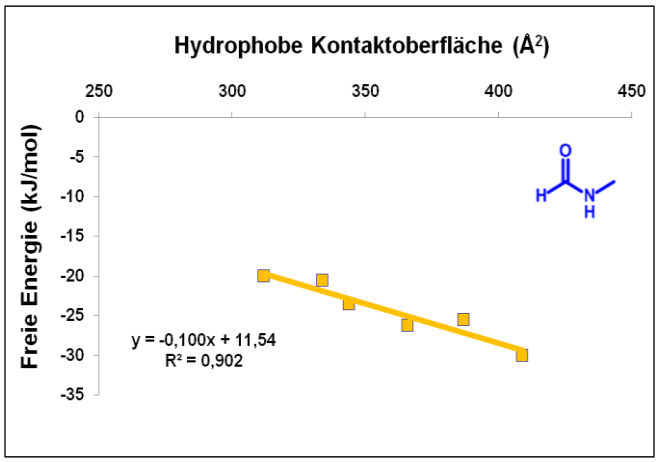
a)



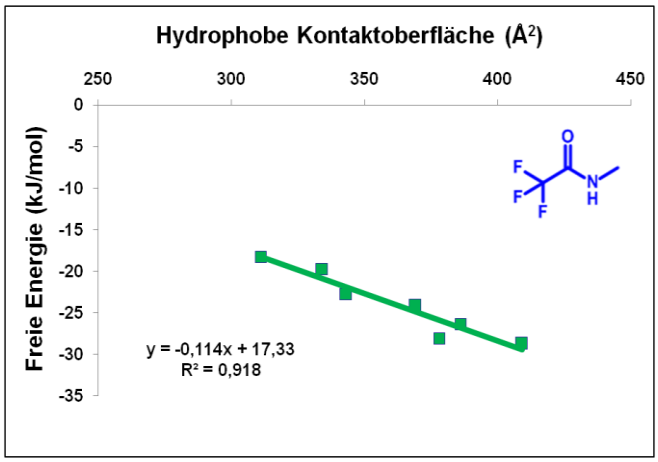
b)



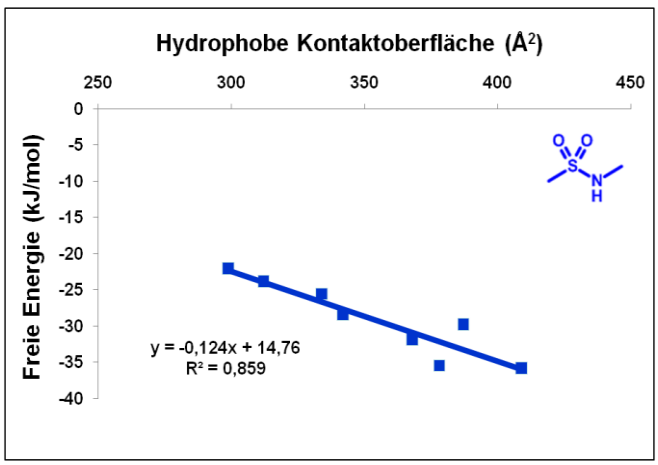
c)



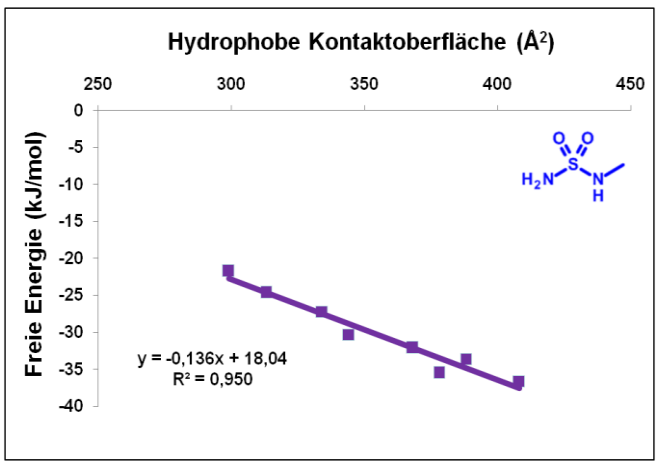
d)



e)

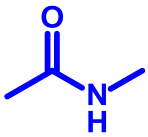
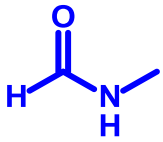
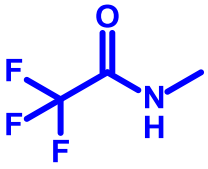
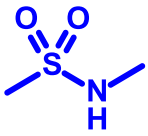
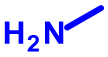
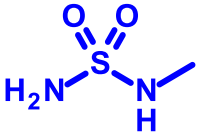


f)



7. Kooperativität

Tabelle 2: Die ermittelten Steigungen der Kurven bei den sieben Serien und die prozentuale Veränderung der Steigung im Vergleich zur Serie 1.

X	Steigung der Gerade ($kJmol^{-1}Å^{-2}$)
	- 0,077
	- 0,083
	- 0,100
	- 0,114
	- 0,124
	- 0,127
	- 0,136

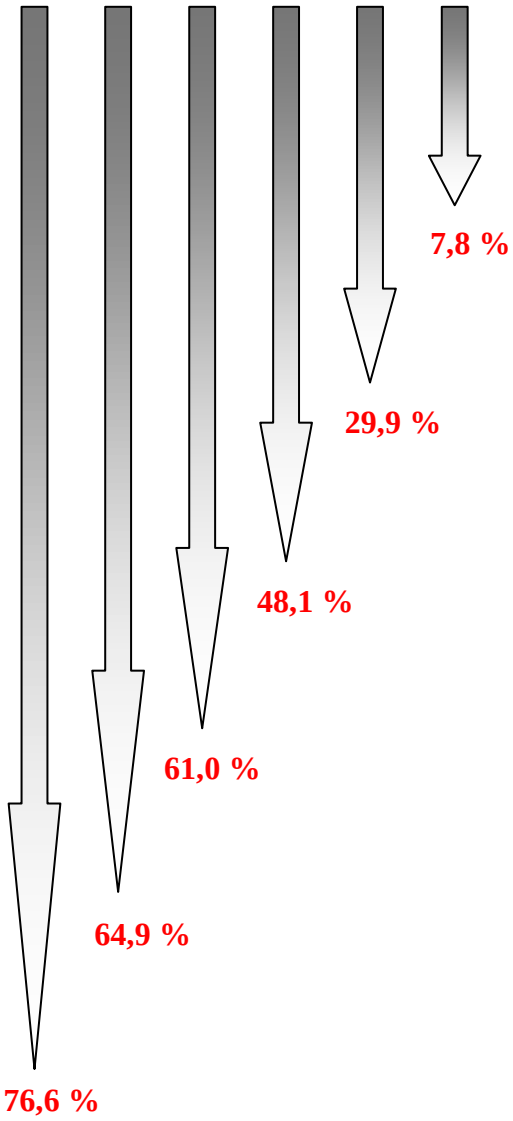
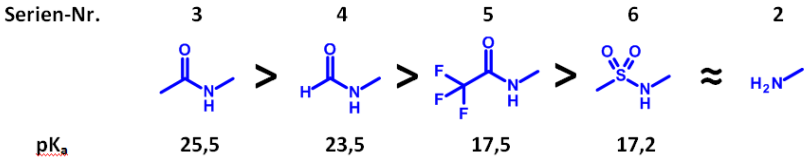


Abbildung 3: Darstellung der Substituenten nach der Höhe der pK_a-Werte in DMSO. Für Serie 2 liegt kein Wert in DMSO vor, jedoch kann er abgeschätzt werden, da der pK_a Wert in Wasser bekannt ist (8.0). Grundsätzlich sind die Werte für DMSO 8-10 Einheiten größer wie die Wasserwerte.

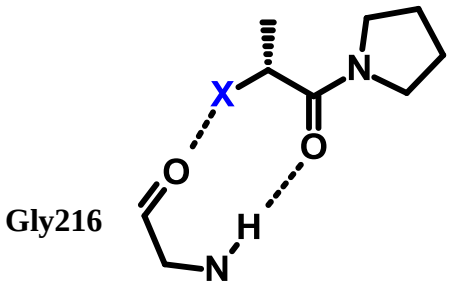
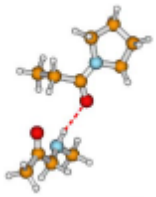


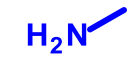
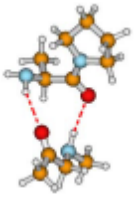
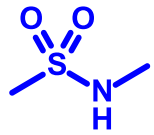
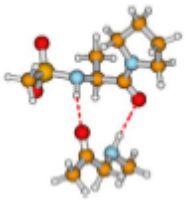
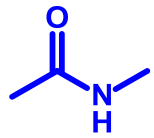
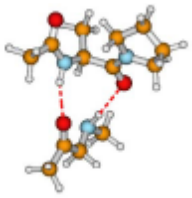
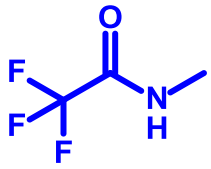
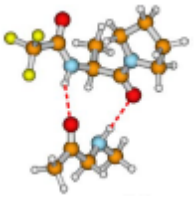
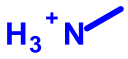
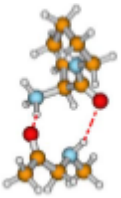
7.3.2 Die Stärke der Wasserstoffbrücke zu Gly216

Um die Hypothese weiter zu untermauern, wurde auf einem reduzierten Gerüst die Stärke der Wasserstoffbrücke mittels quantenchemischer Rechnungen abgeschätzt (siehe Material & Methoden für Details zur Berechnung). Diese Rechnungen wurden freundlicherweise in der Arbeitsgruppe von Prof. Dr. Engels (Julius-Maximilians-Universität Würzburg) durchgeführt. Das Gerüst bestand aus dem Pyrrolidin-Ring des Prolins und aus einem Alanin (Tabelle 3). Die Abschätzung der Energien der Wasserstoffbrücke in Abhängigkeit von den jeweiligen Substituenten wurde im Vakuum durchgeführt. Die Ergebnisse aus diesen Rechnungen liefern kein eindeutiges Bild. Die niedrigste Energie zeigt, wie erwartet, diejenige Serie, die keine Wasserstoffbrücke zu Gly216 aufweist (- 23 kJ/mol für Serie 1). Die Energie steigt sprunghaft an, wenn eine zusätzliche Wasserstoffbrücke ausgebildet wird (- 31 kJ/mol für die unprotonierte Serie 2). Die Protonierung der Amino-Gruppe zeigt dagegen den höchsten Wert aller berechneten Energien. Dies ist verständlich, da ladungsunterstützte Wasserstoffbrücken stets stärker sind als die neutralen Analoga. Der Rest der Werte zeigt jedoch keine eindeutige Korrelation weder zu den beschriebenen pK_a-Werten noch zu den Werten aus Tabelle 2. Die sulfonierte Serie 6 zeigt sogar eine kleinere Energie (- 42 kJ/mol) als die acetylierte (- 45 kJ/mol) und die trifluoracetylierte Serie (- 46 kJ/mol). Dieses ist nicht in Einklang zu bringen mit den Überlegungen aus dem vorherigen Kapitel. Bei näherer Betrachtung lassen sich viele Gründe anführen, wieso man sich nicht auf diese Werte alleine stützen kann. Erstens, die Werte wurden berechnet anhand eines stark reduzierten Systems, das bedingt vergleichbar ist mit dem experimentell untersuchten Fall. Zweitens, wurde wegen der rechenintensiven Abschätzung aus jeder Serie nur die Energie für das jeweilige Alanin-Derivat untersucht. Eine zuverlässige Analyse müsste alle untersuchten P3-Reste miteinbeziehen. So ist es möglich zu untersuchen, ob allein der hydrophobe Rest auch einen

Einfluss auf die Energie der Wasserstoffbrücke ausübt. Drittens, gehen die Berechnungen bei den sulfonierten Serien davon aus, dass die Sulfonylgruppen keine Interaktion mit dem Protein eingehen. Eine in dieser Arbeit untersuchte Serie mit ähnlichen Benzylsulfonylderivaten zeigt aber eindeutig eine zweite Wasserstoffbrücke zu Gly219.[10] Gerade dieser dritte Aspekt könnte der Hauptgrund dafür sein, wieso die berechnete Stärke der Wasserstoffbrücke nicht mit der beobachteten Kooperativität korreliert. Auch würde die zweite Wasserstoffbrücke bei den sulfonierten Serien die Hypothese um die Stärke der Wasserstoffbrücke weiter stützen. Anfangs war schon kritisch angemerkt, dass die Kooperativität bei den Serien 6 und 7 mit der ladungsunterstützten Wasserstoffbrücke in Serie 2 vergleichbar ist. Die zweite Wasserstoffbrücke in Serie 6 und 7 würde den Bereich um Gly216 noch stärker fixieren als eine einzelne Wasserstoffbrücke, auch wenn diese durch eine Ladung verstärkt wird. Zusätzlich ist aber auch noch zu berücksichtigen, dass die einzelnen Liganden mit abweichendem Desolvatationsbeitrag für die Bindung an das Protein freigesetzt werden müssen. Vor allem für die geladene Aminogruppe ist hier ein deutlich stärkerer Beitrag zu erwarten.

Tabelle 3: Die Energien aus den quantenchemischen Rechnungen für die Stärke der Wasserstoffbrücke zu Gly216.

<table>
<tr><td colspan="3"> 
</td></tr>
<tr><td>X =</td><td></td><td>ΔG (kJ/mol)</td></tr>
<tr><td>  </td><td>  </td><td>- 23</td></tr>
</table>

		**- 31**
		**- 42**
		**- 45**
		**- 46**
		**- 102**

7.4 Schlussfolgerung/Conclusion

In dieser Arbeit wurde untersucht, in welchem Ausmaß eine Kooperativität zwischen der hydrophoben Wechselwirkung und der Wasserstoffbrücke zu Gly216 besteht. Die vorliegenden Daten zeigen eindrucksvoll, dass kooperative Effekte in Abhängigkeit von der Substitution am terminalen Amin vorliegen. Die quantenmechanischen Berechnungen anhand eines reduzierten Systems zeigen jedoch nicht den erwarteten Einfluss der Substitution auf die untersuchte Wasserstoffbrücke zu Gly216. Erste Hinweise aus verwandten Bindungsmodi lassen den Verdacht zu, dass die sulfonierten Serien eine zweite Wasserstoffbrücke zum Protein ausbilden. Daher besteht hier die Notwendigkeit, jeweils eine Kristallstruktur aus den Serien 6 und 7 zu bestimmen, um sicher zu stellen, ob tatsächlich eine zweite Wasserstoffbrücke auftritt. Falls diese zweite Wasserstoffbrücke

vorliegt, muss dieses Faktum natürlich in den Rechnungen zur Stärke der Wasserstoffbrücke berücksichtigt werden. Die Hypothese, den Ursprung der Kooperativität auf die Stärke der Wasserstoffbrücke zurückzuführen, bleibt somit noch bestehen, jedoch wird man untersuchen müssen, ob auch die Anzahl der ausgebildeten Wasserstoffbrücken bei den beobachteten kooperativen Effekten eine Rolle spielt.

7.5 Danksagung/Acknowledgments

Wir danken CSL Behring recht herzlichst für die großzügige Bereitstellung von Thrombin-Material aus der Produktion von Beriplast®. Diese Arbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF, Förderkennzeichen 0315161C) gefördert.

7.6 Material & Methoden/Materials & Methods

7.6.1 Bioassay

Der Inhibitionskonstanten (K_i) wurden mit einen photometrischen Assay bei 405 nm gemessen. Der Assay wurde unter Verwendung des chromogenen Substrats Pefachrom tPa (LoxoGmbH, Dossenheim, Germany) wie beschrieben[11] bei folgenden Bedingungen durchgeführt: 50 mM Tris–HCl, pH 7,4, 154 mM NaCl, 5 % DMSO, 0,1 % PEG8000 bei 25°C und verschiedenen Substrat- und Inhibitorkonzentrationen. K_i-Werte ($n \geq 3$) wurden wie beschrieben[12] nach Dixon berechnet.

7.6.2 Berechnung der hydrophoben Kontaktoberflächen

Die Liganden wurden gedockt und energieminimiert mit dem Tripos-Kraftfeld. Die *molcad*-Oberfläche für das Protein wurde generiert. Auf diese Oberfläche wurde nun das entsprechende lipophile Potential abgebildet. Für die Analyse wurden nur die hydrophoben Oberflächenanteile berücksichtigt. Als erstes wurde der Schwerpunkt für den lipophilen Potentialbereich für das Protein berechnet und das Minimum des lipophilen Oberflächenpotentials ausgewählt. Im Folgenden wurde die *molcad*-Oberfläche ein zweites Mal auf die Oberfläche projiziert, diesmal entsprechend der Distanz zwischen Ligand und Protein. Danach ist eine neue Oberfläche mit dem lipophilen Potential entsprechend dem Schwerpunkt der lipophilen Skalierung und der Distanz von 2.6 Å generiert worden, so dass

kein Wassermolekül sich zwischen den Oberflächen aufhalten kann. Nachdem diese neue Oberfläche kreiert wurde, konnte man die gewünschten Oberflächenanteile berechnen.

7.6.3 Berechnung der Stärke der Wasserstoffbrücke zu Gly216

Die Geometrie des Komplexes wurde im Vakuum mit BLYP/TZVP optimiert und dann an Hand einer Einzelpunktberechnung mit CCSD(T)/aVDZ die Energie berechnet. Die Rechengenauigkeit dieser Messungen liegt bei ca. 2-4 kJ/mol.

7.7 Referenzen/References

1 Gohlke H., Hendlich M. & Klebe G. (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol.* **295**, 337-356.

2 Jones G., Willett P., Glen R.C., Leach A.R. & Taylor R. (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol.* **267**, 727-748.

3 Biela A., Khyat M., Tan H., Kong J., Heine, A. *et al.* (2011) Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin. Submitted to *J.Mol.Biol.*

4 Dunitz J.D. (1995) Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol.* **2**, 709-712.

5 Williams D.H., Searle M. S., Mackay J.P., Gerhard U. & Maplestone R.A. (1993) Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. *Proc Natl Acad Sci* **90**, 1172–1178.

6 Baum B., Muley L., Smolinski M., Heine A., Hangauer D. *et al.* (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.

7 Fersht A.R., Shi J.P., Knill-Jones J., Lowe D.M., Wilkinson A.J. *et al.* (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature.* **314**, 235-8.

8 Pace C.N. (1992) Contribution of the hydrophobic effect to globular protein stability. *J Mol Biol.* **226**, 29-35.

9 Baum B. (2009) The well-tempered thrombin: a systematic crystallographic and calorimetric study on the thermodynamics of serine-protease inhibition. PhD thesis.

10 Biela A., Sielaff F., Heine A., Steinmetzer T. & Klebe G. Enthalpic and entropic changes caused by a stepwise disruption of a water network in the S3/4 subsite of thrombin: An example of a classical hydrophobic effect, in preparation.

11 Stürzebecher J., Stürzebecher U., Vieweg H., Wagner G., Hauptmann J. *et al.* (1989) Synthetic inhibitors of bovine factor Xa and thrombin comparison of their anticoagulant efficiency. *Thromb. Res.* **54**, 245–252.

12 Dixon M. (1972) The graphical determination of K_m and K_i. *Biochem. J.* **129**, 197–202.

13 Mark A.E. & van Gunsteren W.F. (1994) Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J Mol Biol.* **240**, 167-176.

8 Preorganisierende Effekte während der Protein-Ligand Interaktion/Preorganizing effects in protein-ligand binding

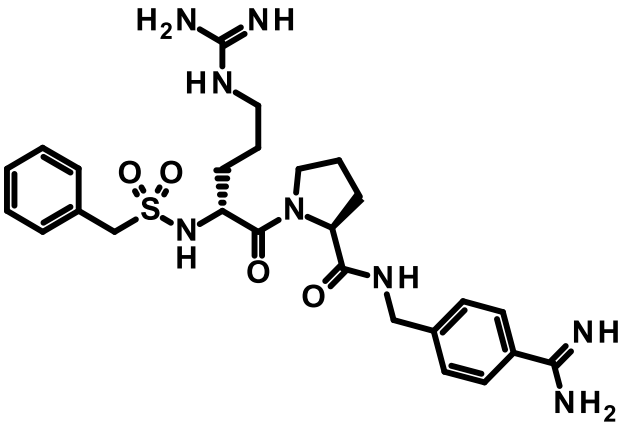
8.1 Kristallographische und thermodynamische Untersuchung von MI001 und MI002

In diesem Kapitel wurden zwei Thrombin-Inhibitoren in Kooperation mit der Arbeitsgruppe von Prof. Dr. Steinmetzer thermodynamisch und kristallographisch charakterisiert. Die beiden Inhibitoren sind mit dem Grundgerüst der Inhibitoren aus dem *Kapitel 4* vergleichbar. In diesem Fall wurde jedoch in P3 keine hydrophobe Aminosäure sondern das polare D-Arginin eingebaut. Ein AMBA- und ein ACB-Anker wurden, wie im Kapitel 4, ausgesucht, um die S1 Tasche zu adressieren (Abbildung 1).

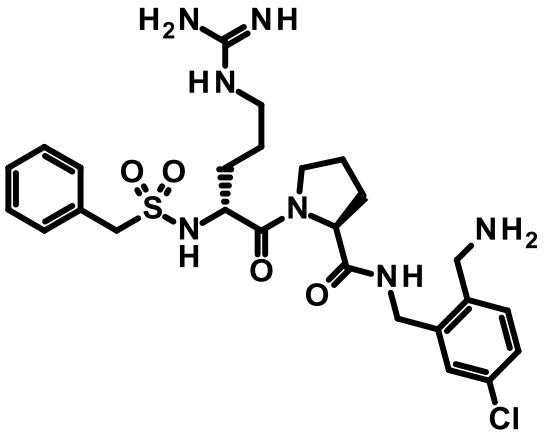
Die Komplexstruktur von **MI001** wurde bei einer Auflösung von 1,45 Å bestimmt (Abbildung 2). Der Bindungsmodus zeigt, wie erwartet, den S1 Anker in der S1 Tasche und das Benzylsulfonylamid in der Nähe einer Disulfidbrücke. Dieser U-förmige Bindungsmodus wurde bisher bei allen Derivaten beobachtet, die in der P3 Position eine chirale Aminosäure enthielten (siehe *Kapitel 4*). Interessanterweise wird die polare Seitenkette des D-Arg nicht in die S3/4 Tasche platziert. Die Guanidin-Gruppe des D-Arg interagiert nur mit dem Glu217 (2.9 Å) in der Bindetasche. Dieses Glutamat ist durch die Interaktion mit dem Inhibitor in einem starken Netzwerk aus Wasserstoffbrücken fixiert. Es interagiert zusätzlich mit dem Lys224 (2.6 Å) und dem Thr172 (2.6 Å). Neben der Interaktion zu dem Glutamat verbrückt ein Wassermolekül eine Interaktion zwischen dem Inhibitor und der NH-Gruppe des Arg173. Interessanterweise ist auffällig, dass die Guanidingruppe stark solvatisiert ist. Insgesamt wurden fünf Wassermoleküle in unmittelbarer Nachbarschaft zur Guanidingruppe (2,9 – 3,5 Å) gefunden. Die S3/4 Tasche behält in diesem Komplex den Solvatationszustand der Apostruktur bei, da kein einziges Wassermolekül durch **MI001** verdrängt wurde. Der Inhibitor **MI002** (PDB-Code 3EQ0) wurde bereits in der Doktorarbeit von Dr. Baum[(1)] kristallographisch untersucht. Die Überlagerung beider Strukturen ist in Abbildung 3 zu sehen. Das Benzylsulfonylamid ist abhängig vom S1 Anker in einer unterschiedlichen

Position. Der Inhibitor mit dem ACB Anker hat im Komplex nicht die Möglichkeit einen U-förmigen Bindungsmodus einzunehmen, da die Aminomethylengruppe sterisch diese Geometrie verhindert. Das *Kapitel 4* behandelt diese Problematik bereits im Detail. Die D-Arg Reste zeigen jedoch ebenfalls unterschiedliche Bindungsmodi. Während in **MI001** eindeutig nur eine Konformation zu sehen ist, werden in der Struktur **MI002** zwei Konformationen beobachtet. Die zweite Konformation in **MI002** ist nur zu 30 % besetzt und zeigt eine ähnliche Position wie das D-Arg in **MI001**. Neben der problematisch niedrigen Besetzung bestehen zusätzlich zwei weitere Punkte, die diese zweite Konformation in Frage stellen.

Abbildung 1: Chemische Strukturen der untersuchten Thrombin-Inhibitoren **MI001** und **MI002**.



MI001



MI002

Abbildung 2: Bindungsmodus des Inhibitors **MI001** im Komplex mit Thrombin. Besonders im Fokus stehen die strukturellen Details in der Region nahe der D-Arg Seitenkette. Die F_o-F_c Differenzelektronendichte ist bei 2σ in grün gezeigt. Günstige Interaktionen mit den jeweiligen Distanzen in Å sind mit gestrichelten Linien angedeutet. Wassermoleküle sind als Kugeln dargestellt. Rote Wassermoleküle interagieren mit dem D-Arg und die hellblauen mit der S3/4 Tasche.

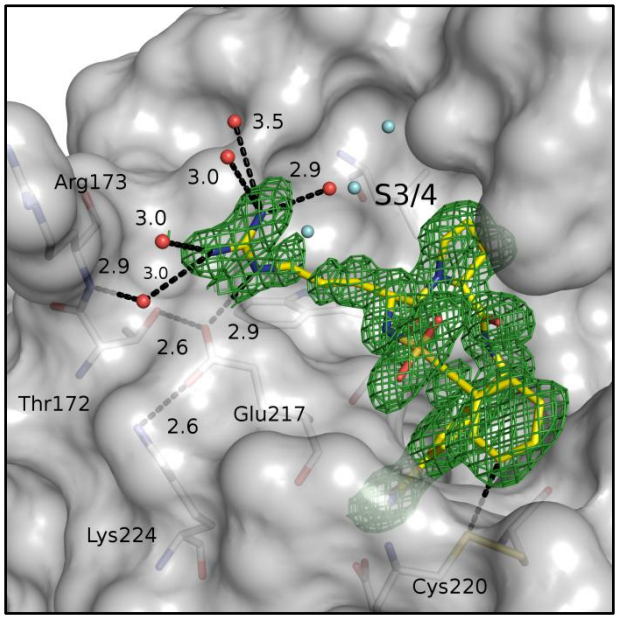
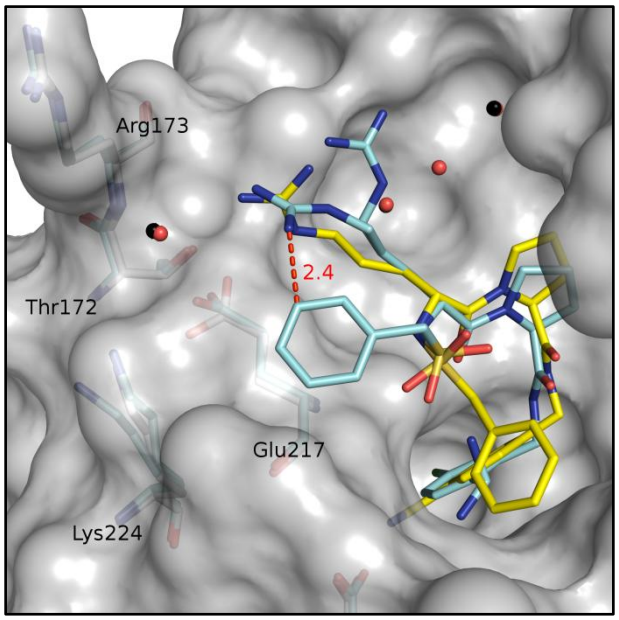


Abbildung 3: Überlagerung der Komplexstrukturen von **MI001** und **MI002** (PDB-Code 3EQ0). Besonders im Fokus stehen die strukturellen Details in der Region nahe des D-Arg. Ungünstige Interaktionen mit den jeweiligen Distanzen in Å sind mit gestrichelten Linien in rot angedeutet. Stickstoff ist in blau, Sauerstoff in rot, Kohlenstoff in gelb (**MI001**) und in hellblau (**MI002**) dargestellt. Wassermoleküle sind entweder als rote (**MI001**) oder als schwarze (**MI002**) Kugeln dargestellt.

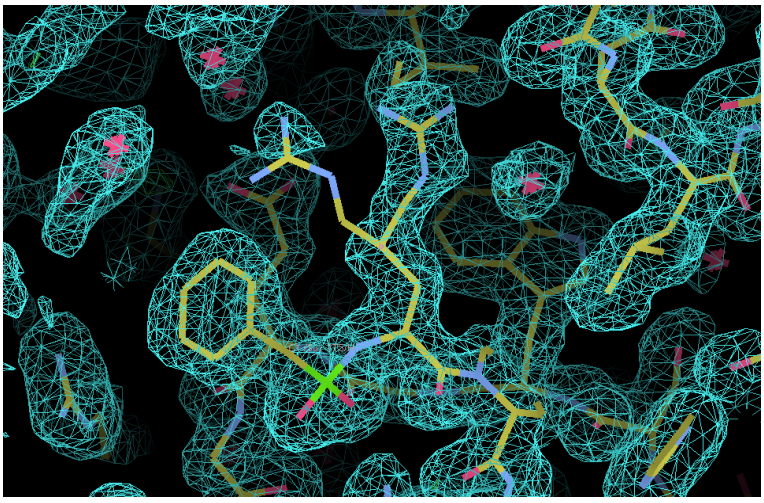


8. Preorganisierende Effekte

Das D-Arg ist in einer ungewöhnlich kurzen Distanz nahe dem Benzyl-Rest des eigenen Inhibitors. Darüberhinaus zeigt die $2F_o$-F_c Elektronendichte auch bei 0,7σ keine überzeugenden Hinweise für eine zweite Konformation (Abbildung 4). Vieles spricht daher in dieser Kristallstruktur für ein zweites Wassermolekül in der Nähe der Guanidin-Gruppe von **MI002**.

Die thermodynamische Charakterisierung (Tabelle 1) zeigt dieselben Trends, die schon im *Kapitel 4* zwischen den zwei S1 Ankern eindeutig zu erkennen waren. Die Enthalpie der Interaktion ist einerseits pufferabhängig (Abbildung 5), was auf die Deprotonierung des His57 zurückzuführen ist. Andererseits bestätigen die thermodynamischen Daten, dass die Interaktion mit dem ACB Anker wesentlich enthalpischer ($\Delta H_{Korr.}$= - 46,5 kJ/mol für **MI002**) abläuft als die Interaktion mit dem korrespondierenden AMBA ($\Delta H_{Korr.}$= - 16,4 kJ/mol für **MI001**). Das ACB-Derivat ist darüberhinaus stärker (ΔG = - 53,3 ± 0,3 kJ/mol für **MI002**) als das analoge AMBA (ΔG = - 46,5 ± 0,1 kJ/mol für **MI001**). Dies bestätigt die Ergebnisse aus dem *Kapitel 4* im Hinblick auf die Unterschiede in der Bindungsaffinität zwischen ACB- und AMBA-Anker.

Abbildung 4: Darstellung des Inhibitors **MI002** (PDB-Code 3EQ0) im Komplex mit Thrombin. Die $2F_o$-F_c Elektronendichte ist in cyan bei 0,7σ gezeigt. Stickstoff ist in blau, Sauerstoff in rot und Kohlenstoff in gelb gezeigt. Wassermoleküle sind als Kreuze zu erkennen.



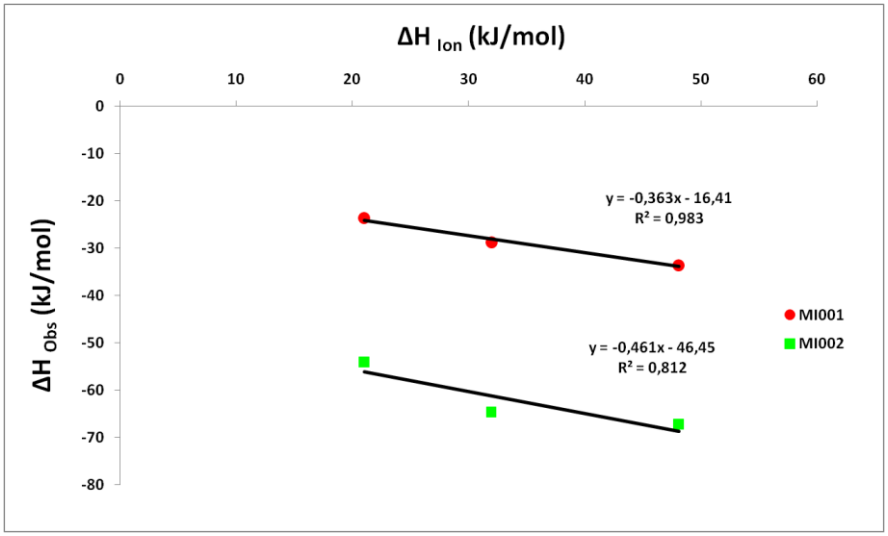
8. Preorganisierende Effekte

Tabelle 1: Bindungsterme ΔG^0, ΔH^0 und - $T\Delta S^0$ (kJ/mol) der untersuchten Liganden aus Kapitel 7.1 und 7.2 gemessen mit ITC.

Ligand-name	ΔG^0 (kJ/mol)	ΔH^0 (kJ/mol) in Tris	ΔH^0 (kJ/mol) in Tricine	ΔH^0 (kJ/mol) in Hepes	ΔH^0 (kJ/mol) Puffer-korrigiert	- $T\Delta S^0$ (kJ/mol)	Anzahl übertragender Protonen (mol)
MI001	- 46,5 ± 0,1	- 33,6 ± 3,2	- 28,8	- 23,6	**- 16,4**	- 32,3	0,36
MI002	- 53,3 ± 0,3	- 67,3 ± 3,2	- 64,7	- 54,1	**- 46,5**	- 6,8	0,46
MI034	- 59,0 ± 0,9	- 63,2 ± 3,1	- 54,6	- 50,5	**- 40,2**	- 18,8	0,47
MI330	- 50,7 ± 0,3	- 44,6 ± 0,2	- 37,9	- 31,2	**- 21,3**	- 29,4	0,49

Die thermodynamischen Daten wurden mithilfe einer ITC-Verdrängungstitration gemessen. Kapitel 4 beschreibt die Vorgehensweise im Detail.

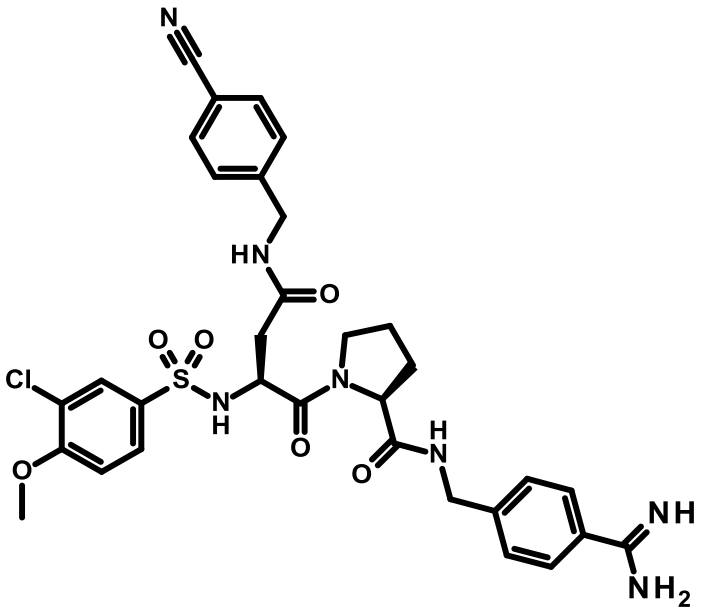
Abbildung 5: Im Diagramm sind die beobachteten Enthalpien (ΔH_{obs}) gegen die Ionisierungsenthalpien (ΔH_{Ion}) des jeweiligen Puffers aufgetragen.



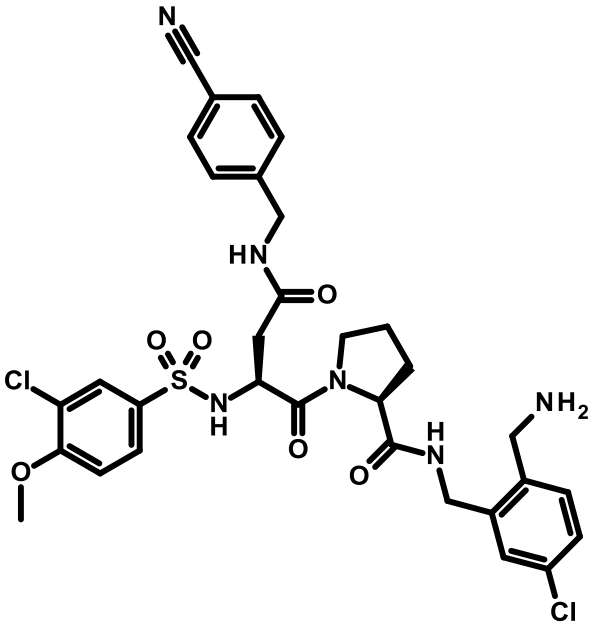
8.2 Kristallographische und thermodynamische Untersuchung von MI034 und MI330

Diese Studie wurde ebenfalls in Kooperation mit der Arbeitsgruppe von Prof. Dr. Steinmetzer durchgeführt. Das Grundgerüst ist in diesem Fall nicht mehr vergleichbar mit dem aus dem vorangegangenen Kapitel. Es wurde ein an zwei Stellen substituiertes Phenylsulfonylamid neben der P3 Position eingebaut (Abbildung 6). Außerdem ist in P3 eine L-Aminosäure (L-Asparagin) eingeführt. Alle anderen in dieser Dissertation untersuchten Thrombin-Inhibitoren waren in P3 mit einer D-Aminosäure ausgestattet. Das Asn ist zudem N-substituiert mit einem 4-Cyanophenylrest. Vergleichbar mit den anderen Studien zur Thrombin-Inhibition ist auch hier der direkte Vergleich der S1 Anker AMBA und ACB wesentlicher Bestandteil dieses Kapitels.

Abbildung 6: Chemische Strukturen der untersuchten Thrombin-Inhibitoren **MI034** und **MI330**.



MI034



MI330

Die kristallographische Untersuchung von **MI330** im Komplex mit Thrombin zeigt, dass der S1 Anker die S1 Tasche und das Prolin in P2 die S2 Tasche ausfüllt (Abbildung 7). Der Chlor-substituierte Phenylsulfonylrest besetzt die hydrophobe S3/4 Tasche besetzt. Das 4-Cyanobenzyl-substituierte L-Asn liegt trotz P3 Position gegenüber der S3/4 Tasche. Die falsche Konfiguration am α-Kohlenstoffatom und die voluminöse Substitution schließen eine Adressierung der S3/4 Tasche durch den P3 Rest aus.

Der Aminomethylen-Anker zeigt eine intramolekulare Wasserstoffbrücke (2.8 Å) mit dem Carbonylsauerstoff des substituierten L-Asn (Abbildung 8a). Diese Interaktion ist zudem durch ein Wassermolekül unterstützt. Der 4-Cyanophenylrest ist in einer Region, die ausschließlich von polaren Aminosäuren (Glu146, Arg221A, Lys224, Glu217) ausgebildet wird (Abbildung 8b). Jedoch kann die Cyanogruppe keine dieser Reste adressieren. In der Kristallstruktur konnten für die Cyanogruppe weder Interaktionen zum Protein noch zum umgebenden Wasser festgestellt werden. Die kleinste Distanz weist 3.9 Å zum Glu217 auf.

Das Interaktionsmuster in der S3/4 Tasche ist durch van-der-Waals Kontakte und Halogen-π-Wechselwirkungen charakterisiert (Abbildung 8c). Das m-ständige Chloratom zeigt dabei auf das Zentrum des Benzolringes des Indols des Trp215 (3.6 Å). Zudem ist das Chloratom in unmittelbarer Nachbarschaft zum Carbonylsauerstoff des Asn98 (3.3 Å).

Abbildung 7: Bindungsmodus des Inhibitors **MI330** im Komplex mit Thrombin. Die F_o-F_c Differenzelektronendichte ist bei 2σ in grün gezeigt. Wassermoleküle sind wegen einer besseren Darstellung ausgeblendet. In dieser Darstellung ist die Oberfläche des Proteins gezeigt.

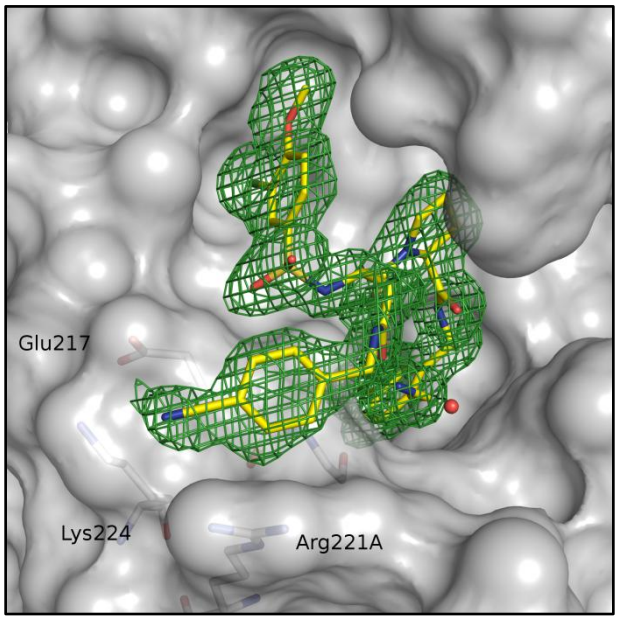
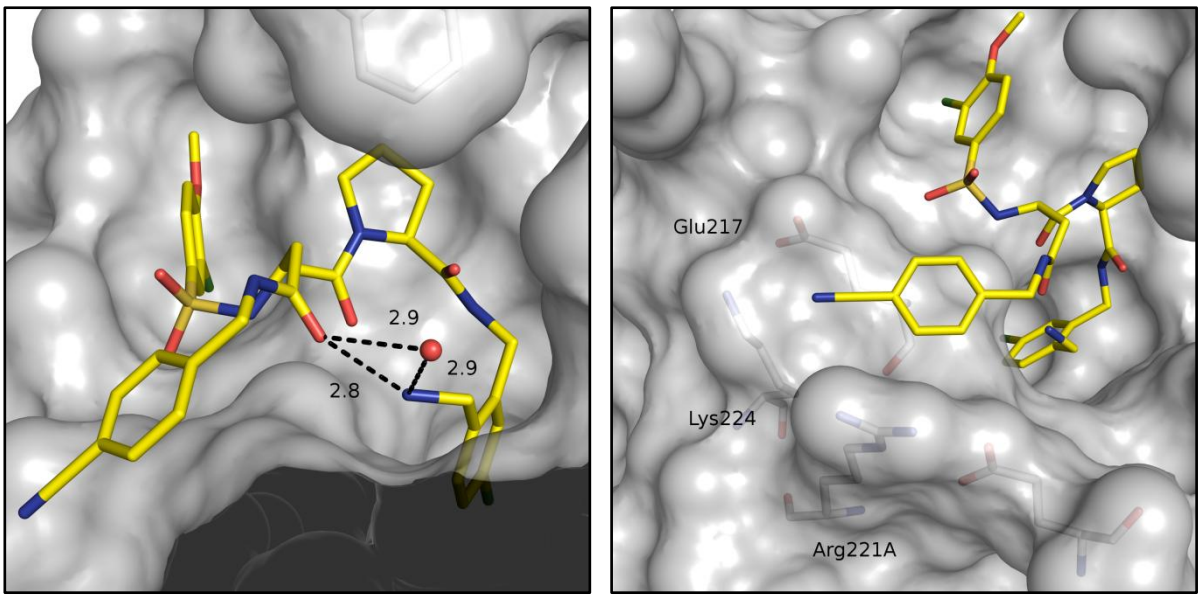
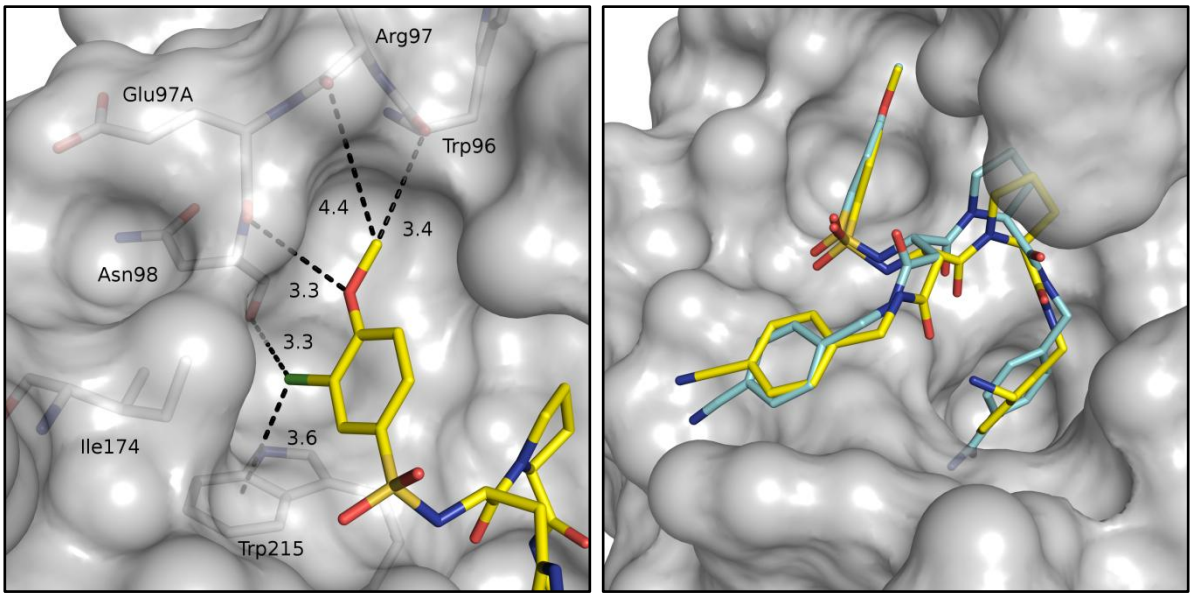


Abbildung 8: MI330 im Komplex mit Thrombin. *a*) Besonders im Fokus stehen die strukturellen Details in der Region nahe des Aminomethylen-Ankers. *b*) Interaktionsmuster zum 4-Cyanophenyl-sulfonylamid. *c*) Bindungsmodus des Inhibitors in der S3/4 Tasche. *d*) Überlagerung von **MI034**[(1)] und **MI330**. Stickstoff ist in blau, Sauerstoff in rot, Kohlenstoff in gelb (**MI330**) und in hellblau (**MI034**) dargestellt.

a) *b*)



c) *d*)



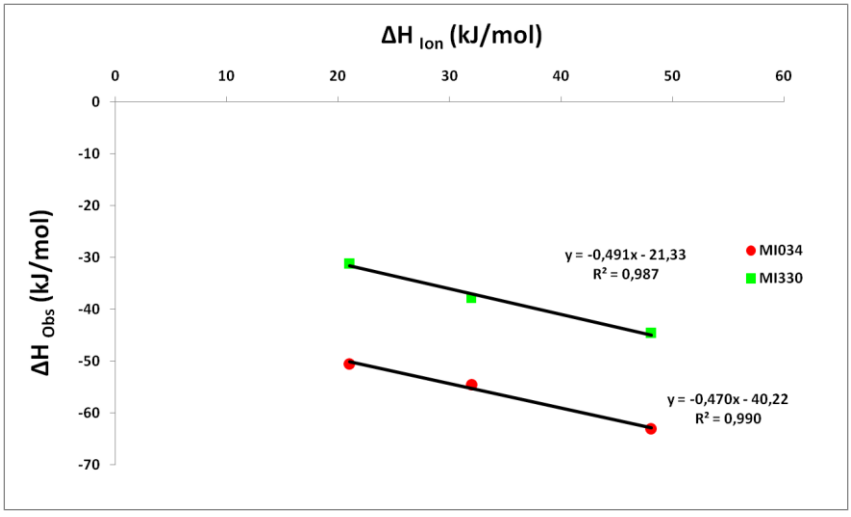
Der Sauerstoff des Methylethers ist jedoch in einer ungünstigen Position, da er nur 3.3 Å vom Carbonylsauerstoff des Glu97A entfernt ist. Daher ergibt sich eine abstoßende Interaktion zwischen einem sp^2 und einem sp^3 Sauerstoff. Die Methylgruppe hingegen liegt in einer Position, die es erlaubt mit den benachbarten Carbonylsauerstoffen van-der-Waals Kontakte auszubilden (3,4 Å, 4,4 Å). Der strukturelle Vergleich zum AMBA zeigt überwiegend einen vergleichbaren Bindungsmodus. Der einzige Unterschied zeigt sich in der Konformation des L-Asn (Abbildung 8d). Das ACB-Derivat offenbart wie oben beschrieben eine intramolekulare Wechselwirkung zwischen dem Carbonylsauerstoff des L-Asn und dem Aminomethylen-Anker. Diese Wechselwirkung fehlt in der AMBA-Struktur. Dort dreht sich die Amidgruppe um 180° und der Carbonylsauerstoff steht zum Wasser hin.

Das neuartige Grundgerüst führt auch zu einem thermodynamischen Profil (Tabelle 1), das nicht vergleichbar ist mit den Benzylsulfonylamiden aus vorangegangenen Studien. In diesem Fall ist das AMBA (ΔG = - 59,0 ± 0,9 kJ/mol für **MI034**) weitaus stärker als das ACB-Derivat (ΔG = - 50,7 ± 0,3 kJ/mol für **MI330**). Bei allen bisher untersuchten Benzylsulfonylamiden wurde für den ACB-Anker stets eine stärkere Bindungsaffinität beobachtet als bei den korrespondierenden AMBA-Liganden. Ebenfalls interessant sind die enthalpischen Beiträge für **MI034** und **MI330**. Während der ACB-Anker bei den Benzylsulfonylamiden immer einen höheren enthalpischen Beitrag aufwies als das AMBA, wurde hier das Gegenteil beobachtet. Nun zeigt das thermodynamische Profil für das AMBA

einen höheren enthalpischen Term ($\Delta H_{Korr.}$= - 40,2 kJ/mol für **MI034**) als das analoge ACB-Derivat ($\Delta H_{Korr.}$= - 21,3 kJ/mol für **MI330**).

Die Kristallstruktur von **MI034** offenbart, dass die preorganisierenden Effekte nicht so ausgeprägt sind wie bei den Benzylsulfonylamiden. **MI330** zeigt mit seinem Aminomethylen-Anker nur schwache intramolekulare preorganisierende Interaktionen. Diese Interaktion ist höchstwahrscheinlich auch nicht sehr entscheidend, da B-Werte der P3-Seitenkette eindeutig zeigen, dass dieser Rest durch eine hohe Unordnung charakterisiert ist. Dieses widerspricht einer starken intramolekularen Interaktion. Der Einbruch in der Bindungsaffinität von **MI330** ist daher wahrscheinlich durch das Fehlen von preorganisierenden Effekten zu erklären. Die Anzahl der übertragenden Protonen war vergleichbar mit allen anderen bisher untersuchten Thrombin-Inhibitoren (0,47 mol für **MI034** und 0,49 mol für **MI330**, Abbildung 9).

Abbildung 9: Im Diagramm sind die beobachteten Enthalpien (ΔH_{obs}) gegen die Ionisierungsenthalpien (ΔH_{Ion}) des jeweiligen Puffers aufgetragen.



8.3 Referenzen

1 Baum B. (2010) The well-tempered Thrombin: A systematic crystallographic and calorimetric study on the thermodynamics of serine-protease inhibition. *Dissertation.*

Tabelle 2: Datensammlung und Verfeinerungsstatistik für die Liganden aus den Kapiteln 7.1 und 7.2.

Structure (PDB entry)	THR-**MI001** complex **(3U98)**	THR-**MI330** complex **(3U9A)**
A. Data collection and processing		
No. Crystals used	1	1
Wavelength (Å)	0.91841	1.00
Space group	C2	C2
Unit cell parameters		
a, b, c (Å)	70.3, 71.5, 72.2	69.9, 71.2, 72.6
β (°)	100.4	100.1
Matthews coefficient ($Å^3$/Da)	2.5	2.5
Solvent content (%)	51	51
B. Diffraction data[a]		
Resolution range (Å)	50 – 1.45	50 – 1.58
	(1.48 – 1.45)	(1.61 – 1.58)
Unique reflections	59.463 (3.072)	46.917 (2.328)
R(I)sym (%)	5.9 (45.3)	3.7 (47.1)
Completeness (%)	95.2 (97.9)	97.6 (95.6)
Redundancy	3.0 (2.9)	3.0 (2.8)
I/σ(I)	17.4 (2.6)	28.8 (2.4)
C. Refinement		
Resolution range (Å)	22.6 – 1.45	35.7 – 1.58
Reflections used in refinement	56.456 / 2.856	44.933 / 2.233
(work/free)		
Final R values for all reflections	16.4 / 18.7	16.4 / 19.6
(work/free) (%)		
Protein residues (L chain/H chain)	28 / 251	28 / 251
Sodium ions	2	2
Inhibitor atoms	39	
Water molecules	344	253
RMSD from ideality		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.117	1.087
Ramachandran plot		
Residues in most favoured regions (%)	84.1	85.0
Residues in additionally allowed		
regions (%)	15.9	15.0
Residues in generously allowed		
regions (%)	-	-
Mean B-factor ($Å^2$)		
Protein (L + H chain)	19.7	24.7
Binding site[b]	15.7	21.0
Inhibitor	19.9	25.0
Water molecules	34.2	35.1

[a] Numbers in parenthesis are for the highest resolution shell.

[b] Definition of the binding site: all amino acids which are 4 Å within the inhibitor.

9 Zusammenfassung & Perspektiven/Summary & Perspektives

9.1 Deutsche Version

Der zentrale Focus dieser Arbeit liegt auf der kristallographischen und thermodynamischen Untersuchung der Rolle von Wassermolekülen bei der Protein-Ligand-Wechselwirkung.

Da die meisten untersuchten Liganden einen zu starken ($K_i \leq 1$ nM) bzw. zu schwachen ($K_i \geq 1$ µM) Affinitätsbereich abdecken, ist die direkte ITC-Titration für die thermodynamische Charakterisierung nicht in Frage gekommen. Es ist schließlich ein Verdrängungsansatz ausgewählt worden, um eine zuverlässige thermodynamische Analyse in diesen Affinitätsbereichen zu garantieren. Der erste Teil dieser Arbeit (Kapitel 2) widmet sich daher der Etablierung und Validierung der Verdrängungstitration für die Charakterisierung von schwach und stark affinen Liganden. Es ist gezeigt worden, dass die Dissoziationskonstanten (K_d) aus der Verdrängungs-ITC sehr gut mit den kinetisch bestimmten Inhibitionskonstanten (K_i) korrelieren. Diese Methode kann auch angewandt werden um Fragmente, die für ihre überaus schwache Bindungsaffinität bekannt sind, thermodynamisch zu charakterisieren. Gerade bei Fragmenten wird in kinetischen Assays offensichtlich, dass ein großer Nachteil darin besteht eine korrekte Bindungsaffinität zu erhalten. Die kinetischen Assays sind nicht sensitiv genug oder entsprechende Assayprotokolle fehlen um eine zuverlässige Abschätzung im mikro- bzw. millimolaren Bereich zu liefern. Die korrekte Bestimmung der Bindungsaffinität ist jedoch im frühen FBDD (Fragment-based drug discovery) von entscheidender Bedeutung, da nur dann eine sinnvolle Aussage getroffen werden kann, welche Fragmente in die weitere Optimierung einbezogen werden können. Die Verdrängungs-ITC stellt hier eine zuverlässige Alternative dar, wie aus der guten Korrelation der Bindungsaffinitäten von verschiedenen Methoden ersichtlich ist. Neben der Bindungsaffinität kann die Verdrängungs-ITC auch den enthalpischen Parameter mit hoher Zuverlässigkeit bestimmen. Dieses stellt ebenfalls ein wichtiges Kriterium dar, um zu entscheiden, welche Fragmente in die nähere Auswahl für die Optimierung fallen. Auch ist es bei der Verdrängungs-ITC möglich die Proteinkonzentration zu reduzieren, wenn Liganden mit einem niedrigen enthalpischen Beitrag untersucht werden. Der oft in der ITC kritisierte

Proteinverbrauch kann um das dreifache reduziert werden. Außerdem ist theoretisch abgeleitet worden, wie die Konzentration eines schwach affinen Liganden sein muss um einen gewissen Grad an Inhibition zu gewährleisten. Die aus diesen Überlegungen resultierende Gleichung ist in ein Excel-Sheet übertragen worden, wo in Abhängigkeit von der erwarteten Affinität und der Proteinkonzentration der Grad der Inhibition bei einer gegebenen Ligandkonzentration berechnet werden kann.

Die nächste Studie befasst sich (Kapitel 3) mit der S1 Tasche von Thrombin. Am Boden dieser Tasche befindet sich das deprotonierte Asp189. Dieses ist in der Apostruktur von Thrombin sehr gut solvatisiert. Ziel war es nun zu untersuchen, wie sich eine Änderung der Aspartat-Solvatisierung durch eine Pyridin-basierte Ligandenserie auf die Bindungsaffinität und das thermodynamische Profil auswirkt. Methodisch basiert die thermodynamische Untersuchung dieser mikromolaren Pyridininhibitoren auf der Validierung der Verdrängungs-ITC. Diese Serie zeigt ein tieferes Verständnis über die Folgen von Desolvationseffekten, die in der Proteinbindetasche und im Liganden auftreten. Die Apoform von Thrombin zeigt ein fest gebundenes und geordnetes Wassermolekül am negativ geladenen Asp189 in der S1 Tasche. Die beobachteten Effekte bei der Adressierung der Solvatationsschicht der S1 Tasche mit Pyridin- und den analogen Methylpyridinium-Derivaten zeigen deutlich die Gründe für die nur mikromolare Aktivität. In allen Pyridin-Derivaten werden Wassermoleküle entdeckt, die in unterschiedlicher Art so gut es geht den ursprünglichen Solvatationszustand zu konservieren versuchen. Die damit verbundenen entropischen Verluste sind immens. In zwei Ligandkomplexen finden sich sogar ungeordnete Ligandbestandteile in der S1 Tasche als Folge von Desolvatationseffekten. Der Preis für die Desolvatation einer geladenen Aminosäure ist einfach zu groß, um in diesem Fall eine komplette Verdrängung des Wassermoleküls zu gewährleisten. So resultiert ein Kompromiss, wo 50 % der Wasser verdrängt werden und 50 % der Wasser am Asp189 verbleiben. Eine zu hohe Desolvatationsenergie kann daher zu ungeordneten Systemen führen, nur um den Preis der Desolvatation so gering wie möglich zu halten. Auch wird deutlich, dass fast alle positiv geladenen Methylpyridinium-Derivate schwächer sind als die neutralen Pyridin-Analoga. Die quantenchemische Berechnung der Solvatationsenergien der Liganden zeigen deutliche Unterschiede zwischen den neutralen und geladenen Derivaten. Die geladenen Verbindungen benötigen eine Größenordnung mehr Energie um desolvatisiert zu werden als die neutralen Liganden. Die aufgeklärten Komplexstrukturen zeigen zudem, dass die

geladenen Methylpyridinium-Derivate nicht optimal das am Boden lokalisierte negativ geladene Asp189 adressieren können. Entweder lässt sich aus sterischen Gründen kein kurzer Abstand erreichen oder die Methylgruppe bietet einen zu guten Schutz, um über eine kurze Entfernung mit dem negativen Asp189 zu interagieren, beispielsweise wie es in den Benzamidinen-Liganden zu realisieren ist. Daher kann die Energie für die Desolvatation nicht durch die freiwerdende Energie aus der Interaktion zwischen dem Methylpyridinium und dem Asp189 aufgebracht werden. Ein Verlust der Bindungsaffinität ist die Folge. So ist ebenfalls gezeigt worden, dass je näher die positive Ladung des Methylpyridiniums am Asp189 platziert ist desto geringer fällt der enthalpische Nachteil aus. Ein anderes Pyridin ist zudem fähig die Solvatationsschicht der S1 Tasche komplett in einen ungeordeneten Zustand zu transferieren. Der Vergleich zur Apostruktur und dem analogen Methylpyridinium enthüllt einen günstigen entropischen Term, der auf die ungeordnete Wasserhülle zurückzuführen ist.

Der vierte, fünfte und sechste Teil befasst sich mit Wassereffekten bei der hydrophoben Wechselwirkung. Eine vorherige Arbeit über die Bindung von hydrophoben Liganden zu der stark hydrophoben S_1‘ Tasche von Thermolysin zeigt einen enthalpisch getriebenen hydrophoben Effekt. Dies wird auf die unzureichende Desolvatation der S_1‘ Tasche zurückgeführt, was durch eine verwandte Arbeit von Homans *et al.* bestätigt wird. Basierend auf der Hypothese, dass das Bindungsprofil einer hydrophoben Wechselwirkung abhängig ist vom Solvatationsgrad der untersuchten Bindetasche, sind zwei Serien zum einen für die S3/4 Tasche von Thrombin, zum anderen für die S_2‘ Tasche von Thermolysin konzipiert worden. Beide adressierten Bindetaschen sind durch ausgeprägte Solvatisierung aber unterschiedliche Form gekennzeichnet.

Die S3/4 Bindetasche hat die Form eines Halbkreises und ist leicht zugänglich für Wassermoleküle. Zwei Serien mit unterschiedlichem S1-Anker sind dann in P3 Position hydrophob modifiziert worden. Die thermodynamische Charakterisierung zeigt einen hydrophoben Effekt, der eindeutig entropisch getrieben ist. In dieser Studie ist die S3/4 Tasche mithilfe der immer größer werdenden hydrophoben Modifizierungen schrittweise desolvatisiert worden. In beiden Serien verbessert sich die Bindungsaffinität um das 40-fache. Ausgehend von einer nanomolaren Aktivität ist die Bindungsaffinität in den niedrigen pikomolaren Bereich hydrophob optimiert worden. Die Benzamidin-Derivate sind sogar durch einen Bindungsmodus mit zwei gebundenen Liganden gekennzeichnet. Der zweite

gebundene Ligand ist zwar nur zu 50 % populiert, aber man erkennt deutlich die Differenzelektronendichte. Überraschenderweise zeigt dieser Bindungsmodus eine bemerkenswerte Ähnlichkeit mit dem Bindungsmodus von Fibrinopeptid A, das ein Spaltprodukt des natürlichen Thrombin-Substrats (Fibrinogen) darstellt.

Trotz der berechtigten Kritik bei der Verfeinerung von nur partiell populierten Wassermolekülen sind auch in dieser Arbeit Wassermoleküle entdeckt worden, die nicht komplett besetzt sind. Die komplette Desolvatation reiche auch hier nicht aus, da der enthalpische Preis für die Verdrängung aus dieser Proteinbindetasche nicht durch neue Interaktionen zum Protein vollständig kompensiert werden kann.

Die im fünften und sechsten Kapitel untersuchte S_2‘ Tasche von Thermolysin ist dagegen weniger gut geformt, aber dafür wegen ihrer Lage auf der Oberfläche optimal solvatisiert.

Im fünften Kapitel unterschieden sich die untersuchten Thermolysin-Liganden nur durch eine terminale Carboxyat- und/oder eine Methylgruppe. Eine überraschende Nicht-Additivität von Beiträgen einer Carboxylat- und/oder Methylgruppe ist festgestellt worden. Das Hinzufügen einer Methylgruppe bewirkte eine relativ kleine Zunahme in der Freien Energie mit kleinen Änderungen in der Enthalpie/Entropie, während das im nächsten Schritt carboxylierte Derivat mit einem starken Affinitätssprung und großen Änderungen in der Enthalpie/Entropie begleitet war. Wurde das Carboxylat jedoch zuerst und dann erst die Methylgruppe eingefügt, wurden umgekehrte Effekte beobachtet: jetzt war die saure Gruppe von geringfügigen Effekten begleitet, während die zusätzliche Methylgruppe enorme Veränderungen verursachte. Die Carboxylat-Gruppe stört in beiden Komplexen das lokale Wassernetzwerk. Die zusätzliche Methylgruppe bietet dagegen günstige Wechselwirkungen für zusätzliche Wassermoleküle an. In allen Komplexen, abgesehen von einem Beispiel, wird ein zusammenhängendes Wassernetzwerk zwischen den funktionellen Gruppen des Proteins und des Liganden beobachtet. Bei der Komplexstruktur mit dem carboxylierten Liganden fehlt die endständige Methylgruppe und dadurch bricht das Wassernetzwerk zusammen. Dieser Zusammenbruch der Wasserkette ergibt die überraschende thermodynamische Signatur mit nur geringen Affinitätsunterschieden bei dem carboxylierten Liganden, dem die Methylgruppe fehlt. Eine zusätzliche Methylgruppe dagegen hat günstige Interaktionsstellen für Wassermoleküle geschaffen und das Netzwerk wurde wieder hergestellt, begleitet von starken Affinitätsänderungen und einer enormen Änderung in der Enthalpie/Entropie.

Im sechsten Kapitel wurde die hydrophobe Bindung zur S_2‘ Tasche mit verschiedenen hydrophoben Liganden untersucht. Es hat sich gezeigt, dass kleine Reste eindeutig bevorzugt werden. Große und sperrige Substituenten haben auf der Proteinseite zu ungünstigen

Konformationen geführt. Ein Benzyl-Rest, zum Beispiel, hat zwar in dieser Position eine Verdrängung von Wassermolekülen bewirkt, aber dieser entropische Gewinn wurde in Teilen zunichte gemacht durch den Verlust der Planarität einer benachbarten Peptid-Bindung. So gingen große Reste mit einem deutlichen Verlust in der Bindungsaffinität einher aufgrund von Konformationsänderungen.

Die Adressierung der S_2‘ Tasche zeigt, wie in den beiden Serien aus der vierten Studie, ebenfalls ein entropisch dominiertes Signal, welches dem klassischen hydrophoben Effekt entspricht. Die vorliegende Reihe von eng verwandten Liganden zeigt, dass beide thermodynamischen Parameter betroffen sind und dass viele kleine strukturelle Änderungen das thermodynamische Profil bestimmen. Wenn ein zusammenhängendes Wassernetzwerk zusammenbricht, zeigen sich deutliche Verluste in der Enthalpie, während die Entropie stark zunimmt. Insbesondere bei der Bindung von Liganden in offene hydrophobe Taschen, die leichten Zugang zu Wasser haben, können sich neue Bindestellen für Wassermoleküle ergeben, wenn Ligandenbereiche zur Wasserphase ausgerichtet sind. In unserer Studie wurde zum Beispiel ein zusätzliches Wassermolekül über dem Carboxylat (*capping position*) und an der Seite des Benzolrings entdeckt. Auch solche Phänomene tragen natürlich auf der molekularen Ebene zum hydrophoben Effekt bei. Abschließend kann man feststellen, dass die hydrophobe Interaktion weder als überwiegend "entropisch" noch "enthalpisch" beschrieben werden kann. Kleine strukturelle Veränderungen auf molekularer Ebene bestimmen schließlich darüber, welche treibenden Kräfte hinter der Bindung von hydrophoben Bindungspartnern stehen.

In dieser Serie hat sich zudem ein weiteres Mal gezeigt, dass wenn eine strukturell wohldefinierte Solvatation auf der Proteinseite vorhanden ist, der klassische hydrophobe Effekt sichtbar wird. Alles spricht also dafür, dass das Bindungsprofil eines hydrophoben Liganden stark von den Solvatationseigenschaften der adressierten Proteinbindetasche abhängt. Andererseits zeigt sich, dass die Qualität der Solvatation einer Proteinbindetasche höchstwahrscheinlich von der Form und Lokalisierung der Bindetasche mitgestaltet wird. Die S_1‘ Tasche hat zum Beispiel eine klar definierte Gestalt in der Form eines abgeschnittenen Kegels. Bisher konnte der tatsächliche Solvatationszustand jedoch nicht beschrieben werden, da die Apostruktur bisher durch die Anwesenheit des Verdauungsprodukts Val-Lys nicht zugänglich ist. Alle Untersuchungen in dieser Arbeit und die Ergebnisse aus der Literatur deuten jedoch darauf hin, dass kristallographisch keine Wassermoleküle in der S_1‘

Bindetasche vorhanden sind. Ein an diese Arbeit anschließendes Projekt versucht dieses Geheimnis zu lüften indem eine Thermolysinmutante produziert wird, die sich nicht mehr selbst verdauen kann, da das problematische Valin am C-Terminus zum Glycin mutiert wird. Nun wird die Kristallstruktur dieser Mutante zeigen, ob in der Tasche tatsächlich Wassermoleküle zu detektieren sind. Auch wenn in der Kristallstruktur dieser Thermolysinmutante ebenfalls keine S_1‘ Wasser gefunden werden, bleibt immer noch eine gewisse Unsicherheit vorhanden, da man letztendlich nicht feststellen kann, ob die Wassermoleküle wirklich fehlen oder einfach in einem ungeordneten Zustand vorliegen, der naturgemäß in der Kristallographie schwer zugänglich ist. Nun muss hier deutlich betont werden, dass einige in dieser Arbeit gefundene Wasserpositionen trotz offensichtlicher Unordnung (insgesamt 6 Strukturen) sichtbar sind, wenn die Strukturen ausreichend hoch aufgelöst werden. In gewissem Maße würde diese Beobachtung für einen S_1‘ Solvatationszustand sprechen, in der Wassermoleküle keine höhere Population als 50 % aufweisen, da man sonst diese Art von Wasser höchstwahrscheinlich detektieren könnte, wie in dieser Arbeit gezeigt. Ein leerer Raum, gleich einem Vakuum, würde dagegen den Gesetzen der Physik folgend kollabieren. Daher wird man zwangsläufig andere Methoden anwenden müssen, die sich besser eignen eventuelle ungeordnete Wassermoleküle zu detektieren. Dazu gehören das NMR und die Neutronenstreuung. Letzteres hätte aber ebenfalls mit denselben Problemen zu kämpfen wie die Röntgenstreuung. Bis dahin stehen zwei Konzepte sich gegenüber: die eine Hypothese geht von einem leeren hydrophoben Raum aus und die andere von Wassermolekülen, die über viele Positionen im Raum verschmiert sind.

Neben den Einblick in die hydrophobe Interaktion ist es gelungen Klarheit über den Ursprung der beobachteten Protonierung bei der Bindung von Phosphonamiden zu bringen. Alle bisher untersuchten Phophonamide zeigen eine Abhängigkeit der Enthalpie von dem untersuchten Puffersystem. Dies ist ein deutliches Zeichen für einen Bindungsprozess, der von einer Protonierungsreaktion überlagert ist. Die Untersuchung an verschiedenen Liganden hat schließlich ergeben, dass weder die NH-Gruppe des Phosphonamids noch die endständige Carboxylgruppe in die Protonierung involviert ist. Eine nähere Untersuchung der Aminosäuren in der gesamten Proteinbindetasche hat ergeben, dass Glu143 eine zentrale Rolle bei der Protonierungsreaktion spielt. In der Apoform ist dieses Glutamat deprotoniert und damit fähig mit dem Zink-Ion zu interagieren. Während der Bindung mit

den phosphonhaltigen Liganden wird das Glutamat vom Zink-Ion weggedrängt und gleichzeitig protoniert, um mit dem Sauerstoff des Phosphonamids eine Wasserstoffbrücke aufzubauen.

Der letzte Teil der Arbeit (Kapitel 7) befasst sich mit der Kooperativität zwischen der Wasserstoffbrücke zu Gly216 und hydrophoben Kontakten zur S3/4 Tasche von Thrombin. Unterschiedliche Substitutionen an der endständigen Aminogruppe, welche die Wasserstoffbrücke zu Gly216 aufbaut, zeigen tatsächlich deutliche Unterschiede im Ausmaß der Kooperativität. Diese kooperativen Effekte zeigen eine Korrelation mit den jeweiligen pK_a Werten der Substituenten, was darauf schließen lässt, dass die Stärke der Wasserstoffbrücke beeinflusst wird. Eine quantenchemische Abschätzung der Energie dieser Wasserstoffbrücke zeigt kein einheitliches Bild. Hier ist jedoch anzumerken, dass die Rechnungen einige Schwächen aufweisen. Die Abschätzung erfolgt auf einem reduzierten Gerüst und nur für das jeweilige Alanin-Derivat in P3. Überdies deuten die Kristallstrukturen aus der vierten Arbeit darauf hin, dass einige untersuchte Substitutionen eine zusätzliche Wasserstoffbrücke zum Protein ausbilden, die nicht in der MD-Simulation berücksichtigt wurde. Eine zuverlässige Aussage über den Ursprung dieses Effekts ist leider zu diesem Zeitpunkt nicht möglich. Aus jeder Serie ist eine Kristallstruktur nötig um im Anschluss eine zuverlässige Berechnung der Stärke der Wasserstoffbrücke durchzuführen. Diese Abschätzung könnte somit alle bisher nicht einbezogenen Interaktionen berücksichtigen.

9.2 English version

The presented thesis focuses on the crystallographic and thermodynamic analysis of the role of water molecules in protein-ligand binding.

Since most of the investigated ligands are found in an affinity range, which is too strong ($K_i \leq 1$ nM) or too weak ($K_i \geq 1$ µM) for direct ITC titration, this approach was not applicable for a thermodynamic characterization. Therefore, ITC displacement titrations were chosen to enable reliable thermodynamic analyses in these affinity ranges. The first part of this work (Chapter 2) is therefore devoted to the establishment and validation of the displacement titration for the characterization of low and high affinity ligands. It has been shown that the dissociation constants (K_d) of the displacement ITC correlate very well with kinetically determined inhibition constants (K_i). This displacement method can also be applied to

fragments, which are known for their very weak binding affinity, in order to record accurate binding constants. The kinetic determination of binding constants for fragments is due to their unexceptionally weak binding not applicable. The latter kinetic assays are not sensitive enough or no appropriate assay protocols have been reported to provide reliable values in micro- to millimolar range. Nevertheless, the correct estimation of the binding affinity is in the early stage of FBDD (fragment-based drug discovery) of utmost importance, since only then a reliable decision can be made which fragments should be followed-up for further optimization. In such cases, the displacement ITC represents a suitable alternative. In addition to the binding affinity, the displacement ITC is also able to determine the enthalpic properties with high reliability. This value also represents an important indicator for the decision which fragments should be further optimized. Furthermore, it is also possible to reduce the protein concentration in an ITC displacement titration to save valuable material when working with ligands showing only tiny enthalpic signal. The often criticized protein consumption in ITC can be reduced at least threefold. A protocol is given to estimate theoretically, most appropriate concentration of the low-affinity ligand to achieve a certain degree of inhibition. The equation to estimate this concentration has been implemented into an Excel spreadsheet, to easily calculate the degree of inhibition as a function of the expected affinity and protein concentration for a given ligand concentration.

The following study (Chapter 3) examined the positively charged Asp189 at the bottom of the S1 pocket of thrombin. It is well-hydrated in the apo structure of thrombin. The aim was now to investigate the impact of a change of the Asp189 solvation state by a series of pyridine-based ligands on the binding affinity and the thermodynamic profile. The thermodynamic study of these micromolar pyridine-based inhibitors is methodically realized by applying the displacement ITC presented in chapter 2. This series provides a deeper understanding of the consequences of desolvation effects performed on the protein as well as ligand side. The apo form of thrombin shows an ordered and tightly bound water molecule next to the negatively charged Asp189 in the S1 pocket. The effects observed for the perturbance of the solvation shell of the S1 pocket by pyridine and methylpyridinium analog inhibitors clearly explain why only micromolar activity is accomplished. In all complex structures with pyridine, it is evident that preserving the original solvation state of Asp189 is a crucial and a common feature upon binding of the pyridine inhibitors. However, the associated entropic losses are immense. In two ligand complexes even disordered ligand

portions are found in the S1 pocket, which evade full desolvation of Asp189 compared with the apo form of thrombin. The price for the desolvation of a charged amino acid is simply too large to ensure in this case a complete displacement of all waters. The result is a compromise, where 50 % of the water is displaced and 50 % remains at Asp189. As a consequence, too high desolvation can lead to disordered systems just in order to keep the price of desolvation as low as possible. It has also been found that almost all positively charged methylpyridinium derivatives are weaker than the neutral pyridine analogues. The quantum-chemical calculation of the solvation of the ligands shows significant differences between the neutral and charged derivatives. The charged compounds require an order of magnitude more energy for desolvation than the neutral ones. The determined complex structures reveal that the charged methylpyridinium derivatives do not optimally address the negatively charged Asp189 at the bottom of the S1 pocket. A short distance to the deprotonated Asp189 cannot be achieved either due to steric reasons or the bulky methyl group provides a good protection to interact in a proper way with the negatively charged Asp189. The optimal interaction geometry to Asp189 as it is found in the benzamidine complexes cannot be realized in this series. Therefore, the energy released from the suboptimal interaction between methylpyridinium and Asp189 is not high enough to compensate for the large desolvation price required for the charged ligands. The inhibitors lose binding affinity as a result. The closer the positive charge of the methylpyridinium is placed to Asp189, the lower the enthalpic penalty turned out. Interestingly, the pyridine derivative with nitrogen in para position remains deprotonated and is capable of transferring the ordered solvation shell of the S1 pocket into a largely disordered state. The comparison to the apo structure and the analog methyl-pyridinium reveals a favorable entropic term, which is due to the disordered water shell in the complex structure with the *p*-pyridine derivative.

The fourth, fifth and sixth part of this thesis deals with water effects in hydrophobic interactions. A previous study on the binding of hydrophobic ligands to the hydrophobic S_1' pocket of thermolysin shows surprisingly an enthalpy-driven hydrophobic effect. This is likely attributed to the insufficient desolvation of the S_1' pocket as evidenced by a related work of Homans *et al.* Based on the hypothesis that the binding signature of a hydrophobic interaction is dependent on the solvation of the investigated binding pocket, two hydrophobically modified series were designed: One for the S3/4 pocket of thrombin and

the other for the S_2' pocket of thermolysin. Both addressed binding sites are hydrophobic and marked by pronounced solvation but exhibiting different shape.

The S3/4 binding pocket is shaped like a crescent, which is easily accessible to water molecules. Two series with different S1 anchor were hydrophobically modified in P3 position. The thermodynamic characterization shows a hydrophobic effect which is clearly entropically driven. In the study, the S3/4 pocket has been gradually desolvated using increasing hydrophobic modifications in P3. In both series, the binding affinity improved by about 40-fold. The binding affinity has been optimized hydrophobically from nanomolar to low picomolar affinity. The benzamidine derivatives are even characterized by a binding mode showing two ligands to be bound simultaneously. The second bound ligand is only populated to 50 % but the difference electron density is well defined. Surprisingly, the additionally bound ligand traces remarkable well the recognition area that accommodates fibrinopeptide A, which is a cleavage product of the natural thrombin substrate (fibrinogen).

Despite the substantiated criticism with respect to the refinement of partially populated water molecules in protein structures, also in the fourth study water molecules have been discovered which indicate no full occupancy. Exhaustive desolvation could not be achieved, since the enthalpic price for full water repulsion from these protein binding pockets cannot sufficiently be compensated by the new interactions formed between protein and ligand.

The examined S_2' pocket of thermolysin in the fifth and sixth study is less well shaped, but ideally solvated because of its exposure to the protein surface.

The ligands differ only by a terminal carboxylate and/or methyl group in the fifth study. A surprising nonadditivity of functional group contributions for the carboxylate and/or methyl groups is detected. Adding first the methyl and then the carboxylate group results in a small Gibbs free energy increase and minor enthalpy/entropy partitioning for the first modification, whereas the second involves strong affinity increase combined with huge enthalpy/entropy changes. Adding however first the carboxylate and then the methyl group yields reverse effects: now the acidic group attachment causes minor effects whereas the added methyl group provokes huge changes. The added COO^- groups perturb the local water network in both carboxylated complexes and the attached methyl groups provide favorable interaction sites for water molecules. In all complexes, apart one example, a contiguously connected water network between protein and ligand functional groups is observed. In the complex with the carboxylated ligand, still lacking the terminal methyl group, the water

network is unfavorably ruptured. This results in the surprising thermodynamic signature showing only minor affinity increase upon COO- group attachment. Since the further added methyl group provides a favorable interaction site for water, the network can be re-established and strong affinity increase with huge enthalpy/entropy signature is then detected.

In the sixth study several hydrophobic S_2' occupants were investigated. Surprisingly, it has been found that small residues are clearly preferred. Large and bulky groups on the ligand side induce unfavorable conformations of protein residues. A benzyl group, for example, has resulted in the displacement of all water molecules, but this entropic gain is partially compensated by the loss of planarity of an adjacent peptide bond. Thus, large ligand portions are associated with a significant loss in binding affinity due to conformational changes on the protein side.

Addressing the S_2' pocket of thermolysin generates also an entropically dominated signal similarly to the series in the fourth chapter.

The present series of closely related complexes shows that both thermodynamic properties are involved and many detailed structural phenomena determine the final signature. If a contiguously connected water network ruptures, an enthalpic loss and entropic gain is experienced. Particularly, in case of accommodation of ligand portions in pockets opening to the bulk solvent and exposing parts of the placed ligand to the water phase, new binding sites for water molecules can be generated, e.g. as in our study at the capping position above the carboxylate group or the site on top of the benzyl ring. Also such phenomena contribute on the molecular level to the finally determined hydrophobic effect. In summary, there are no arguments why the hydrophobic effect should be predominantly “entropic” or “enthalpic”; small structural changes on the molecular level determine whether hydrophobic binding to hydrophobic pockets results in a more enthalpy or entropy-driven signature.

Additionally, it has been shown for the third time that the binding profile of hydrophobic ligands also strongly depends on the presence of a structurally well-defined solvation in the hydrophobic pocket of the apo form. On the other hand, the form and location of the binding pocket determines the properties of its solvation shell. The S_1' pocket, for example, has a clearly defined shape in the form of a truncated cone. So far, the actual solvation pattern of the apo structure could not be described as the uncomplexed state is hardly accessible due to the presence of the digestion product Val-Lys in the active site. All studies

in this work and the results from the literature indicate, however, that no crystallographic water molecules in the S_1' binding pocket can be located. An on-going project tries to unravel the actual S_1' solvation state by creating a thermolysin mutant, which cannot digest itself because it lacks the crucial valine residue at the C-terminus. The crystal structure of this mutant will reveal whether there are localized water molecules detectable or not. If this thermolysin mutant also does not indicate any S_1' waters, there are still some uncertainty in the water question since X-ray crystallography is here not able to differentiate whether the water molecules are really missing or just in a highly mobile state.

It must be emphasized, that some water positions have been found although obvious disorder (total of 6 structures, 50 % occupancy) is indicated, provided that the structures are well enough resolved. To some extent this observation would speak for a S_1' solvation state, where the water molecules are populated to less than 50 % occupancy since beyond this threshold water seems rather easily accessible by X-ray crystallography as shown in this work. In contrast, space without any water molecules (basically a vacuum) is likely to collapse according to the laws of physics. Hence, it will be inevitable to use other methods that are better suited to detect highly mobile water molecules. These methods include NMR and neutron scattering. The latter would have probably the same problems as the experiments with X-ray scattering because only coherent scattering will lead to a diffraction pattern. So far, there are two concepts how to explain the so-called non-classical hydrophobic effect: one hypothesis is based on an empty hydrophobic space and the other by water molecules, which are scattered over many positions in space.

In addition to insights into the hydrophobic interaction, an unresolved question about the origin of the observed protonation effect in case of phosphonamidate binding to thermolysin could be clarified. All previously studied phophonamidates show a dependence of the binding enthalpy on the applied buffer system. This is a clear indication for a binding process, which is superimposed by a protonation reaction. The study of various ligands has finally revealed that neither the NH group of the phosphonamidate nor the ligand's terminal carboxyl group are involved in the pick-up of a proton. A closer examination of the amino acids throughout the protein binding site has shown that Glu143 plays a central role in the protonation reaction. This glutamate is deprotonated in the apo form and thus able to interact with the zinc ion. Glu143 is pushed away upon phosphonamidate binding and

simultaneously becomes protonated to build an H-bond with the oxygen of the phosphonamidate.

The last part of the work deals with the cooperativity between a hydrogen bond to Gly216 and hydrophobic contacts formed in S3/4 pocket of thrombin. Different substitutions at the terminal amino group, which builds up the hydrogen bond to Gly216, exhibit significant differences in the extent of cooperativity. These cooperative effects correlate well with the respective pK_a values of the substituents, which suggest that the strength of the hydrogen bond is affected. A quantum-chemical estimation of the energy of this hydrogen bond reveals a mixed picture. However, it has to be noted that some limitations have to be regarded while interpreting these values. The calculations are performed only on a reduced scaffold and are limited to derivatives showing an alanine in P3. Moreover, the crystal structures from the fourth study suggest that for some of the considered substitutions the formation of an additional hydrogen bond to the protein has to be taken into account which is not regarded in the calculations. A consistent interpretation of the origin of the observed cooperative effect is unfortunately not yet possible. Crystal structures are necessary in each series to perform a reliable calculation of the connection strength of the hydrogen bond. A new estimation by MD simulations could then take into account all interactions previously not included.

10 Anhang/Annex

10.1 Abbreviations

ABP	L-arabinose binding protein
ACB	2-(aminomethyl)-5-chlorobenzylamide
ACE	Angiotension-converting enzyme
Aces	N-(2-acetamido)-2-aminoethanesulfonic acid
AMBA	4-amidinobenzylamide
Cbz	Carboxybenzyl
Cha	Cyclohexylalanine
CSD	Cambridge structural database
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
FBDD	Fragment-based drug discovery
FBLD	Fragment-based lead discovery
FPA	Fibrinopeptide A
Hepes	2-(4-(2-Hydroxyethyl)- 1-piperazinyl)-ethansulfonsäure
HIV	Human immunodeficiency virus
HTS	High-throughput screening
ITC	Isothermal titration calorimetry
K_a	Assoziationskonstante
K_d	Dissoziationskonstante
K_i	Inhibitionskonstante
MD	Molecular dynamics
MR	Molecular replacement
MUP	Major urinary protein
MW	Molecular weight
NMR	Nuclear magnetic resonance
PDB	Protein data bank
PEG	Polyethylene glycol
SAD	Single-wavelength anomalous dispersion
SPR	Surface plasmon resonance
THR	Thrombin
TLN	Thermolysin
TLS	Translation/libration/screw
Tricine	N-(2-hydroxy-1,1-bis(hydroxymethyl)ethyl)glycine
Tris	Tris(hydroxymethyl)-aminomethan

10.2 Publikationen/Publications

Biela A., Park A., Groh M. & Klebe G. (2012) Validation of ITC displacement titration and implications to thermodynamic fragment characterization. Ready for Submission.

Biela A., Nasief N., Heine A., Hangauer D. & Klebe G. (2012) Water displacement from a flat and wet binding pocket: Molecular details determining the hydrophobic effect. Ready for submission.

Biela A., Sielaff F., Steinmetzer T. & Klebe G. (2012) Enthalpic and entropic changes caused by a stepwise disruption of a water network in the S3/4 subsite of thrombin: An example of a classical hydrophobic effect. Ready for submission.

Biela A., Khyat M., Tan H., Kong J., Heine A., Hangauer D. & Klebe, G. (2012) Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin. Submitted to JMB.

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 S_1'-pocket of thermolysin. *Biochim Biophys Acta.* **1800**, 1192-1202.

Fuchs J.E., Spitzer G.M., Javed A., **Biela A.**, Kreutz C., Wellenzohn B. & Liedl K.R. (2011) Minor Groove Binders and Drugs Targeting Proteins Cover Complementary Regions in Chemical Shape Space. *J Chem Inf Model.* **51**, 2223-2232.

10.3 Beiträge auf Konferenzen/Conference contributions

Biela A., Englert L., Zayed M., Hangauer D. & Klebe G. (2008) Thermodynamic and crystallographic analysis of closely related Thermolysin inhibitors. Poster auf der Konferenz *"Applications of BioCalorimetry (abc6)"* in Heidelberg

Biela A., Englert L., Zayed M., Hangauer D. & Klebe G. (2009) Thermodynamic and crystallographic analysis of closely related Thermolysin inhibitors. Poster auf der Konferenz *"Frontiers in Medicinal Chemistry"* in Heidelberg

Biela A., Nasief N., Englert L., Hangauer D. & Klebe G. (2010) Two Hydrophobic Interactions Associated With Different Thermodynamic Profiles In A Single Binding Event. Poster auf der Konferenz *"Developments in Protein Interaction Analysis (DiPIA)"* in Barcelona

Biela A. (2010) Two Hydrophobic Interactions Associated With Different Thermodynamic Profiles In A Single Binding Event. Vortrag auf der Konferenz *"ITC User Day"* in Hamburg

10.4 Lebenslauf/Curriculum Vitae

Persönliche Details

Name	**Adam Jozef Biela**
Beruf	**Apotheker, Diplom-Pharmazeut**
Geburtsdatum	**15.07.1981**
Geburtsort	**Groß Strehlitz**
Staatsangehörigkeit	**deutsch**

Akademischer und schulischer Werdegang

Seit 01/2008	**Philipps-Universität Marburg** **Promotion am Institut für Pharmazeutische Chemie** Arbeitskreis Prof. Dr. Klebe
11/2007 – 12/2007	**Approbation** zum Apotheker
05/2007 – 10/2007	**Industrie-Apotheke, Lollar** Teil des Praktischen Jahres
10/2006 – 04/2007	**Boehringer Ingelheim Pharma GmbH, Biberach an der Riß** **Diplomarbeit in der Abteilung für ZNS Forschung** Arbeitskreis PD Dr. Ferger *„Neurochemische und verhaltenspharmakologische Untersuchung der Wirkung von Proteasom-Inhibitoren auf das dopaminerge Nervensystem der Maus“*
03/2006 – 04/2006	**Max-Planck-Institut für Biophys. Chemie, Göttingen** **Praktikum, NMR basierte Strukturbiologie** Arbeitskreis PD Dr. Reinscheid
10/2002 – 09/2006	**Philipps-Universität Marburg** **Studium der Pharmazie**
11/2001 – 07/2002	**Wehrdienst,** Sondershausen und Stadtallendorf
08/1998 – 07/2001	**Wilhelm-von-Oranien-Schule, Dillenburg** **Abitur**

10.5 Acknowledgments

Diese Arbeit konnte nur mit Hilfe und Unterstützung zahlreicher Menschen entstehen. Besonders danken möchte ich:

Herrn **Prof. Dr. Gerhard Klebe** für seine kontinuierliche Unterstützung und Förderung, insbesondere für das Ermöglichen der Teilnahme an internationalen Tagungen sowie das sorgfältige Korrekturlesen besonders meiner äußerst anstrengenden englischen Werke. Seine stete Gesprächs- und Diskussionsbereitschaft, seine Zuversicht und die gewährten Freiheiten haben eine sehr konstruktive und freundschaftliche Arbeitsatmosphäre ermöglicht. Ein ganz großer Dank außerdem für sein Engagement bei der Entwicklung der Wassernetzwerk-Theorie, die maßgeblich auf Prof. Dr. Klebe zurückgeht.

Herrn **Prof. Dr. Steinmetzer** für seine Geduld mich in die teils sehr dunklen Geheimnisse der Reaktionskinetik einzuführen, für die ausgezeichnete Unterstützung bei der Synthese der untersuchten Thrombininhibitoren und die Bereitschaft diese Arbeit als Zweitgutachter zu bewerten.

Herrn **Prof. Dr. David Hangauer** und seinen Doktoranden **Maan Khayat** und **Nader Nasief** für die Synthese der nunmehr fast 300 Inhibitoren. Insbesondere **Maan Khayat** für die sehr freundschaftliche Zusammenarbeit und die vielen Skype-Konferenzen, in denen sehr ausgiebig über alle Facetten der Thrombin-Inhibition wie auch über die arabische Kultur diskutiert wurde. **Nader Nasief** für die teils sehr emotional geführten Diskussionen per email, die nicht selten in Streit ausgeartet sind. Trotz der vielen Meinungsverschiedenheiten war auch er ein stets verlässlicher Kooperationspartner, der wesentlich zu dieser Arbeit beigetragen hat.

Herrn **Dr. Andreas Heine** für seine ausgezeichnete Einführung in die Kristallographie, für das Korrekturlesen dieser Arbeit und für seine ausdauernde Bereitschaft bei allen möglichen kristallographischen Problemen zu helfen. Ein ganz großer Dank gilt ihm außerdem wegen seiner Bereitschaft an der Umstrukturierung des Thermolysin-Kapitels mitzuwirken. Auch dass er sehr viel Zeit investiert hat, um einige der Thermolysin-Strukturen nochmals zu überprüfen.

Herrn **Michael Betz** ebenfalls für seine Hilfe bei der Entwicklung der Wassernetzwerk-Theorie, die viel Zeit in Anspruch nahm. Auch für sein Engagement bei der Erstellung der Bilder zu den Wassernetzwerken, die er mir freundlicherweise für diese Arbeit zur Verfügung gestellt hat.

Herrn **Dr. Frank Sielaff** für die Synthese einiger Thrombininhibitoren.
Herrn **Felix Terwesten** für seine Bereitschaft die MD-Simulationen zu einigen Thrombin-inhibitoren durchzuführen.
Herrn **Christian Sohn** für seine Hilfe bei allen röntgenkristallographischen Arbeiten und vor allem für seine Unterstützung bei der kinetischen Untersuchung der Thrombininhibitoren.

Den Praktikanten **Serghei Zaychyk** und **Matthias Groh** und dem Bachlorstudenten **Andre Wuttke** für die angenehme Arbeitsatmosphäre und die Entlastung bei einigen harten Projekten.

Herrn **Manuel Neeb** für die interessanten Diskussionen über das Eigenleben des ITC. Auch war er stets ein verlässlicher und treuer Mensagänger.

Dem Personal des **BESSY** und des **SLS** für die freundliche und stets hilfreiche Unterstützung bei den kristallographischen Experimenten.

10. Anhang

Dem **BMBF** für die finanzielle Förderung.

Frau **Dr. Lisa Englert** und Herrn **Dr. Bernhard Baum** für die umfangreiche und geduldige Einführung in die verschiedenen Projekte.

Meinen Nachfolgern **Eggert Rühmann** und **Stephan Krimmer** für das Korrigieren dieser Arbeit.

Der gesamten **AG Klebe** für die tolle Arbeitsatmosphäre. Insbesondere **Barbara Wienen**, **Stephan Jakobi**, **Helene Krüger**, **Martin Stieler**, **Felix Gut** und **Ina Lindemann**.

Herrn **Dr. Oleg Schmidt** für die vielen Diskussionen über das harte Leben in der Forschung und die vielen Gespräche über Gott und die Welt aus der Sicht eines Apothekers und eines Physikers.

Meiner Familie, insbesondere meinen Eltern **Maria** und **Joachim Biela** für die jahrelange Unterstützung, für die Gewissheit, den Glauben an mich und die aufbauenden Gespräche. Natürlich geht auch an Dank an meine Großmutter und meine Brüder **Marius** und **Marcel,** die während dieser Zeit für die nötige Aufmunterung und Ablenkung sorgten.

Abschließend möchte ich mich besonders bei **Inna Schmidt** für ihr Verständnis und ihre unerschöpfliche Motivation bedanken. Dafür dass sie mir stets zur Seite stand und mich in jeder Hinsicht unterstützt hat, gebührt ihr ein Dankeschön vom tiefsten Herzen.

ERKLÄRUNG

Ich versichere, daß ich meine Dissertation

„*The Role of Water in Protein-Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry*"

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

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.....

.....