Optimization of the Drug Design Process by Ligand Pre-organization and Occupancy of Transient Binding Pockets Based on the Model Proteins Thrombin and Aldose Reductase

Dissertation

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

dem
Fachbereich Pharmazie der
Philipps-Universität Marburg
vorgelegt

von

Anna Sophie Sandner

aus Aßlar

Marburg/Lahn 2022

Erstgutachter	Prof. Gerhard Klebe Institut für Pharmazeutische Chemie Philipps-Universität Marburg
Zweitgutachter	Prof. Torsten Steinmetzer Institut für Pharmazeutische Chemie Philipps-Universität Marburg

Eingereicht am 01.03.2022

Tag der Mündlichen Prüfung am 12.04.2022

Hochschulkennziffer: 1108

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn PROF. GERHARD KLEBE am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Januar 2016 bis August 2020 durchgeführt

Für meine Eltern.

„Sich zu mühen und mit dem Widerstande zu kämpfen ist dem Menschen Bedürfniß, wie dem Maulwurf das Graben. Der Stillstand, den die Allgenugsamkeit eines bleibenden Genusses herbeiführte, wäre ihm unerträglich. Hindernisse überwinden ist der Vollgenuß seines Daseyns; sie mögen materieller Art seyn, wie beim Handeln und Treiben, oder geistiger Art, wie beim Lernen und Forschen: der Kampf mit ihnen und der Sieg beglückt.“

A. SCHOPENHAUER, Aphorismen zur Lebensweisheit,1851

Summary

The design and development of a drug is complex and usually extends over many years. Different approaches can facilitate lead optimization. This work focuses on specific aspects that could lead to advances in drug discovery.

After an introduction to drug design and the two studied target proteins, the first section of this thesis (Chapter 2) deals with the analysis and understanding of the function of transient binding pockets and the resulting perspective to improve the optimization of a lead structure. For this project, the already well-studied specificity pocket of human aldose reductase (ALR-2) was chosen. This enzyme is considered a key player of the polyol pathway. When glycolysis, an important degradation pathway of energy metabolism, is saturated, e.g., due to increased glucose levels in diabetics, the mechanism of the polyol pathway is activated. As a result, ALR-2 is stimulated to convert D-glucose to D-sorbitol. A series of inhibitors with low steric demand and functional groups of different electronic nature at the terminal aromatic moiety or terminal substituent, as well as their structural and thermodynamic properties, were studied in detail by isothermal titration calorimetry (ITC) and X-ray crystallography. In addition, the electrostatic potentials and charge distribution at the terminal aromatic groups of the inhibitors and their effects on binding to the transient pocket were analyzed by electronic surface potential (ESP) calculations. These analyses confirmed the previously established theory that terminal aromatic systems of the inhibitors with an electron-deficient aromatic group can trigger the opening of the specificity pocket and lead to preferential π-interaction with the electron-rich indole moiety of W111. The results also indicate that a certain volume of inhibitor seems to be a prerequisite for pocket opening, since too small substituents lead to more complex binding positions with increased residual mobility. It was further shown that a shift in pH between pH 5 and 8 has no effect on the binding position of the inhibitors in the crystal with respect to the opening of the specificity pocket. This allows a comparison between thermodynamic and crystallographic data obtained at different pH values.

The following chapter of this thesis (Chapter 3) focuses on the strategy of pre-organization used in late-stage drug design for ligand optimization without having to resort to a completely different scaffold. Here, the S_1 pocket of the also well-studied serine protease thrombin, whose conversion of fibrinogen to fibrin plays an important role in blood clotting, served as a model. In this work, we first present a series of congeneric thrombin ligands with a variety of functional groups that trigger preorganization prior to binding. The resulting structural

fixation of a ligand in its bioactive conformation, either by fixation of the bound conformation in a suitable ring system or by fixation via intramolecular hydrogen bonds (H-bonds), has positive effects on affinity for entropic reasons. Fixation in solution and complex formation were characterized by crystallography, ITC, and molecular dynamics (MD) simulations. Clearly, these preorganizational modifications do not affect the overall binding mode, provided that the required bound conformation of the molecule with the archetypal binding mode is well reproduced by the modified molecules in the bound state. Thus, the most important interactions are preserved. However, the results of this study also show that the presence of an additional intramolecular contact, preferably in the form of a dominantly populated conformer, is equally important to achieve the expected affinity enhancement. At worst, the modified scaffold in solution adopts a new, strongly preferred conformation, and the preorganization effect, expected to enhance the affinity for the target, is lost. Based on these results, it can be shown that the thermodynamics of preorganization are largely governed by enthalpy rather than entropy. Furthermore, in this work, an important salt bridge is shielded by actively reducing its surface exposure, resulting in an improved enthalpic binding profile.

Part three of this thesis (Chapter 4) discusses the selectivity-determining features in the S_1 pockets of the serine protease thrombin, the previously described blood clotting factor, and the related enzyme trypsin, which is responsible for digestion. Human trypsin is a proteolytic enzyme produced by the duodenum whose function is to degrade larger proteins into smaller components in the intestine. Thus, it represents one of the most important digestive enzymes. In these studies, a series of ligands were evaluated for their selectivity toward both peptidases using X-ray and neutron crystallography along with ITC. The results revealed that the local geometry of the two S_1 pockets is highly conserved. However, thrombin cleaves peptide chains only after arginine, whereas trypsin cleaves after lysine and arginine. Thrombin has a Na^+ binding site near D189 which is not present in trypsin. This suggests that simple steric features cannot explain the selectivity difference. Although the E192Q exchange at the edge of the S_1 pocket has very little effect on the steric and dynamic properties of ligand binding and the local geometry of the two serine proteases, the different partitioning into enthalpy and entropy contributions is a clear indication of a given difference. The analyses of this study suggest that E192, together with the thrombin-specific sodium ion, contribute to generate an electrostatic gradient in the S_1 pocket. This feature is not present in trypsin and is therefore important, together with the differences in solvation patterns in the S_1 pocket, for the selectivity of both enzymes. The observation of protonation effects induced in this context is the first evidence for significant charge attenuation at the carboxylate group of D189 in thrombin compared to trypsin. From these studies, it appears that this phenomenon is one of the most important selectivity-determining features between the two proteases and thus influences and controls the other selectivity-distinguishing features that are less obvious at first glance, such as the differences in the solvation pattern and in the arrangement of

water molecules in the two enzymes. In summary, it can be deduced from this work that it is not straightforward to assign the selectivity-determining features in the S_1 pockets of thrombin and trypsin to a single dominant factor.

The last part of the thesis (Chapter 5) is an abbreviated subproject and deals with eight thrombin inhibitors that differ in the P_1 residue. Due to the lack of time, only the crystal structures obtained by X-ray crystallography were used to analyze the binding behavior of the inhibitors. Their analysis shows that a P_1 thiophene changes its binding behavior in the S_1 pocket of thrombin once a halogen substituent is attached. The comparably electron-rich sulfur of a pure thiophene residue turns toward the inhibitor's own P_2 and P_3 carbonyl groups. In contrast, the sulfur of a P_1-thiophene, which is rather electron-poor due to the halogen substituents, can serve as an H-bond acceptor and thus interact water-mediated with the π-system of Y228. However, a halogen atom in a P_1-furan is not capable of recruiting such a water. Due to the shallower angle of oxygen compared to sulfur and the resulting change in geometry, the distance between the halogen and the ring of Y228 is optimized for a direct halogen-π-interaction. While a methoxy substituent in the *para*-position of a P_1 6-ring does not exhibit a particularly strong interaction, the same substituent in the *metha*-position can also interact directly with Y228. However, compared to other P_1 6-rings, this ring occupies a slightly different position. A hydroxy group in the *ortho*-position forms a water-mediated interaction with the π-system of Y228. Due to lack of time and the late takeover of data analysis by a former PhD student, additional thermodynamic and kinetic measurements of the binding process were not possible in the present work. Therefore, no definitive conclusions on the affinity and selectivity of thrombin can be derived based on this chapter.

Zusammenfassung

Die Entwicklung von Arzneimitteln ist komplex und erstreckt sich in der Regel über viele Jahre. Verschiedene Ansätze können die Optimierung von Leitstrukturen erleichtern. Diese Arbeit konzentriert sich auf bestimmte Aspekte, die zu Fortschritten in der Arzneimittelforschung führen könnten.

Nach einer Einleitung in das Thema Wirkstoffdesign und die Vorstellung der beiden Zielproteine, befasst sich der erste Abschnitt dieser Arbeit (Chapter 2) mit der Analyse und dem Verständnis der Funktion von transienten Bindetaschen und der daraus resultierenden Perspektive zur Verbesserung der Optimierung einer Leitstruktur. Für dieses Projekt wurde die bereits gut untersuchte Spezifitätstasche der menschlichen Aldose Reduktase (ALR-2) gewählt. ALR-2 gilt als Schlüsselenzym des Polyolweges. Ist die Glykolyse, ein zentraler Abbauweg des Energiestoffwechsels, beispielsweise durch erhöhte Glukosespiegel bei Diabetikern überlastet, wird der Mechanismus des Polyolweges in Gang gesetzt. Auf diese Weise wird ALR-2 dazu angeregt, D-Glukose in D-Sorbit umzuwandeln. Zunächst wurde eine Reihe von Inhibitoren mit geringen sterischen Anforderungen und funktionellen Gruppen unterschiedlicher elektronischer Natur am terminalen aromatischen Teil oder terminale Substituenten sowie ihre strukturellen und thermodynamischen Eigenschaften mittels isothermaler Titrationskalorimetrie (ITC) und Röntgenkristallographie eingehend untersucht. Darüber hinaus erfolgte eine Analyse der elektrostatischen Potenziale und der Ladungsverteilung an den terminalen aromatischen Gruppen der Inhibitoren und deren Auswirkungen auf die Bindung in die transiente Tasche durch Berechnungen des elektronischen Oberflächenpotenzials (ESP). Diese Analysen bestätigten die zuvor aufgestellte Theorie, dass terminale aromatische Systeme der Inhibitoren mit einer elektronenarmen aromatischen Gruppe eine Öffnung der Spezifitätstasche auslösen und zu einer bevorzugten π-Wechselwirkungen mit der elektronenreichen Indoleinheit der Aminosäure W111 führen können. Außerdem deuten die Ergebnisse darauf hin, dass ein bestimmtes Volumen des Inhibitors eine Voraussetzung für die Taschenöffnung zu sein scheint, da zu kleine Substituenten zu komplexeren Bindungspositionen mit erhöhter Restmobilität führen. Es hat sich weiterhin erwiesen, dass eine Verschiebung des pH-Werts zwischen pH 5 und 8 keine Auswirkungen auf die Bindungsposition der Inhibitoren im Kristall in Bezug auf die Öffnung der Spezifitätstasche hat. Dies ermöglicht einen Vergleich zwischen thermodynamischen und kristallographischen Daten, die bei verschiedenen pH-Werten erhoben wurden.

Das folgende Kapitel dieser Arbeit (Chapter 3) befasst sich mit der Strategie der Vororganisation, die in der späten Phase des Wirkstoffdesigns zur Liganden Optimierung eingesetzt wird, ohne auf ein komplett anderes Gerüst zurückgreifen zu müssen. In diesem Fall diente die S_1 Tasche der ebenfalls gut untersuchten Serinprotease Thrombin, die durch die Umwandlung von Fibrinogen in Fibrin eine wichtige Rolle bei der Blutgerinnung spielt, als Modell. In dieser Arbeit wird zunächst eine Reihe kongenerischer Thrombin Liganden mit einer Vielzahl funktioneller Gruppen vorgestellt, die vor der Bindung eine Vororganisation auslösen. Die daraus resultierende strukturelle Fixierung eines Liganden in seiner bioaktiven Konformation, entweder durch Fixierung der gebundenen Konformation in einem geeigneten Ringsystem oder durch Fixierung über intramolekulare Wasserstoffbrückenbindungen (H-Bindungen), hat aus entropischen Gründen positive Auswirkungen auf die Affinität. Die Fixierung in Lösung und die Komplexbildung wurden durch Kristallographie, ITC und Molekulardynamiksimulationen (MD) charakterisiert. Es wird deutlich, dass diese präorganisatorischen Modifikationen den gesamten Bindungsmodus nicht beeinträchtigen, sofern die erforderliche gebundene Konformation des Moleküls mit dem archetypischen Bindungsmodus durch die modifizierten Moleküle im gebundenen Zustand gut reproduziert wird. Die wichtigsten Wechselwirkungen bleiben somit erhalten. Die Ergebnisse dieser Studie verdeutlichen jedoch auch, dass das Vorhandensein eines zusätzlichen intramolekularen Kontakts, vorzugsweise in Form eines dominant besiedelten Konformers, ebenso wichtig ist, um die erwartete Affinitätsverbesserung zu erzielen. Im ungünstigsten Fall nimmt das modifizierte Gerüst in Lösung eine neue, stark bevorzugte Konformation an, und der Präorganisations-Effekt und damit die dadurch gewonnene Affinität für das Target geht verloren. Anhand dieser Ergebnisse lässt sich demonstrieren, dass die Thermodynamik der Präorganisation weitgehend von der Enthalpie und weniger von der Entropie dominiert wird. Darüber hinaus wird in dieser Arbeit eine wichtige Salzbrücke durch aktive Verringerung ihrer Oberflächenexposition abgeschirmt, was zu einem verbesserten enthalpischen Bindungsprofil führt.

Abschnitt drei der vorliegenden Arbeit (Chapter 4) befasst sich mit den selektivitätsbestimmenden Merkmalen in den S_1 Taschen des bereits beschriebenen Blutgerinnungsfaktor Thrombin und der für die Verdauung verantwortliche Serinprotease Trypsin. Das menschliche Trypsin ist ein proteolytisches Enzym, das vom Zwölffingerdarm produziert wird und dessen Aufgabe es ist, größere Eiweißverbindungen im Dünndarm in kleinere Bausteine zu zerlegen. Damit stellt es eines der wichtigsten Verdauungsenzyme dar. Im Rahmen dieser Studien wurde eine Reihe von Liganden auf ihre Selektivität bezüglich beider Peptidasen mittels Röntgen- und Neutronenkristallographie sowie ITC untersucht. Die Ergebnisse zeigen, dass die lokale Geometrie der beiden S_1 Taschen in hohem Maße konserviert ist. Allerdings spaltet Thrombin Peptidketten nur nach Arginin, während Trypsin nach Lysin und Arginin spaltet. Thrombin verfügt über eine Na^+-Bindungsstelle in der Nähe von D189, die bei Trypsin nicht vorhanden ist. Daraus ergibt sich die Schlussfolgerung, dass einfache sterische Merkmale den

Selektivitätsunterschied nicht erklären können. Obwohl der E192Q-Austausch am Rand der S_1 Tasche nur einen sehr geringen Einfluss auf die sterischen und dynamischen Eigenschaften der Ligandenbindung und der lokalen Geometrie beider Serinproteasen hat, ist die unterschiedliche Aufteilung in Enthalpie- und Entropiebeiträge ein starker Hinweis auf einen gegebenen Unterschied. Die Analysen dieser Studie legen nahe, dass E192 zusammen mit dem Thrombin-spezifischen Na^+-Ion dazu beiträgt, einen elektrostatischen Gradienten in der S_1 Tasche zu erzeugen. Diese Eigenschaft ist in Trypsin definitiv nicht vorhanden und ist daher zusammen mit Unterschieden im Solvatationsmuster in der S_1 Tasche für die Selektivität beider Enzyme von Bedeutung. Die Beobachtung der in diesem Zusammenhang induzierten Protonierungseffekte stellt den ersten Hinweis auf eine signifikante Ladungsabschwächung an der Carboxylatgruppe von D189 in Thrombin im Vergleich zu Trypsin dar. Aus den Studien geht hervor, dass diese Eigenschaft eines der wesentlichen selektivitätsbestimmenden Merkmale zwischen den beiden Proteasen ist und somit die anderen, auf den ersten Blick weniger offensichtlichen Selektivitätsunterscheidungsmerkmale, wie die Unterschiede im Solvatationsmuster und der Anordnung der Wassermoleküle in den beiden Enzymen, beeinflusst und steuert. Schlussendlich geht aus dieser Arbeit hervor, dass es nicht ohne weiteres möglich ist, die selektivitätsbestimmenden Merkmale in den S_1 Taschen von Thrombin und Trypsin einem einzigen dominanten Faktor zuzuordnen.

Der letzte Abschnitt dieser Arbeit (Chapter 5) stellt ein verkürztes Teilprojekt dar und beschäftigt sich mit acht Thrombininhibitoren, die sich im P_1-Rest unterscheiden. Aus Zeitgründen wurden nur die mittels Röntgenkristallographie erhaltenen Kristallstrukturen zur Analyse des Bindungsverhaltens der Inhibitoren herangezogen. Ihre Analyse zeigt, dass ein P_1-Thiophen sein Bindungsverhalten in der S_1-Tasche von Thrombin ändert, sobald ein Halogensubstituent angefügt wird. Der vergleichsweise elektronenreiche Schwefel eines reinen Thiophen-Restes wendet sich den Inhibitoren eigenen Carbonylgruppen P_2 und P_3 zu. Im Gegensatz dazu kann der Schwefel eines P_1-Thiophens, der aufgrund der Halogensubstituenten eher elektronenarm ist, als H-Brückenakzeptor dienen und somit dem π-System von Y228 wasservermittelt interagieren. Ein Halogenatom in einem P_1-Furan ist jedoch nicht in der Lage, ein solches Wasser zu rekrutieren. Aufgrund des flacheren Winkels des Sauerstoffs im Vergleich zum Schwefel und der daraus resultierenden Änderung der Geometrie ist der Abstand zwischen dem Halogen und dem Ring von Y228 für eine direkte Halogen-π-Wechselwirkung optimiert. Während ein Methoxysubstituent in der *para*-Position eines P_1-6-Rings keine sonderlich starke Interaktion auszubilden scheint, ist derselbe Substituent in *metha*-Position ebenfalls in der Lage, direkt mit Y228 zu interagieren. Im Vergleich zu anderen P_1-6-Ringen nimmt dieser Ring damit jedoch eine etwas andere Position ein. Eine Hydroxygruppe in *ortho*-Position bildet eine wasservermittelte Wechselwirkung mit dem π-System von Y228. Aufgrund von Zeitmangel und der späten Übernahme der Datenanalyse durch einen ehemaligen Doktoranden in diesem Kapitel waren zusätzliche thermodynamische und kinetische Messungen des

Bindungsprozesses in der vorliegenden Arbeit nicht möglich. Daher können auf der Grundlage dieses Abschnitts keine endgültigen Schlussfolgerungen zur Affinität und Selektivität von Thrombin gezogen werden.

Contents

Publications Arising from this Thesis

Articles

Parts of this work have been published or prepared for submission to a scientific journal. Chapter 2, 3 and 4 are already published, while Chapter 5 has been prepared for possible publication at a later date. Therefore, each chapter is preceded by an introductory remark, which illustrates the contribution of the author of this work.

Schiebel J, Gaspari R, Sandner A, Ngo K, Gerber HD, Cavalli A, Ostermann A, Heine A & Klebe G, Charges Shift Protonation: Neutron Diffraction Reveals that Aniline and 2-Amino-pyridine Become Protonated Upon Binding to Trypsin, *Angew. Chem. Int. Ed.* **2017**, 56, 4887 – 4890, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Sandner A, Hüfner-Wulsdorf T, Heine A, Steinmetzer T & Klebe G, Strategies for late-stage optimization: Profiling the thermodynamics by preorganization and salt bridge shielding, *Journal of Medicinal Chemistry* **2019**, 62, 9753 – 9771, ACS Publications, Washington, USA.

Sandner A, Ngo K, Schiebel J, Marca Pizarroso AI, Schmidt L, Wienen-Schmidt B, Wenzel B, Steinmetzer T, Ostermann A, Heine A, & Klebe G, How a Fragment Draws Attention to Selectivity Discriminating Features between the Related Proteases Trypsin and Thrombin, *Journal of Medicinal Chemistry* **2021**, 64, 3, 1611 – 1625, ACS Publications, Washington, USA.

Sandner A, Ngo K, Sager CP, Scheer F, Daude M, Diederich WE, Heine A & Klebe G, Which Properties Allow Ligands to Open and Bind to the Transient Binding Pocket of Human Aldose Reductase?, *Biomolecules* **2021**, 11, 12, 1837, MDPI, Basel, Switzerland.

Talks

Sandner A, Thermodynamic and Crystallographic Characterization of Preorganized Thrombin Inhibitors, West Coast Structural Biology Workshop, Asilomar Conference Center, Pacific Groove, CA, USA, 19.03.2019

Sandner A, Preorganization of Inhibitors as a Boost for Binding Affinity, PMC Seminar Series, Universität Greifswald, Greifswald, Germany, 04.06.2021

Posters

Sandner A, Steinmetzer T, Heine A & Klebe G, Thermodynamic and Crystallographic Characterization of Preorganized Thrombin Inhibitors, 8th Joint BER II and BESSY II User Meeting, Berlin, Germany, 2016 – 2017.

Sandner A, Steinmetzer T, Heine A & Klebe G, Thermodynamic and Crystallographic Characterization of Preorganized Thrombin Inhibitors, West Coast Structural Biology Workshop, Asilomar Conference Center, Pacific Groove, CA, USA, 17.03. – 20.03.2019

Abbreviations

Å	Ångström (1 Å = 1 – 10 m)
ADMET	Absorption, distribution, metabolism, excretion, and toxicity
ALR-1	Aldehyde Reductase
ALR-2	Aldose Reductase
APS	Ammonium persulfate
AT	Antithrombin III
B-factor	Debye-Waller-Factor
Cryo-EM	Cryogenic Electron Microscopy
CSD	Cambridge Structural Database
d	Distance
Da	Dalton
DAHC	Diammonium hydrogen citrate
DMSO	Dimethyl sulfoxide
DP	Differential Power
DTT	Dithiothreitol
E. coli	Escherichia coli
EC	Enzyme Commission numbers
EDTA	Ethylene diamine tetra acetate
ESP	Electronic surface potential area
F	Factor
FBDD	Fragment based drug discovery
F_C	Calculated structure amplitudes
Fig.	Figure
F_O	Observed structure amplitudes
g	Gram
G°	Gibbs free energy
GIST	Grid inhomogeneous solvation theory
GoI	Glycerol molecule
GSH	Glutathione
GSSG	Oxidized Glutathione
H°	Enthalpy
H-bond	Hydrogen bond
HEPES	4-(2-Hydroxyethyl)-1-piperazine ethane sulfonic acid
HIT	Heparin-induced thrombocytopenic syndrome
HTS	High Throughput Screening
IC_{50}	Half-maximum inhibitory concentration

IPTG	Isopropyl-*β*-D-thiogalactopyranosid
ITC	Isothermal titration calorimetry
K	Kelvin
k	Kilo
K_a	Association constant
K_d	Dissociation constant
K_i	Inhibition constant
LMWH	Low molecular weight heparins
M	Molarity (mol/L)
MD	Molecular dynamics
min	Minute
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced
NAPAP	N^{α}-(*β*-naphthylsulphonyl-glycyl)-D,L-*p*-amidino-phenyl-alanyl-piperidine
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
occ.	Occupancy
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
pH	Potentialis hydrogenii
pK_a	Logarithmic acid dissociation constant
PMF	Potential of mean force
R	Ideal gas constant
R-factor	Reliability factor
RMSD	Root mean square determination
rpm	Revolution per minute
S°	Entropy
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SMILES	Simplified Molecular Input Line Entry System
T	Absolute temperature
TEMED	N,N,N′,N′-Tetramethylethan-1,2-diamin
TIM	Triose-phosphate isomerase
TLS	Translation, rotation, and screw-rotation
tPA	tissue Plasminogen Activator
TRIS	Tris(hydroxamethyl)aminomethane
UFH	Unfractionated heparin
vWF	von Willebrand Factor
WHAM	Weighted histogram analysis method
WT	Wild type
X-ray	Röntgen radiation

List of Figures

Chapter 1

General Introduction

1.1 Drug Design

In this thesis, different topics will be discussed that could lead to progress in drug design research. In the following, the current state of science from the early days of drug design up to the present day is briefly sketched in order to start further research from this point.

Over time, drug therapy has become more and more complex as the knowledge about the various causes of a given disease has increased. However, the success of drug research is steadily progressing. At the beginning of human history, Nature was used to treat diseases. Without any a priori knowledge based on observations and experience, various plants, or parts and extracts thereof were used. This procedure was mainly based on the trial of already existing recipes. Later it was possible to isolate active substances from plants, even if the discovery of active substances was still mainly due to coincidence. The best-known example of this is the discovery of penicillin by ALEXANDER FLEMMING in 1928, who observed a tainted staphylococcus culture and found that no bacteria seemed to grow around the fungus.[1] It was not until 1980 that classical rational drug research began to develop slowly, by being able to elucidate the causes of diseases and understand at the molecular level what led to them.[2] These are usually either endogenous proteins, which for various reasons trigger a pathogenic mechanism in the body, or exogenous proteins from bacteria or viruses. Once the target is identified, the search for a suitable lead structure that fits chemically and sterically into the binding pocket of the target protein is the starting point for the development of a drug. Even today, the search for active ingredients is largely based on Nature and the compounds often imitate the natural substrates of the target. Besides the use of natural substances, de novo structure-based inhibitor design is also performed to retrieve a putative lead compound. A first possible approach came up with high-throughput screening (HTS), in which entire libraries of up to 10^7 substances are tested for their activity in relation to the target protein. However, this method has some non-negligible disadvantages. Since the number of synthesizable substances has been estimated to about 10^{60}, only a very small amount of the possible compounds can be tested at all.[3,4] Furthermore, the hit rate for an appropriate lead structure, which is suitable for further, usually very elaborate optimization, is only about 1% which is associated with high costs.[5] Another promising approach is fragment-based drug discovery (FBDD). It has now emerged as an alternative to HTS.[6,7] It was established more than 30 years ago. In contrast to HTS, compounds exhibiting a maximum of 250 Da and containing less than 16 heavy atoms are considered in the first stage.[8] Another approach toward the design and selection of fragments is the rule of three. According to this rule, ideal fragments should have a molecular weight below 300 Da, a *c* log (P) below 3, and the number of hydrogen bond donors and acceptors should each be less than 3, as should the number of rotatable bonds.[9] These rules limit the number of synthesizable compounds drastically and thus limits the search space. Structures of lower complexity, such as fragments, yield higher hit rates in screenings than

more complex compounds, such as drug-sized molecules. A reason for this is that more complex molecules tend to mismatch with the receptor due to suboptimal interactions or clashes. Another option nowadays is computer-aided drug design for which X-ray crystal structure analysis or structures determined by NMR techniques are used as input. This allows to design a lead compound based on the three-dimensional structure of the protein using structure-based drug design. However, to design a drug, further optimization is required, as more than solely good binding of the lead to the target is necessary. Rather, it is important to adapt the compounds to the target in order to achieve a certain effect. Since proteins usually occur in families with many structurally related isoforms, a lead that has been discovered must be optimized, for example regarding its selectivity, to avoid undesired side effects. The so-called ADMET rules also play an important role in lead optimization. Based on these rules, a promising lead compound is optimized both in terms of absorption, i.e., solubility and oral bioavailability, and in terms of distribution in the body. Optimization is also necessary regarding human metabolism and excretion. Finally, the pharmaceutical substance to be developed still needs to be tested and optimized with respect to its toxicity.[10] Nevertheless, the design of a drug based solely on the structural properties of the compound and protein is not sufficient. Instead, other effects such as thermodynamic binding profile, water displacement, inhibitor preorganization and thus the restriction of the inhibitor's degrees of freedom need to be considered. For example, it is possible that a binding pocket changes its structure during the binding process, here a target-based strategy is clearly more suitable. More sophisticated methods such as X-ray crystallography and isothermal titration calorimetry (ITC) are used to provide a more detailed insight into these factors. After the lead structure has been optimized, it is tested for efficacy and safety in preclinical trials using animal experiments. The described design cycle for lead optimization (Figure 1), in which, among other aspects, the quantitative structure-activity relationships (QSAR) are repeatedly optimized, until a suitable compound is found. Finally, before a drug can be introduced to the market, it has to undergo clinical phase I − III studies after the development of investigational medicinal products, in which safety, tolerability and efficacy are tested through its use in humans.[2] The development of a drug accordingly takes 12 − 15 years on average.[11]

Different approaches can facilitate lead optimization. Focusing on designing a preorganized drug molecule from a lead, for example, can greatly improve the binding behavior and affinity. By focusing on addressing a transient binding pocket, the selectivity of the drug can be influenced. With investigations of transient binding pockets and optimization of the inhibitor binding behavior by preorganization, further progress in the research area of drug design, and thus in future drug development, can be achieved.

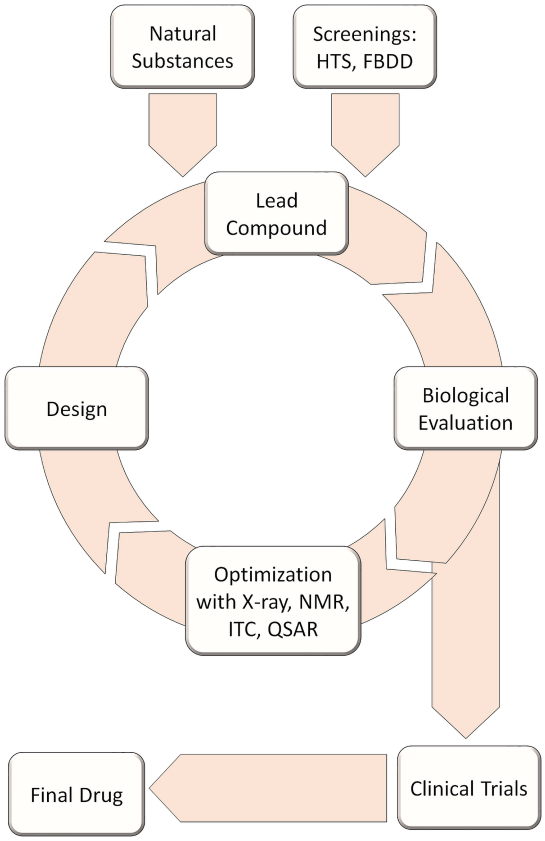


Figure 1: Drug design cycle.[2] The drug design consists of a varying number of cycles in which the initial lead is evaluated and continuously improved. Methods such as X-ray structure analysis, NMR, ITC, and computer-aided methods such as virtual screening are used.

1.2 Known Target Proteins as Milestones in Drug Development

1.2.1 Human Aldose Reductase

One of the focal points of this thesis is the analysis and understanding of the function of transient binding pockets, and the resulting perspective to improve optimization of a lead. To accelerate our investigations, we have chosen human aldose reductase (ALR-2) as target protein, since this enzyme is already well known and heavily studied. By understanding the functional role of transient binding pockets, it might be possible to better control selectivity with respect to other related enzymes and, in the case of aldose reductase, to reduce late-stage complications of diabetes. It is therefore important to know in advance the structure and function of the target protein ALR-2 and to summarize the complications caused by this protein in the late stages of diabetes.

1.2.1.1 Diabetes – a Global Disease

The target protein ALR-2 is largely responsible for the long-term complications of diabetes. In order to understand which effects an improved control of the transient binding pocket of this protein during lead optimization might have, a short introduction to diabetes will be given below.

Diabetes mellitus describes a group of metabolic disorders of the carbohydrate metabolism, whose common finding is an increase in the blood sugar level (hyperglycemia) caused by absolute or relative insulin deficiency. Severe hyperglycemia leads from classic symptoms such as polyuria, polydipsia, and visual disorders to ketoacidosis. Chronic hyperglycemia is associated with long-term damage and functional disorders of various tissues and organs (eyes, kidneys, nerves, heart and blood vessels).[12]

Today's standard of living in the first world is characterized by an oversupply of high caloric nutrition paralleled by reduced physical effort and activity. As a result, the society is steadily aging more and more, not only in economically well represented countries. In 2017, the worldwide prevalence of diabetes mellitus in adults between the ages of 20 and 79 was estimated at 425 million, mainly due to type 2 diabetes mellitus.[12] This number is expected to increase by 48% to 629 million by 2045, while social inequality in the development and detection of diabetes is increasing as well.[13]

Diabetes mellitus is classified into four types.[14] Insulin is a hormone that is essential for cells to use blood sugar and keep blood glucose levels within acceptable limits, (Figure 2). It is secreted by the pancreas in response to high glucose levels in the blood and is in delicate balance with its natural antagonist glucagon.[15] About 5 – 10% of diabetics suffer from the chronic autoimmune disease diabetes mellitus type I, which is caused by a destruction of the *β*-cells of the islets of Langerhans in the pancreas and leads to an absolute lack of insulin.[16] As a result, glucose can no longer be absorbed by the insulin-dependent tissues. Glucose is missing as an energy supplier in the cells and accumulates in the blood. In addition, the production of new glucose in the liver is running out of control since there is no negative feedback. The glucose is then released into the blood stream. However, since it cannot be utilized by the cells, it remains in the blood and the blood sugar level rises (Figure 2). Patients with type I diabetes are dependent on insulin treatment. Diabetes mellitus type II is a disease where insulin is available but cannot act properly at its target site, the cell membranes (so called insulin resistance). It ranges from (genetically determined) insulin resistance with a relative insulin deficiency to an absolute insulin deficiency in the later course of the disease. This type of diabetes is associated with overweight and poor nutrition. It represents the most common form of diabetes with 90% of all diabetes cases.[14] Besides the types already mentioned, there are two further, much rarer types: Diabetes mellitus type III (other specific types of diabetes), which includes diabetes due to genetic effects or hormonal disorders, and type IV, also called

gestational diabetes.[12] However, all forms of diabetes have in common that patients suffer from so-called long-term complications, which is a fundamental part of the research in this thesis.

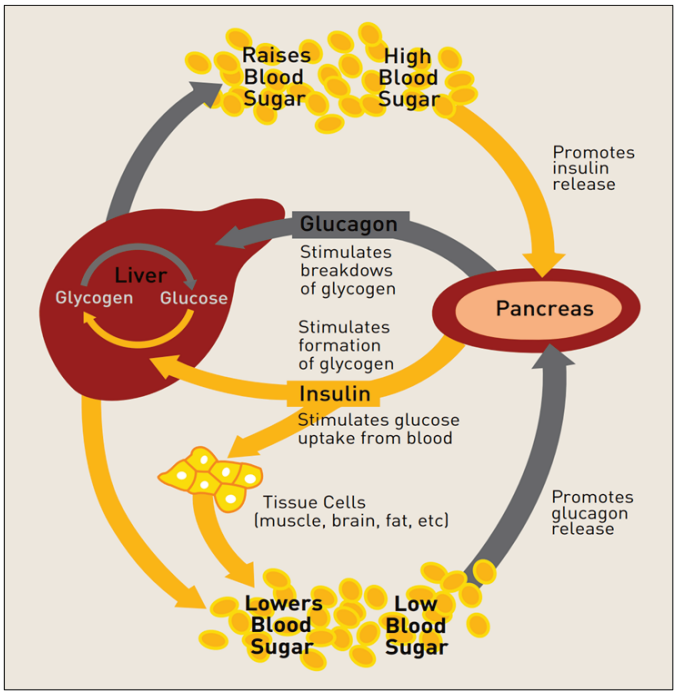


Figure 2: Normal insulin production and action, according to IDF.[15]

1.2.1.2 The Role of human Aldose Reductase in Combating Long-Term Diabetic Complications

In addition to many acute metabolic complications of diabetes, such as diabetic ketoacidosis and hyperglycemic hyperosmolar non-ketotic coma, almost all diabetes patients eventually develop characteristic long-term complications, mainly caused by the protein ALR-2. The complications include retinopathy, which can occur in all cases of diabetes. As with all diabetes-specific long-term complications, the occurrence and extent of retinopathy depends on the duration of the disease and the quality of the patient's treatment.[17,18] The ultimate consequence of diabetes, which is associated with the highest mortality, is nephropathy. However, it occurs in only about 35 – 45% of all type I diabetes patients and in less than 20% of all type II diabetes patients.[19-21] Long-term diabetes can also lead to neuropathies, usually as a manifestation of peripheral symmetrical sensorimotor neuropathy, leading to an increased incidence of cardiovascular disease, especially affecting women.[22] The occurrence of these and other late effects can generally be divided into the following three categories of pathological mechanisms:

(**a**) glucose-related, including abnormalities in polyol metabolism,[23]

(**b**) vascular mechanisms, including endothelial abnormalities, hyperfiltration and intrarenal hypertension in the kidney,[24]

(**c**) other mechanisms, including abnormalities in platelet function and growth factors and genetic influences.[25-27]

Since the glucose-dependent pathways mentioned in (**a**) seem to be the most important ones in the development of late-stage complications, the research in this thesis focuses on this pathway. As early as 1966, GABBAY *et al*. were able to prove that the polyol pathway plays an important role in the development of diabetes.[28]

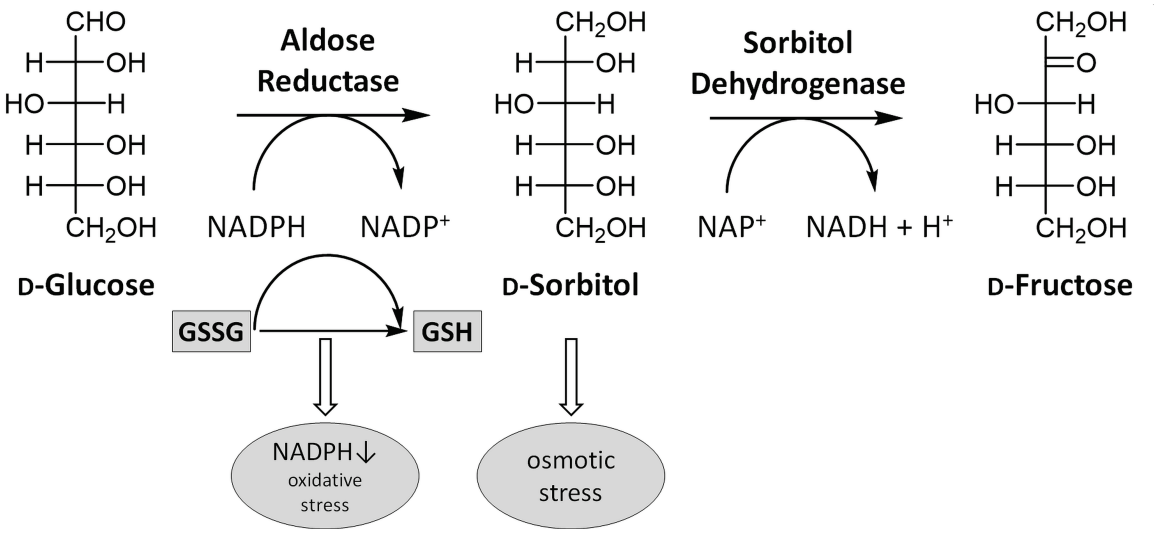


Figure 3: Polyol pathway. First step: Conversion of D-glucose to D-sorbitol catalyzed by aldose reductase using the cofactor NADPH. Second step: Conversion of D-sorbitol to D-fructose catalyzed by the sorbitol dehydrogenase using the cofactor NAD^+.[29]

Free glucose is normally converted during glycolysis into energy in the form of ATP, reduction equivalents and later into the final products CO_2 and lactate. Glycolysis is a central degradation pathway of energy metabolism and occurs in all organisms. In the first step, the hexokinase phosphorylates glucose to glucose-6-phosphate. However, due to an increased glucose content in diabetics, this pathway can quickly become saturated, whereupon the polyol pathway becomes increasingly activated as (Figure 3). In the first and rate-limiting step, D-glucose is converted into D-sorbitol by the oxidoreductase aldose reductase (ALR-2, EC 1.1.1.21) with consumption of NADPH. The sorbitol dehydrogenase is then responsible for further conversion to D-fructose.[30] For a long time, accumulating sorbitol alone was considered responsible for increasing osmotic stress, but today we know that NADPH is not only required for the conversion of glucose in the polyol pathway, but it is also a cofactor involved in the production of the important antioxidant glutathione (GSH). Osmotic stress due to the accumulation of additional sorbitol and oxidative stress due to the decrease in the $NADPH/NADP^+$ ratio and reduced NAD^+ level are the main causes of various complications of diabetes.[31] The

development of aldose reductase inhibitors blocks the polyol pathway and thus prevents the development of cell damage caused by oxidative and osmotic stress. Inhibitors can thus contribute to the reduction of long-term diabetic complications. However, caution is advised as ALR-2 has additional functions in the organism. It also plays an important role in detoxifying the metabolism by converting harmful aldehydes and ketones, which arise as by-products of metabolism, into less problematic substances.[30-32]

1.2.1.3 Human Aldose Reductase as a Target for Drug Design

To study the transient binding pocket of ALR-2 and to gain a progressive understanding of the drug design process, it is necessary to understand the structure of this protein.

The primary structure of aldose reductase was first determined in 1987/1989 by CARPER *et al*.[33] and in 1992 the three-dimensional structure of the enzyme was determined using X-ray crystallography.[34] ALR-2 is a monomeric 36 kDa cytoplasmic NADPH-dependent oxidoreductase belonging to the aldo-keto reductase superfamily. It consists of a total of 315 amino acids that form an α/*β*-barrel motif, consisting of eight and two additional smaller *α*-helices connected by eight parallel *β*-sheets (Figure 4). This so-called TIM-barrel was named after the enzyme triose-phosphate isomerase (TIM), since the folding motif was first discovered in this enzyme. It is a unique structural motif for oxidoreductases, as most reductases exhibit a Rossman folding. The cofactor $NADP^+$ is located at the top of the barrel, with the nicotinamide ring pointing downwards in the center of the barrel and pyrophosphate spanning the barrel lip. The binding pocket is located close to the flexible C-terminal loop at the end of the TIM-barrel in the barrel core. This loop, shown in blue, is responsible for substrate binding, while the so-called safety belt loop, shown in cyan, regulates cofactor binding.[35,36] The active site of the ALR-2 is in the upper part of the barrel structure[2] and can be classified into a flexible and a rigid part. Y48, H110 and W111 are positionally very stable and form the edge of the anionic binding site. In this pocket ALR-2 is able to bind and stabilize negatively charged substrates or inhibitors due to the positive charge of the adjacent cofactor (pink). The highly flexible C-terminal loop is located opposite. It allows the enzyme to open an additional sub-pocket and to adapt very well to substrates of different sizes. L301 and F122 (red) are the so-called gatekeeper residues for entering the specificity pocket, which is additionally bordered by T113, L301 and Y309 (Figure 5A and B).[37] During catalysis, the cofactor NADPH is bound first, followed by substrate binding of a sugar or other aldehydes. C1 of the substrate is reduced by a hydride transfer from the cofactor to the carbonyl carbon of the substrate and a concerted proton transfer of Y48 via H110 to the substrate.[38,39] Due to a conformational change in the safety loop of the protein, this is a slow process.[40]

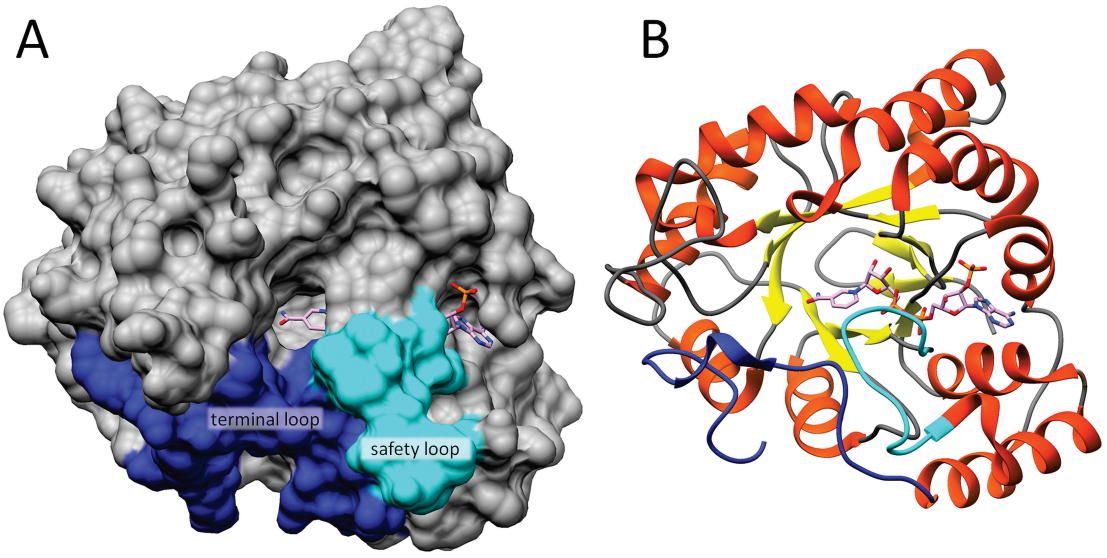


Figure 4: Crystallographic structure of human aldose reductase (ALR-2). ALR-2 in complex with $NADP^+$ (PDB-Code: 2J8T, 0.82 Å).[35] A) The surface of the ALR-2 is shown in grey, while the very flexible C-terminal loop is highlighted in blue and the safety belt loop in cyan. The cofactor is shown in pink. B) Ribbon representation of ALR-2. The α-helices (red) and the parallel β-sheets (yellow) shape the ALR-2 into a conserved protein fold called the TIM-barrel, which contains the catalytic site in the barrel core.[41] The loops are shown in grey and the cofactor in pink.

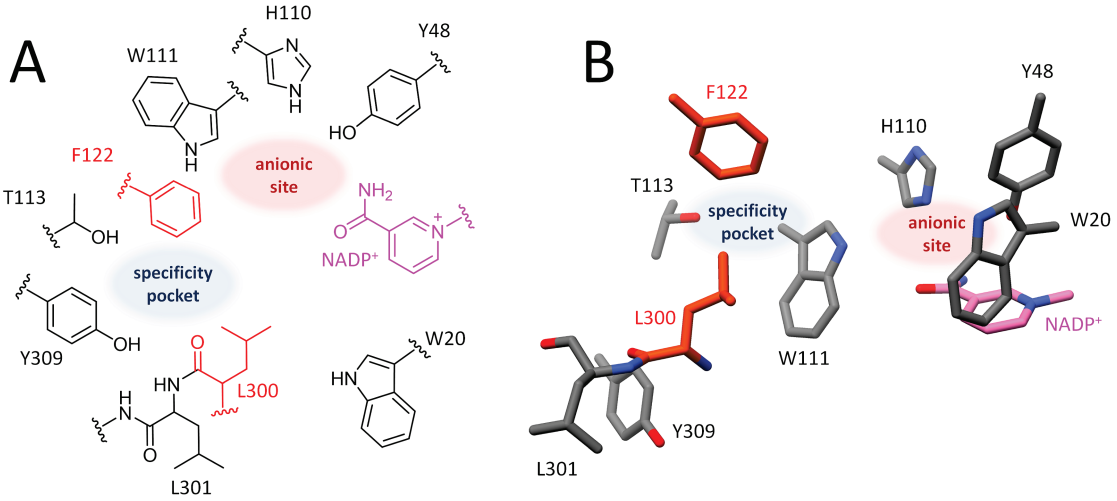


Figure 5: Binding site of ALR-2. A) Open specificity pocket in schematic. B) structural view. The gatekeeper residues L300 and F122, which open the access to the specificity pocket, are shown in red, while $NADP^+$ is shown in pink. The area of the specificity pocket is highlighted in blue, the anionic binding site in red.

1.2.2 The Serine Proteases Human α-Thrombin and Trypsin

A further focus of this thesis concerns preorganization to improve the binding behavior of inhibitors and the investigation of selectivity discrimination between the related proteases trypsin and thrombin. With thrombin, we have also chosen a well-studied protein as a target for these two investigations. This allows us to concentrate our research on the actual topic of preorganization or selectivity without elucidating the importance and relevance of the function of the protein. Consequently, it is essential to understand the effects of an improved thrombin inhibitor and the important role of thrombin in blood clotting, as well as the function of this target enzyme and serine proteases in general, in order to make progress in drug design using e.g., the concept of preorganization. Additionally, to discuss the topic of selectivity between thrombin and trypsin, the role of trypsin in the human body should be explained in advance.

1.2.2.1 Serine Proteases and their Functions

Both α-thrombin and trypsin are serine proteases. In the following a brief introduction to serine proteases in general is given to understand the function of both enzymes.

Serine proteases are a very common type of enzyme in the human body and in other mammals and accordingly thus account for one third of all proteases.[42,43] Altogether there are four different types of proteases, classified into the chymotrypsin-like serine protease families, subtilisin, carboxypeptidase Y and CLP like serine proteases. The largest group, the chymotrypsin-like serine proteases, to which the proteins trypsin and thrombin studied in this thesis belong, is present in many different physiological processes such as digestion, fibrinolysis, immune response, signal transduction or apoptosis.[42,44] Proteases are enzymes that cleave the amide bond between two amino acids of a substrate with different sequence selectivity. In a serine protease, this is achieved by the catalytic triad consisting of serine, histidine, and aspartic acid. In the case of chymotrypsin-like proteases it is S195, which nucleophilically attacks the functional group (abbreviated as group in the following) of an amide to be cleaved with its hydroxy group, (Figure 6a). After deprotonation of the alcohol, the nucleophilic attack is amplified by the imidazole side chain of H57, while the adjacent D102 compensates the resulting positive charge by taking up a proton from H57. A tetrahedral transition state is formed which is stabilized by the oxyanion hole (O-hole), a small pocket next to the S195 side chain consisting of two backbone NH groups (b). The next step is the collapse of the transition state with release of the amine portion from the active pocket (c). Subsequently, the formed acyl-enzyme intermediate is hydrolyzed with release of the C-terminal cleavage product (d).[2]

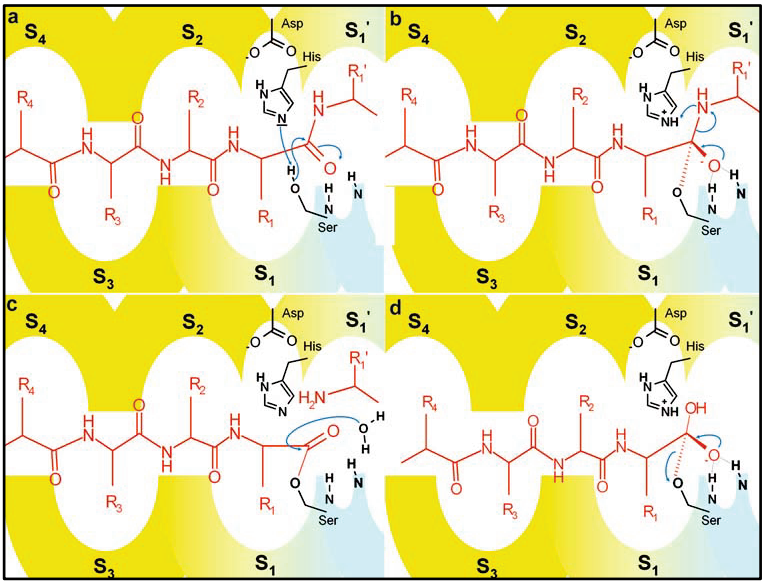


Figure 6: Mechanism of protein cleavage by the catalytic triad (H57, D102 and S195) of serine proteases.[2]

Both serine proteases trypsin and thrombin follow the described mechanism, but there are some significant differences between proteases that are actually highly similar. For example, while thrombin is highly selective in its substrate selection and only cleaves after arginine, trypsin cleaves almost all peptide sequences following either arginine or lysine. The high selectivity of thrombin is crucial due to its leading role in the blood clotting cascade.

1.2.2.2 α-Thrombin and its Role in the Blood Clotting Cascade

An example of a protease function is the conversion of a biologically inactive enzyme, called zymogen, into its active form by cleavage.[45] The active protein factor II (FIIa) used as a target in this work has a weight of 39 kDa and is also known as thrombin. Its development starts with the production of its precursor factor II (FII), known as prothrombin, whose gene is [2,46] and plays an important role in all processes of thrombosis and hemostasis. In particular at the end of the blood coagulation cascade formed by a number of structurally related serine proteases, it has an essential function (Figure 7). In a recent injury, the initial stage is primary hemostasis, in which the platelet thrombus is formed by the binding of collagen and platelets with the support of von Willebrand factor (vWF). Subsequently, the process of secondary hemostasis and thus wound healing is initiated in two different pathways. Extrinsic blood coagulation is caused by a lesion of the vascular system, whereas in the intrinsic path it is caused by reduced blood flow or pathogenically altered vascular walls. In both pathways, inactive factors, mostly serine proteases (box, light grey), are activated by a cleavage through other serine proteases. The factors activated in this way (ellipse, dark grey) subsequently activate

the next factor. This occurs until both pathways finally end up in the activation of FX which promotes the formation of thrombin (ellipse, red) after autocatalytic and proteolytic cleavage of prothrombin. At this point, thrombin triggers the cleavage of insoluble fibrinogen for blood clotting by aggregating soluble fibrin and cross-linking it to form a thrombus.[2,47] Consequently, a thrombin inhibitor optimized by preorganization in the required bound geometry, which selectively binds to thrombin but no other serine proteases, can be used to control diseases caused by increased or falsely induced blood clotting.

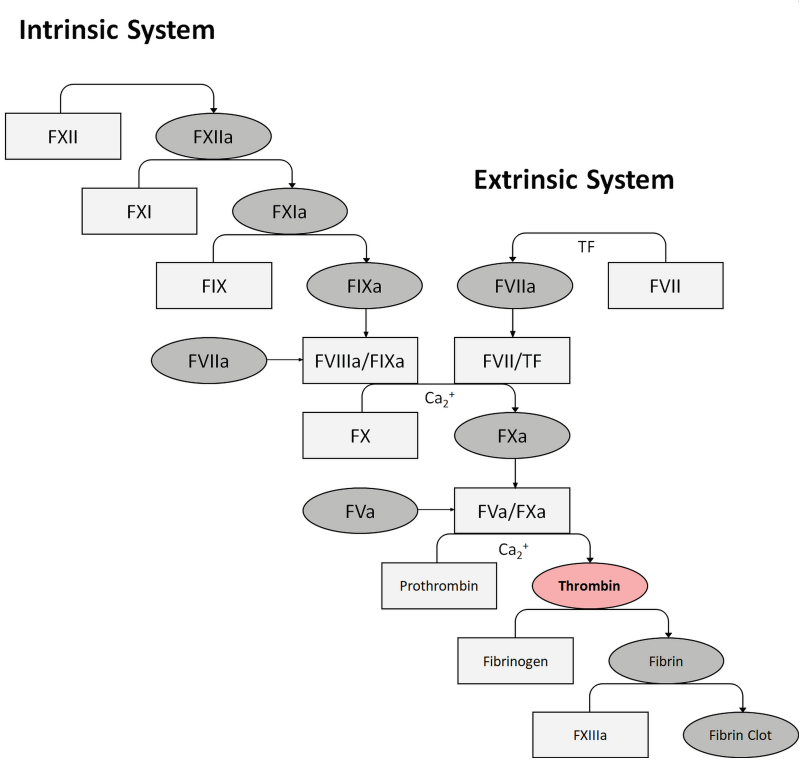


Figure 7: Schematic view of the secondary hemostasis. Extrinsic (caused by a lesion of the vascular system) and intrinsic system of blood clotting (caused by reduced blood flow or pathogenically altered vessel walls).[47,48] Inactive zymogens are shown as light grey boxes, while active proteases are presented as dark grey ellipses. The key enzyme thrombin is highlighted in a red ellipse.

1.2.2.3 Physiological and exogenous thrombin inhibitors

The mechanisms of hemostasis are crucial to prevent blood loss after external trauma. If the factors that control clotting are not precisely balanced, it may lead to thrombosis, stroke or abnormal bleeding. To prevent excessive and therefore pathological coagulation, human blood contains some physiological anticoagulants. These multiple factors are controlled by a highly sensitive and well-controlled balance.[40,44] The main anticoagulant circulating in the bloodstream is antithrombin III, which inhibits not only thrombin but also a number of other serine proteases. When the inhibitor heparin binds to antithrombin III, the conformation changes and the affinity to all substrates increases.[49] Another physiological inhibitor is heparin cofactor II (HCII), which is specific for thrombin and is released in the event of vascular injury, unlike antithrombin III. Other potential diseases are strokes and myocardial infarctions, which are caused by venous occlusion and lead to poor nutrient and oxygen supply to the downstream tissues. These cardiovascular events are responsible for one third of all mortality worldwide.[50] Due to its special role in the blood coagulation mechanism, thrombin is an attractive target for drugs to prevent pathological blood coagulation events.[51]

The most effective and common compounds include traditional anticoagulants, such as 3 − 30 kDa polysaccharide unfractionated heparin (UFH) and its derivatives, low molecular weight heparins (LMWH), such as enoxaparin. The negatively charged protein UFH gained from the intestinal mucosa of pigs is able to interact with the positively charged exosite II of thrombin (Figure 8B). It promotes the inhibition of thrombin and FXa by antithrombin III, resulting in a reduction of the blood clotting process. However, this heparin has a low bioavailability due to its size and high charge density. Consequently, it requires intravenous administration. In addition, its interaction with platelet FIV is critical as it leads to heparin-induced thrombocytopenic syndrome (HIT), the reduction in platelet count. To address these complications, LMWHs with weights from 2 to 10 kDa have been developed that specifically inhibit FXa but still cause HIT and have a high risk of bleeding (C).[50,52] The vitamin K antagonists, such as warfarin, belong to the most frequently used anticoagulants, similar to heparin. However, these traditional anticoagulants can be administered orally. They exert their effect by inhibiting the vitamin K-dependent carboxylation of various factors involved in blood clotting. Despite a narrow therapeutic window, these drugs have multiple interactions with food and other drugs.[50]

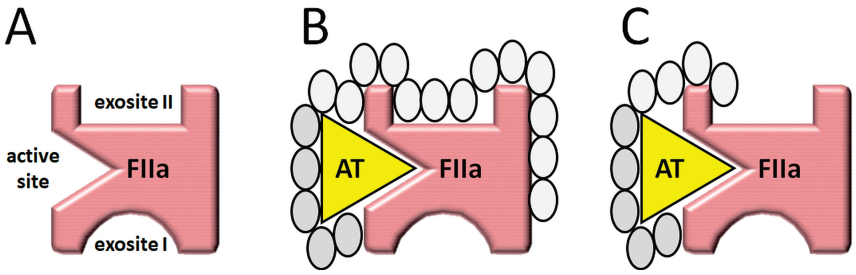


Figure 8: Systematic representation of thrombin. Thrombin (A, pink) is capable of binding UFH (B) and LMWH (C) using the co-factor antithrombin III (AT, yellow) according to LEE and ANSELL.[52]

Heparin and vitamin K antagonists belong to the group of indirect anticoagulants. However, the protein hirudin isolated from the saliva of the leech *hirudo medicinalis* does not require a cofactor. It therefore belongs to the group of direct anticoagulants. Direct thrombin inhibitors can reduce the risk of HIT. For example, hirudin interacts with thrombin in a bivalent binding mode, binding simultaneously to both the active site and exosite I in a 1:1 stoichiometric complex. Due to this binding mode, hirudin is considered a high-affinity inhibitor.[50] Hirudine is available in a recombinant form as lepirudin and desirudin, called *r*-hirudins. Both are high-affinity inhibitors with a narrow therapeutic window and bind irreversibly in the same bivalent mode as hirudin, (Figure 9A) *r*-hirudins are currently approved for the treatment of HIT. However, it is known that both drugs are associated with a high risk of bleeding and should be administered parenterally.[53]

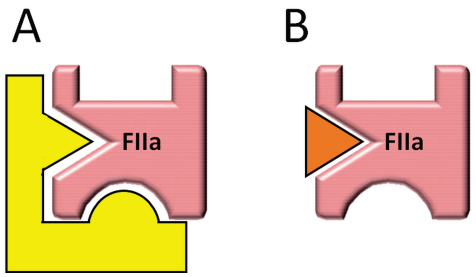


Figure 9: Schematic representation of hirudin binding in thrombin. A) Thrombin (pink) with the *r*-hirudins lepirudin and desirudin (yellow). B) Thrombin bound to the small direct anticoagulants melagatran and dabigatran (orange) adapted to the active center of thrombin, according to LEE and ANSELL.[52]

The group of small direct anticoagulants forms another class of thrombin inhibitors. One of the first highly effective inhibitors in this category is N^{α}-(*β*-naphthylsulphonyl-glycyl)-D,L-*p*-amidino-phenyl-alanyl-piperidine (NAPAP), (Figure 10). It was developed based on the knowledge that benzamidine inhibits thrombin. Linking the benzamidine group to a peptide structure led to a significant improvement. For a long time NAPAP was the strongest

representative of the class of small direct thrombin inhibitors (K_i = 6 nM). However, it has only a small selectivity advantage compared to trypsin binding.[2] Based on the active center of thrombin, the peptide-like inhibitor Melagatran was designed to bind univalently to thrombin and mimics almost perfectly the Phe-Pro-Arg structure of the natural substrates of thrombin.[54] The schematic representation of the binding of melagatran to thrombin is (Figure 9B). The inhibitor, which was administered as the double prodrug Ximelagatran, was the first promising oral direct thrombin inhibitor, but quickly disappeared from the market due to the risk of hepatotoxicity.[52] Based on the NAPAP-thrombin complex, the negatively charged and therefore highly polar drug Dabigatran was developed. It acts as a direct inhibitor of thrombin.[53] Due to its poor bioavailability, a double prodrug has also been developed for this thrombin inhibitor, which is called Dabigatran etexilate and can be administered orally. In contrast to Ximelagatran, no risk of liver toxicity has been observed so far. Dabigatran etexilate is characterized by the absence of any cross-reactions and a broad therapeutic window. It consequently represents an important development in the research of direct thrombin inhibitors.[55]

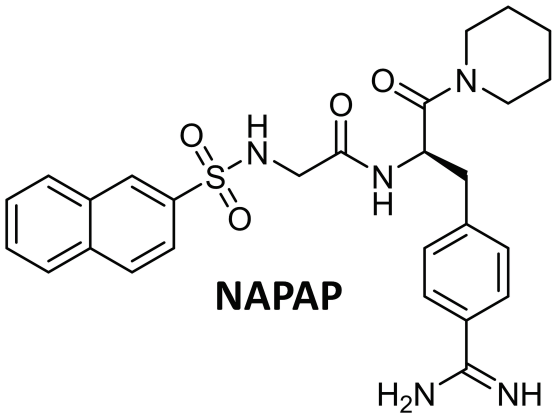


Figure 10: First small direct and highly effective thrombin inhibitor N^{α}-(β-naphthylsulphonyl-glycyl)-D,L-*p*-amidino-phenyl-alanyl-piperidine (NAPAP) based on benzamidine.[2]

The development of anticoagulants and the associated ability to treat thromboses or similar disorders is therefore subject to constant change. There are already some promising drugs on the market. However, the disadvantages of these drugs, such as parenteral administration or the high risk of HIT, cannot be dismissed. Therefore, the development of new optimized drugs in this area is essential.

1.2.2.4 *α*-Thrombin as a Target for Drug Design

Improvements in lead optimization in terms of inhibitor preorganization and selectivity on the target protein thrombin require prior detailed knowledge of its structure. In the following, its structure as well as sites in the protein that are important for drug design are presented.

Unlike many other serine proteases, thrombin which consists of a light chain of 36 amino acids (L-chain) and a heavy chain of 259 amino acids (H-chain) linked by a covalent disulfide bridge,[56] is highly selective and requires an arginine in the S_1 pocket while accepting only glycine or serine in the P_1' position.[2,44] Due to the important role, thrombin plays in the blood clotting cascade, its excellent selectivity toward inhibitors is extremely useful for thrombin and is induced by various insertions with respect to trypsin and chymotrypsin, including the 60's loop, the *γ*-loop, and the two binding sites exosite I and II.

One of these specific structural motifs is the so-called 60's-loop, which is shown in yellow (Figure 11A). It shapes and narrows the upper rim to the entrance of the active site and is responsible for selectivity in the S_2 pocket. This hydrophobic insertion of nine residues from Y60a to T60i leads to a pocket that is selective for small residues such as valine or proline,[44] and shields H57 and S195 from unsuitable substrates, limiting the accommodation of non-specific substrates in the active site. Besides the 60's-loop, the *γ*-loop, or autolysis loop, is another insertion important for specificity, which is also shown in yellow (A). This very flexible loop, which consists of five residues, forms the lower edge of the entrance of the active center and seems to be important for selectivity in terms of binding protein C and fibrinogen.[44,57] The loop at exosite I, centered around the residue K70, is the main factor for the recognition of negatively charged substrates such as fibrinogen, thrombomodulin, hirudin and the thrombin receptors (B). Accordingly, it enhances the correct position of the substrates in the protein and is also important for allosteric modulations of the active center. Exosite II, an anion-binding C-terminal helix, is located on the opposite side of exosite I (B).[44,45] As already mentioned in Chapter 1.2.2.3, it hosts particularly positively charged amino acid residues that interact with polyanionic inhibitors such as heparin. This enables them to allosterically influence the active site as well.[44]

Since the described structure of thrombin and its binding pocket, more precisely the S_1 binding pocket, which is responsible for the catalytic cleavage and specificity of thrombin, is so well known, we decided to use this enzyme for our studies on preorganization and selectivity.

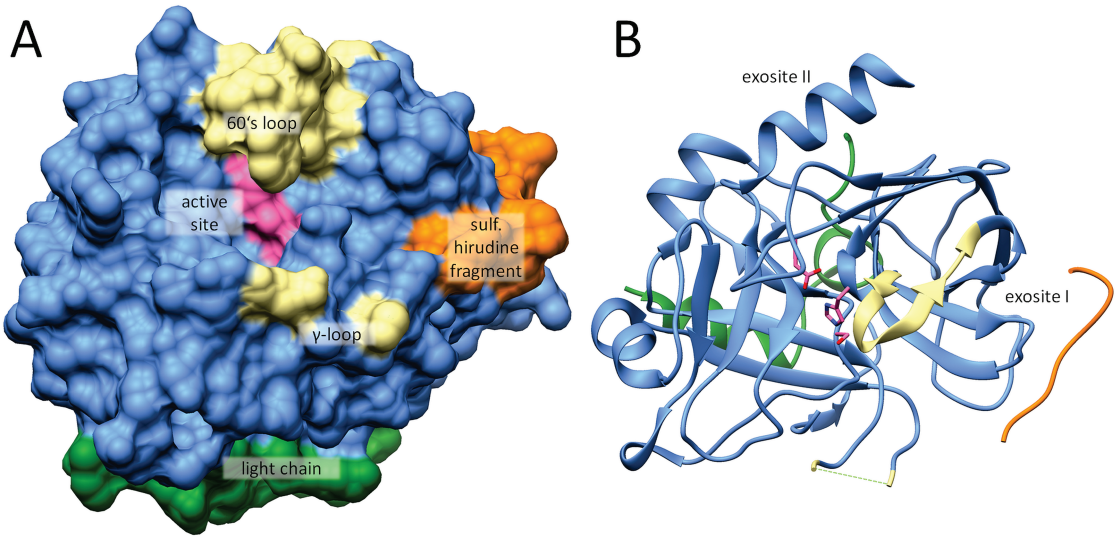


Figure 11: *Apo* structure of α-Thrombin. A) surface and B) ribbon representation (PDB-code: 3D49, 1.50 Å). The heavy chain contains two six-stranded β-barrels, between which the active site is located. The sulfated hirudin fragment (54 – 65, Bachem, Bubendorf, Switzerland), binding to exosite I, is shown in orange. The 60's-Loop and γ-loop are depicted in yellow. The heavy chain is represented in blue, the light chain in green. The magenta-colored part in the center of the protein represents the catalytic triad.

1.2.2.5 The omnivorous Trypsin

To investigate the substrate specificity of serine proteases in this work, it is important to have basic knowledge of trypsin (Enzyme Commission number (EC) 3.4.21.4), an exocrine pancreatic hydrolase of about 26 kDa with strong proteolytic properties. Trypsin was first discovered by WILHELM KÜHNE in 1876.[58] Like thrombin, trypsin is a commonly well-studied protein that is very well suited as a test case to further improve the methods of drug development. However, trypsin consists of only one protein chain. It is secreted as zymogenic trypsinogen in the exocrine pancreas and then activated in the duodenum by the enzyme enteropeptidase or by already activated trypsin molecules via proteolysis.[59] The activation of trypsinogen to trypsin occurs after cleavage of the peptide bond between K15 and I16, which causes a conformational change leading to activation.[60] Human trypsin is a mixture of three different isoforms: cationic trypsin, anionic trypsin and mesotrypsin, with the former accounting for the largest proportion.[61] The protein also belongs to the group of serine proteases and is responsible for the digestive system of many vertebrates as it transforms different dietary proteins into smaller, more easily digestible proteins.[62] Therefore, its active site has to be capable to accommodate a wide range of different substrates.[44] To ensure this, trypsin is much less selective with respect to its substrates, although it is very similar to the serine protease thrombin in both structure and mechanism. This is partly due to the absence of the 60's-loop described

in Chapter 1.2.2.4, which is present in thrombin and shields the S_1 pocket from unsuitable substrates. It is important to consider selectivity as a crucial parameter in the development of new anticoagulant drugs in order to avoid undesired therapeutic side effects. For example, orally active thrombin inhibitors need to be selective against trypsin to avoid digestive tract disorders.

Due to its high structural similarity to the related highly selective serine protease thrombin, trypsin is well suited for the investigation of potential specificity mediating factors in this work.

1.3 Methods of Modern Drug Design

1.3.1 The Significance of Isothermal Titration Calorimetry for Drug Design

Although X-ray crystallography can be used to decipher the structure of a protein and to support the design of a lead structure, this is not sufficient for the development of a suitable drug candidate. To achieve a desired effect of a molecule on a target, it is not only important to achieve perfect fit into the binding pocket. The kinetic and thermodynamic parameters of a binding process also play an essential role and can influence the therapeutic effect. These data are often determined by assays or isothermal titration calorimetry (ITC) and combined with information from X-ray structure analysis to obtain an accurate overall picture of the investigated compound.

Inhibitor affinities are often obtained by cheap and widely used methods, such as the determination of the half-maximum inhibitory concentration IC_{50} or the inhibition constant K_i by various assay methods, such as enzyme-kinetic fluorescence assays. However, in advanced stages of drug design, different hits may have similar affinity values, making it difficult to decide which compound should be used for further consideration.[63] Thermodynamic fingerprinting of a binding event can therefore provide more accurate information. The binding process of an inhibitor to a protein corresponds to a chemical reaction that is associated with a more or less strong heat exchange. Accordingly, heat can either be released (exothermic) or consumed (endothermic) during the binding process. This heat signal can be detected by ITC to determine the binding of an inhibitor to a protein.[2] The instrument consists of a sample cell which is normally filled with the protein solution and a reference cell containing water (Figure 12). The syringe is filled with the inhibitor solution being gradually injected into the protein solution for measurement with an evenly rotating syringe. Due to the coating of both cells with an adiabatic shield and while maintaining a constant temperature through a compensation heating, slight temperature changes caused by the binding event in the sample cell

are detected and quantified by analyzing the difference in thermal power required to keep the sample cell at the same temperature as the reference cell.[64]

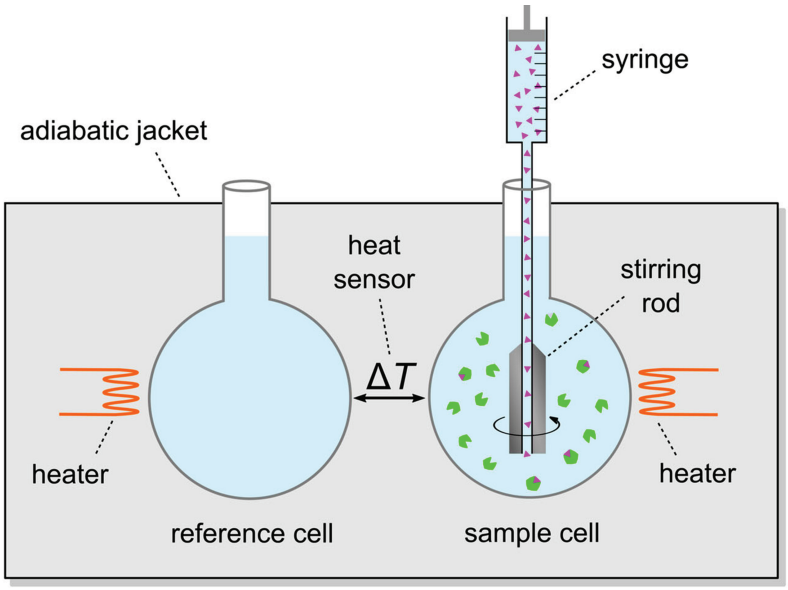


Figure 12: Schematic depiction of an ITC device according to KRIMMER *et al.*[64] The inhibitor solution (magenta) is stepwise injected into the protein solution (green) in the sample cell. The temperature difference resulting from the binding is measured relative to the reference cell, filled with water.

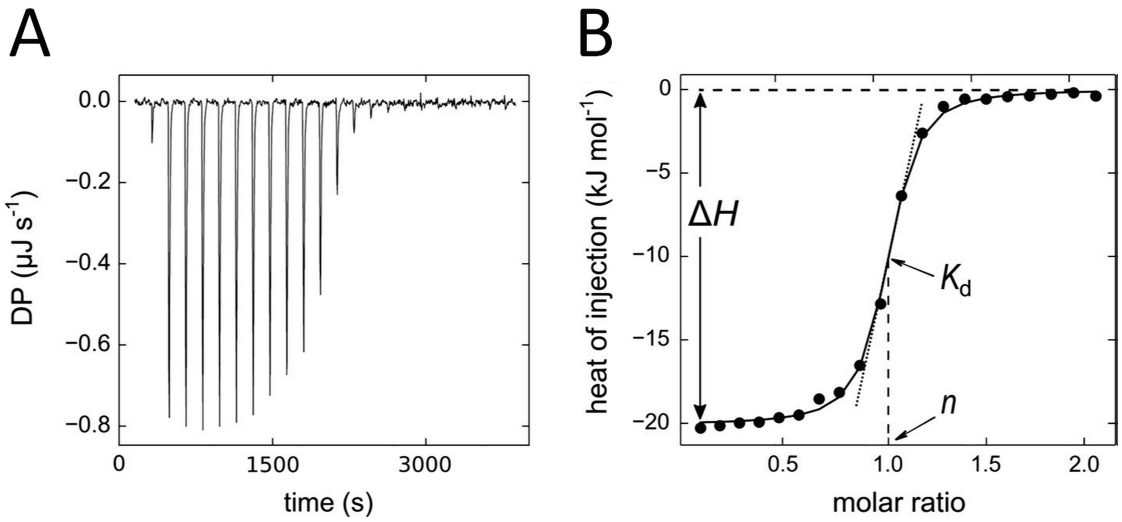


Figure 13: Exemplary representation for the result of an ITC measurement.[64] A) Raw thermogram of an ITC measurement device. The differential power (DP) in µJ/s is plotted against time. B) Integrated raw data and isotherm. The molar change in enthalpy is plotted against the molar ratio of the binding reaction. K_d corresponds to the slope in the inflection point of the sigmoidal curve. $\Delta H°$ equals the difference between the plateaus. The stoichiometry *n* can be extracted as the abscissa at the inflection point.

The differential power is shown in a raw thermogram (Figure 13a) and can be expressed as an isotherm by integrating each individual peak (Figure 13b). As the saturation of the protein increases, the heat signal decreases, so that the binding constant of the inhibitor to the protein can be derived from the curve.[2] The affinity of an inhibitor that binds to its target protein is described by the change in Gibbs' free energy Δ*G°* of the system before and after the binding event, which cannot be derived directly from the integrated data. According to the Gibbs-Helmholtz-Equation (Equation 1) it consists of a change in enthalpy Δ*H°* (kJ/mol) and a change in entropy −*T*Δ*S°* (kJ/mol), the latter weighted by the absolute temperature. Δ*H°* represents the difference between the lower and upper plateau of the resulting fitted regression curve and can be used as an indicator for the quantity and quality of non-covalent interactions. The dissociation constant (K_d), which is the inverse of K_a according Equation 2, corresponds to the slope at the inflection point of the sigmoidal curve. The Gibbs-free binding energy Δ*G°* (kJ/mol) is logarithmically related to the association constant K_a, weighted with the ideal gas constant *R* (8.314 J/mol*K) and the absolute temperature T (K). It can hence be calculated as

$$\Delta G° = \Delta H° - T\Delta S° \qquad \text{Equation 1}$$

$$K_d = \frac{1}{K_a} \qquad \text{Equation 2}$$

$$\Delta G° = -RT \ln K_a \qquad \text{Equation 3}$$

described in Equation 3. The entropy change Δ*S°* describes the change of the order parameters and the distribution of the system over several accessible states. The change in entropy is mainly associated with conformational changes of the inhibitor and the protein. When an inhibitor flexible in solution binds to a protein and thereby fixes both its own and the conformation of the protein binding pocket, degrees of freedom are lost, resulting in negative Δ*S°*. However, other processes also play a role, for example the release of water molecules that leave the inhibitors or proteins hydration shell and thus a previously fixed position during the binding process.[64] A classic example is the displacement of well-ordered water molecules from apolar surfaces and the associated increase in entropy, which is considered the driving force of association in the so-called hydrophobic effect.[65] Displacement titrations may be necessary due to a very high or very low affinity of the inhibitors, which may lead to a too high or too low *c*-value. This value indicates the sigmoidal shape of the titration curve and is typically between 10 and 100 to obtain an exact value for the association constant K_a and thus a reliable evaluation. As the *c*-value increases, the curve approaches a rectangular shape, which leads to an inaccurate determination of K_a. The same applies if the curve is flattened out by a too low *c*-value. It can be calculated according to Equation 4 and is defined as the product of the stoichiometry *n* and the protein concentration *C* in the cell, divided by the dissociation

constant K_d. Displacement titration allows the titration curve to be shifted into a *c*-value range, which leads to proper sigmoidal isotherms. However, an ITC measurement can be influenced by several other aspects, such as the impact of the buffer used for the reaction when protonation states change, the experimental settings, protein and inhibitor impurities, sources of systematic errors and inappropriate assumptions, e.g., about the applied model in the analysis software. In this way, the thermodynamic data of a binding event obtained by ITC can allow further interpretation of the driving forces of inhibitor binding and thus support the data of high-resolution structures during drug development or the evaluation of new approaches in drug development.[64]

$$c = K_a \cdot n \cdot C = \frac{n \cdot C}{K_d} \qquad \text{Equation 4}$$

1.3.2 3D Structural Analysis of Proteins

The elucidation of the structural features of various enzymes plays an important role in drug development, as already described in chapter 1.1. This allows to determine how a drug binds to the target protein and what effects it has on the structure of the protein. In combination with e.g., thermodynamic, and kinetic data, this is important to subsequently obtain information on the effect of the drug.

Currently, there are three established methods to obtain a model of a protein. The most recent method is cryogenic electron microscopy (Cryo-EM), which allows imaging of the studied object close to its original state. In this method, an aqueous sample solution is applied to a grid and flash frozen in liquid ethane or a mixture of liquid ethane and propane. A cryo-electron microscope is then used to obtain many thousands of images of the frozen sample solution from different perspectives. By analyzing the individual images and combining their data, an electron density and thus a 3D model of the object can be created. This method has the advantage of much smaller sample sizes and the fact that a protein does not need to be crystallized, as an aqueous solution is sufficient for the study. However, Cryo-EM also has disadvantages, such as a poor signal-to-noise ratio which makes it difficult to detect features of a particular sample when only a few samples are examined. As a result, data acquisition can take much longer than with synchrotron X-ray structure analysis, and the resolutions currently achievable are in the range of 3 Å, far below those of protein crystallography. [66,67]

Determining the structure of proteins by NMR also offers the advantage that the protein can be studied in aqueous solution, thus under reasonably physiological conditions. For this purpose, several thousand atomic distances are determined with an accuracy of 0.2 – 0.5 nm. However, this procedure is quite laborious and is therefore rarely used.

Depending on the wavelength, however, X-ray crystallography can achieve resolutions of less than 1 Å if the crystal is of sufficient quality. The method of X-ray crystal structure analysis used in this work dates back to an experiment by MAX LAUE in 1912 in which he demonstrated the wave character of X-rays. He found out that due to the three-dimensionally ordered lattice-like structure of the crystals, interferences are to be expected. The process by which radiation at the crystal lattice is deflected by interference to numerous reflections observable in different directions in space is called X-ray diffraction.[2,68] The method has been developed over the years and is now an important tool for the drug design process, allowing the exact spatial arrangement of all atoms of a crystalline chemical molecule to be determined.

The principle of X-ray structure analysis is shown in Figure 14. Data collection can be performed with modern high-performance synchrotrons, but also with in-house equipment of smaller research institutions. In any case, a linear X-ray beam first hits a usually deep-frozen protein crystal. The X-ray radiation scattered by the crystal is then recorded by a detector and a diffraction image is generated. Using computer-aided methods, electron density maps can be calculated based on this diffraction image, which are used for model building, refinement, and the completion of the final structure.[69] From today's point of view, however, it can be assumed that the method of X-ray structure analysis will become less important as it may increasingly be replaced by Cryo-EM.[66,67]

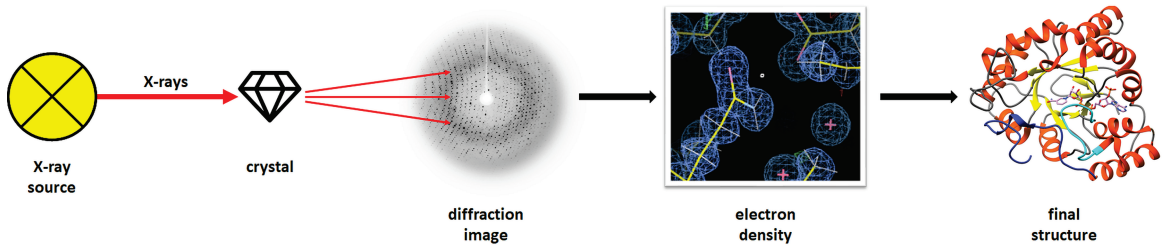


Figure 14: The principle of X-ray structure analysis. Linear X-rays emerge from an X-ray source and hit the protein crystal. A detector is used to obtain a diffraction image of the rays scattered by the crystal. The electron density used to model the final structure is calculated by computer-aided methods.

1.3.2.1 The Electron Density

Usually, two common types of electron density maps are used to build the model and to further refine it. One of them is the difference electron density map, in which the electron density of the "known" atoms is excluded. Thus, to calculate this map, the structure factor amplitudes F_C calculated from the structural model at a given refinement level are subtracted from the observed structure factor amplitudes F_O. This highlights the density peaks of the missing

atoms and allows an easier interpretation of the density map. In addition, the $F_O - F_C$ map can also be used to show the disorder of the atoms. Peaks appear in a difference map at locations in the unit cell where the model did not explain enough electron density (usually shown as green meshes) whereas valleys appear at locations where the model contained too much electron density (usually shown as red meshes).[69,70] Additionally, a factor of the atomic position is created in the difference electron map, resulting in an approximation to the true atomic position. However, if the deviation of the model from the true atomic position is too large, a negative density is caused. Another special form of the difference electron density map is the so-called omit map. Here, parts of the structure in defined areas of the unit cell are excluded. In the case of a polder map, the background noise generated by the solvent is also considered. In areas where the standard difference electron density is not sufficiently defined to determine an exact atomic position, these maps can be used as an aid.[69] The $2F_O - F_C$ map, on the other hand, includes the $F_O - F_C$ map and the electron density around the model and is usually shown in blue.[70] Typically, hydrogen atoms are difficult to detect in such density maps because of their minor contribution to the structure factors due to their low electron density. However, at good resolution below less than 1 Å and when all other atoms have been correctly detected, well visible peaks for hydrogen atoms often appear in the $F_O - F_C$ density. The resolution of a crystal structure plays an important role in the analysis of the exact atomic positions. It is defined by the distance between the center of an X-ray diffraction image and the point where the diffraction data are just visible. At low resolution, the electron density is therefore more diffuse, and it is difficult to find the arrangement of the atoms. In contrast, at high resolutions, the positions of the atoms are in excellent agreement with the electron density. The polypeptide trace, for example, is detectable at 3.5 − 2.8 Å, provided the sequence is known. At much better resolutions above 2.8 Å, not only the positions of the backbone but also those of the individual amino acid side chains can be determined. If sections of a structure can only be detected poorly, this is often due to a weak resolution, which is usually caused by an insufficient quality of the crystal.[69] However, resolution is only one factor that affects the quality of electron density.

1.3.2.2 The *R*-Factor as a Quality Criterion for the Crystallographic Model

As a further aid to refinement, the *R*-value is used to evaluate how well the calculated structure factors match the experimentally observed structure factors and how well the model created ultimately reflects the electron density. The value is calculated from F_O and F_C, under additional consideration of a scaling factor *k* as shown in Equation 5.[68] A completely random set of atoms gives an *R*-value of about 0.63, while a perfect fit would have a value of 0.

To reduce errors and avoid bias of data, R_{free} is used additionally. Before refinement is started,

$$F = \frac{\sum ||F_o| - k|F_c||}{\sum |F_o|} \qquad \text{Equation 5}$$

a small part, usually about 5% experimental observations, is removed from the data set and refinement is performed with the remaining data. The R_{free} value is then calculated by checking how well the model predicts the 5% that were not used in the refinement. The R_{free} value is usually only slightly higher than the *R*-value, so that overinterpretation of the data can be detected and avoided during model building.[69]

1.3.2.3 Dynamic Disorder - The Debye-Waller Factor

Today, crystallographic protein structure analysis also allows us to analyze the molecular motion of individual atoms and thus determine the strength of a molecule's bond to its site. The *B*-factor (also called the temperature factor or Debye-Waller factor) is measured in units of Å^2 and can be specified for each atom or group of atoms. It is a consequence of the dynamic disorder in the crystal caused by temperature-dependent vibrations of the atoms in the structure and is therefore a measure of the positional deviation of an atom oscillating about the given coordinates (Equation 6). The value u_i^2 (mean square displacement) corresponds to the average deviation of the atomic vibration of the atom number *i*.[69,71,72]

$$B_i = 8\pi^2 \cdot u_i^2 \qquad \text{Equation 6}$$

The higher the *B*-factor, the greater the spatial extent of the electron cloud. It thus reflects the mobility of an atom and can indicate errors in the modeling. The *B*-factors, as well as the positions of the atoms and their occupation fractions, can be adjusted to reflect empirical values during refinement. For simplicity, the vibrations of the atoms are considered isotropic in the refinement. If sufficient data are available, individual *B*-factors can also be refined anisotropically instead of isotropically. [69,71,72]

1.3.2.4 Static Disorder - The Occupancies

At high resolutions, high *B*-factors of protein crystals are often accompanied by static disorder. Molecules or parts of molecules can adopt different orientations in different unit cells. The resulting electron density is an average of the situation in all unit cells of the crystal,

resulting in the electron density indicating multiple positions, but weakening at a single position. However, if all the conformations of the side chain are clearly visible, they can all be assigned in the model. However, the positions are rarely exactly equal in occupancy, so the electron density of one of the conformations is usually stronger and easy to interpret, while the density of the others is so weak that the position of the side chain cannot be adequately determined. Based on the difference electron density, the individual occupancies can be adjusted during refinement allowing analysis of the different occupancies of entire protein regions. This can result in entire loops of a protein being in a double conformation, or a binding pocket being simultaneously open to a certain percentage but also closed to a corresponding percentage.[69,71,72]

Despite good crystal quality and high resolution, it is possible that especially the arrangement of amino acid side chains localized on the surface of the protein cannot be unambiguously identified or even no electron density can be detected at all. In many cases, the reason for this is an increased flexibility of the amino acid side chains. Instead of a single fixed position, these amino acids adopt a different conformation in each unit cell. The occupancy is correspondingly low, which attenuates the difference electron density to the point where it is no longer visible in the density map. As a result, often only half or none of the side chain of the amino acid can be assigned in the model structure. However, not only the individual amino acids can be simultaneously present at several positions. When a ligand is soaked into the crystal, this does not automatically mean that it binds with 100% occupancy in the protein. In particular, inhibitors with lower affinity tend to have lower occupancies. Even in this case, it is possible that the same inhibitor occupies different positions in the unit cells. It often happens that part of the inhibitor is conserved in the binding pocket, but another part is not fixed by strong interactions and therefore remains very flexible. When an inhibitor is only partially inserted into a model, this indicates increased flexibility of the other inhibitor part; an exact binding position could not be determined.

1.3.2.5 Correlation of Crystal Structures and Thermodynamics

However, these refinement tools are not only used to determine structures. In combination with other analytical methods such as isothermal titration calorimetry, X-ray protein crystallography allows the investigation of molecular biological processes. In particular, water molecules are becoming increasingly important in drug development. With high resolutions it is possible to analyze the water network and to link the resulting information with thermodynamic data.

Binding of an inhibitor in a binding pocket of a protein usually requires breaking a water network. During the formation of a protein-ligand complex well-ordered water molecules are

displaced from the binding pocket into the main solvent. Such a process costs energy, which is reflected in terms of an unfavorable enthalpic contribution. However, the release of many previously fixed water molecules has a positive effect on the entropic contribution to binding. Due to this classical hydrophobic effect, a poor binding enthalpy can be compensated by a favorable entropy. Consequently, the binding affinity of a ligand is often strongly influenced by the desolvation energy.[73] Similar strategies are used in the development of inhibitors. If a ligand is already strongly constrained in its conformation before complex formation, it barely loses any degrees of freedom upon binding in the protein, which means that the entropic contribution to binding is only minimally negatively affected by this.

In this way, models of protein structures are created and binding angles and lengths, as well as the binding pockets of a protein, can be visualized and analyzed. In drug design, it is now possible to design a lead that is structurally exactly adapted to the protein.

Chapter 2

Which Properties Allow Ligands to Open and Bind to the Transient Binding Pocket of Human Aldose Reductase?

Anna Sandner[1], Khang Ngo[1], Christoph P. Sager[1], Frithjof Scheer[2], Michael Daude[3], Wibke E. Diederich[2,3], Andreas Heine[1] and Gerhard Klebe[1]

[1] Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, Marbacher Weg 6, 35037 Marburg, Germany

[2] Institut für Pharmazeutische Chemie, Zentrum für Tumor und Immunbiologie, Philipps-Universität Marburg, Hans-Meerwein-Straße 3, 35032 Marburg, Germany

[3] Zentrum für Tumor und Immunbiologie, Core Facility Medicinal Chemistry, Philipps-Universität Marburg, Hans-Meerwein-Straße 3, 35043 Marburg, Germany

Biomolecules 2021, *11*, *1837*

2.1 Introductory Remarks

This project was elaborated in cooperation with Dr. KHANG NGO, DR. CHRISTOPH P. SAGER (group of PROF. GERHARD KLEBE, Philipps Universität Marburg, Germany), Dr. FRITHJOF SCHEER, and MICHAEL DAUDE (group of PROF. WIBKE E. DIEDERICH, Philipps Universität Marburg, Germany). Laboratory procedures (ca. 80%), including crystal preparation, X-ray crystallographic data collection, structure solution, assay and ITC measurements and their evaluation were performed by the author of this thesis. Synthesis of the inhibitors (10%) was performed by Dr. KHANG NGO, DR. FRITHJOF SCHEER, and MICHAEL DAUDE. All computational ESP-calculations (ca. 10%) were performed by Dr. CHRISTOPH P. SAGER.

PROF. ANDREAS HEINE (Philipps-Universität Marburg, Germany), PROF. WIBKE E. DIEDERICH and PROF. GERHARD KLEBE helped to interpret the data.

The manuscript of the publication was collaboratively written and edited by the author of this thesis, Dr. KHANG NGO, Prof. W. DIEDERICH, PROF ANDREAS HEINE and PROF. GERHARD KLEBE. This Chapter is published in Biomolecules.

2.2 Abstract

The transient specificity pocket of aldose reductase only opens in response to specific ligands. This pocket may offer an advantage for the development of novel, more selective ligands for proteins with similar topology that lack such an adaptive pocket. Our aim was to elucidate which properties allow an inhibitor to bind in the specificity pocket. A series of inhibitors that share the same parent scaffold but differ in their attached aromatic substituents were screened using ITC and X-ray crystallography for their ability to occupy the pocket. Additionally, we investigated the electrostatic potentials and charge distribution across the attached terminal aromatic groups with respect to their potential to bind to the transient pocket of the enzyme using ESP calculations. These methods allowed us to confirm the previously established hypothesis that an electron-deficient aromatic group is an important prerequisite for opening and occupying the specificity pocket. We also demonstrated from our crystal structures that a pH shift between 5 and 8 does not affect the binding position of the ligand in the specificity pocket. This allows for a comparison between thermodynamic and crystallographic data collected at different pH values.

2.3 Introduction and Scientific Objective

A large number of proteins bind substrates and endogenous inhibitors in near-surface pockets with clearly defined cavities and exhibit no major conformational differences between the *apo*- and *holo*- forms of the protein.[74] However, many proteins possess transient binding pockets which arise from functional adaptations. For example, in the case of aldose reductase (ALR-2, EC 1.1.1.21), an oxidoreductase capable of processing a large scale of rather structurally diverse substrates of varying size and hydrophobicity, such a pocket only opens in the presence of particular substrate molecules. The opening creates an additional transient pocket volume of about 340 Å^3 and involves changes in hydrophobic contacts of the pocket-flanking residue side chains along with a plane-flip of a peptide bond once an inhibitor penetrates the pocket.[75] Structural data and molecular dynamics simulations suggest that the closed conformation of the enzyme is the more stable form and at room temperature the probability of an intermediate pocket opening seems almost barred.[76,77] Yet, for binding kinetics, the opening of the pocket is not the rate-determining step of the binding process. Until now, very little was known about the thermodynamics and kinetics of the opening and closing of such transient binding pockets.[77] Nevertheless, their role in protein-protein interactions as well as in orthosteric and allosteric modulations is known to be important in many biological processes.[78,79]

Importantly for drug discovery, such a transient binding pocket may offer a selectivity advantage over proteins that have similar topology, but no such adaptive pocket. For example, this becomes apparent in the family of aldo-keto reductases. The sequence of ALR-2 matches, with 65% identity, the sequence of the closely related aldehyde reductase (ALR-1, EC 1.1.1.2) from the same protein family. The structure of the catalytic sites of these proteins is very similar, especially in the region of the rigid residues Y48, H110, and W111 (Figure 15A and B). However, the flexible loop at the C-terminus differs significantly (Figure 15D).[80] Due to the similar residues lining the active sites of these proteins, scaffolds used to target ALR-2 may also bind to ALR-1, causing undesired side effects.[81] Therefore, it is important to identify features of the transient pocket in ALR-2 that allow for the design of more selective and discriminating inhibitors. As mentioned, ALR-2 has a transient pocket which is not observed in ALR-1 since, in this isoform, the opening of a similar pocket would require the rupture of a strong salt bridge (Figure 15C). Thus, inhibitors binding to the opened specificity pocket should selectively bind to ALR-2.[82] They could consequently serve as a promising starting point for the development of novel drugs designed to treat the effects of late-stage diabetes. Such inhibitors have the potential advantage of exhibiting reduced side effects due to the inability to bind to ALR-1, which lacks an equivalent transient binding pocket.

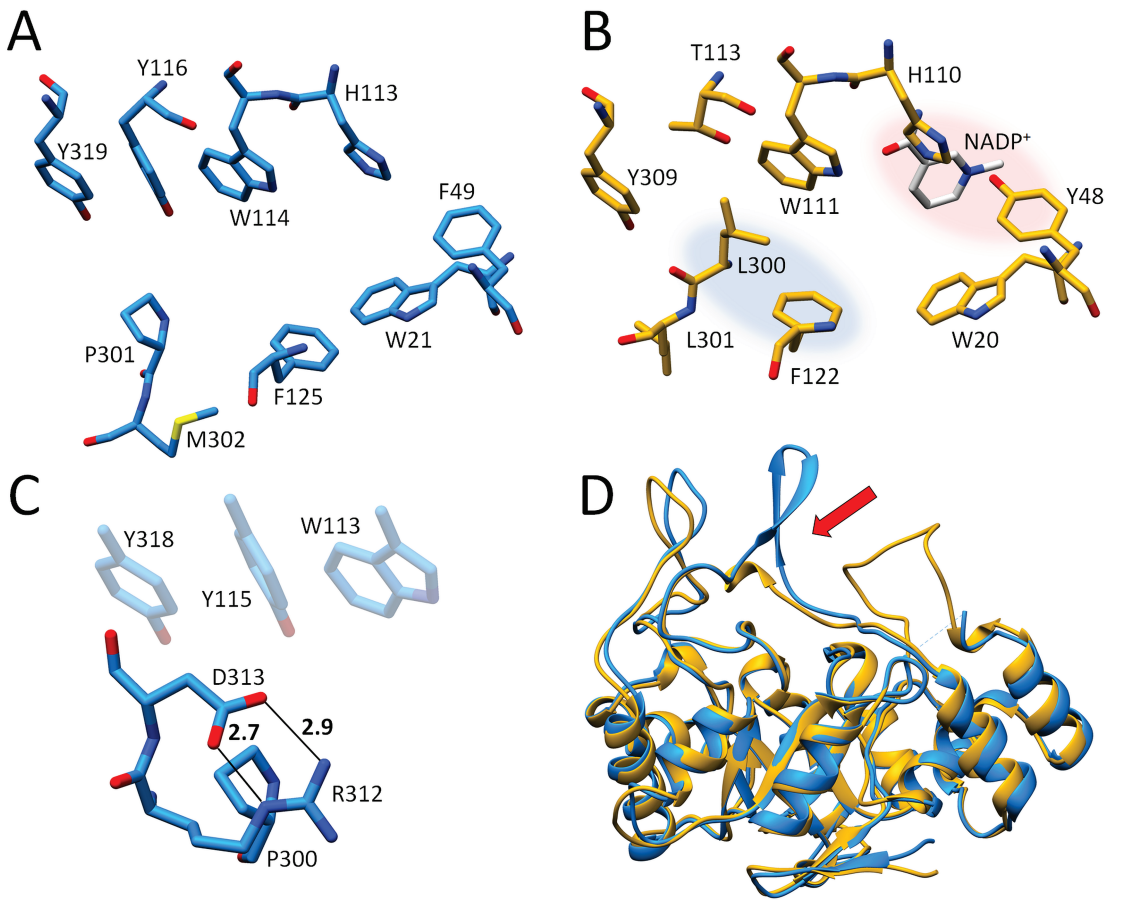


Figure 15: Comparison of ALR-1 and ALR-2. A) Active site of ALR-1, EC 1.1.1.2 (1HQT), carbon atoms light blue B) Active site of ALR-2 EC 1.1.1.21 (not deposited), carbon atoms in gold. The anion binding site is highlighted in red, the specificity pocket in blue. C) Salt bridges between R312 and D313 that would have to be ruptured upon a putative opening of a specificity pocket of ALR-1. D) Comparison of the flexible C-terminal loops of ALR-1 and ALR-2 (highlighted with a red arrow). Superposition of ALR-1, shown as a golden ribbon, with ALR-2, shown as a blue ribbon. In this and all following figures, carbon atoms are always colored in a way to distinguish and highlight particular structures, whereas oxygen atoms are displayed in red, nitrogen atoms in blue, fluorine atoms in light green and sulfur atoms in yellow to indicate their atom type.

To better understand the structural and thermodynamic binding features driving the binding and accommodation in the transient pocket of ALR-2, we investigated nine inhibitors (**2.1** – **2.9,** Figure 16) in conjunction with their potential to open and occupy this pocket. Inhibitors **2.1** and **2.2** were previously reported in a study by Rechlin *et al.*[77] and serve as a reference in the present contribution. All nine inhibitors share a common binding motif to occupy the anion binding pocket, the previously described high affinity scaffold R_1 (Figure 16), composed of the well-studied 2-arylcarbamoyl-phenoxy acetic acid moiety.[77,83]

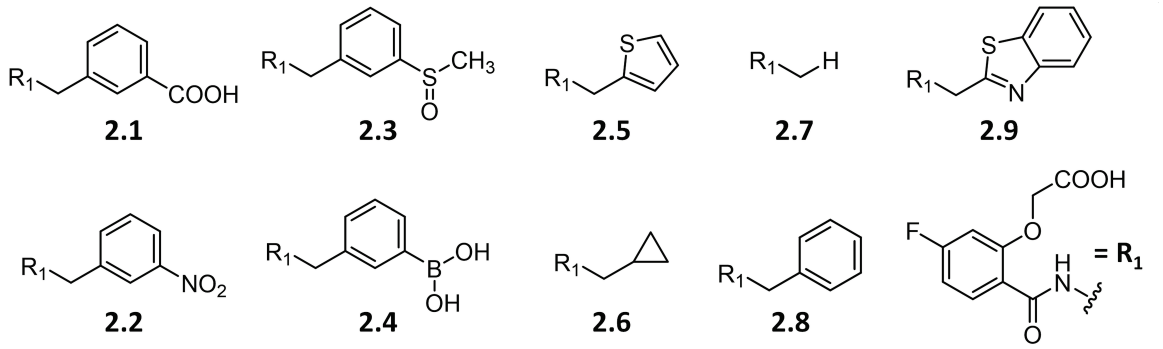


Figure 16: Schematic representation of ALR-2 inhibitors **2.1** – **2.9** with the parent scaffold R_1 (lower right). These inhibitors were used to study the binding features of transient binding pocket of ALR-2. Inhibitors **2.1** – **2.2** and **2.7** – **2.9** were previously investigated in a study by us (RECHLIN *et al.).*[77]

2.4 Results

2.4.1 Enzyme Kinetic and Thermodynamic Data

To gain more detailed insights into the binding behavior of aldose reductase inhibitors, the affinities of **2.3** – **2.6** were determined using an enzyme kinetic fluorescence assay. To further elucidate the binding behavior of aldose reductase inhibitors, we additionally applied ITC. As the dissociation constants K_d show, inhibitors **2.3** and **2.4** have single-digit micromolar affinities while **2.5** and **2.6** bind more weakly, in the two to three-digit micromolar range (Table 1). As inhibitors **2.5** and **2.6** showed low potency, the Gibbs free energy of binding ($\Delta G°$) (derived from the dissociation constant K_d, determined by ITC) and the enthalpy ($\Delta H°$) could not be determined with high accuracy by direct titrations. To record data at larger *c*-values[64,84-86] and thus to obtain more accurate K_d values, displacement titrations were performed with the reference inhibitor **2.9** (Figure 65, Appendix). Displacement titrations with the same reference, inhibitor **2.9**, were similarly performed for inhibitor **2.1**, which also showed very low affinity as previously characterized by RECHLIN *et al.*[77] In contrast, inhibitor **2.2** showed a remarkably high affinity. Since the *c*-value[64,84-86] from a direct titration of inhibitor **2.2** was too large and thus no reasonable K_d value could be extracted, the thermodynamic signature of this ligand was determined by a displacement titration using the weak reference inhibitor **2.7**.

A common phenomenon when measuring the thermodynamic parameters of a complex formation between an inhibitor and a protein is the putative superposition of changes in protonation states during protein binding.[64] Such changes can mask and thus obscure the thermodynamic signature of the binding event itself. Thus, before any reasonable analysis of the recorded thermodynamic data can be performed, such changes in the protonation states of the

inhibitor and protein need to be considered and corrected. Previous measurements for the same class of ligands showed that the binding event was accompanied by an uptake of 0.8 moles of protons on average, which remains constant across the series.[87] Therefore, it is reasonable to assume that the relative difference in the thermodynamic profile remains unaffected for the whole series.[87-89] For this reason, all inhibitors in this series were measured solely in 10 mM HEPES (4-(2-Hydroxyethyl)-1-piperazine ethane sulfonic acid) buffer at pH 8, which allowed a relative comparison of thermodynamic binding data across the compound series. Furthermore, dimethyl sulfoxide (DMSO), added as solubility enhancer, has been described to bind to ALR-2.[90] As weak inhibitor, DMSO will be displaced from the binding site upon accommodation of our more potent ligands. Since this contribution will likely be the same for all studied compounds, it will cancel out in our relative comparison.

The binding enthalpy $\Delta H°$ of the ALR-2 inhibitors was extracted from the ITC data using Equation 12 as described in Chapter 6.10.2. According to the Gibbs−Helmholtz equation (Equation 1), the entropic contribution $-T\Delta S°$ to binding was calculated as the difference between $\Delta G°$ and $\Delta H°$. The results of the enzyme kinetic and thermodynamic measurements are presented in Table 1. Figure 17 shows the thermodynamic profiles of the inhibitors investigated, including those from our previous work for comparison.[77]

The ITC measurements were performed by different researchers. Since different protein charges and ITC parameters, such as the ambient temperature and local humidity, may influence the ITC measurements, the comparability of the thermodynamic data collected in this work and the previously recorded data had to be assessed. To first compare the relative thermodynamic values, the high-affinity reference inhibitor **2.9** used for displacement titrations was measured by direct titration, and the results were compared with previous data (Table 2). Regarding the measurements of RECHLIN *et al.*, the K_d value differs by a factor of eight compared to the data measured herein. However, the thermodynamic signature of inhibitor **2.9** seems to be very similar considering the results of both measurements (Figure 18). The enthalpy contribution of the complex formation with ALR-2 is equal (−53.3 and −54.0 kJ/mol); however, a closer look reveals a difference of $-T\Delta\Delta S°$= 4.4 kJ/mol in the entropy range of the previously measured inhibitor **2.9**, also reflected in the affinity and Gibbs free energy.[77] In this work, we repeated the original method as closely as possible, but the measurements did not lead to identical results in the overall amount of enthalpy/entropy compensation. Thus, a direct comparison of the data from both studies should be treated with some caution. Although an absolute comparison is not possible, a relative comparison of the thermodynamic signatures of the inhibitors is, however, justified.

Table 1: Kinetic and thermodynamic parameters of the investigated ALR-2 inhibitor series.

Inhibitor	IC_{50} [a] [nM]	K_d [µM]	$\Delta G°$ [kJ/mol]	$\Delta H°$ [kJ/mol]	$-T\Delta S°$ [kJ/mol]	titration mode
2.1	-	36.2 ± 11.4 [b]	–25.4 ± 0.7 [b]	–36.4 ± 2.9 [b]	11.0 ± 3.6 [b]	displ. [b, c]
2.2	47.3 [b]	0.018 ± 0.003 [b]	–44.3 ± 0.5 [b]	–65.6 ± 0.5 [b]	21.3 ± 0.9 [b]	displ. [b, d]
2.3	198.5 ± 8.8	5.0 ± 0.6	–30.3 ± 0.3	–39.1 ± 0.9	8.8 ± 0.9	direct
2.4	190.0 ± 9.7	2.3 ± 0.3	–32.2 ± 0.3	–45.0 ± 0.7	12.8 ± 0.8	direct
2.5	1518.0 ± 241.1	77.6 ± 15.9	–23.4 ± 0.5	–30.0 ± 1.4	6.6 ± 1.9	displ. [c]
2.6	1440.0 ± 154.3	107.7 ± 16.6	–22.6 ± 0.4	–37.5 ± 0.8	14.9 ± 1.2	displ. [c]
2.7	-	6.6 ± 0.4 [b]	–29.6 ± 0.2 [b]	–35.4 ± 0.3 [b]	5.8 ± 0.4 [b]	direct [b]
2.7	-	7.2 ± 0.9 [b]	–29.4 ± 0.3 [b]	–39.0 ± 1.0 [b]	9.6 ± 1.2 [b]	displ. [b, c]
2.8	-	7.5 ± 1.4 [b]	–29.3 ± 0.4 [b]	–29.9 ± 0.4 [b]	0.6 ± 0.4 [b]	direct [b]
2.8	-	9.1 ± 0.7 [b]	–28.8 ± 0.2 [b]	–30.6 ± 0.5 [b]	1.8 ± 0.7 [b]	displ. [b, c]

The thermodynamic data were determined by ITC in HEPES buffer pH 8.0. The data are not corrected for the putatively overlaying protonation effect. [a] IC_{50} was determined by an enzyme kinetic fluorescence assay. [b] Data were collected by RECHLIN *et al.*[77] [c] Thermodynamic data were determined with **2.9** as strong reference inhibitor. [d] Thermodynamic data were determined with **2.7** as weak reference inhibitor.

It has been previously demonstrated with inhibitors **2.1** and **2.2** that differences in affinity can differ significantly by exchanging a single substituent at the terminal aromatic ring. RECHLIN *et al*. discovered that inhibitor **2.2**, with a nitrophenyl moiety, has a nanomolar K_d, while the same scaffold decorated with an isosteric carboxylate group (inhibitor **2.1**) binds in the micromolar range Table 1.[77] Yet, the inhibitors **2.3** and **2.4** as well as **2.5** and **2.6** are similar, especially with regard to their free enthalpy of binding.

Although **2.3** binds much less enthalpically than **2.4**, the sulfoxide shows a slight entropic advantage, resulting in a similar affinity of both inhibitors. The K_d values of the ITC measurements reflect the high potency of **2.3** and **2.4**, which matches with the half-maximum inhibitory concentration IC_{50} values determined in the kinetic fluorescence assay, as summarized in Table 1. The same is evident for the pair **2.5** and **2.6**. Both inhibitors have a sole hydrophobic terminal substituent and are significantly less potent than the other inhibitors in the series. Inhibitor **2.5** has an entropic advantage over **2.6**; however, it exhibits a similar Gibbs free energy as **2.6** due to a less negative enthalpy value.

Table 2: Thermodynamic parameters of **2.9** measured in this and a previous study of RECHLIN *et al.*[77]

Inhibitor	K_d [µM]	$\Delta G°$ [kJ/mol]	$\Delta H°$ [kJ/mol]	$-T\Delta S°$ [kJ/mol]	titration mode
2.9 (from this study)	0.42 ± 0.08	–36.4 ± 0.5	–53.3 ± 0.5	16.9 ± 0.7	direct
2.9 (derived from [77])	0.05 ± 0.01	–41.5 ± 0.4	–54.0 ± 1.0	12.5 ± 1.1	direct

The thermodynamic data were determined by ITC in HEPES buffer pH 8.0. The data are not corrected for the overlaying protonation effect.

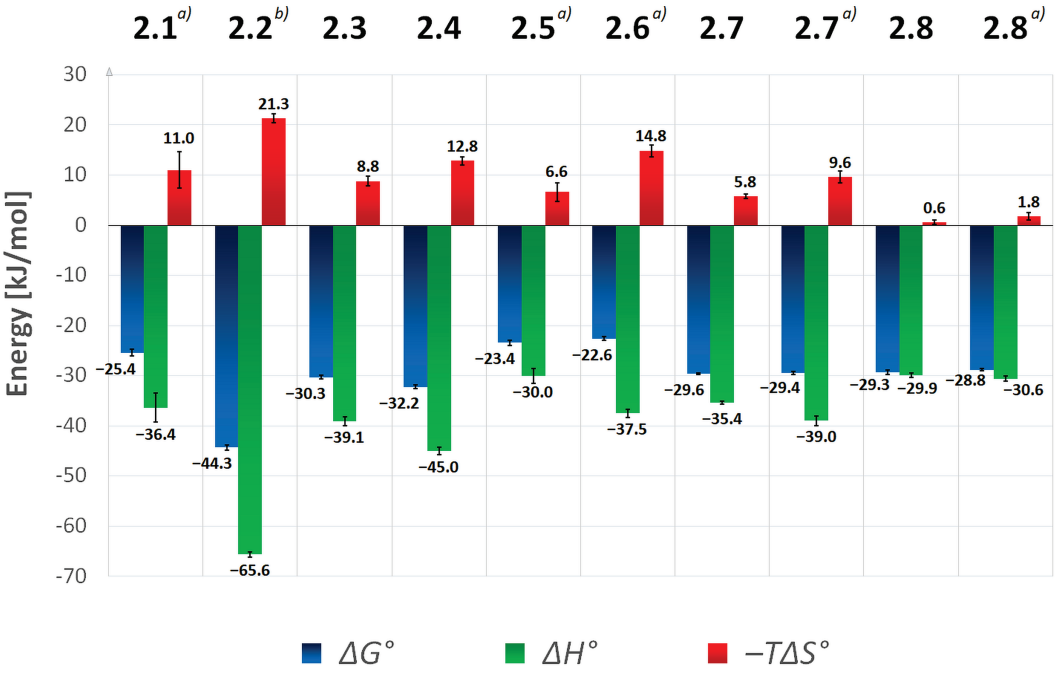


Figure 17: Thermodynamic parameters of ALR-2 inhibitors determined by ITC. Data for inhibitors **2.1**, **2.2**, **2.7** – **2.8** were previously collected by RECHLIN *et al.*[77] All measurements were performed in HEPES buffer pH 8.0 and are not corrected for overlaying protonation effects, thus only relative differences should be interpreted. Standard deviations of at least three measurements were averaged and standard deviations (SD's) are displayed as error bars. [a)] Analyzed by displacement titrations using **2.9** as a strong inhibitor. [b)] Analyzed by displacement titrations using **2.7** as a weak inhibitor.

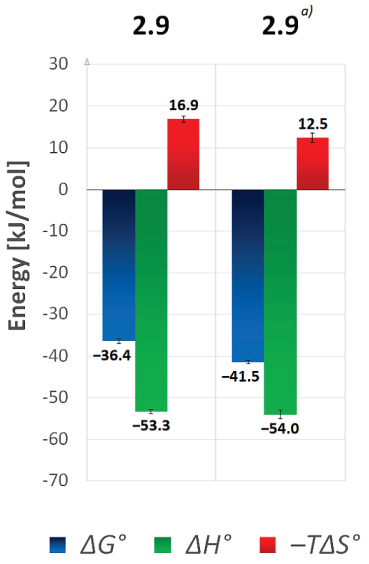


Figure 18: Thermodynamic parameters of **2.9** determined by direct ITC titrations. The thermodynamic data determined in this work (left) are compared with those [a)] previously determined by RECHLIN *et al.*[77] (right). Measurements were performed in HEPES buffer and are not corrected for superimposed protonation effects. SD's of triplicate measurements are displayed as error bars.

2.4.2 Crystal Structures of Human Aldose Reductase Inhibitor Complexes

Here, eight crystal structures of complexes of wild type ALR-2 with inhibitors **2.1** – **2.8** are presented and analyzed. The obtained X-ray structures for each complex ranged from the very high resolution of 0.93 to 1.19 Å, which makes interpretation of many structural details possible (Table 3). All structures, except for the ALR-2 • 5 complex, were deposited in the PDB. The crystal structure of inhibitor **2.1** was previously determined by soaking at a pH of 8. In order to validate whether soaking at pH 5 or 8 causes any impact on structure, the analysis with this ligand was repeated at the lower pH.

Table 3: Summary of the crystallographic data of investigated ALR-2 inhibitor complexes, their PDB-codes and the opening status of the transient specificity pocket.

Inhibitor	PDB code	Resolution [Å]	Pocket state
2.1	4YS1 [a)]	1.07	closed
2.2	4QBX [a)]	0.98	open
2.3	6TUF	1.15	hybrid
2.4	6TUC	1.06	closed
2.5	not deposited	0.93	closed
2.6	6SWY	0.93	closed
2.7	4PUU [a)]	1.14	closed
2.8	4Q7B [a)]	1.19	open

[a)] Crystallographic data determined by RECHLIN *et al.* [77]

2.4.2.1 Effects of Different pH Conditions on the Terminal Acetic Acid Carboxy Group at the Parent Scaffold

A comparison of inhibitor **2.1** in complex with ALR-2, soaked at both pH 5 and 8, revealed no differences in the position of the inhibitor between the two structures (Root mean square determination (RMSD) value of 0.12 Å in the position of the backbone atoms (Figure 25)). The terminal acetic acid carboxylate group bound the anion binding pocket in two orientations (see below) and did not differ in occupancy when crystallized at both pH values. Thus, no dependence on the pH value was observed. The pK_a value of the acetic acid moiety is below 4.[91] However, its spatial proximity to the positively charged side chain of H110 likely supports the presence of a deprotonated carboxylate group in this region. Therefore, it can be assumed that this group of the ligand is charged in the complex at both pH 5 and 8, so a difference in the structural geometry is rather unlikely.

2.4.2.2 Comparison of the Binding Poses of Inhibitors 2.1 – 2.6

A close comparison of ALR-2 inhibitors **2.1** – **2.6** revealed that they all share a similar binding mode in the anionic binding pocket (see Chapter 1.2.1.3). Inhibitors **2.1** and **2.2**, investigated in our previous study,[77] were used as a reference to characterize and compare the binding modes of the candidates **2.3** – **2.6** studied here.

With regards to scaffold R_1 (Figure 16), which is shared by all inhibitors, the terminal carboxylate group can adopt two alternative conformations (or poses in the following). This is shown, for example, by the ALR-2 complex with inhibitor **2.1**, for which the structure refines to 77% inhibitor occupation (Figure 19B). In the first pose (occupancy 28%), the carboxylate group forms H-bonds simultaneously to H110, W111, and Y48, whereas in the second pose (b) (occupancy 49%) H110, Y48, and the water molecule O1b (d = 2.8 Å) are in contact. O1b interacts additionally with the inhibitor's amide carbonyl oxygen and ether oxygen, and thereby stabilizes the bound conformation of inhibitor **2.1**. With this geometry, the inhibitor's terminal benzoic acid moiety turns outward, and the specificity pocket remains in the closed state. The inhibitor protrudes from the surface of the protein (Figure 19A). The so-called gatekeeper residues, L300 and L301, already described in chapter 1.2.1.3, adopt an ordered conformation and seal the specificity pocket. The terminal carboxylic acid function at the phenyl ring is not resolved in the $F_O - F_C$ difference in electron density at the 3σ level, likely due to two competitive orientations of the *meta*-attached acid group along with the enhanced residual mobility of this ligand portion. Therefore, it was not assigned in the final deposited structure.[77] In summary, this inhibitor binds in a well-defined conformation outside the specificity pocket, which remains in the closed state.

In comparison to **2.1**, the binding pose of **2.2** refines to full occupancy, as it lacks the presence of the water molecule O1b. This is accompanied by the formation of a weak, intramolecular H-bond between the inhibitor's carboxylate and amide groups (d = 3.0 Å, Figure 19D). The latter group is flipped over by 180° and shifts the phenyl ring of the terminal group towards the former positions of the gatekeeper residues. Therefore, L300 and L301 give way and the specificity pocket fully opens to accommodate inhibitor **2.2** (Figure 19C). The nitro group of **2.2** interacts with L300, supporting its spatial fixation in the specificity pocket. The observed non-classical secondary H-bond between the C-H dipole of the phenyl ring of Y309 and the negatively polarized oxygen atom of the nitro group has been described previously.[88] Additionally, the phenyl ring of the terminal group is stabilized by a π-π-stacking with the aromatic system of W111 in the opened specificity pocket (d = ~3.4 Å).[37] In the case of **2.2**, the amide carbonyl group points in the opposite direction when compared to **2.1** and interacts with the water molecule O4 (d = 2.7 Å). Remarkably, the carboxylate group at the acetic acid moiety of **2.2** binds with only one well-defined orientation in the anion binding pocket. While the complex with inhibitor **2.1** gives rise to a closed specificity pocket, **2.2** mirrors a completely

opened pocket. For this reason, the poses of both inhibitors were used as reference extremes to compare the binding modes of **2.3** – **2.6** in detail.

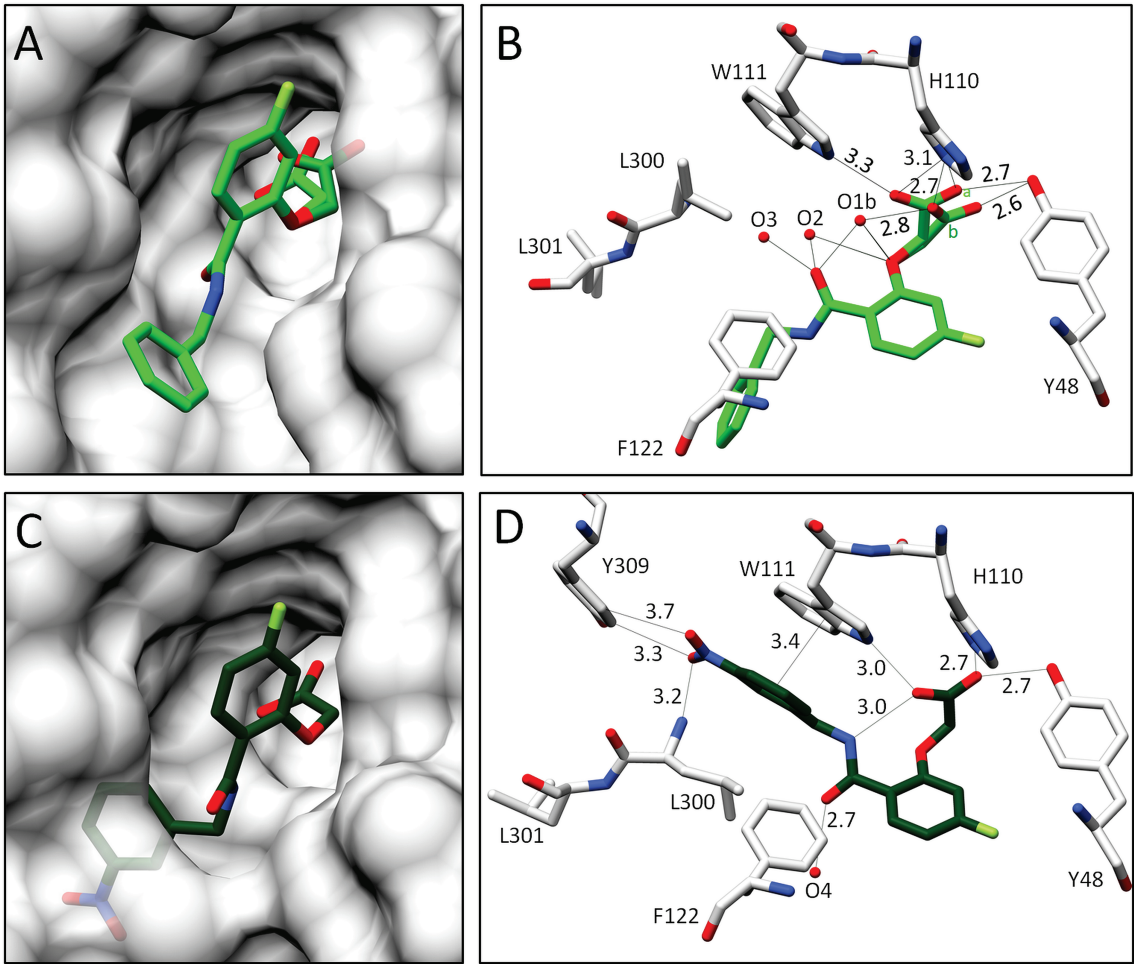


Figure 19: Crystal structures of inhibitors **2.1** and **2.2** in complex with ALR-2 wild type. Inhibitors **2.1** (4YS1)[77] in light green (A, B) and **2.2** (4QBX)[77] in dark green (C, D) bound to the active site of ALR-2. To distinguish the position of both conformations, b is highlighted by slightly darker color. On the left, the protein is depicted by its transparent solvent-accessible surface (light gray) whereas on the right, the interactions are indicated as black lines. Selected residues are displayed for better orientation. Oxygen atoms are shown in red, nitrogen atoms in blue, and fluorine atoms in light green.

Inhibitor **2.3** seemed to bind in a complex with ALR-2 in a similar fashion to inhibitor **2.2**. However, a more careful examination reveals that the gatekeeper residues L300 and L301 can be refined in two different orientations (Figure 20B). Additionally, the terminal sulfoxide portion and the adjacent phenyl ring of **2.3** adopt two alternative conformations, while the placement of the R_1 scaffold adopts identical geometry for both conformations. The confirmation b of **2.3** binds into the opened specificity pocket with 40% occupation. Similar to **2.2**, the phenyl ring of the terminal moiety is able to stabilize the position of the inhibitor within the opened specificity pocket by π-π-stacking with the aromatic system of W111 (d = ~3.4 Å).

However, on closer inspection of the $2F_O - F_C$ density (Figure 21A), L300 and L301 are also visible in an orientation that seals the specificity pocket. In this orientation, the inhibitor's terminal portion would collide with the gatekeeper residues. The observed electron density distribution suggests that **2.3** adopts a second conformation (a) and remains 60% outside of the specificity pocket, which keeps the closed state (Figure 21A). Similar to the binding pose of **2.1**, any electron density indicating the placement of the terminal *meta*-attached phenylsulfoxide group in the closed state is missing. Therefore, it could not be added to the finally deposited structural model. Since the terminal portion of **2.3** in conformation a is not involved in any strong directional interaction, it is likely that it remains with much higher residual mobility compared to conformation b. The split binding mode of **2.3** agrees with the double conformations of the gatekeeper residues.

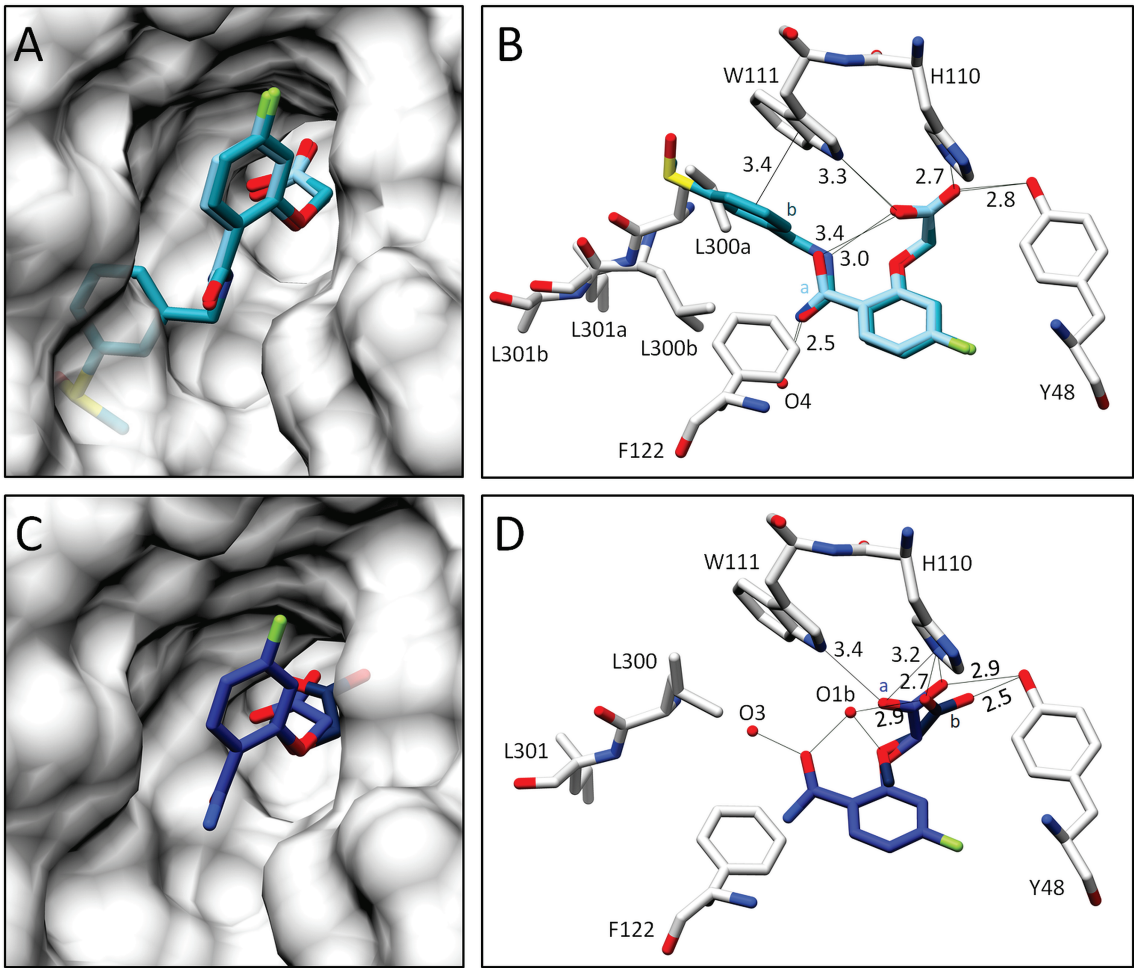


Figure 20: Crystal structures of inhibitors **2.3** and **2.4** in complex with ALR-2 wild type. Inhibitors **2.3** (6TUF) in light blue (A, B) and **2.4** (6TUC) in blue (C, D) bound to the active site of ALR-2. To distinguish the position of both conformations, b is highlighted by slightly darker color. On the left, the protein is depicted by its transparent solvent-accessible surface (light gray) whereas on the right, the interactions are indicated as black lines. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow and fluorine atoms in green.

The boronic acid derivative **2.4** refines to full occupancy and adopts an orientation outside the specificity pocket. The gatekeeper residues keep the specificity pocket closed and suggest that **2.4** is unable to open the transient binding pocket (Figure 20D). Hence, the terminal part of the inhibitor, including the boronic acid function, is not visible in the electron density and therefore could not be modeled. Next to the anion binding pocket, a large portion of the positive difference electron density (Figure 21B) is observed indicating that this part of the inhibitor binds with high residual mobility and likely adopts at least two alternative orientations as already described for **2.1**. Furthermore, the presence of O1b next to the carboxylate group in orientation b (36% occupancy) further suggests a binding mode outside the specificity pocket.

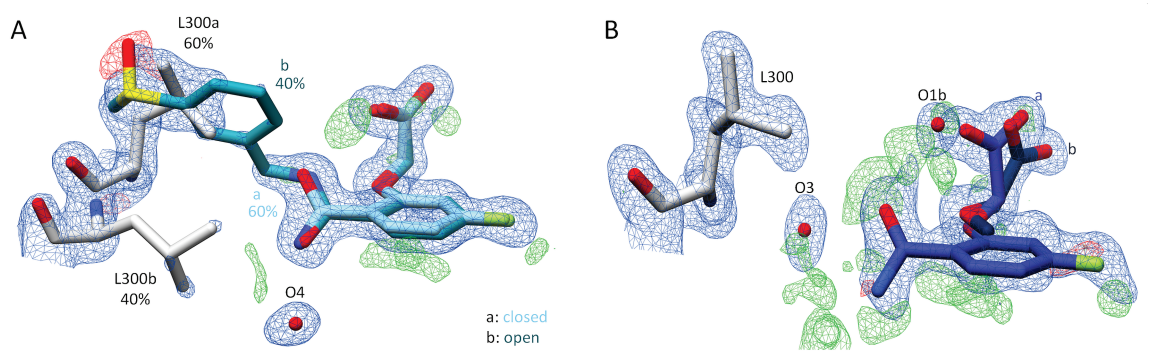


Figure 21: Representation of **2.3** and **2.4** in complex with ALR-2. A) Representation of **2.3** and its electron density. Inhibitor **2.3** (6TUF) in light blue. To distinguish the conformations of **2.3**, conformation b is highlighted by a slightly darker color. Conformations a and b of the gatekeeper residue L300 are shown in grey. The respective occupancy is indicated in each case. The difference electron density map ($F_O - F_C$) is depicted as red (negative), and green (positive) meshes at the 3σ level. The $2F_O - F_C$ density is depicted as blue mesh at the 1σ level. B) Representation of inhibitor **2.4** (6TUC) in blue and its electron density. To distinguish the position of the inhibitor conformations, conformation b is highlighted by a slightly darker color. The gatekeeper residue L300 is shown in grey. The difference electron density map ($F_O - F_C$) is depicted as red (negative), and green (positive) meshes at the 3σ level. The $2F_O - F_C$ density is depicted as blue mesh at the 1σ level.

As described above, the binding mode of **2.3** showed two orientations, either in or outside the specificity pocket. A detailed analysis of the difference electron density in the anion binding pocket (Figure 22) suggests also here presence of water molecule O1b, which agrees well with the situation in the complexes in **2.3** and **2.4**. Furthermore, the remaining density next to the carboxylate group can be assigned to a second placement of this group in the anion binding pocket, in agreement with the binding poses found for the inhibitors binding outside the specificity pocket.

Inhibitors **2.5** and **2.6** exhibit a significantly smaller terminal substituent compared to other inhibitors (Figure 16) and consequently require less space to bind in the protein. This might

explain why these inhibitors deviate in their binding pose from the pattern seen for the reference inhibitors **2.1** and **2.2**.

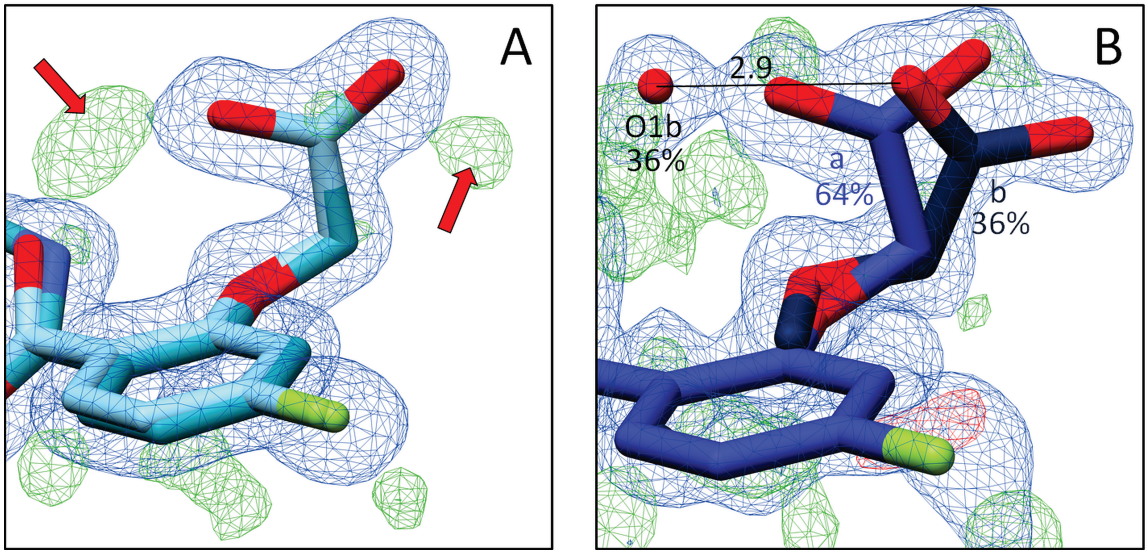


Figure 22: Representation of the carboxylate group of the parent scaffold of inhibitors **2.3** and **2.4**. A) Inhibitor **2.3** (6TUF) in light blue. The red arrows indicate the likely positions of the additional water molecule O1b and the second conformation of the acetic acid moiety in the complex with **2.3**. B) Inhibitor **2.4** (6TUC) in blue in complex with ALR-2. To distinguish the conformations, b is highlighted by a slightly darker color. Interactions are indicated as black lines. The difference electron density map ($F_O - F_C$) is depicted as red and green meshes at the 3σ level. The $2F_O - F_C$ density is depicted as blue mesh at the 1σ level. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow and fluorine atoms in green.

In the complex with **2.5**, the specificity pocket remains in closed state (Figure 23A). This is well defined by the electron density assigned to the gatekeeper residues (Figure 24A). Nevertheless, a closer inspection of the electron density next to the refined position shows that the modeled inhibitor conformation, refined to an occupancy of 87%, is supposedly not the only assignable one. However, the residual density was not sufficient to identify and model an additional conformer of **2.5**. Thus, any conclusions on a split orientation of the carboxylate group in the anion binding pocket as observed for **2.1** and **2.4** along with the presence a partially occupied water molecule O1b is difficult to assign (Figure 24A). The terminal thiophene moiety, which must bind outside the specificity pocket, is not detectable in the residual density. Despite multiple data set collections of this complex and extended analysis, e.g., by generating polder maps, it was not possible to determine the position of the thiophen moiety of **2.5** or the assignment of a second conformation of the entire inhibitor. Because of these uncertainties, the crystal structure of this complex has not been deposited in the PDB.

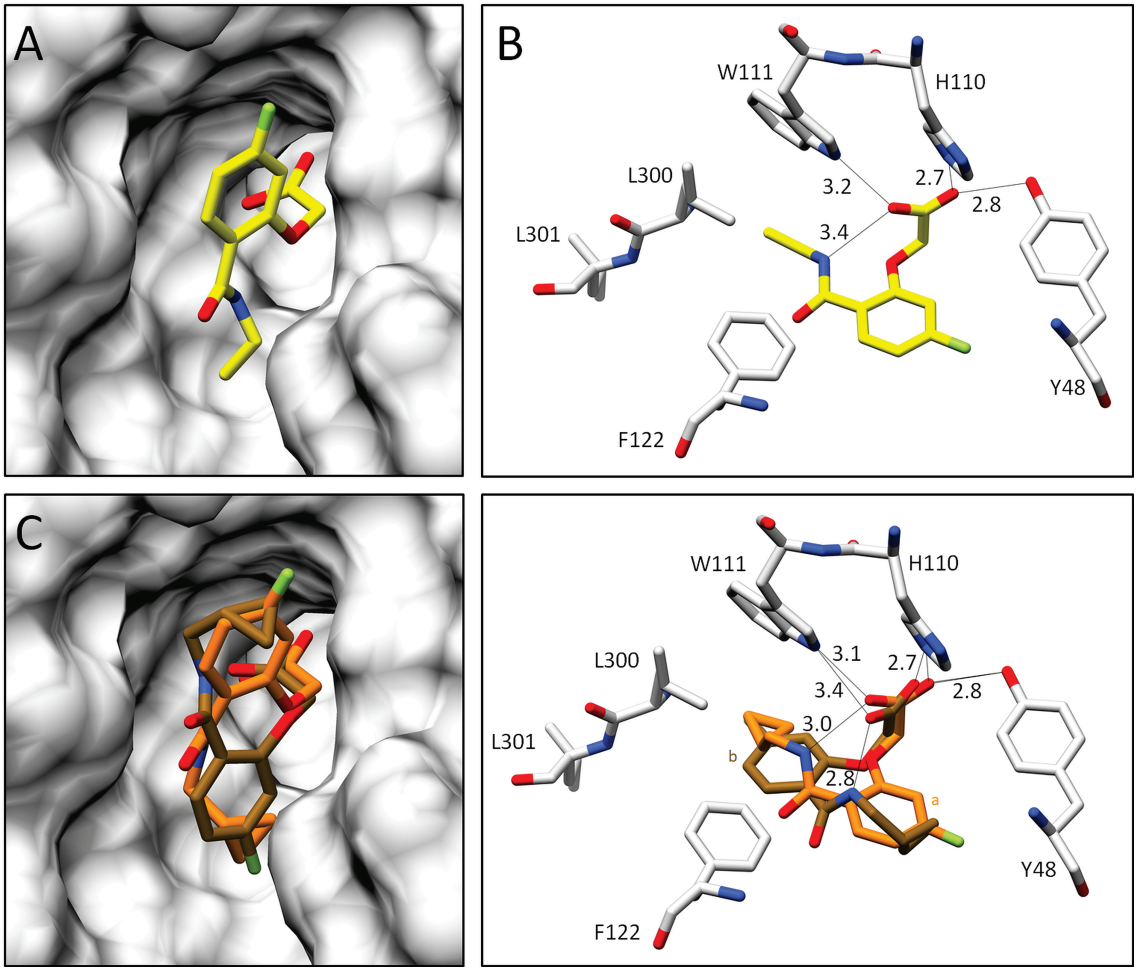


Figure 23: Crystal structures of inhibitors **2.5** and **2.6** in complex with ALR-2. Inhibitors **2.5** (not deposited) in yellow (A, B) and **2.6** (6SYW) in orange (C, D) bound to the active site of ALR-2. To distinguish the position of the conformations, b is highlighted by a slightly darker color. On the left, the protein is depicted by its transparent solvent-accessible surface (light gray) whereas on the right, the interactions are indicated as black lines. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow and fluorine atoms in green.

Additionally, in the ALR-2 in complex with inhibitor **2.6**, the transient specificity pocket remained closed, and no crystallographic water molecules could be assigned in the binding pocket next to the inhibitor (Figure 23C, D). As in the case of **2.5**, the $2F_O - F_C$ electron density assigned to the gatekeeper residues undoubtedly suggests binding to the closed state (Figure 24B). However, the $F_O - F_C$ density indicates an alternative placement of **2.6** even in the closed state. In the first orientation (conformation a, 42% occupancy) the amide bond occupies a position similar to that of **2.2**. This is likely possible because the cyclopropyl ring does not require much space and L300 can remain in a position usually found for the transient binding pocket in the closed state. In addition, conformation b with 58% occupation could be successfully refined (Figure 24B). Again, likely due to the minor spatial requirements of the cyclopropyl moiety, inhibitor 6 is able to flip over and bind with its fluorophenyl ring toward the

gatekeeper residues. The cyclopropyl ring instead adopts the original position of the latter aromatic portion. However, with inspection of the difference electron density ($F_O - F_C$), it can be suggested that **2.6** still exhibits high residual mobility and adopts additional conformations within the binding pocket. Nonetheless, the still unexplained difference electron density difference does not allow modeling of further placements. The previously described inhibitor **2.5** has a short terminal substituent similar to **2.6**, and the residual difference electron density ($F_O - F_C$) also suggests presence of further conformers. Therefore, a flipped orientation of the inhibitor may also be given in this case. Since the assigned geometry already explains 87% occupancy, no further poses were modeled for **2.5**.

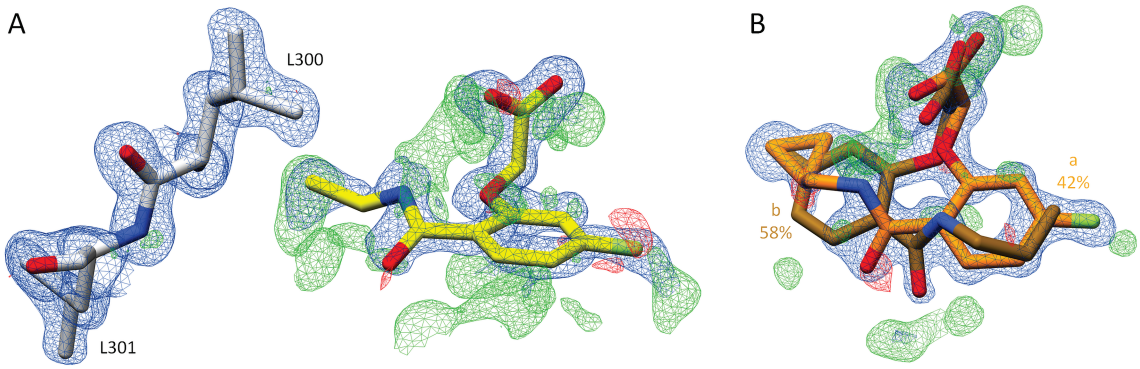


Figure 24: Representation of **2.5** and **2.6** in complex with ALR-2. A) Crystal structure of ALR-2 • **2.5** (not deposited) and the electron density around the ligand. Inhibitor **2.5** is shown in yellow; gatekeeper residues L300 and L301 are indicated in grey. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow and fluorine atoms in green. B) Crystal structure of ALR-2 • **2.6** (6SYW) and the corresponding electron density. Inhibitor **2.6** is shown in orange. To distinguish the conformations of **2.6**, b is highlighted by a slightly darker color. The respective occupancy is indicated in each case. Difference electron density map ($F_O - F_C$) is depicted in both cases as red, and green meshes at a contour level of 3σ. The $2F_O - F_C$ density is depicted as blue mesh at the 1σ level.

In summary, the new inhibitors presented here, apart from **2.3**, bind outside the specificity pocket, which remains in the closed state. For **2.3**, both placements were observed. Binding of the terminal acetic acid moiety to the anionic binding pocket was well agreed, and a split binding pose with two orientations was supposedly found in the case of binding to the closed state. The fluorophenyl ring of all inhibitors binds to the active binding site almost exclusively at the same position, except for orientation b in the ALR-2 complex with inhibitor **2.6**, where this ligand is flipped over. The *meta*-attached functional groups at the terminal phenyl ring (**2.1**, **2.3**, and **2.4**) remain undefined in the electron density maps, likely due to residual mobility and scatter over at least two orientations. Once placed in the transient pocket, the terminal phenyl ring is well defined in the electron density maps (**2.2** and **2.3**).

2.5 Discussion

1.1.1 Effects of Differences in pH on the Terminal Carboxy Group of the Parent Scaffold

In the analysis of all investigated crystal structures, the carboxylate group of the acetic acid group at inhibitor parent scaffold attracts attention, as it can occur in two conformations. This phenomenon has already been discovered by RUIZ *et al*.[91] The second conformation of the carboxylate group was found for the first time in this study in association with a crystal structure soaked at pH 8 instead of the previously applied pH 5. Based on this investigation, RUIZ *et al*. concluded that despite different crystallization conditions and space groups, the complex geometry is nearly the same and the most prominent structural difference is the double conformation of the acetic acid group. However, for **2.1** – **2.6**, which were all crystallized and soaked at pH 5, except **2.1** and **2.2**, we hypothesize that the double conformation and the occurrence of water O1b can also be detected at soaking conditions of pH 5. To exclude a significant difference in the complex geometry, the ALR-2 complexes studied by RECHLIN *et al.* must be compared with the complexes **2.3** – **2.6** soaked at pH 5. For this purpose, an additional structure of complex **2.1**, now crystallized and soaked at pH 5, was determined.

A comparison of inhibitor **2.1** in complex with ALR-2, soaked at both pH 8 and pH 5 (Figure 25) reveals no differences in the position of the inhibitor between the two structures. The terminal carboxylate group is also present in two orientations and does not differ in occupancy from the complex soaked at pH 8. This leads to the conclusion that the double conformation of the acetic acid moiety is not dependent on the applied pH conditions. The carboxylic acid derivatives have a pK_a value below 4.[91] The spatial proximity to the positively charged side chain of H110 further supports the presence of a deprotonated carboxylate group in the inhibitor. Therefore, it can be assumed that this group of the ligand is charged in the complex either at pH 5 and 8, so that a difference in the structural geometry is rather unlikely for this reason.

To confirm the equivalence of the entire protein structure at both pH values not only visually, but also to ensure that the altered pH does not cause further differences in the structural geometry, an RMSD (root-mean-square-deviation of atomic positions) value of the two complexes with **2.1** was calculated. The RMSD is the mean distance between the atoms of two or more superimposed proteins, which is usually based on the atoms of the backbone. The similarity is determined by the RMSD of the Cα atomic coordinates according to an optimal rigid body superposition.[92,93] RMSD values were calculated by Equation 7;[93] d_i is the distance between atom *i* and either a reference structure or the mean position of the *n*-equivalent atoms. In this case it was calculated for the Cα atoms of the residues 0 to 315 of the pH 5 and the pH 8 structures of **2.1**. Fitting was performed using the McLachlan algorithm[94] as implemented in the program ProFit.[95]

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}$$ Equation 7

An RMSD value of 0.12 Å, (Figure 25), indicates that both structures are highly similar and show only minor differences in the position of the backbone atoms. To assess this result, the value of the same sequence stretch of the complex with **2.1** and **2.2** soaked at pH 8 was determined. As already indicated in Chapter 2.4.2, the two complex structures differ significantly in the area of the gatekeeper residues. While L300 and L301 are closed the transient pocket in the case of **2.1**, they take a completely different spatial position in the complex with **2.2**, in consequence of the opening the specificity pocket upon inhibitor binding. As expected, the RMSD value of 0.31 Å is significantly higher, indicating a major structural difference.

Comparing only the gatekeeper motif (A299-A302), the structural differences in atomic positions become even more obvious resulting in a value of 1.01 Å. An RMSD of 0.08 Å computed for the same region between the structures of **2.1** soaked at pH 5 and pH 8 confirms that the shifted pH value does not cause any structural differences. Consequently, despite different crystallization conditions, the structures are well comparable.

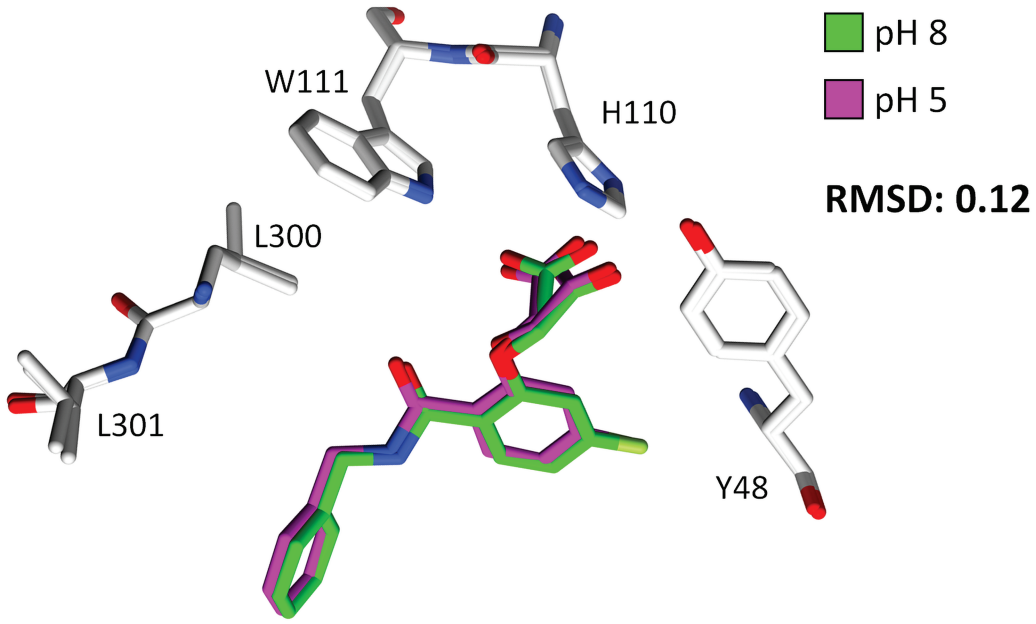


Figure 25: Comparison of crystallographic structures of **2.1** soaked at different pH values. Inhibitor **2.1** in complex with ALR-2, soaked at pH 8 (4YS1, green)[77] and pH 5 (not deposited, magenta). Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, and fluorine atoms in green.

2.5.1 Structural and Thermodynamic Comparison

Previously, it was suggested that the ability of the terminal aromatic group of inhibitors **2.1** and **2.2** to penetrate and accommodate the specificity pocket does not only depend on its ability to form enthalpically favorable interactions in this pocket, but rather it also relies on the energy contribution necessary for desolvating this group upon binding.[78] In fact, to occupy the specificity pocket, either the charged carboxylate group of inhibitor **2.1** or the uncharged nitro group of inhibitor **2.2** must fully discard their hydration shell. Desolvating a charged group is by far more expansive. Thus, inhibitor **2.1** does not open the specificity pocket and therefore remains outside and protrudes from the protein. There, it remains partially exposed to the solvent. In contrast, the nitro group of **2.2** is energetically less costly to desolvate, and a 1000-fold more potent binding into the pocket is observed. The significant decrease in the entropic contribution of **2.2** compared to **2.1** (Figure 17) was related to a remarkably stronger fixation of the gatekeeper residues and adjacent residues. This was concluded from a *B*-factor analysis.[77] Furthermore, the significantly higher affinity of inhibitor **2.2** over **2.1** may additionally result from the formation of an H-bond to one of the oxygen atoms of the nitro group by the amide group L300 within the specificity pocket. A π-π-stacking of the nitrophenyl ring to the aromatic indole moiety of W111, described in chapter 2.4.2, is also partly responsible for the enthalpic advantage of **2.2**.

To validate the hypothesis that the terminal group should not bear a charged moiety, but a functional group, to undergo hydrogen bonding within the transient pocket, the sulfoxide **2.3** and the boronic acid derivative **2.4** were synthesized and analyzed with respect to their power to open and bind the specificity pocket. As shown in Figure 20A – D (Chapter 2.4.2), **2.4** is unable to address the specificity pocket, whereas the sulfoxide **2.3** can adopt a conformation to bind into the specificity pocket. However, it is likely that this geometry is energetically very similar to that with the terminal phenylsulfoxide outside the pocket, as this geometry is also populated in the crystal structure.

With a pK_a value of about 9,[96] the boronic acid group of **2.4** in complex with ALR-2 is protonated and thus uncharged. In that respect, it resembles **2.2**. At first glance, it is surprising that the terminal boronic acid does not bind in the specificity pocket, although it would be able to form an H-bond to L300 through its hydroxyl groups, similar to the nitro moiety. The structure of **2.4** in complex with ALR-2 resembles that of **2.1** in terms of its binding mode and thermodynamic signature. Its improved potency results from a more favorable enthalpic contribution, possibly explained by the enthalpically less costly desolvation of the uncharged boronic acid group.

2.5.2 Comparison of the Electronic Properties of the Terminal Aromatic Substituent

Closer inspection of the electronic properties of the preceding phenyl ring attached to the different functional groups (COOH, NO_2, SOMe, $B(OH)_2$) may provide an explanation as to why a terminal boronic acid does not match with the properties of the nitro group in **2.2**, but better agrees with those of a negatively charged carboxylate group in **2.1** (Figure 26). It is well known that a nitro group has a strong inductive electron-withdrawing effect. This leads to strong electron-accepting properties and transforms the adjacent phenyl ring into an electron-deficient aromatic portion.[97] Figure 26C visualizes the electronic surface potential area (ESP) of the nitro-phenyl moiety. While the substituent itself is rather electron-rich, the potential across the preceding ring is reduced (green color).

Considering the electronic properties of the indole moiety in tryptophan, the aromatic system tends to be electron-rich due to the lack of electron-withdrawing functional groups. This is indicated by the yellow color across the ring system (Figure 26H).

This electron enrichment of tryptophan (here W111), in combination with the electron deficient nitro-phenyl ring of the ligand, is an ideal prerequisite for the stacking interactions of the two aromatic systems found in the complex of **2.2** (Chapter 2.4.2, d = 3.4 Å). Apart from the high desolvation costs of the carboxylate group of **2.1**, the strongly enhanced charge distribution on the phenyl ring adjacent to the carboxylate group, particularly if this group is present in its deprotonated state, may prevent opening of the transient pocket along with the establishment of a stacking interaction with W111 in the complex with inhibitor **2.1**.

Boronic acid, although polar and uncharged similar to the nitro group in **2.2**, exerts quite different electronic effects on the adjacent phenyl ring, turning it into a rather electron-rich aromatic moiety (Figure 26D, yellow color). This charge distribution may still be detrimental for a favorable π-π-stacking with the indole moiety of W111. In consequence, the transient pocket remains sealed, and the inhibitor binds outside of the pocket

Steuber *et al.* investigated other aldose reductase inhibitors with similar scaffolds but different terminal aromatic portions.[87] There, the terminal phenyl ring was decorated by an *ortho*-fluoro and *para*-bromo substituent. Analysis of the ESP of this substituted phenyl ring (Figure 26F) reveals similar properties to a nitrophenyl moiety established by the strong electron-withdrawing effect of the attached halogen atoms (green color). Remarkably, inhibitors with this aromatic portion open the specificity pocket and bind with high potency. This supports the previously stated hypothesis that an electron-withdrawing substituent leading to an electron-deficient aromatic ring to establish a stacking geometry with the indole ring of W111 is another prerequisite for binding to the specificity pocket. [83]

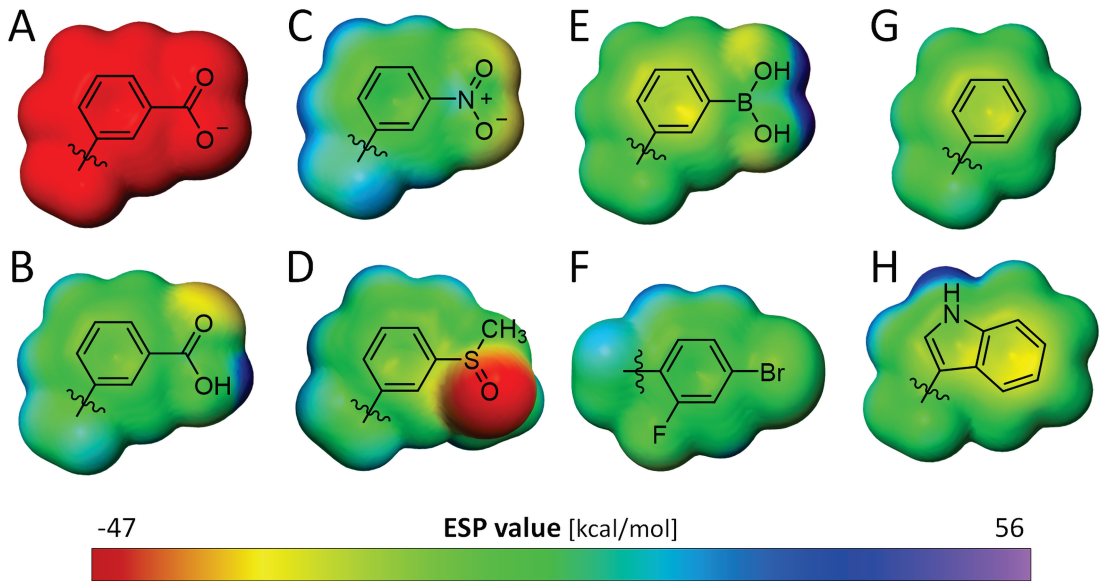


Figure 26: Electronic surface potential area (ESP) of the terminal aromatic inhibitor moieties and an indole ring. A) benzoate moiety, B) benzoic acid moiety, C) nitrophenyl moiety, D) phenyl sulfoxide moiety, E) phenylboronic acid, F) 4-bromo-2-fluoro phenyl moiety, G) phenyl moiety, and H) indole moiety of a tryptophan calculated with the CC-PVTZ(-F)++ basis set and the M06-2X-D3 theory level, Jaguar, Schrödinger, LLC, New York, NY, USA, 2020.[98] The graphical representation of the potential ranges from −47 kcal/mol (red) to +56 kcal/mol (purple).

This hypothesis can be applied to the analog **2.8**, which is able to bind to the transient specificity pocket and to interact with W111.[77] Obviously, an unsubstituted phenyl ring is still electron-deficient enough to establish the required stacking with W111, even though **2.8** is only a micromolar inhibitor of ALR-2.

The sulfoxide derivative **2.3** in the crystal structure binds at 40% to the transient pocket whereas 60% remains outside. The lack of charge on the sulfoxy group and the required electron-deficiency of the aromatic ring seem to match the necessary conditions for binding to the transient pocket. It is difficult to estimate which factor is responsible for the fact that no full occupancy of the transient pocket was observed. Factors such as the steric demand of the non-planar sulfoxy group and the non-ideal geometry of the group to interact favorably through hydrogen bonds with the pocket residues must be taken into consideration. Furthermore, it should be noted that the S = O bond is not coplanar with the phenyl ring in the adopted binding mode, breaking electronic conjugation with the π-electrons of the phenyl ring. This will definitely impact the electron-withdrawing properties of the sulfoxy group. In our calculations, coplanar geometry was assumed. As the structure of the **2.2** complex shows, the electron-withdrawing nitro group also remains in coplanar geometry within the transient binding pocket. These conformational effects of the attached groups may have an important influence on the electronic properties of the adjacent phenyl ring of the inhibitors.

2.6 Conclusion

To determine with which conditions and energy expenditure the specificity pocket of human ALR-2 opens, two essential factors were investigated. First, the structural and thermodynamic properties of inhibitors with functional groups of different electronic nature at the terminal aromatic moiety or terminal substituents of smaller steric demand were elucidated. Second, an investigation of the electrostatic potential and charge distribution across the terminal aromatic groups of the inhibitors, and their effects on binding to the transient pocket of the enzyme, was performed.

Regarding the quantum chemical analysis of the electrostatic potential of the terminal phenyl ring of inhibitors **2.1** – **2.4** modulated by the attached substituents, the hypothesis emerged that an electron deficient aromatic group is necessary for binding into the specificity pocket.[83] In the transient pocket, an electron-rich indole moiety W111 is exposed, which creates an interaction site for π-stacking with the terminal aromatic group of the inhibitor. It is possible that an electron-withdrawing substituent such as a nitro group or, as stated previously by STEUBER *et al.,* an appropriate pattern of halogen substituents, can correctly adjust the electron density distribution on the terminal phenyl ring of the inhibitor. This way, the terminal aromatic substituent may undergo favorable stacking interactions with the neighboring tryptophan residue.[87] Seemingly, the unsubstituted phenyl ring in **2.8** or the attachment of a sulfoxy group in **2.3** generates a charge distribution just sufficient to allow stacking in the opened transient pocket. As a result, the opened and closed binding pose for **2.3** are both populated in the crystal structure. Inhibitor **2.8** binds with its unsubstituted phenyl ring into the transient pocket, but it is equally as potent as **2.7**, which lacks the terminal phenyl ring and leaves the transient pocket in closed state.[77] Thus, apart from the favorable desolvation costs of the group to be accommodated in the transient pocket, an electron-deficient terminal aromatic group supports the opening of and binding to the specificity pocket of ALR-2. It is likely that the group needs a certain volume, as smaller substituents such as a cyclopropyl group give rise to more complex binding poses with enhanced residual mobility.

Due to the complexity of the adopted binding poses along with differences in desolvation costs and residual mobility, the correlation of differences in the thermodynamic signatures with changes in binding poses is impossible. For example, a phenyl and thiophene ring are assumed to be isosteric. Nevertheless, **2.8** and **2.5** differ by the placement of this group. The thermodynamic profiles of both are quite different, likely also reflecting that the phenyl derivative occupies the transient pocket, whereas the thiophene analog remains outside.

Additionally, using X-ray crystallography, we demonstrated that a shift in pH between pH 5 and 8 does not impact the binding pose of **2.1** with respect to the opening of the specificity pocket. This finding is important because it proves that the crystallographic results from previous studies performed using co-crystallization are comparable to the measurements

presented in this study completed by soaking. The findings also suggest that the structural data are relevant across a certain pH range. While the enzyme shows optimal enzymatic activity at a pH of 6, the value applied in previous studies during the enzyme kinetic measurements, the ITC measurements were performed at pH 8. Inhibitor soaking, on the other hand, works best with a citrate buffer at a pH of 5. To compare the data and draw conclusions across enzyme activity, thermodynamic signatures, and crystal structure, it is necessary to validate that the large pH shift between 5 and 8 does not affect the properties of the active site and thus inhibitor binding.

Chapter 3

Strategies for Late-Stage Optimization: Profiling the Thermodynamics by Preorganization and Salt Bridge Shielding

Anna Sandner, Tobias Hüfner-Wulsdorf, Andreas Heine,
Torsten Steinmetzer, and Gerhard Klebe*

*Institut für Pharmazeutische Chemie, Philipps-Universität Marburg,
Marbacher Weg 6, 35032 Marburg, Germany

Journal of Medicinal Chemistry 2019, 62, 9753 – 9771

3.1 Introductory Remarks

This project was elaborated in cooperation with DR. TOBIAS HÜFNER-WULSDORF (group of PROF. GERHARD KLEBE, Philipps Universität Marburg, Germany. All laboratory procedures (ca. 70%), including crystal preparation, X-ray crystallographic data collection, structure solution, assay and ITC measurements and their evaluation were performed by the author of this thesis. All computational calculations and MD simulations (ca. 30%) were performed by DR. TOBIAS HÜFNER-WULSDORF.

PROF. ANDREAS HEINE (Philipps-Universität Marburg, Germany), PROF. TORSTEN STEINMETZER (Philipps-Universität Marburg, Germany) and PROF. GERHARD KLEBE helped to interpret the data.

The manuscript of the publication was collaboratively written by the author of this thesis, Dr. TOBIAS HÜFNER-WULSDORF and PROF. GERHARD KLEBE. This Chapter is published in Journal of Medicinal Chemistry.

3.2 Abstract

Structural fixation of an inhibitor in its bioactive conformation may, due to entropic reasons, improve affinity. We present a congeneric series of thrombin inhibitors with a variety of functional groups triggering preorganization prior to binding. Fixation in solution and complex formation have been characterized by crystallography, ITC, and molecular dynamics (MD) simulations. First, we show why these preorganizing modifications do not affect the overall binding mode and how key interactions are preserved. Next, we demonstrate how preorganization thermodynamics can be largely dominated by enthalpy rather than entropy because of the significant population of low-energy conformations. Furthermore, a salt bridge is shielded by actively reducing its surface exposure, thus leading to an enhanced enthalpic binding profile. Our results suggest that the consideration of the inhibitor solution ensemble by MD simulation is necessary to predict preorganizing modifications that enhance the binding behavior of already promising binders.

3.3 Introduction and Terms of Reference

Fixation of an inhibitor in its bioactive conformation is a popular late-stage optimization strategy during pre-clinical drug development. This strategy is commonly used to improve the binding affinity of an already promising binder by stabilizing the bioactive conformation. By following this concept, binding affinity is effectively enhanced due to the reduction of the

entropic penalty to be paid upon binding resulting from inhibitor fixation.[99] Quite contrary, preorganization can also result in a dominating enthalpic binding behavior. This observation caused some controversy about the underlying physical principles of this design strategy.[100,101] Even though the concept of preorganization is well appreciated in the drug discovery community, it is surprisingly little studied, for example, by thermodynamic investigations to elucidate the driving forces of how it contributes to affinity enhancement.

To accomplish conformational fixation, different strategies for inhibitor modification emerged. Since all of these strategies ultimately change the chemical composition of the inhibitor molecule, the introduction of new chemical features must be adjusted to those already present in the parent scaffold. If this is not considered, it potentially leads to the population of non-bioactive conformations, and therefore prevents the intended gain in binding entropy. Furthermore, it is necessary that the protein-inhibitor complex is able to readily tolerate the introduced chemical modifications with respect to preservation of the binding mode obtained for the parent scaffold. If this is not successful, the inhibitor is forced to rearrange itself into an energetically more favorable conformation with regard to binding, which leads to a loss of available degrees of freedom. Conformational fixation can be achieved by incorporating cyclic elements or fusing adjacent rings.[101,102] This strategy effectively aims at rigidifying the inhibitor by reducing the number of freely rotatable bonds. However, this strategy is not always straightforward and possible entropic advantages upon binding can be overcompensated by other detrimental contributions.[101,103,104] Further known strategies for the introduction of non-covalent preorganization are the enhancement of steric effects,[105,106] intramolecular hydrogen bonding,[107] ion-pairing[108-110] or π-π-stacking.[111] Although conformational fixation by cyclisation appears to be a straightforward design strategy, it may require significant synthetic effort. Moreover, it is crucial to match the spatial arrangement of the cyclic portions with the steric requirements at the binding pocket. The concept of triggering preorganization by the introduction of additional non-covalent interactions can seemingly be accomplished with less synthetic effort as only few additional chemical features must be introduced, as demonstrated in this study. Furthermore, within this strategy the inhibitor remains with a higher degree of residual flexibility and may thus better tolerate minor design inaccuracies.

As an example, RÜHMANN *et al.* reported about a congeneric series of peptidomimetic thrombin inhibitors that comprise either an open-chain or rigid proline moiety in P_2 position.[99] Here the tailored preorganization within this inhibitor series resulted in a 1000-fold boost in affinity for the proline derivative, which was mainly explained by an entropic advantage, as confirmed by a detailed thermodynamic and structural analysis. This observation is in agreement with other studies.[100] While the binding mode of the inhibitors was highly conserved, residual mobility in the bound state, order / disorder phenomena of side-chains from active-site residues and the surrounding surface water network formed along the protein-inhibitor interface

differ slightly. Most importantly, as MD simulations confirmed, the huge affinity difference among the congeneric inhibitors results from the deviating number of conformers accessible in solution, which are lost upon binding.

The phenomenon of preorganization can already be found in nature, consequently the discovery of drugs can be inspired by nature. Alkaloids and steroids are examples of natural products with rigid chemical scaffolds and therefore limited conformational variability.[112] This property endows them with optimal structural preorganization for binding to a given target protein.[99] Comparing a flexible inhibitor with a similar restrained inhibitor, both capable of forming the same pairwise interactions with the protein or the solvent, the latter will sacrifice a smaller number of conformational degrees of freedom during complex formation. This results in a lower entropic penalty. However, this theory does not seem to be generally valid, since these effects will coincide with solvation / desolvation contributions of the functional groups from the inhibitor and protein.[101]

For the inhibitor fixation we selected the serine protease thrombin as well-established target protein, often used for model studies in drug research.[43,113] As described in Chapter 1.2.2.2 thrombin plays an important role in the regulation of blood clot formation and is therefore a pharmacologically important target to prevent cardiovascular diseases.[43,114] Meanwhile, several synthetic thrombin inhibitors have been approved for therapeutic use.[115] The protease comprises three differently shaped pockets (S_1, S_2 and $S_{3/4}$)[116,117] of which S_1 is the most relevant regarding specificity. The proximal S_2 pocket distinguishes thrombin from other chymotrypsin-like serine proteases exhibiting the in Chapter 1.2.2.4 described 60' loop. Finally, the distal $S_{3/4}$ pocket opens widely toward the solvent and accommodates larger hydrophobic inhibitor portions.[118]

In the current chapter, we will follow the design option of introducing an intramolecular hydrogen bond as decisive element to achieve inhibitor preorganization. We designed and synthesized a set of seven peptidomimetic inhibitors (Figure 27) which were studied by crystal structure analysis, enzyme kinetics, ITC and MD simulations. Due to their unique fit and activity in thrombin, the natural substrates fibrinopeptide A and HCII served as a template for the inhibitor design (Figure 28).

All inhibitors consist of P_1, P_2, and P_3 building blocks, which bind in the corresponding S_1, S_2, and $S_{3/4}$ binding pockets of the enzyme (*cf.* the nomenclature of Schechter and Berger), respectively.[119] Whereas inhibitor **3.1** lacks at the aromatic P_1 portion, a substituent to undergo intramolecular hydrogen-bond formation, the inhibitors **3.2** – **3.7** are decorated at the P_1 aromatic group with a hydroxylmethylene or amino-methylene substituent that is potentially available for intramolecular preorganization (Figure 27).

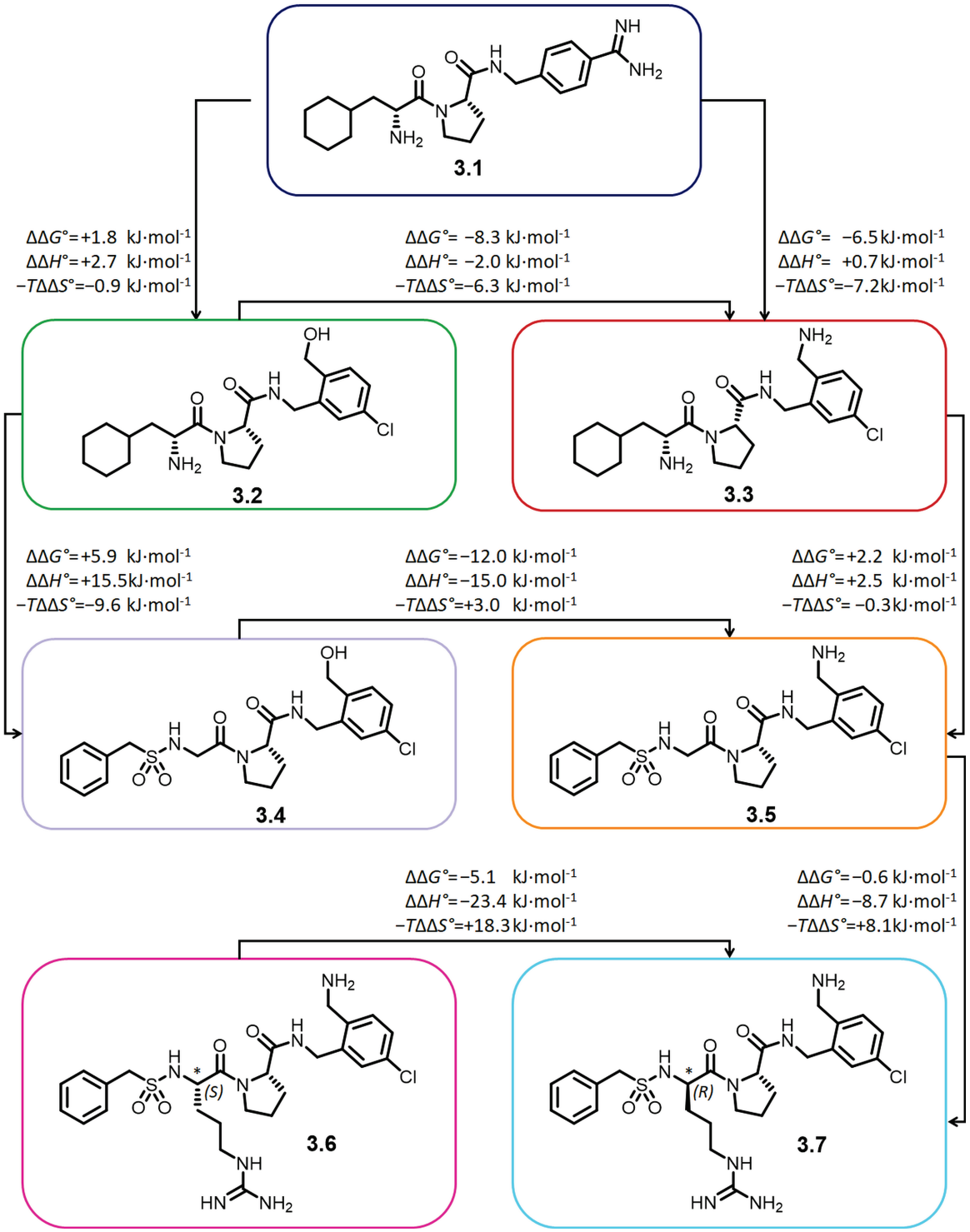


Figure 27: Peptidomimetic thrombin inhibitors used in the current study for the matching pair analysis and their corresponding ITC profiles reported for various pairs of inhibitors.[120]

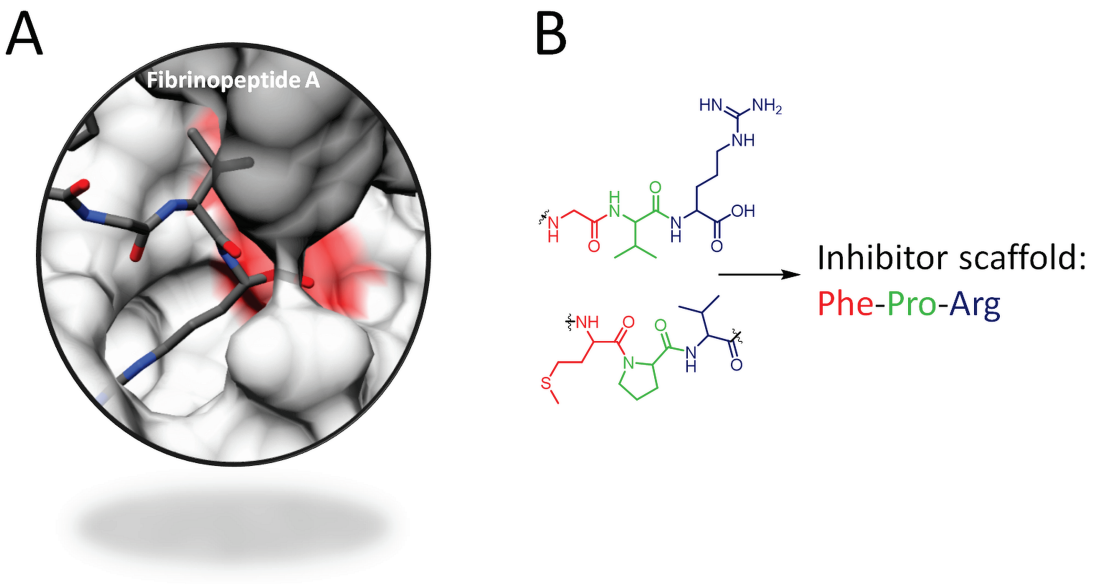


Figure 28: Origin of the thrombin inhibitor scaffold. A) Binding mode of the natural substrate Fibrinopeptide A in the S_1 pocket of human Thrombin (PDB-Code: 1FPH[121], 2.5 Å). B) Based on the C-terminal end of fibrinopeptide A (top) and the peptide sequence of HCII (bottom), which naturally bind into the S_1 pocket of thrombin, our inhibitor template Phe-Pro-Arg (right) was developed.

The P_2 proline of the parent D-Phe-Pro-Arg template fits well below the 60's-loop and exhibits adequate replacement for the valine side chain present in the natural substrate fibrinogen.[99,121] In P_1, we initially started with a benzamidine moiety (inhibitor **3.1**). However, as poor pharmacokinetic properties are usually associated with this group, we considered an *meta*-chlorobenzyl anchor to occupy the S_1 pocket (**3.2** – **3.7**).[122] For intramolecular preorganization, an *ortho*-hydroxy or amino methyl group was introduced at the P_1 residue (**3.2**, **3.4** and **3.3**, **3.5** – **3.7**). Potent substrate-analogue thrombin inhibitors require a hydrophobic P_3 side chain such as phenyl, naphthyl or cyclohexyl. For further improved potency, an additional terminal sulfonyl group has been frequently attached as considered in **3.4** – **3.7**.[123] We studied as matching pairs **3.2**/**3.3**, **3.4**/**3.5** and, to investigate the impact of stereochemistry of an attached arginine-like side chain at P_3, the pair **3.6**/**3.7**. Inhibitor **3.5** (3RML), was previously described in a study of Biela *et al.*[73] The structural, thermodynamic and conformational properties of the different matched molecular pairs were compared relative to each other. Even though the series appears quite congeneric, some very surprising effects were discovered, which strongly modulate affinity (scatter in $\Delta G°$ = ±14.2 kJ/mol) and the partitioning in the enthalpic and entropic contributions (scatter in $\Delta H°$ = ±23.7 kJ/mol and $-T\Delta S°$ = ±18.4 kJ/mol). Remarkably, these differences can hardly be explained based on bound state adopted by the different inhibitors, as these mostly vary only marginal. Instead, the deviating conformational properties of the inhibitors in bulk water phase prior to thrombin binding have to be considered.

3.4 Enzyme Kinetic and Thermodynamic Data

The affinity of the seven inhibitors was determined by an enzyme kinetic fluorescence assay according to Dixon (Figure 64, Appendix) **3.1** – **3.3** and **3.5** – **3.7** fall into low nano to picomolar range (Table 4). Only **3.4** deviates and shows significantly reduced affinity. To determine the thermodynamic signatures, we applied ITC. As the investigated inhibitors show high potency, the Gibbs free energy of binding $\Delta G°$ (derived from the dissociation constant K_d) could not be determined accurately enough by direct titrations. To obtain lower *c*-values[64,84-86] and thus a more accurate K_d, displacement titrations with the reference inhibitor **3.8** were performed (Figure 65, Appendix).[124] The enthalpic value $\Delta H°$ was taken from a separate direct titration of the strong binding inhibitor showing a virtually rectangular shape.[64] According to the Gibbs-Helmholtz equation, the entropic contribution $-T\Delta S°$ to binding was calculated as the difference between $\Delta G°$ and $\Delta H°$. Figure 29 (top) shows that all inhibitors exhibit strong exothermic binding signals.

Table 4: Enzyme kinetic and thermodynamic parameters of the investigated inhibitor series.

inhibitor	K_i[a)] [nM]	$\Delta G°_{Assay}$ [kJ/mol]	$\Delta G°_{ITC}$[b)] [kJ/mol]	$\Delta H°$ [kJ/mol]	$-T\Delta S°$ [kJ/mol]	titration mode
3.1	0.15	–56.1	–39.1 ± 0.1	–55.3 ± 0.6	16.2 ± 0.6	direct
3.2	9.20	–45.9	–37.3 ± 0.2	–52.6 ± 0.2	15.3 ± 0.2	displ.
3.3	0.07	–58.1	–45.6 ± 0.1	–54.6 ± 0.3	8.9 ± 0.1	displ.
3.4	483.63	–36.0	–31.4 ± 0.2	–37.1 ± 0.2	5.7 ± 0.3	direct
3.5	0.63	–52.5	–43.5 ± 0.2	–52.1 ± 1.1	8.7 ± 0.2	displ.
3.6	3.51	–48.3	–38.9 ± 0.6	–37.4 ± 1.3	–1.4 ± 0.6	displ.
3.7	0.29	–54.4	–44.0 ± 0.9	–60.9 ± 0.3	16.8 ± 0.9	displ.

[a)] K_i was determined by a Dixon plot of a kinetic enzyme assay. $\Delta G°_{Assay}$ was calculated from K_i using the equation $\Delta G° = -RT \ln K_i$. [b)] The thermodynamic data were determined by ITC in HEPES buffer pH 7.8. The data are not corrected for the putatively overlaying protonation effect.

The thermodynamic data of **3.1**, determined by displacement titrations, indicate that the concentration of the inhibitor solution appears to be significantly lower than expected. Based on ^{1}H NMR measurements in DMSO-d_6, inhibitor **3.1** contains a cis-content of approximately 17%. Despite its high *c*-value, this led to **3.1** being characterized only by the values determined by direct titrations (Figure 68, Appendix). Because **3.4** showed lower affinity to thrombin with a *c*-value in an acceptable range, its binding was also recorded by direct titrations. All inhibitors exhibit high potency toward thrombin with **3.3** being the strongest and the structurally related **3.2**, the second weakest binder.

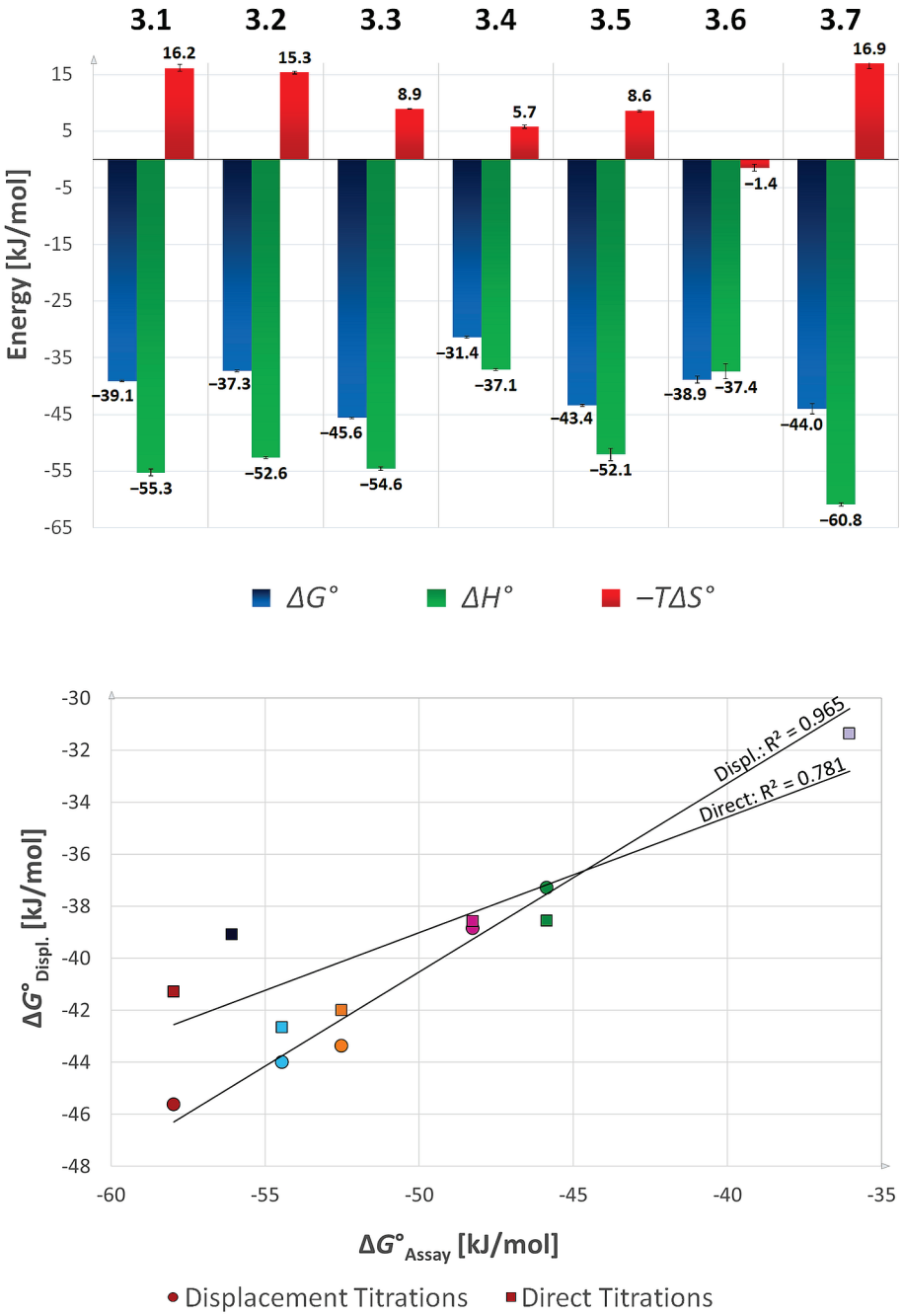


Figure 29: ITC results of **3.1** – **3.7**. Top: Thermodynamic parameters determined by ITC. **3.1** and **3.7** were analyzed by direct titrations and **3.2** – **3.6** by displacement titrations. The measurements were performed in HEPES buffer and are not corrected for overlaying protonation effects. Standard deviations of three measurements are indicated as error bars. Bottom: Correlation between Gibbs free energies of binding ΔG° of direct and displacement ITC with ΔG° as calculated from the kinetically determined K_i. The correlation with the data of the direct titrations is given in squares. The correlation with the data of the displacement titrations is represented in circles. The colors correspond to the inhibitors shown in Figure 31, Figure 32 and Figure 33.

The affinities determined by ITC and by the kinetic enzyme assay deviate in absolute values; however, they show similar trends, particularly when the enzyme kinetic data are correlated with those from the displacement titrations (97%, Figure 29, bottom). Less accurate is the agreement using the data of the direct titrations. This is related to the fact that dissociation constants in the one-digit nanomolar to sub-nanomolar range can only be determined by ITC with reduced accuracy.

Heat signals recorded by ITC can be affected by superimposed changes in protonation states of the binding partners. Such effects need to be detected and accordingly corrected by measuring the system from buffers with deviating heat of ionization.[64] From our multiple ITC experiments with thrombin, we observed that H57 releases approximately 0.5 mol of protons upon inhibitor binding.[73,99,118,125] Furthermore, in a previous study, we determined that the terminal P_3 amino group in inhibitors based on the scaffold used in **3.1** − **3.3** entraps about 0.5 mol of protons during binding. Thus, overall, both changes compensate each other, and no buffer dependence is found. We also studied related sulfonamides; however, they did not show any buffer dependence apart from the usual proton release of H57.[73] Finally, the amino group of the attached P_1 amino-methylene group can potentially change protonation state, as its pK_a value can be estimated to ≈ 8.5 − 9.0. However, a previous study considering this scaffold did not disclose any significant buffer dependency resulting from a change in protonation at this group.[73] Likely, already in aqueous solution, this functional group is protonated and charged (*cf.* conformational analysis described in the following section). Accordingly, for **3.1** − **3.3**, no buffer dependence will be detectable, whereas **3.4** − **3.7** are expected to show the usual ≈ 0.5 mol protons released by H57. We therefore measured the current inhibitor series in HEPES buffer at pH 7.8 only, and no corrections for changes in protonation states have been accomplished. Nevertheless, because of described superimposed effects, only the data of **3.1** − **3.3** and **3.4** − **3.7** can be compared relative to each other in the following. Interestingly, the benzamidine derivative **3.1** and the chlorobenzyl analogue **3.2** with the $-CH_2-OH$ substituent possess similar thermodynamic signatures, whereas the derivative with the $-CH_2-NH_2$ side chain shows a remarkably stronger Gibbs free energy mainly for entropic reasons (Figure 27). A similar observation is found for the matching pair **3.4**/**3.5** where again the $-CH_2-OH$ analogue **3.4** loses significant affinity, however here for predominantly enthalpic and lesser entropic reasons. Finally, the matching pair of the two epimers **3.6**/**3.7** discloses the latter as the more potent inhibitor with a huge enthalpic advantage, strongly compensated by an opposing entropic contribution.

3.5 Crystal Structures of Human Thrombin Inhibitor Complexes

The inhibitors **3.1** – **3.6** adopt a conformation in the binding pocket of thrombin, in which the P_3 side chain or P_4 sulfonyl portion is oriented toward the $S_{3/4}$ pocket. We will refer to this characteristic conformation as the S-shape conformation (according to the shape of the atoms in the backbone of the peptidomimetic inhibitors **3.1** – **3.6**). The S-shape conformation must be distinguished from the U-shape or horseshoe-like backbone conformation that is adopted by **3.7** (Figure 30) and previously reported inhibitors.[99] In all complex structures, the inhibitor's P_2 proline is accommodated in the S_2 pocket covered by Y60a and W60d of the previously described 60's-loop (Chapter1.2.2.4). Compared to the S_2 pocket, the more flat and solvent-exposed $S_{3/4}$ pocket above W215 is bordered by L99 and I174 and has a hydrophobic character. It hosts the cyclohexyl group of the P_3 side chain of compounds **3.1** – **3.3** or the N-terminal benzyl sulfonamide moiety of **3.4** – **3.6**. In the U-shaped complex with **3.7**, the sulfonamide oxygen atoms are placed at a different position compared to **3.4** – **3.6**. This small perturbation of the orientation of the sulfonamide linker is accompanied by a different binding mode of the P_3 and P_4 portions. In **3.2** – **3.7**, the highly specific S_1 pocket is accommodated by a chlorobenzyl moiety (**3.2** – **3.7**), whereas **3.1** comprises a benzamidine moiety closely mimicking the arginine of natural substrates.[44]

In all structures, a hydrogen bond is formed between the P_1 amido-NH and the S214 carbonyl oxygen with d = 2.9 Å. It fixes the inhibitor at the boundary between the S_1 and S_2 pockets [Figure 31 – Figure 33, on the left the binding mode is shown and on the right the difference electron densities ($F_O - F_C$)].

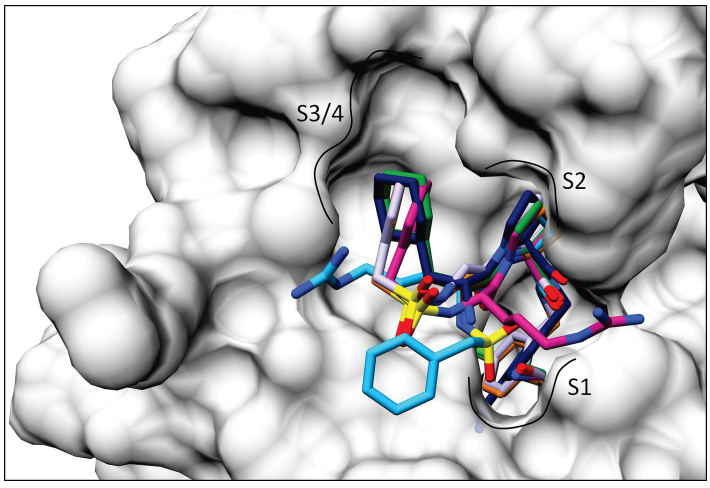


Figure 30: Crystallographic superposition of the thrombin inhibitors **3.1** – **3.7**. Inhibitor **3.1** (5JZY, dark blue), **3.2** (5LCE, red), **3.3** (5LPD, green), **3.4** (6ROT, pale purple), **3.5** (3RML, orange), **3.6** (6GBW, pink), and **3.7** (5JFD, light blue), as bound to the active site of thrombin. Oxygen atoms are displayed in red, nitrogen atoms in blue, and sulfur atoms in yellow. The protein is depicted by its transparent solvent-accessible surface (light gray), and the subpockets are bordered by black lines.

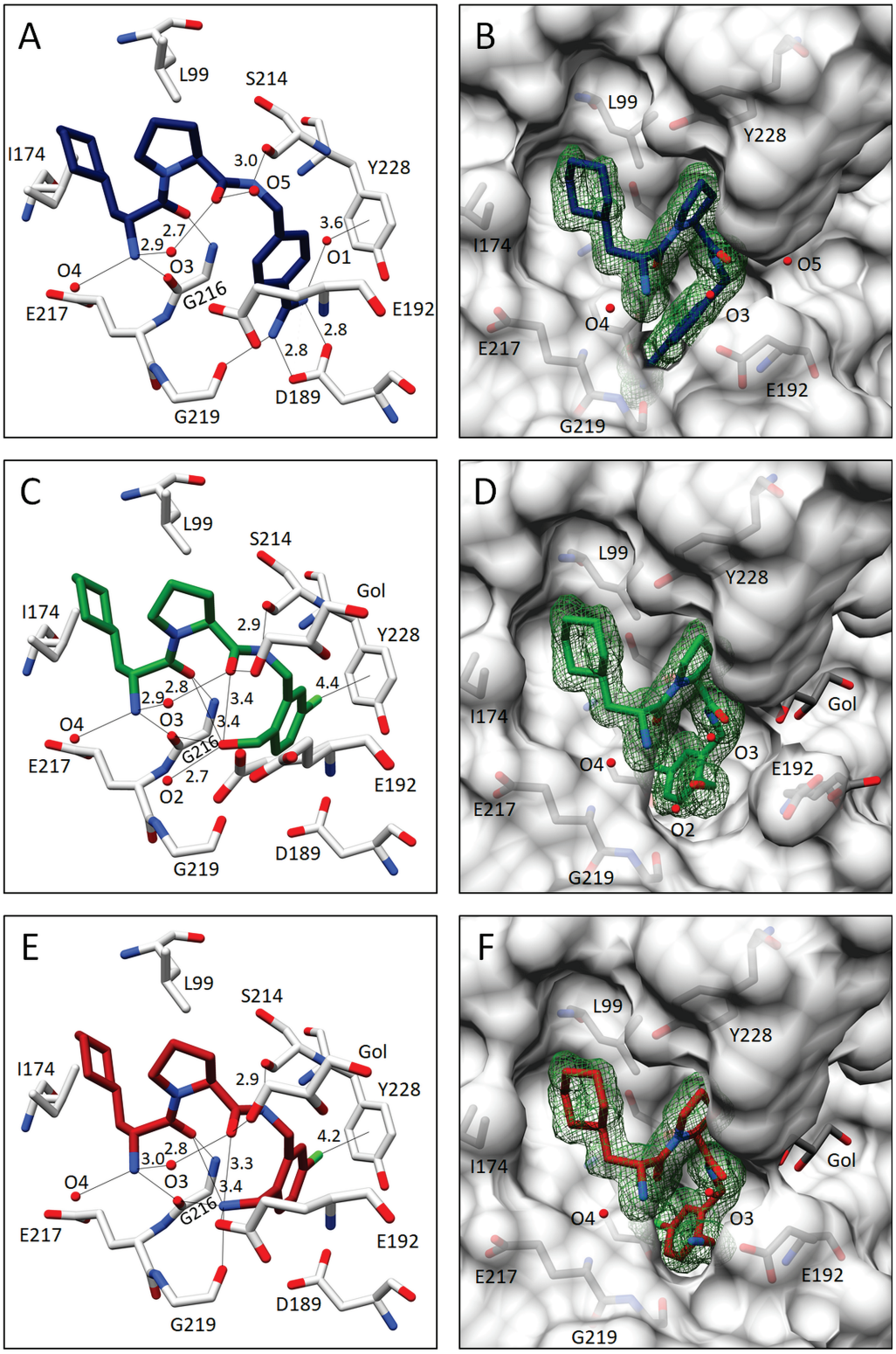


Figure 31: Inhibitors **3.1** – **3.3** bound to the active site of thrombin. Inhibitor **3.1** (5JZY) [(A, B) dark blue], **3.2** (5LCE) [(C, D) green], and **3.3** (5LPD) [(E, F) red]. Water molecules are shown as red spheres within a radius of 5 Å around the inhibitor. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, and sulfur atoms in yellow. On the left, the interactions are indicated as black lines, whereas on the right, the protein is depicted by its transparent solvent-accessible surface (light gray) and the electron density difference map ($F_O - F_C$) as green mesh at the 3σ level.

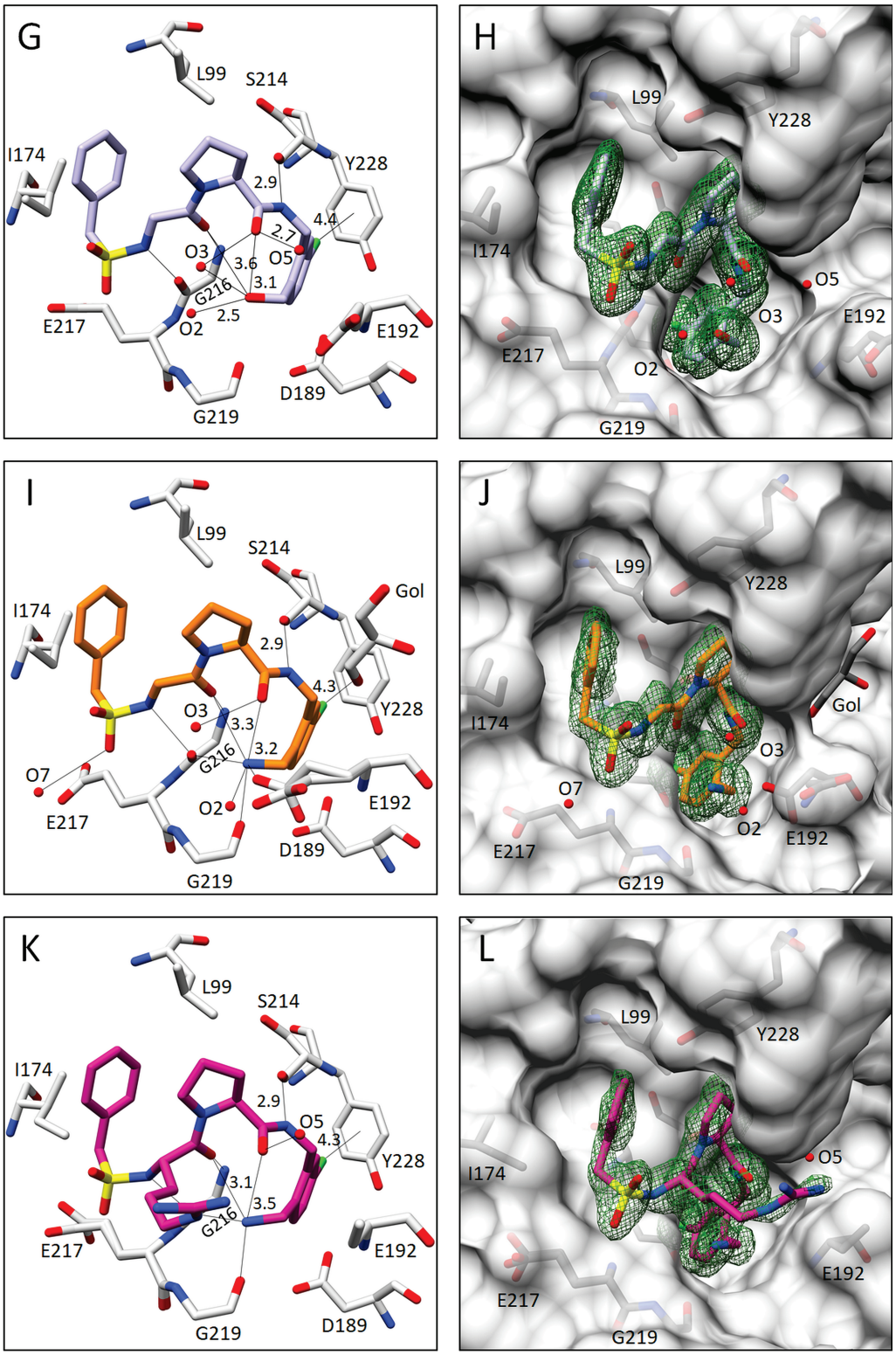


Figure 32: Inhibitors **3.4** – **3.6** bound to the active site of thrombin. Inhibitor **3.4** (6ROT) [(G, H) pale purple], **3.5** (3RML) [(I, J) orange], and **3.6** (6GBW) [(K, L) pink] Water molecules are shown as red spheres within a radius of 5 Å around the inhibitor. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, and sulfur atoms in yellow. On the left, the interactions are indicated as black lines, whereas on the right, the protein is depicted by its transparent solvent-accessible surface (light gray) and the electron density difference map ($F_O - F_C$) as green mesh at the 3σ level.

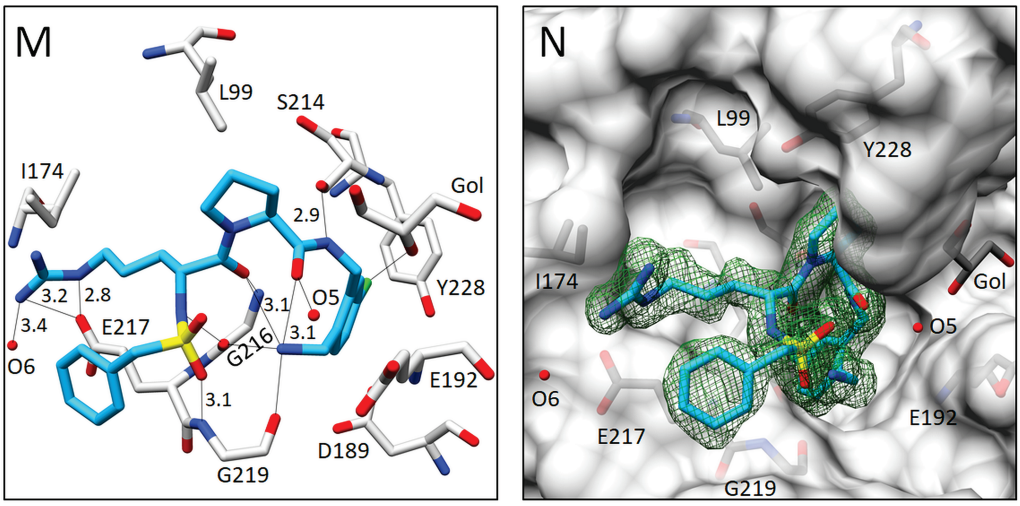


Figure 33: Inhibitor **3.7** bound to the active site of thrombin. Inhibitor **3.7** (5JFD) [(M, N) light blue]. Water molecules are shown as red spheres within a radius of 5 Å around the inhibitor. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, and sulfur atoms in yellow. On the left, the interactions are indicated as black lines, whereas on the right, the protein is depicted by its transparent solvent-accessible surface (light gray) and the electron density difference map ($F_O - F_C$) as green mesh at the 3σ level.

The benzamidine moiety of **3.1** establishes a strong salt bridge to the deprotonated D189 at the bottom of the S_1 pocket with d = 2.8 Å. Therefore, it is the only inhibitor in the series that addresses this amino acid via direct H-bonds; its binding mode is shown in Figure 31A, whereas the difference electron density ($F_O - F_C$) is given in the panel on the right (Figure 31B). Furthermore, the benzamidine interacts with G219 and water O1. In **3.2** – **3.7**, the chlorobenzyl moiety replaces the benzamidine anchor along with water O1 and interacts with Y228 via a halogen-π interaction with d = 4.2 – 4.4 Å.[126] Furthermore, they form the expected intramolecular H-bonds between their $-CH_2-NH_2$ or $-CH_2-OH$ substituent at the aromatic P_1 ring and the inhibitor's own P_2 and P_3 carbonyl oxygen atoms with d = 3.1 – 3.6 Å (Figure 31C, E, Figure 32G, I, K, and Figure 33M). Obviously, in both cases, the P_1 anchor remarkably influences the finally adopted inhibitor geometry at the binding site. The obtained binding mode is triggered by intramolecular H-bonding using the attached $-CH_2-NH_2$ or $-CH_2-OH$ substituents, thus confirming our initial design hypothesis. The orientation of the chlorobenzyl moiety is further stabilized by these groups through a contiguous pattern of H-bonds to the carbonyl oxygen of G216 and the NH group of either the terminal P_3 amino group of **3.2** and **3.3** or the sulfonamide-NH of **3.4** – **3.7**. Corresponding contacts cannot be built by **3. 1**, which lacks the facility to establish similar intramolecular H-bonds.

The latter motif of inhibitors **3.2** – **3.7** (Figure 31C, E, Figure 32G, I, K, and Figure 33M) is also part of the typical antiparallel β-ladder-binding motif, found for all studied inhibitors, and being observed in many serine proteases with their bound substrates or peptidelike inhibitors.[121,118,73] In addition, **3.1** – **3.3** exhibit, intramolecularly, a water-mediated H-bond (via O3

with d = 2.7 − 3.0 Å) between the terminal, likely charged, P_3 ammonium group and the P_2 backbone carbonyl oxygen. Although this water is also present in two of the sulfonamide complexes (**3.4** − **3.5**), it only forms an H-bond to the P_2 carbonyl group. Accordingly, its influence on the adopted binding mode appears minor as the distance to the sulfonamide moiety amounts to more than d = 4.2 Å, far beyond a usual H-bond distance. Additionally, **3.1** − **3.3** recruit water molecule O4 with d = 2.7 Å. Because of the steric demand of the sulfonyl group, this water is expelled from the complexes with **3.4** − **3.7**. Moreover, the chlorobenzyl derivatives with the $-CH_2-NH_2$ substituent form a hydrogen bond to the carbonyl oxygen of G219 with d = 3.0 Å, whereas the $-CH_2-OH$ analogues **3.2** and **3.4** lack this contact (Figure 31C and Figure 32G). This suggests that also the $-CH_2-NH_2$ group binds in the charged state, which is further supported by a calculated pK_a value of 9.1 for this group according to a local pK_a estimation using the pK_a plugin of ChemAxon.[127] However, the hydroxyl derivatives recruit water O2, which is also observed in the case of **3.5** (Figure 32I), where it is spatially shifted by 1.9 Å compared to the complexes with **3.2** and **3.4**.

Only the amino-methylene anchor of **3.3** (Figure 31E) and **3.5** (Figure 32I) involves E192 into H-bond contacts, a residue usually described as highly flexible. This property is confirmed when comparing its side-chain orientation in the seven structures studied here and the plethora of deposited thrombin complexes in the Protein Data Bank (PDB).[125,128] For **3.6**, E192 could not be assigned satisfactorily to the difference electron density, and the blurred density suggests a pronounced scatter of the side chain over multiple orientations (Figure 32K, L). As a control, we collected further data sets using different crystals of the complex with **3.6**; however, refinement of these data suggested the same disorder for E192 in all cases. While in **3.1**, **3.4**, **3.6**, and **3.7**, O5 interacts, shifted by 2.5 Å, with the inhibitor's P_2 carbonyl oxygen atom, a glycerol molecule (Gol), picked up from the soaking buffer, is observed at the same site in the complexes with **3.2** and **3.3** (Figure 31C, E). In this region, usually the natural substrate fibrinogen binds. Inhibitor **3.5** also hosts a glycerol molecule in the S_1' pocket (Figure 32I). In a previous study on inhibitor binding to thermolysin, we studied the impact of glycerol on the observed binding poses.[129] Glycerol molecules are well-known to replace three water molecules as a kind of rigidified surrogate in a crystal structure and help to stabilize the crystal packing. We subsequently reperformed the crystal structures without glycerol from another buffer; however, the structures and the binding modes remained unchanged. Therefore, supposedly, binding modes are not significantly affected by the use of glycerol; however, this additive can help to stabilize the crystal packing. The epimers **3.6** and **3.7** differ structurally only at the stereo center of the P_3-arginine residue of the inhibitor, which leads to an entirely different binding mode of the benzylsulfonyl group (Figure 32K and Figure 33M). While the latter group of the (*S*)-arginine derivative (**3.6**) fits perfectly well into the $S_{3/4}$ pocket as for **3.4** and **3.5**, the epimer **3.7** with the (*R*)-arginine is forced to adopt an orientation largely exposing both the arginine side chain and the benzylsulfonyl moiety toward the bulk solvent (Figure

33N). Interestingly, **3.7** does not adopt the back-folded S-shape geometry. As commonly observed for other inhibitors of similar congeneric series, the sulfonamide of **3.7** still interacts with the backbone amide in G219 (3.1 Å) and is thus able to adopt the usual U-shape.[99,73] Also, the arginine side chain of the inhibitor recruits water molecule O6 and is fixed by a bidentate salt bridge with the carboxylate group of E217. The (*S*)-arginine side chain of **3.6** is oriented toward the bulk solvent (in the crystal packing the closest contact is to W60D with 4.2 Å) and apparently does not form any direct interaction with thrombin. The difference electron density of this side-chain moiety is only partly defined compared to the remaining inhibitor (Figure 32L). We created a polder map (Figure 34) which allows placement of the as side chain (Figure 32L). The P_4 group orients into the $S_{3/4}$ pocket, prohibiting the characteristic U-shaped backbone conformation of bound inhibitors with (*R*)-configured P_3 residues and N-terminal benzylsulfonyl group.

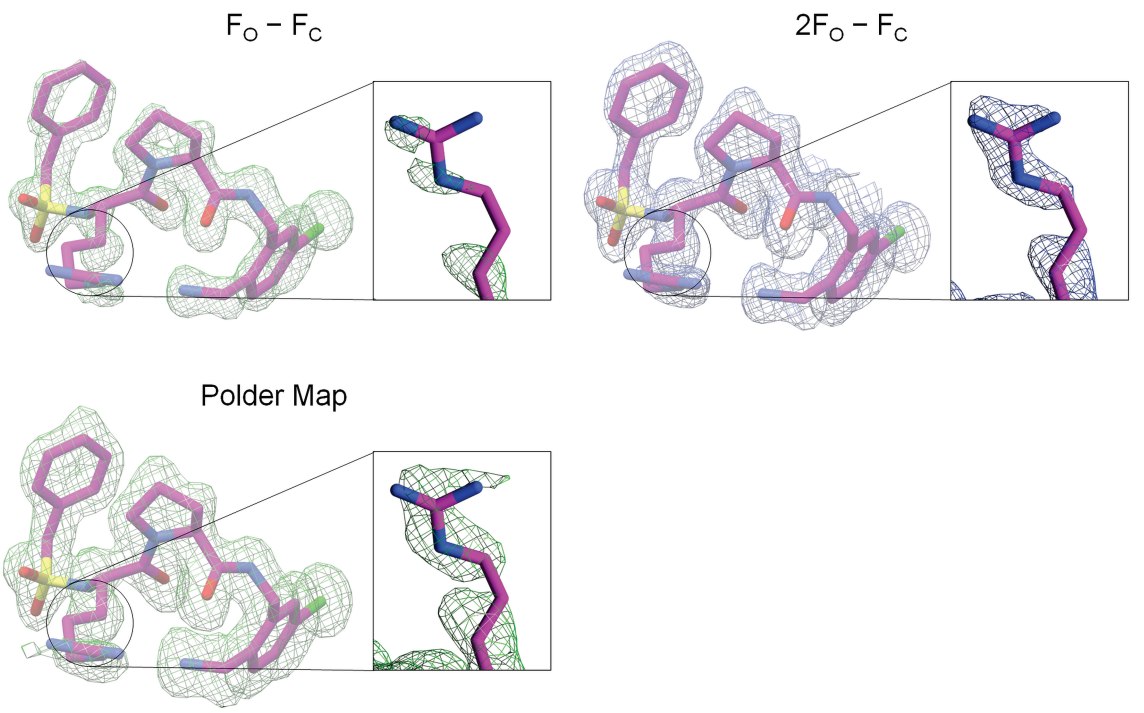


Figure 34: Summary of the different density maps of the arginine side chain of **3.6** (PDB code: 6GBW). Top left: $F_O - F_C$ difference electron density map at the 3σ level prior to the inhibitor's insertion into the model. Top right: $2F_O - F_C$ density map at the1σ level after the inhibitor's insertion into the model. Bottom left: Polder map at the 3σ level.

3.6 Molecular Dynamics Simulations

As shown in the crystal structures, in agreement with our design hypothesis to create inhibitors potentially capable to bind in a preorganized fashion, **3.2** − **3.7** all bind to the protein with intramolecular H-bonds stabilizing the bound state. This feature results from a well-placed $-CH_2-\underline{\mathbf{X}}H_n$ (X = N, O) substituent at the aromatic P_1 anchor. Because of the lack of a similar substituent, **3.1** is not capable to establish such an intramolecular restraint. This observation calls upon the question, whether in aqueous solution, prior to thrombin binding, the inhibitors adopt preferentially the same or a similar conformation as in the binding pocket. In such case, preorganization could contribute to binding and boost affinity. However, this common paradigm of drug design solely holds if not only the quantity of bioactive conformations is increased, but also the properties of these conformations are similar to those of the molecule that does not contain the preorganization feature (in the present case, compound **3.1**). Because chemical modifications are introduced in order to trigger the preorganization of a molecule, it effectively has become a different molecule. Consequently, it has altered physicochemical properties, which must be investigated and taken into account when modifying the conformations of the unbound inhibitor. In order to shed light on these properties on an atomistic scale, especially whether the introduced chemical modifications not only lead to enhanced entropic binding but also cause a shift in the enthalpy of binding, we performed extensive MD simulations (overall 14 µs accumulated simulation time) of the thrombin systems. In a recent contribution, we could show that MD simulations of inhibitors in aqueous solution provide reliable insights into the inhibitor's conformational properties as convincing agreement with Nuclear Overhauser Effect (NOE) data determined by NMR spectroscopy on the same inhibitors was found.[103] Also, recent studies by WITEK *et al.* successfully applied MD simulations in conjunction with NMR in order to investigate the solution ensemble of inhibitor molecules.[130,131]

To study the conformational behavior of **3.1** − **3.7**, we performed umbrella sampling MD simulations of the bound as well as the unbound inhibitor in aqueous solution (for further details, see https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.9b01196, Supporting Information for SANDNER *et al. J. Med. Chem.*, **2019**, 62, 21, 9753–9771). In order to obtain proper sampling along conformations that are relevant for the investigation of important conformational preferences, we choose the distance d_{O-X} between the $-CH_2-\underline{\mathbf{X}}H_n$ (X = N, O, considering in the calculations for N, an ammonium group) and the P_3 carbonyl-**O** to be the reaction coordinate for **3.2** − **3.7** (Figure 35B).

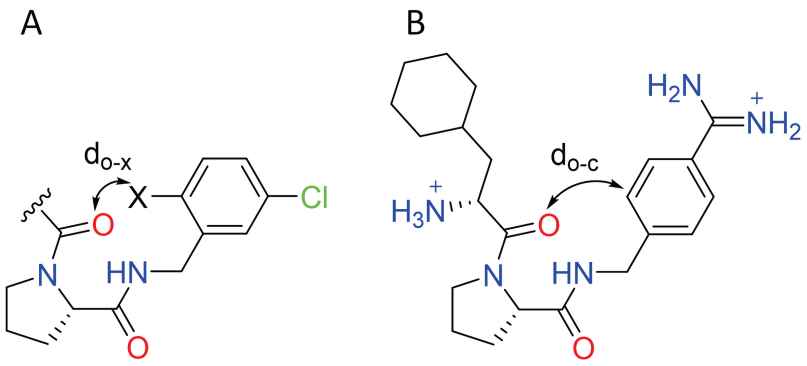


Figure 35: Schematic representation of the reaction coordinates of **3.2** – **3.7**. Reaction coordinates d_{O-X} and d_{O-C} (with X = N, O) were used for umbrella sampling simulations of **3.2** – **3.7** A) and **3.1** B), respectively. The protonation states are depicted as employed in the MD simulation. Note that the depicted 2D representations must not be confused with conformations in a three-dimensional space.

Analogously, for compound **3.1**, we used the distance between the ring carbon atom at position 2 of the P_1 moiety and as before the carbonyl-**O** of the P_3 residue (Figure 35A). The reaction coordinate was sampled in windows with a width of 0.1 Å ranging from 2.5 to 9.0 Å. Each window was 10 ns in length and was recorded in triplicates, thus allowing for wellconverged potential of mean force (PMF) estimates. In order to keep the inhibitor in the binding site and effectively to make the PMF comparable between the different inhibitors, we restrained the inhibitor in the binding pocket using a harmonic restraining potential. The potential was applied between the aromatic portion of the P_1 moiety on the inhibitor and the carboxylic acid function of D189 in the S_1 binding pocket with a flat bottom ranging from 0 to 12 Å. The potential was only barely encountered during all umbrella sampling calculations and was found to have an impact on the PMF estimations (for further details, see https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.9b01196, Supporting Information for SANDNER *et al. J. Med. Chem.*, **2019**, 62, 21, 9753–9771).

The resulting free energy profiles along the reaction coordinate d_{O-C} (for **3.1**) as well as d_{O-X} (for **3.2**/**3.3**, **3.4**/**3.5**, and **3.6**/**3.7**) are summarized in Figure 36A – F. We identified all relevant minima and their occupancies along the reaction coordinate using the probability distribution obtained from the weighted histogram analysis method (WHAM).[132,133] Also, we estimated the preorganization energy for the unbound state, ΔE_{PO}, from an additional energy analysis on top of the umbrella sampling simulations. This energetic contribution is interpreted as the change in energy of the system in order to adapt the preorganized state (i.e., the bound conformation). If this value is negative, preorganization is energetically preferable, and if it is positive, then it is unfavorable. The minima, occupancies, and preorganization energies are summarized in Table 5. A detailed description of the algorithm, which was used to estimate these properties of the unbound ensemble, is outlined in the Experimental Section.

Table 5: Summary of the number of preferred conformational states of **3.1** – **3.7** in the unbound and bound states.

inhibitor	Number of preferred states unbound [a]	Relative occupancy unbound [b]	Number of preferred states bound [a]	Relative occupancy bound [b]	ΔE_{PO} [kJ/mol] [c]
3.1	2 (3.5, 7.1)	46, 54	1 (4.7)	100	0.0 [d]
3.2	2 (3.7, 7.4)	54, 45	2 (2.9, 5.3)	18, 82	–2.1 ± 0.2
3.3	2 (2.8, 8.4)	67, 33	1 (3.9)	100	–0.8 ± 0.1
3.4	4 (3.3, 4.4, 5.7, 7.6)	20,21,25,34	2 (2.9, 5.4)	16, 84	0.2 ± 0.1
3.5	4 (2.8, 4.5, 6.1, 7.6)	39,18,16,27	1 (2.9)	100	3.7 ± 0.1
3.6	3 (2.8, 4.7, 7.4)	51,24,26	1 (2.9)	100	0.1 ± 0.1
3.7	4 (2.8, 4.5, 6.1, 7.6)	37,20,17,26	1 (3.1)	100	3.0 ± 0.1

[a] Reaction coordinate values of minimal free energy obtained from the probability distribution according to the WHAM estimator. Units are Å.

[b] Relative probabilities of the detected local minima normalized on the sum over all detected minima in percentage.

[c] The standard errors were estimated from 10 000 bootstrapping attempts and error propagation.

[d] Due to reasons of internal symmetry, the preorganization energy of this molecule was set to 0.0 kJ/mol. The intrinsic mirror symmetry of the P_1 portion of 3.1 along the reaction coordinate results in no distinguishable conformational states.

Inhibitor **3.1** served as point-of-reference in order to evaluate the degree of self-interaction promoted preorganization that is directly induced by polar interactions along the reaction coordinate. It can be seen that the first free energy minimum of the unbound state (3.5 Å) of the reaction coordinate matches perfectly well with the experimentally observed distance in the bound state (d = 3.5 Å). In addition to this minimum, we found another one at 7 Å, which reflects the same geometry of **3.1** because of the intramolecular mirror symmetry of the P_1 group taking flipped orientation (Figure 36A). For inhibitor **3.1**, the second major minimum at 7 Å found in the bulk water phase is not reproduced in the protein-bound state (Figure 36B). We believe this is due to the fact that reaching the flipped orientation of the benzamidine moiety requires passing through a sterically highly strained geometry, which would also force the protein-binding pocket to adopt highly unfavorable conformations. The inhibitor is (smoothly) forced to adopt the individual reaction coordinates, starting from the reaction coordinate value and conformation found in the crystal structure. We assume that after passing these unfavorable conformations, the protein environment is not able to fully adopt a relaxed conformation for inhibitor **3.1** at the anticipated minimum at 7 Å within our umbrella sampling approach. For the unbound inhibitor in the unconstraining bulk water phase, this flip-over of the aromatic portion is not restricted. In contrast to the other inhibitors in the present series, the preorganization energy as well as the conformational occupancies of **3.1** inevitably inherents the intrinsic mirror symmetry of the P_1 group. Consequently, it is apparent that the preorganization energy of **3.1** must be set to 0 kJ/mol because it already adopts the preorganized configuration in solution.

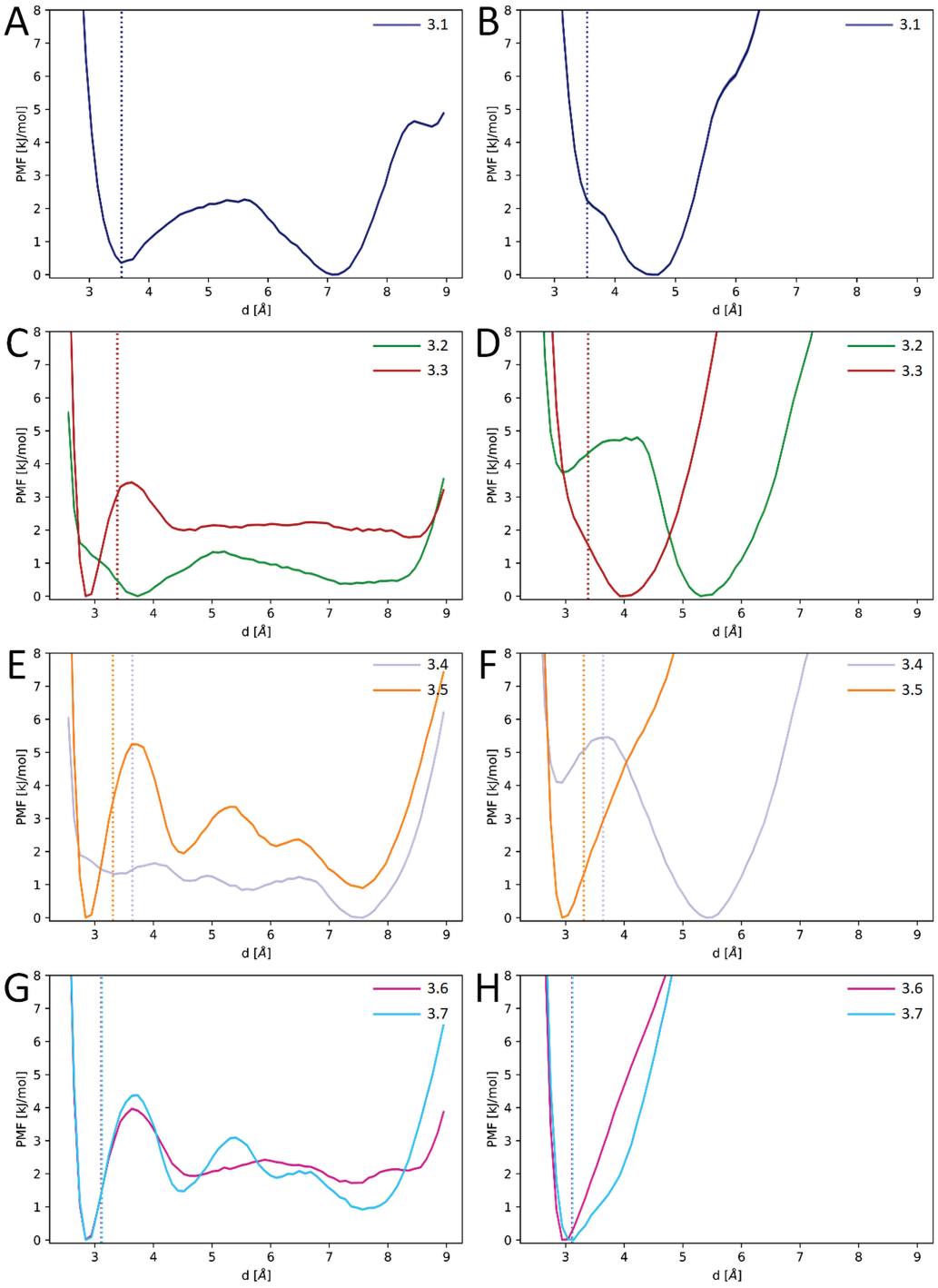


Figure 36: PMF plots for different inhibitor molecules in the unbound (left) as well as in the bound state (right). The dashed line indicates the distance (i.e. the value of the reaction coordinate) found in the crystal structures of the respective protein-inhibitor complex. Note, that for **3.2** and **3.3** the reaction coordinate was found to have similar values in the crystal structures, therefore only the value of **3.3** is visible in the plot. Similarly, **3.6** and **3.7** have identical values, therefore only the dashed line for **3.7** is visible. For all PMF curves, upper and lower bound error estimates based on 10 000 bootstrapping attempts are overlaid but are effectively not visible due to the small error ranges. A, C, E, G: PMF for the unbound state of inhibitor **3.1** and inhibitor pairs **3.2**/**3.3**, **3.4**/**3.5** and **3.6**/**3.7**, respectively; B, D, F, H: PMF for the bound state of inhibitor **3.1** and inhibitor pairs **3.2**/**3.3**, **3.4**/**3.5** and **3.6**/**3.7**, respectively.

From the PMF profile of inhibitor **3.2**, it is apparent that in the unbound state (Figure 36C) the hydroxyl group leads to a free energy minimum at a similar position as that found for **3.1**. However, the depth of the minimum is lower and shallower than that for **3.1**, and by that, it is accompanied by a small population shift toward the correctly preorganized first minimum at 3.7 Å (3.4 Å in the crystal structure). In the bound state of **3.2** (Figure 36D), two PMF minima are found. This can be due to a distribution between two rotamers of the $-CH_2-OH$ group: one involving the terminal OH in an intramolecular H-bond and the other orienting the OH group toward neighboring water molecules as well as to the amide hydrogen atom of E192 (for further details, see https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.9b01196, Supporting Information for SANDNER *et al. J. Med. Chem.*, **2019**, 62, 21, 9753–9771). From these two rotamers, only the one with an intramolecular hydrogen bond (2.9 Å, see Figure 36D) corresponds to the populated conformer that resembles the one found in the crystal structure. Inhibitor **3.3** shows in the unbound state only a single major minimum at 2.8 Å (Figure 36C), indicating a short intramolecular hydrogen bond. This PMF minimum corresponds to a slightly extended value of 3.9 Å in the bound state (3.4 Å in the crystal structure). Remarkably, it does not have any additional minimum in the bulk water phase. This indicates a correctly preorganized inhibitor molecule and ultimately explains the experimentally found entropy-driven boost in the binding free energy of **3.3** compared to that of **3.1** and **3.2**. The preorganization energies of **3.2** and **3.3** are both negative, −2.1 and −0.8 kJ/mol, respectively. The less favorable preorganization energy of **3.3** compared to that of **3.2** can be due to the higher desolvation cost of the charged ammonium group at **3.3**.

Inhibitor molecule **3.4** exhibits in the unbound state (Figure 36E) multiple shallow PMF minima along the reaction coordinate, thereby leaving **3.4** in multiple conformations with a global minimum at 7.6 Å (34% populated). Also, inhibitor **3.4** has the largest reaction coordinate value (3.6 Å) in the crystal structure among all inhibitors. In contrast, **3.5** shows four clearly defined PMF minima, with the deepest minimum at 2.8 Å (3.3 Å in the crystal structure) also being the most populated one (39%). In the bound state (Figure 36F), **3.4** experiences two minima very similar to **3.2**, which possibly indicates the exaggerated handling of interactions between solvent molecules and the OH group of the inhibitor. Similar to the comparison of **3.2**/**3.3**, compound **3.5** (ΔE_{PO} of 3.7 kJ/mol) exhibits a more unbeneficial preorganization energy than inhibitor **3.4** (0.2 kJ/mol). However, when comparing **3.2**/**3.4** and **3.3**/**3.5**, it is apparent that both **3.2** and **3.3** have favorable preorganization energies with respect to their benzylsulfonyl-glycin analogues **3.4** and **3.5**. This is caused by a special *β*-turn conformation (see Figure 37C for a representative conformation of inhibitor **3.5**), which is mediated by an intramolecular hydrogen-bond interaction between the oxygen atoms of the sulfonamide group and the amide hydrogen atom of the P_1 group. Thus, this energetically stable feature can most likely only be adapted by molecules that bear the benzylsulfonyl-glycine moiety. The conformation is adopted at extended values of the reaction coordinate and is very low in

energy (see Figure 37A, B for energy profiles). A very similar PMF profile as for the unbound and bound states of **3.5** was found for inhibitor **3.7** (Figure 36G, H), as it shows four distinct minima in the unbound form and one major minimum in the bound form. Also, the major PMF minimum at 2.8 Å for these inhibitors in the water phase is similar to that found for **3.3** in the water phase. Thus, the charged ammonium group seems to trigger the occurrence of short reaction coordinate values d_{O-X}, which are in a typical range for hydrogen bonds. Quite contrary, the P_1 portions bearing a hydroxyl group (such as in **3.2** and **3.4**) are found to have the corresponding PMF minimum at much wider distances in their unbound states (d = 3.7 and 3.3 Å, respectively). From the comparison of **3.5** and **3.7**, we conclude that the addition of the (*R*)-arginine side chain does not influence the adopted conformational ensemble in the bulk water phase drastically. In contrast to **3.7**, inhibitor **3.6** exhibits a shallow PMF profile in the unbound state, however with a major minimum at 2.8 Å (3.1 Å in the crystal structure for both **3.6** and **3.7**). This indicates that inhibitor **3.6** is better preorganized than inhibitor **3.7**. Consequently, the (*S*)-arginine group improves the preorganization of the inhibitor by destabilizing the states, which do not correspond to the bound state. The preorganization energy for inhibitor **3.7** is less favorable with respect to **3.6** (3.0 vs 0.1 kJ/mol, respectively), which indicates for **3.7** stronger intramolecular interactions experienced in the conformational states at larger values of the reaction coordinate (67% of the time) than for **3.6** (50% of the time) in the unbound state prior to protein binding.

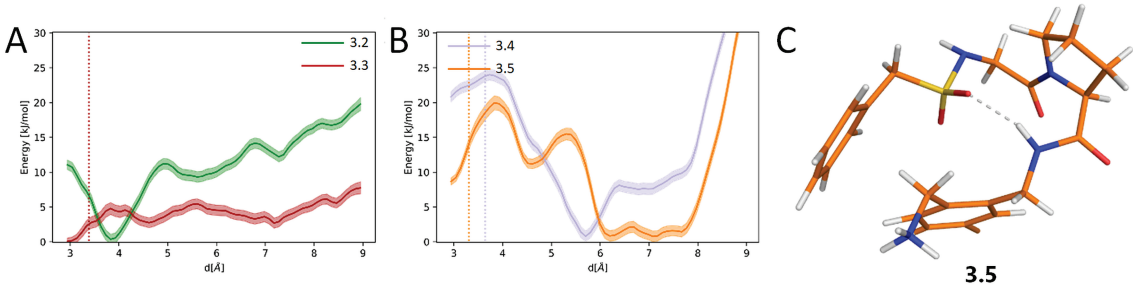


Figure 37: Energies and representative *β*-turn conformation. The standard error in the energy diagrams were obtained from 10 000 bootstrapping attempts. A) Energy profile along the reaction coordinate for inhibitor molecules **3.2** and **3.3** in their unbound states. B) Energy profile along the reaction coordinate for inhibitor molecules **3.4** and **3.5** in their unbound states. C) Representative *β*-turn conformation found for inhibitor **3.5** at d_{O-X} = 7.0 Å. The dashed line indicates a hydrogen bond of 1.9 Å between the hydrogen atom at the amide group and one of the oxygen atoms in sulfonamide group.

3.7 Discussion and Conclusion

Crystal structure analysis confirms similar binding modes for all studied inhibitors with efficient filling of the S_1, S_2, and $S_{3/4}$ pockets apart from epimer **3.7** containing the polar (*R*)-arginine in P_3 position which leaves the hydrophobic $S_{3/4}$ pocket partly unoccupied. All chlorobenzyl derivatives, exhibiting a well-placed $-CH_2-\underline{\mathbf{X}}H_n$ (X = N, O) substituent at the aromatic P_1 anchor, fix the bound conformation via an intramolecular linkage between the amino or hydroxy-methylene group at P_1 and the oxygen atom of the carbonyl group from the P_3 portion. In some cases, a recruited water molecule (O3) is involved in mediating the contact. Remarkably, **3.1**, which lacks the $-CH_2-\underline{\mathbf{X}}H_n$ substituent at the P_1 anchor, also adopts this geometry in the crystal structure. This underlines that our design correctly fixed the required conformation at the binding site. This per se allows for preorganization; however, it will boost affinity only if this geometry is already adopted in the unbound state.[134]

At first, the inhibitors **3.1**, **3.2**, and **3.3** will be compared. All exhibit a terminal amino group, which, as previous studies with this scaffold have shown,[125] is present as the charged ammonium group. As described above, because of compensating effects, no buffer dependence has been found for this inhibitor scaffold.[125] In **3.1**, the benzamidinium group contributes a second positive charge and, as shown in other studies with this P_1 anchor,[118] the group is already protonated in aqueous solution prior to thrombin binding. This molecule is well preorganized in the unbound state, adopting conformations that resemble the one found in the protein-bound state as indicated by our MD simulations. The benzamidinium group forms a strong dual salt bridge (2.8 Å) to D189. A water molecule (O1), already present in the *apo* structure,[118] is located above the centroid of the aromatic ring of Y228 (3.6 Å) and forms an H-bond to one of the amidino-NHs. The displacement of this water molecule from the *apo* enzyme by the chlorine atoms of **3.2** − **3.7** is likely accompanied by a negative enthalpic signal. This is explained by the residual mobility of this water molecule in the S_1 pocket, and it does not gain additional degrees of freedom once released to the bulk phase, but there it benefits from establishing more favorable interactions with adjacent water molecules.[118,135]

The inhibitors **3.2** and **3.3** differ only by the replacement of the primary amino (**3.3**) with a hydroxy group (**3.2**) providing a very different thermodynamic binding profile (Figure 29, top). Assuming that the amino-methylene group is protonated, it will serve as a three-fold H-bond donor, whereas the hydroxy-methylene group can only donate one proton (Figure 31C, E). As a result, **3.2** cannot interact with G219 and E192. Instead, **3.2** accepts an H-bond from water O2 with d = 2.7 Å. In total, **3.3** can form more and, because of its charged state, stronger contacts than the structurally related **3.2**. Overall, **3.3** is by $\Delta\Delta G^\circ_{3\rightarrow2}$ = −8.3 kJ/mol more potent than **3.2** (Figure 27). The MD simulations suggest that **3.3** adopts, prior to protein binding, preferentially (67% of the time, see Table 4) a conformation similar to that in the bound state, and this effect of better preorganization boosts the affinity of **3.3** over **3.2**. However, the

energetic cost of depopulating the extended conformations of **3.3** is higher than that of **3.2** (see Table 2). Thus, a gain in binding energy for **3.3** over **3.2** in the bound state is lowered by this unfavorable contribution. Enthalpically, both inhibitors experience a similar signature ($\Delta\Delta H°_{3\rightarrow2}$ = −2.0 kJ/mol); thus, the affinity difference is mainly of entropic nature ($-T\Delta\Delta S°_{3\rightarrow2}$ = −6.3 kJ/mol). This is in agreement with other examples, where correct preorganization provides an entropic benefit and accordingly boosts affinity.[99-101]

When comparing **3.4** and **3.5**, structurally similar properties are observed for both sulfonamide derivatives. Also here, the differences result from the exchange of the amino (**3.5**) to the hydroxy group (**3.4**). Nevertheless, the affinity loss of the hydroxy derivative is in this case larger ($\Delta\Delta G°_{5\rightarrow4}$ = 12.0 kJ/mol) and factorizes in an enthalpic loss of $\Delta\Delta H°_{5\rightarrow4}$ = 15.0 kJ/mol and an entropic advantage of $-T\Delta\Delta S°_{5\rightarrow4}$ = −3.0 kJ/mol (Figure 27). On first glance, this opposing signature to the previous matching pair **3.2**/**3.3** argues against the beneficial effect of preorganization. In solution, **3.5** adopts to 39% of the time a conformation also found in the binding pocket. However, additional conformational states that correspond to *β*-turn-type geometries are populated in the water phase. Obviously, upon protein binding resulting in one preferred conformation, **3.5** sacrifices a fair number of populated conformational states in solution (Figure 36E), which reduces its binding entropy compared to its D-phenylalanine analogue **3.3**, which populates less conformers prior to binding. Interestingly, **3.4** is hardly found in solution in the S-shaped geometry required at the binding site. Instead, several extended conformations are predominantly populated and energetically preferred (see Figure 37B). Nonetheless, we must note that extended conformations for **3.2** and **3.4** are favored in the bound state as well. We assume that this is partly caused by a force-field artifact that leads to non-optimal treatment of the hydroxyl group, as was also noted elsewhere.[136,137]

Ideally, one would need to use the force fields with optimized parameters for OH groups. Also, NMR investigations could be carried out in order to examine the solution ensemble of these inhibitors and eventually correct for shortcomings in the parameters of the force field. Moving from unbound to bound state indicates that for **3.4** entropically only minor losses are there in the number of conformational states; however, adopting the S-shaped geometry is here energetically less favored, particularly considering that several *β*-turn-type geometries are populated in solution. Likely, this contributes to the less favorable $\Delta G°$ and $\Delta H°$ signals of binding.

Finally, the two epimers **3.6** and **3.7** differ not only in their thermodynamic binding signatures but also in the adopted binding modes. Both bear a central P_3 arginine residue, which, according to its pK_a value, is present in the charged state. The (*S*)-configured **3.6** binds similar to **3.4** and **3.5**, except that it exposes its arginine side chain toward the solvent. The poorly defined electron density suggests that the side chain is scattered over several arrangements. Compared to **3.3** and **3.5**, **3.6** loses binding affinity by about 1 order of magnitude. Likely, this is related to detrimental effects of the surface exposure of the arginine side chain in the bound

state. In a recent study, we detected that charged inhibitor groups still binding close to the protein surface can easily reduce the binding affinity by one or two orders of magnitude.[138]

For steric reasons, the (*R*)-analogue **3.7** orients its central arginine side chain toward the $S_{3/4}$ pocket, which remains, however, partly unoccupied as proper placement of the terminal guanidinium group into this hydrophobic site would be energetically quite unfavorable. Some filling of the pocket is achieved by the hydrophobic $-(CH_2)_3-$ portion of the arginine side chain. On first glance, it is surprising that **3.7** is by $\Delta\Delta G^\circ_{6\rightarrow7}$ = −5.1 kJ/mol more potent than **3.6** (Figure 27). The benzyl sulfonamide group binds along the surface of the complex shielding E217 from direct exposure to solvent. Interestingly, the inhibitor's guanidinium group at the central arginine residue involves the carboxylate function of E217 in a salt bridge. As noted earlier, the contribution of a surface-exposed salt bridge to affinity is often negligible.[139] However, here, because of this inhibitor-induced self-shielding effect of the adjacent benzyl group, the formed salt bridge gets buried and likely contributes now significantly to the enthalpic signal of the protein binding process. This effect likely even overcompensates the energetically unfavorable preorganization energy induced by the (*R*)-arginine moiety on this inhibitor. To some extent, the enthalpic gain $\Delta\Delta H^\circ_{6\rightarrow7}$ = −23.4 kJ/mol is compensated by an entropically unfavorable contribution $-T\Delta\Delta S^\circ_{6\rightarrow7}$ = +18.3 kJ/mol spent to produce the significantly higher ordered and more rigid binding pose of **3.7**. Both inhibitors **3.6** and **3.7** exhibit the $-CH_2-NH_2$ side chain at the P_1 anchor, which was made responsible for the favorable inhibitor preorganization, particularly of **3.3** over **3.2**. In the unbound state, MD assigns similar conformational properties to **3.6** and **3.7** with respect to the reaction coordinates, and they resemble the bound situation (**3.6**: 50%, **3.7**: 37%). Inhibitor **3.7** shows a slightly enhanced population of additional conformational states (63%) than **3.6** (50%). The latter aspect further contributes to the huge entropic difference $-T\Delta\Delta S^\circ_{6\rightarrow7}$ = +18.3 kJ/mol observed between the two inhibitors and reflects the deviating conformational properties of **3.6**/**3.7** in the unbound situation. In late-stage drug optimization, medicinal chemists seek for design concepts to optimize their inhibitors without swapping to an entirely different scaffold. One promising strategy is preorganization. This can either be accomplished by fixing the bound conformation in an appropriate ring system. An alternative concept with usually easier to synthesize molecules is the fixation via intramolecular H-bonds. The current study demonstrates that such a strategy can be successful, if the required bound conformation of the molecule with the archetype binding mode (in our study, inhibitor **3.1**) is well reproduced by the modified molecules in the bound state (in our study, **3.2** – **3.6**). However equally important, the added intramolecular contact must be present, best as overwhelmingly populated conformer (*cf.* **3.3**), in aqueous solution prior to protein binding. Only then the expected affinity boost will be experienced. In the worst case, the modified scaffold adopts a novel, highly preferred conformation in solution and the preorganization effect is lost (*cf.* **3.4**). In the current case, replacement of the charged N-terminal P_3 residue in **3.1** – **3.3** by a neutral benzylsulfonyl-glycine segment in **3.4** – **3.5**

significantly populates geometries which resemble preferred *β*-turn-type conformers in solution prior to protein binding and explains some of the observed affinity losses.

Furthermore, this study provides unexpected insights into two other effects, often overlooked in inhibitor optimization. They become obvious comparing the two studied epimers **3.6** and **3.7**. One of them, even though showing the same optimal filling of all specificity pockets as the other members of the congeneric series, loses significant affinity. Likely, this results from the fact that an attached charged side chain, even though it is oriented away from the protein toward the bulk solvent, still has sufficient detrimental impact on binding so that the affinity is reduced by an order of magnitude. Instead, the other epimer **3.7** gains affinity and shows a huge enthalpic benefit. Even though it fills one specificity pocket only partly, it succeeds to involve the carboxylate group of a surface-exposed glutamate residue in a salt bridge. Usually, the formation of such surface-exposed protein-inhibitor salt bridges contributes nothing to binding affinity. However here, these inhibitors succeed to adopt a binding pose that shields and therefore "self-buries" the formed salt bridge to some extent. By this, the formed surface-exposed contact is removed from water access and contributes then significantly to inhibitor binding.

Chapter 4

A Fragment Points to Selectivity Discriminating Features between the Related Proteases Trypsin and Thrombin

Anna Sandner[1], Khang Ngo[1], Johannes Schiebel[1], Angela Ilse Marca Pizarroso[1], Linda Schmidt[1], Barbara Wienen-Schmidt[1], Benjamin Wenzel[1], Torsten Steinmetzer[1], Andreas Ostermann[2], Andreas Heine[1], Gerhard Klebe[1]

[1] Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, Marbacher Weg 6, 35032 Marburg, Germany

[2] Heinz Maier-Leibnitz-Zentrum, Technische Universität München, Lichtenbergstraße 1, 85748 Garching, Germany

Journal of Medicinal Chemistry **2021**, 64, 3, 1611 – 1625

4.1 Introductory Remarks

This project was elaborated in cooperation with Dr. Khang Ngo (group of Prof. Gerhard Klebe, Philipps-Universität Marburg, Germany) and Dr. Johannes Schiebel (group of Prof. Gerhard Klebe, Philipps-Universität Marburg, Germany).

The thrombin variant E192Q was expressed, purified, and crystallized by the author of this work. In addition, the author performed most of the enzyme kinetic measurements and other ITC measurements. The estimated own contribution is approximately 40%.

Further expression, crystallization, and crystal structure analyses as well as various ITC and enzyme kinetic assay measurements (approx. 30%) were performed by Dr. Khang Ngo. Dr. Johannes Schiebel performed the additional expression, crystallization, and crystal structure analysis as well as the preparation of the neutron structures (approx. 20%). The protein variant E192Q was mutated, expressed, and prepared for the enzyme kinetic assay (approx. 5%) by Angela Ilse Marca Pizarroso (group of Prof. Gerhard Klebe, Philipps-Universität Marburg, Germany). Linda Schmidt (group of Prof. Gerhard Klebe, Philipps-Universität Marburg, Germany) mutated and expressed the thrombin protein variants D221aA − D222K and Y225P and prepared them for the enzyme kinetic assay (approx. 5%).

Angela Ilse Marca Pizarroso, Linda Schmidt, Dr. Benjamin Wenzel (group of Prof. Torsten Steinmetzer, Philipps-Universität Marburg, Germany) and Prof. Torsten Steinmetzer helped to carry out the enzyme kinetic measurements. Dr. Andreas Ostermann (Heinz Maier-Leibnitz Zentrum, Technische Universität München, Germany) helped with the neutron diffraction experiments. Prof. Andreas Heine (Philipps-Universität Marburg, Germany) and Prof. Gerhard Klebe helped to interpret the data.

The manuscript was collaboratively written by the author of this thesis, Dr. Khang Ngo, Dr. Johannes Schiebel and Prof. Gerhard Klebe.

4.2 Abstract

In the S_1 pocket, the serine proteases thrombin and trypsin commonly feature D189 and a A190S and E192Q exchange. Nevertheless, thrombin cleaves peptide chains solely after arginine, and trypsin after lysine and arginine. Thrombin exhibits a Na^+-binding site next to D189, which is missing in trypsin. The fragment benzylamine shows direct H-bonding to D189 in trypsin, while in thrombin, it forms an H-bond to E192. A series of fragments and expanded inhibitors were studied against both enzymes and mutated variants by crystallography and ITC. The selectivity-determining features of both S_1 pockets are difficult to assign to one dominating factor. The A190S and E192Q replacements may be regarded as highly conserved as

no structural and affinity changes are observed between both proteases. With respect to charge distribution, E192, together with the thrombin-specific sodium ion, helps in creating an electrostatic gradient across the S_1 pocket. This feature is definitely absent in trypsin but important for selectivity along with solvation-pattern differences in the S_1 pocket.

4.3 Introduction and Terms of Reference

The success of a drug molecule in therapy strongly depends on its potency, efficacy, and selectivity toward a given target. At first, high potency is a prime prerequisite for its suitability along with uncritical toxicity. Efficacy is correlated with binding kinetics, which depends on the individual association (k_{on}) and dissociation (k_{off}) rate constants and influences the residence time of a drug molecule on its target. In addition, high selectivity for a target protein may be prerequisite for effective therapy without unwanted side effects; however, a mixed profile toward several closely related isoforms of a target family can also be desired. The question of which profile of a drug molecule is optimal for the action toward a given pharmacological target is difficult to answer in advance and makes it problematic to assess the therapeutic value of drug-development candidates prior to clinical trials. However, as long as the molecular determinants of already known selectivity profiles are hardly understood, it is difficult to come up with novel compounds mastering unseen selectivity challenges. An obvious criterion for high potency and often also for sufficient selectivity is the optimal shape complementarity of the drug and binding site; however, steric fit is perhaps the most intuitive but, by far, not the only criterion to achieve selectivity.

An important class of target structures are serine proteases as they fulfill numerous key functions within the human body (Chapter 1.2.2.1). Trypsin, for instance, plays a decisive role during digestion due to its ability to catalyze the cleavage of proteins' peptide bonds C-terminally of an arginine or lysine residue.[140] The family of trypsin-like serine proteases, of which trypsin is one of the most prominent and best studied enzymes, does also contain many blood clotting factors such as thrombin and factor Xa, two enzymes that cleave substrates exclusively after an arginine residue. For both, approved anticoagulants have been developed that enable the treatment of life-threatening thromboembolic diseases.[141,142]

Trypsin-like serine proteases are structurally closely related. In order to avoid undesired therapeutic side effects, it is therefore essential to consider selectivity as a determining parameter during the development of new anticoagulant drugs. For instance, orally active thrombin and factor Xa inhibitors must be selective against trypsin to avoid being trapped in the intestine where trypsin is present in high concentrations of up to 6 µM.[143] A high selectivity is also required against their physiological opponents, the fibrinolytic proteases plasmin and tissue Plasminogen Activator (tPA), which are responsible for the dissolution of blood clots.[144] The

ability of an inhibitor to discriminate the intended target, e.g., thrombin, from trypsin can be used as a first indicator to tailor selectivity.[145]

Substantial selectivity discrimination results already from the portion of an inhibitor binding to the S_1 site, which, in both proteases, hosts an aspartate residue at the bottom of this pocket.[146] In light of this hypothesis, structural variations between trypsin and thrombin within this pocket are made responsible for the fact that trypsin cleaves substrates C-terminally of arginine and lysine while thrombin specifically cleaves substrates following an arginine residue only.[147] A change from an alanine to a serine residue in position 190 constitutes the most obvious and prominent difference between the S_1 pockets of thrombin and trypsin and therefore was not only used to explain the different substrate specificities but also for the design of thrombin-selective inhibitors. Although both proteases have a highly similar S_1 pocket, our results suggest a difference in the electrostatic properties of D189, which may contribute to their selectivity differences.[135,148] It can be hypothesized that the observed differences may be at least partially caused by a sodium ion binding close to D189 in thrombin, which is absent in trypsin. This electrostatic attenuation of D189 may be further pronounced by the occurrence of a charged E192 residue at the rim of the S_1 pocket in thrombin that is replaced by an uncharged Q192 in trypsin.

The current study tries to shed more light on the selectivity-determining features between trypsin and thrombin. Our investigations were initially stimulated by an intriguing observation while studying the binding of benzylamine as a probe fragment versus both proteases. In contrast to the obvious assumption that such a simple fragment would bind with its basic functional group to the carboxylate group of D189 at the bottom of the S_1 pocket of both proteases, this binding mode is only found for trypsin. In thrombin instead, the fragment avoids interacting with D189, and seeks contact to E192 to which it likely forms a salt bridge. Such findings could be important in a fragment-based lead optimization project for thrombin and may suggest an alternative strategy to follow as in the trypsin case. We believe that the result in thrombin is even more surprising as E192 is rather solvent-exposed at the rim of the thrombin-binding pocket and, according to its pK_a value in aqueous solution, glutamic acid should be less acidic than aspartic acid and thus the less probable residue to get involved in a salt bridge with a basic residue.[149]

To rationalize this surprising difference, which might further underscore our hypothesis about selectivity-discriminating features, high-resolution neutron and X-ray crystallography, a fluorescence assay, ITC, and site-directed mutagenesis to improve our knowledge about the selectivity features of both proteases. We believe that the collected insights may help in the development of novel drug molecules targeting selectively the important class of trypsin-like serine proteases of which thrombin possesses a cation-binding site next to the S_1 aspartate whereas other family members lack such a close-by ion. In general, our data will improve the understanding of the features in operation to determine selectivity in proteins.

4.4 Results and Discussion

4.4.1 Binding Data of Fragments with Basic S_1-Binding Head Group

At first, we characterized the binding properties of benzylamine (**4.3F**) in comparison to the related basic S_1-binding head groups benzamidine (**4.1F**) and *N*-amidinopiperidine (**4.2F**) against trypsin and thrombin using a fluorescence-based kinetic enzyme inhibition assay.[135,148] We included two additional thrombin variants in this analysis where A190 was replaced by serine in the S_1 pocket and E192 by Q192 next to the rim of this pocket. The idea of the two mutated thrombin variants was to make the S_1 pocket of thrombin more similar to that of trypsin. In Figure 38, the obtained K_i values are depicted.

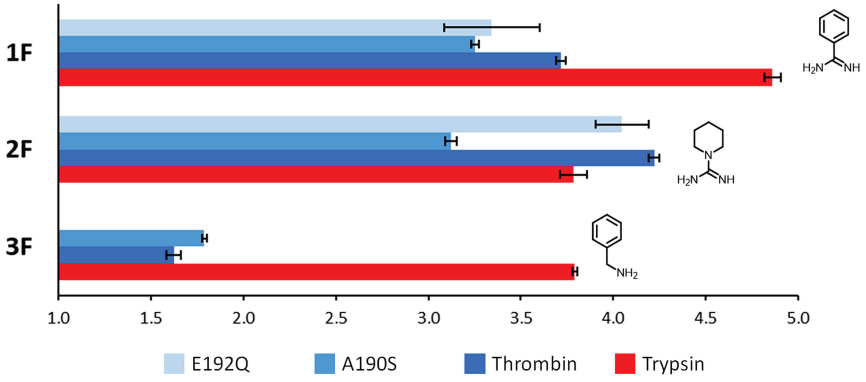


Figure 38: Inhibitory potency of **4.1F** – **4.3F**. Inhibitory potency expressed as pK_i (negative decadic logarithm of the binding constant) of three fragment-like inhibitors benzamidine (**4.1F**), *N*-amidinopiperidine (**4.2F**), and benzylamine (**4.3F**) against trypsin (red), thrombin (blue), A190S thrombin (light blue), and E192Q thrombin (pale blue). For the latter mutated variant, no binding of benzylamine could be detected.

The most potent fragment against trypsin is benzamidine (**4.1F**) followed by *N*-amidinopiperidine (**4.2F**) and benzylamine (**4.3F**) with the latter two being nearly equally potent. Against thrombin, the two amidine fragments swap their order of affinity,[135] whereas benzylamine surprisingly loses more than two orders of magnitude in potency. This trend is also reflected by the two mutated forms. Moreover, the A190S thrombin variant reduces more strongly the affinity of *N*-amidinopiperidine than of benzamidine. Benzylamine inhibits the A190S variant similarly to the wild type, whereas the E192Q variant shows weak or even no affinity against this inhibitor in that no binding could be recorded.

4.4.2 Crystal Structures with Benzylamine as S_1-Binding Head Group

Structurally, we could recently characterize **4.1F** and **4.2F** in complex with trypsin and thrombin to a very high resolution.[128,135] In the case of trypsin, even a neutron diffraction study could be accomplished, which allows structural characterization of protons (and deuterons).[150] Therefore, in the trypsin case, the orientation and the rotational degrees of freedom of water molecules are also observed.[135] The residual solvation patterns of the complexes formed in the S_1 pocket of both proteases, also in comparison to the uncomplexed *apo* protein, deviate significantly and take impact on the binding kinetics.[135,151,152] This explains the affinity difference for both inhibitors. However, the large affinity difference of benzylamine (**4.3F**) against both enzymes stimulated us to investigate its binding modes by crystal-structure analysis.

In the case of trypsin, we succeeded to determine a combined neutron and X-ray structure in complex with **4.3F** (Figure 39). To obtain the most reliable solvation structure of the residual water molecules in the S_1 pocket with respect to their rotational and orientational degrees of freedom, we applied combined XN refinement making simultaneous use of the collected neutron and X-ray data.[153]

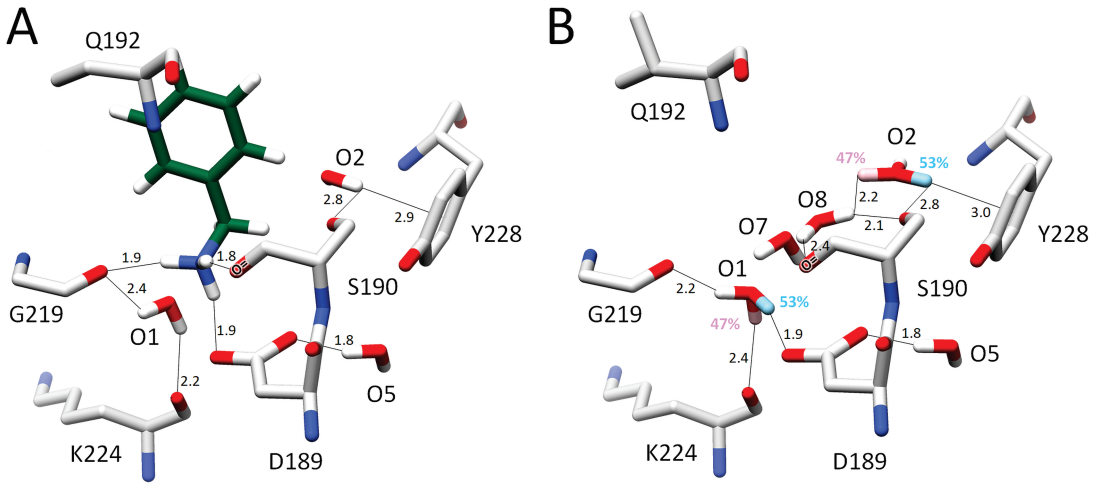


Figure 39: XN structure of the benzylamine-trypsin complex. Selected protein residues are depicted as gray sticks. **4.3F** is shown as green sticks including all hydrogens. A) Binding pose of **4.3F** within the S_1 pocket (PDB code: 5MOR), distances between hydrogen-bond donor deuterium and acceptor atoms are given.[150] B) Dependency of the rotational state of waters O1 and O2 of the uncomplexed *apo* structure (PDB code: 5MOP).[135] O1 and O2 were found to exist in two approximately equally populated rotational states, each represented by an individually colored hydrogen (pink and cyan) with the indicated refined occupancies.

As anticipated, **4.3F** binds in the cationic state with its protonated ammonium group directed toward the carboxylate group of D189. It forms a direct hydrogen bond with one of the carboxylate oxygens (1.9 Å). The other two hydrogens of the NH_3^+ head group are hydrogen-bonded to the backbone carbonyl oxygens of S190 and G219 ($d_{O\cdots H}$: 1.8, 1.9 Å). O1 adopts a fixed geometry with weak H-bonds to the backbone carbonyl groups of G219 (2.4 Å) and K224 (2.2 Å). In the *apo* structure (Figure 39B), O1 is found with multiple orientations and mediates contacts between the carboxylate group of D189, and the backbone carbonyl oxygens of K224 and G219.[135],Water molecule O2, which is located on top of Y228 on the opposite site of the pocket is also found in the complex with **4.3F** with ordered geometry and forms a weak contact to S190Oγ (2.8 Å). Apparently, the residual solvation pattern seems to support the fixation of the inhibitor in the S_1 pocket. Remarkably, the latter water molecule O2 is also disordered in the *apo* enzyme and assumed to swap dynamically between two orientations. Thus, upon **4.3F** binding, O1 and O2 experience major changes in their rotational degrees of freedom compared to the situation in the uncomplexed protein. This observation is particularly surprising, as in the complex with aniline, O2 remains distributed over two states whereas O1 adopts an ordered geometry.[154] This contrasts to the trypsin complex with 2-aminopyridine where, now, O1 remains with multiple orientations and O2 transforms into an ordered state.[154] Obviously, the residual solvation pattern and its intrinsic dynamics depend on the bound inhibitor, which possibly also influences the binding thermodynamics and the selectivity-discriminating properties of accommodated inhibitors.

Surprisingly, benzylamine binds to thrombin with an entirely different orientation as it recruits the carboxylate group of the spatially rather flexible E192 for the interaction (Figure 40A). Due to its pK_a value of 9.3 in aqueous solution,[155] it likely binds also in a protonated state forming a salt bridge. In addition, the ammonium group is hydrogen-bonded to the backbone carbonyl oxygens of G216 and 219 ($d_{N\cdots O}$: 3.1 Å; $d_{N\cdots O}$: 3.2 Å) and to water molecule W6 (3.0 Å). To cross-validate this rather unexpected result and see whether steric demand may be important, we determined the crystal structure of the highly similar 4-methyl benzylamine, which, however, adopts the same binding pose in the S_1 pocket. O2, found to occupy the complex with **4.3F**, is displaced by the attached methyl group for steric reasons (Figure 40A). Likely, the loss of water O2 impacts the stabilization of the adjacent O1 in the complex with the 4-methyl derivative as the difference density does not disclose the presence of this water molecule in the complex as well.

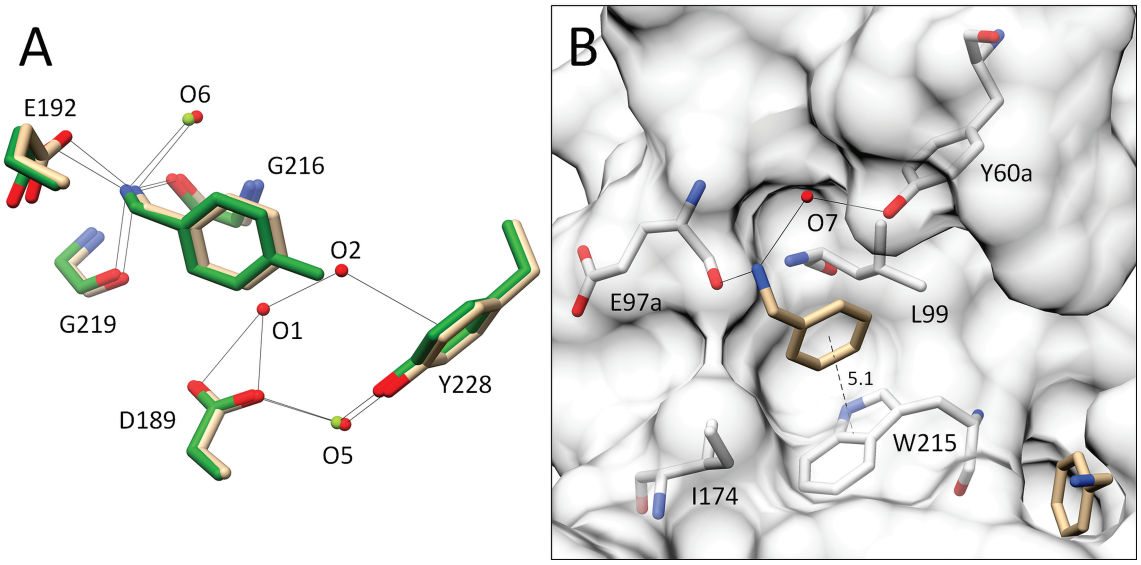


Figure 40: Fragment **4.3F** in complex with thrombin. A) X-ray crystal structures of benzylamine (**3F**) (beige, water molecules red spheres, PDB code: 6T56) and 4-methyl benzylamine (green, water molecules green spheres, PDB code: 6T55) in the S_1 pocket of thrombin. Selected protein residues and the ligand **3F** are depicted as beige and green sticks, respectively. O2 is replaced from the complex by the 4-methyl group of the extended derivative. B) A second copy of **3F** is found in the $S_{3/4}$ pocket of thrombin forming an edge-to-face interaction with W215.

In the case of **4.3F**, a second copy of the fragment is found in the distal $S_{3/4}$ aryl binding site of thrombin with a slightly inclined edge-to-face geometry on top of W215 (Figure 40B). The amino group is hydrogen-bonded to the backbone carbonyl oxygen of E97a, and a further contact is mediated via a water molecule to the hydroxy group of Y60a. Interestingly, the 4-methyl analog of **4.3F** does not accommodate the $S_{3/4}$ pocket, likely because the attached methyl group hampers the formation of an efficient edge-to-face packing. Additionally, we tried to soak further 4-alkyl benzylamine analogs with longer substituents than a 4-methyl group into thrombin crystals; however, of these, none could be observed as bound to the protease. Likely, the basic head group preferentially selects the carboxylate group of E192 instead of D189. Supposedly, it forms more beneficial interactions with the acid group of this residue (see discussion below). For steric reasons, derivatives with larger 4-alkyl substituents than methyl are no longer bound to thrombin's S_1 pocket, preventing a similar binding mode. Obviously, even with the larger side chains, D189 is not selected as the interaction partner. In comparison to the previously studied basic head groups benzamidine and *N*-amidinopiperidine, both showing good affinity toward both proteases, our structural data for the **4.3F** thrombin complex explain the strongly deviating assay results obtained for trypsin and thrombin. Supposedly, the second copy of **4.3F** in the $S_{3/4}$ pocket of thrombin is only weakly binding as no salt bridges are formed. The refinement, however, indicates full population of **4.3F** in the $S_{3/4}$ pocket. Replacing E192 by glutamine in thrombin results in an even stronger or

complete loss of binding affinity, which falls below the detection limit of the applied assay (Figure 38).

4.4.3 Binding Data and Crystal Structures of Extended Tripeptide-Like D-Phe-Pro-NH-CH_2-R_2 Inhibitors

The rather surprising binding-mode difference of benzylamine in trypsin and thrombin stimulated us to study also the properties of elongated D-Phe-Pro-NH-CH_2-R_2 inhibitors with the three P_1 basic R_2-groups benzamidine (**4.1L**), *N*-amidinopiperidine (**4.2L**), and benzylamine (**4.3L**) (Figure 41) to assess whether the extended inhibitor scaffold enforced a binding mode with the benzylamine head group oriented toward D189. The synthesis of the compounds is described in the Experimental Section.

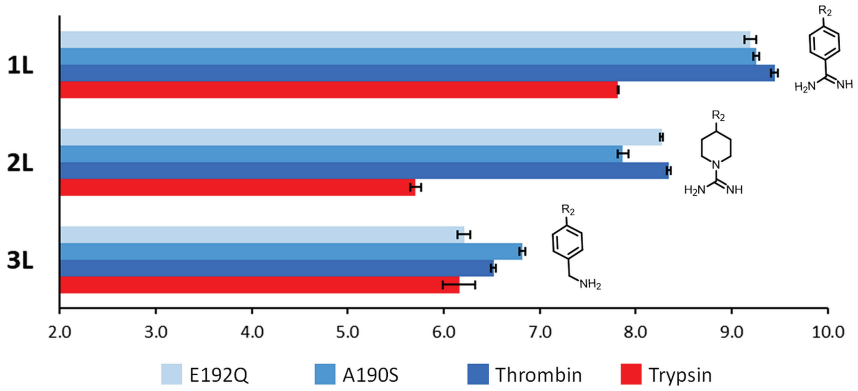


Figure 41: Inhibitory potency of **4.1L** – **4.3L**. Inhibitory potency expressed as pK_i of three D-Phe-Pro-CH_2-R_2 inhibitors **4.1L**, **4.2L**, and **4.3L** against trypsin (red), thrombin (blue), A190S thrombin (light blue), and E192Q thrombin (pale blue).

The affinities of **4.1L – 4.3L** deviate not only in total potency but also in the individual trends from those of the pairwise matching fragment-like P_1 inhibitors. Overall, an affinity improvement of roughly 1.5 – 2.5 orders of magnitude is observed compared with the P_1 fragments, which results from additional interactions experienced by the extended inhibitor scaffolds in the S_2 and $S_{3/4}$ pockets of both enzymes. The expanded inhibitors are more potent against thrombin than trypsin, especially in the case of both amidine derivatives **4.1L** and **4.2L**. This likely correlates with the fact that the corresponding specificity pockets are structurally better developed in thrombin and allow to form more favorable interactions with this protein.[148] Remarkable are the binding data for the D-Phe-Pro-NH-CH_2-benzylamine **4.3L** as this inhibitor seems to recover a major part of the affinity difference found between trypsin and thrombin for the P_1 fragment **4.3F**. Compound **4.3L** is now nearly equally potent against both enzymes and the two mutated thrombin variants.

The affinity data suggest that **4.3L** must adopt a different geometry in thrombin as **4.3F**, its fragment-like head group analog. The crystal structures confirm this assumption (Figure 42A).

A superposition of the complexes with **4.3L** in trypsin and thrombin shows very similar binding modes (Figure 43). Accordingly, all three inhibitors (**4.1L**, **4.2L**, and **4.3L**) now bind with their basic P_1 head groups into the S_1 pocket toward D189. As a mutual superposition shows, identical inhibitors establish virtually the same binding modes in either trypsin or thrombin. In thrombin, the basic P_1 group of **4.3L** is now pointing toward D189 and it forms, similarly as the corresponding P_1 fragment in trypsin, H-bonds to D189, G219, and the carbonyl oxygen of residue 190. Additionally, D189 forms via O5 a water-mediated contact to Y228. The side chain of E192 in the thrombin complex adopts two alternative conformations. The lower populated one (II, 37%) orients away from the bound inhibitor. Instead, the higher occupied orientation (I, 63%) bends back toward the S_1 pocket and forms via O3 an H-bonded contact to G219NH. The corresponding Q192 residue adopts in trypsin a single orientation, which aligns well with the back-folded conformer II in thrombin. Nevertheless, **4.3L** is still lower in affinity against thrombin in comparison to **4.1L** and **4.2L**. Considering that the extended tripeptide-like analogues collect a major part of their affinity enhancement with the P_2 and P_3 portions, the contribution of the benzylamine head group of **4.3L** cannot keep up with that experienced by the P_1 groups of **4.1L** and **4.2L**. The latter amidine-substituted head groups are definitely more appropriate to achieve strong binding to thrombin. Likely, the situation is different in trypsin where **4.3L** is even slightly more potent than **4.2L**.

Furthermore, the large affinity difference of **4.1L** and **4.2L** is surprising against trypsin and thrombin. It matches for trypsin with the data already found for the P_1 fragment-like head groups **4.1F** and **4.2F** where benzamidine is preferred. Remarkably, it is reversed in the case of thrombin (Figure 38 and Figure 41) where **4.2F** is somewhat more potent than **4.1F**. To shed some light on the affinity difference between the benzamidine and *N*-amidinopiperidine head groups toward both enzymes, we performed in another contribution[128] elaborate molecular dynamics simulations along with GIST (grid inhomogeneous solvation theory) evaluations[156-158] of the trajectories to estimate on the solvation thermodynamic parameters.[156] This study showed that the affinity difference between benzamidine and *N*-amidinopiperidine fragments is explained, at least to some part, by deviating solvation patterns in the S_1 pocket. They result from differing solvation barriers for the rehydration of both proteins.[151] Likely, such effects, found for the P_1 head groups, cannot be neglected in discussing the affinity and selectivity differences of **4.1L** and **4.2L**.

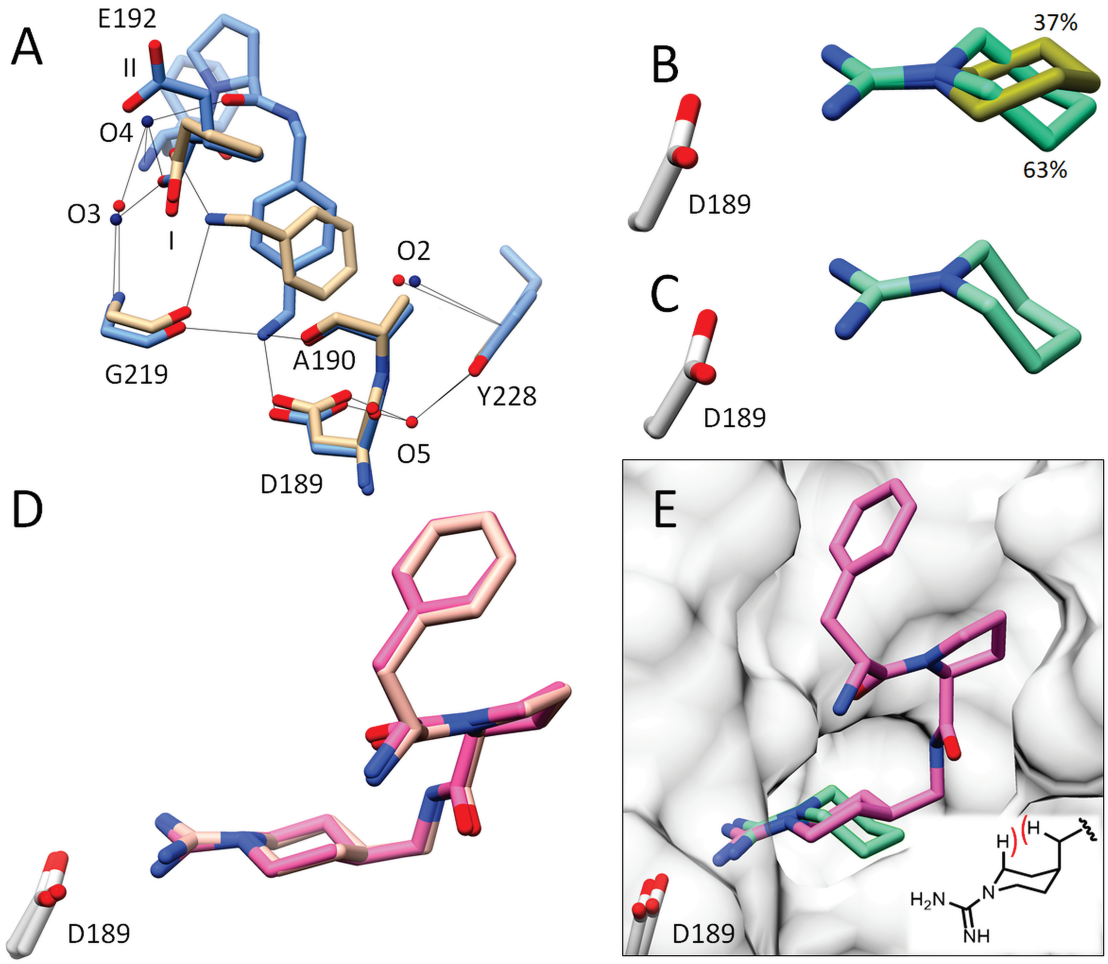


Figure 42: Representation of **4.2L**, **4.3L, 4.2F** and **4.3F** in complex with serine proteases. A) X-ray crystal structure of **4.3L** with thrombin (blue, PDB code: 6T53). The corresponding complex with **4.3F** (beige, PDB code: 6T56) is superimposed. E192 adopts two alternative conformations, I (63% populated) and II (37%), in the complex with **4.3L**. The former corresponds to the geometry found with **4.3F**. B) In the complex with trypsin, *N*-amidinopiperidine (**4.2F**) adopts two different configurations at the ipso nitrogen, showing planar (turquoise, PDB code: 5MNN[135]) and pyramidal geometry (olive). C) In the corresponding complex with thrombin, only the planar geometry at the ipso nitrogen is found (turquoise, PDB code: 4UE7[128]). D) The extended inhibitor **4.2L** occurs only with pyramidal geometry and adopts in both proteins (trypsin, light pink (PDB code: 5MNQ[128]) and thrombin, magenta (PDB code: 6T57)) a virtually identical geometry. E) The planar geometry found for **4.2F** (turquoise, PDB code: 4UE7[128]) cannot be established by **4.2L** (magenta, PDB code: 6T57) in the binding pocket (shown for the thrombin complex) for steric reasons. In the bound structure of **4.2L**, only an equatorial substitution at the C4 atom of the piperidine ring is accepted, whereas an axial arrangement is avoided due to unfavorable 1,3-diaxial interactions (see insert).

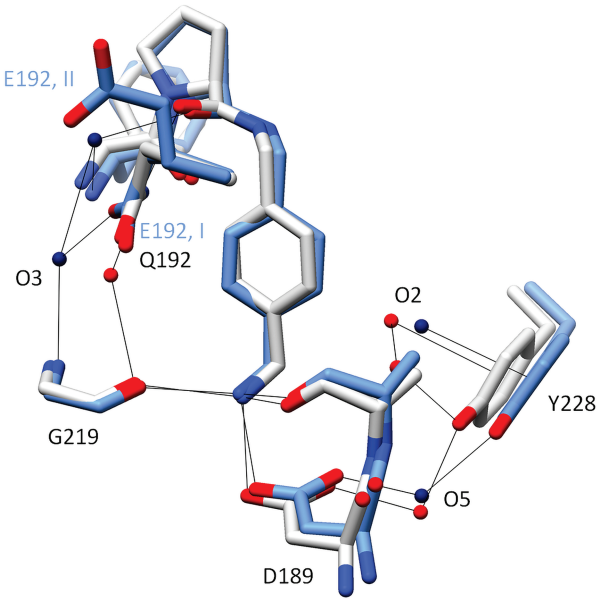


Figure 43: Superposition of the complexes with **4.3L** in trypsin (gray, waters red spheres, PDB code: 6YDY) and thrombin (blue, waters dark blue spheres, PDB code: 6T53). The side chain of E192 adopts two alternative conformations in thrombin (I: 63% toward the inhibitor and II: 37%, away from the bound inhibitor). The corresponding Q192 in trypsin is found in only one single orientation toward the inhibitor.

We studied the geometries of **4.2L** with both proteins in more detail. Again, virtually, the same binding poses are observed. This is in contrast to the geometries found for the fragment-like **4.2F**, as we reported previously.[135,148] The fragment-like inhibitor **4.2F** binds to trypsin with two alternative configurations at the ipso nitrogen of the guanidinium group (Figure 42B). It adopts either a planar (63%) or pyramidal geometry (37%). The second arrangement breaks up the conjugation of the delocalized electron system and concentrates the positive charge on the terminal nitrogens of the guanidinium group directly involved in the salt bridge with the carboxylate group of D189. Remarkably, in thrombin, **4.2F** adopts full planar geometry (Figure 42C).[128] In our previous study,[135] we performed a search in the Cambridge Structural Database (CSD), which showed that pyramidalization at N occurs predominantly in cases where the guanidinum group is involved in strongly polarizing bidentate salt bridges to a carboxylate ion in the crystal structures. In contrast, examples lacking this contacting salt bridge remain much closer to a planar arrangement. These results suggest that the D189 carboxylate group has a stronger polarizing effect in trypsin than in thrombin.[147]

Interestingly, the extended D-Phe-Pro analogue **4.2L** is exclusively found in a pyramidal geometry at the designated *ipso* nitrogen in both proteins. This is explained by the spatially restricting wall of the binding pockets, whereas a planar geometry would force the attached substituent to occupy the unfavorable axial position, creating sterically inconvenient 1,3-diaxial interactions across the six-membered ring (Figure 42D, E). To avoid such unfavorable geometry, the restrictions force **4.2L** to adopt a pyramidal configuration at its *ipso* nitrogen. This leads

to a stronger localization of the electron density on the exocyclic nitrogens and thus enhances the electrostatic interactions with the carboxylate group of D189. This observation correlates with a stronger relative affinity increase of nearly 2 orders of magnitude in the case of thrombin compared with trypsin while expanding the P_1 head group *N*-amidinopiperidine to the tripeptidic analogue **4.2L** with the simultaneous transition from a planar (**4.2F**) to pyramidal geometry (**4.2L**).

4.4.4 ITC Titrations of 4.3L

We recorded ITC experiments with **4.3L** and both enzymes from three different buffers at a pH of 7.8 (Table 6) along with those for **4.1L** and **4.2L**.[125,135] All experiments show hardly any buffer dependence. We can therefore conclude that **4.3L**, having a pK_a value of 9.3 in aqueous solution, binds to the protein without changing its charged protonation state. The same is valid for the more basic head groups of **4.1L** and **4.2L**. As we could show in previous studies, the terminal P_3 amino group becomes fully protonated upon binding. With respect to the overall protonation inventory, this is compensated by the similar amount of protons released from H57 upon inhibitor binding.[125,145,159]

Table 6: Thermodynamic data of $\Delta G°$, $\Delta H°$, and $-T\Delta S°$ for binding of **4.1L – 4.3L** to trypsin and thrombin at pH 7.8.

	thrombin				trypsin			
inhibitor	$\Delta G°$ [a] [kJ/mol]	$\Delta H°$ [b] [kJ/mol]	$-T\Delta S°$ [c] [kJ/mol]	Δn_{H^+} [d] [mol]	$\Delta G°$ [a] [kJ/mol]	$\Delta H°$ [b] [kJ/mol]	$-T\Delta S°$ [c] [kJ/mol]	Δn_{H^+} [d] [mol]
4.1L	–45.2	–38.8	–6.4	0.04	–42.6	–38.2	–4.4	0.17
4.2L	–42.9	–39.5	–3.4	–0.15	–31.1	–21.8	–9.3	0.12
4.3L	–34.8	–19.2	–15.6	–0.04	–33.4	–28.9	–4.6	0.20

[a] $\Delta G°$ is given as the mean of titrations in three different buffers (HEPES, Tricine and Tris).
[b] All $\Delta H°$ results are presented as buffer-corrected values.
[c] $-T\Delta S°$ ° has been calculated as the difference between $\Delta G°$ and the buffer-corrected $\Delta H°$ value.
[d] The molar number of transferred protons results from the slope regression line obtained for the buffer correction.

The formation of the five complexes of **4.1L**, **4.2L**, and **4.3L** with trypsin and **4.1L** and **4.2L** with thrombin is enthalpically favored, which agrees with the observation that a salt bridge is formed to D189. The binding of inhibitor **4.3L** to thrombin falls somewhat out of this series as its binding exhibits a remarkably stronger entropic signature compared to the other complexes. This observation is striking, although the latter complex involves, similarly as with trypsin (Figure 43), the inhibitor's basic head group in a salt bridge with D189. All the more, the entropically strongly favored profile is surprising. However, there is an important

difference considering the interaction inventory of O2 in the six complexes (Figure 44). In the case of **4.1L** and **4.2L**, the charged amidino function establishes an H-bond to O2 in both enzymes. In contrast, the inhibitor's basic head group of **4.3L** orients away from O2; thus, any H-bonding network between this water to the inhibitor is missing (Figure 44C, F). However, since trypsin exhibits a serine at position 190 and thrombin an alanine, O2 can only form an H-bond to S190Oγ in the trypsin complex, whereas in the thrombin complex, this water molecule remains without a similar hydrogen-bonding interaction. It solely contacts F227C═O via a proper hydrogen bond, and a similar contact is found to V227 in the trypsin complex. The expanded contact to W215 and the interaction with the aromatic ring of Y228 help in keeping this water molecule in position. Thus, the significantly reduced inventory of polar and directional H-bonds around O2 in the complex of **4.3L** with thrombin suggests enhanced residual mobility of O2 in the latter complex compared to all other ones, which embed O2 in a network of several hydrogen bonds. Supposedly, this fact contributes to the enhanced entropically favored thermodynamic profile of the **4.3L**-thrombin complex.

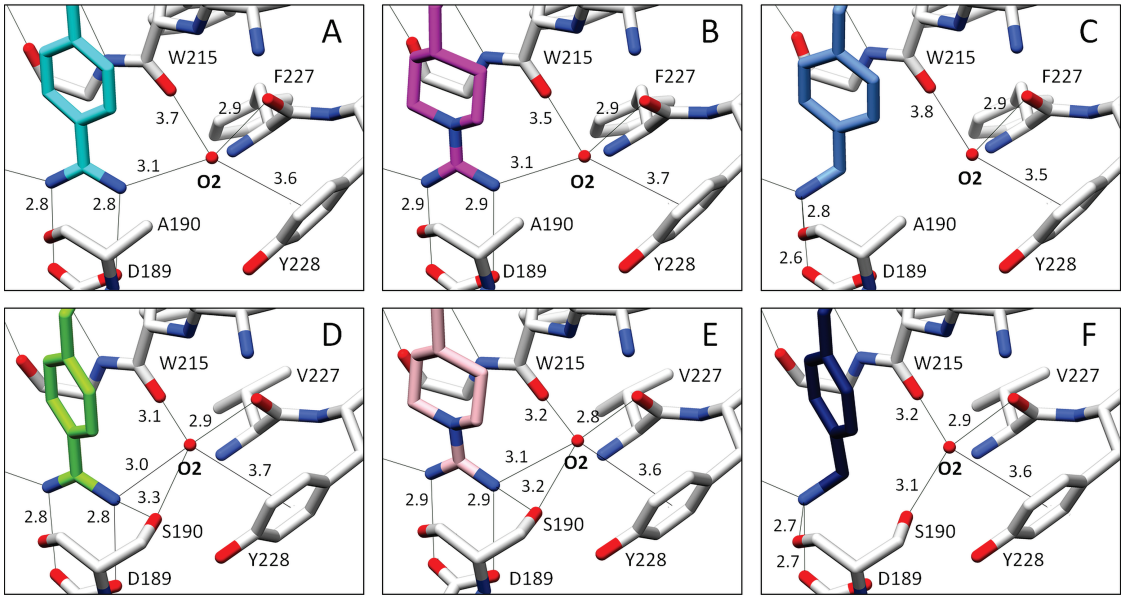


Figure 44: Crystallographically determined binding modes of **4.1L**, **4.2L**, and **4.3L** in complex with trypsin and thrombin. Binding modes of with thrombin (upper row, A – C) and trypsin (lower row, D – E) in the S_1 pocket next to water molecule O2. A) **4.1L** (cyan) with thrombin (PDB-Code: 2ZDA22,25) B) **4.2L** (magenta) with thrombin (PDB-Code: 6T57), C) **4.3L** (light blue) with thrombin (PDB-Code: 6T53) 3, D) **4.1L** (light green) with trypsin (PDB-Code: 6ZQ2), E) **4.2L** (pale) with trypsin (PDB-Code: 5MNQ8), F) **4.3L** (dark blue) with trypsin (PDB-Code: 6YDY). Of the bound inhibitors only the polar head group is shown. In all complexes, O2 forms interactions with the surrounding amino acids, mostly as hydrogen bonds or expanded polar contacts. With Y228, O2 interacts with the aromatic moiety of this residue, all distances in Å. The interaction inventory in thrombin is lower due to the S190A replacement. Particularly, in the **4.3L**-thrombin complex C), the interaction pattern is weakened as the bound inhibitor does not interact with O2.

4.4.5 Affinity and Structural Data of the Mutated A190S Variant of Thrombin

As mentioned, the only obvious difference within the S_1 pockets of thrombin and trypsin is the replacement of alanine by serine in position 190. This exchange has been made responsible for the selectivity discrimination of both proteases.[125,146,147] To make thrombin more trypsin-like, we exchanged alanine by serine via site-directed mutagenesis. In order to assess whether the A190S replacement can be made responsible for the selectivity difference, we determined binding affinities with our fluorometric assay and solved the crystal structures of two inhibitors in the complex with wild-type thrombin and the mutated variant. The affinity difference with respect to the wild type of the studied inhibitors (Figure 38 and Figure 41) appears minor apart from *N*-amidinopiperidine **4.2F** where a difference of approximately 2 orders of magnitude is surprisingly experienced.

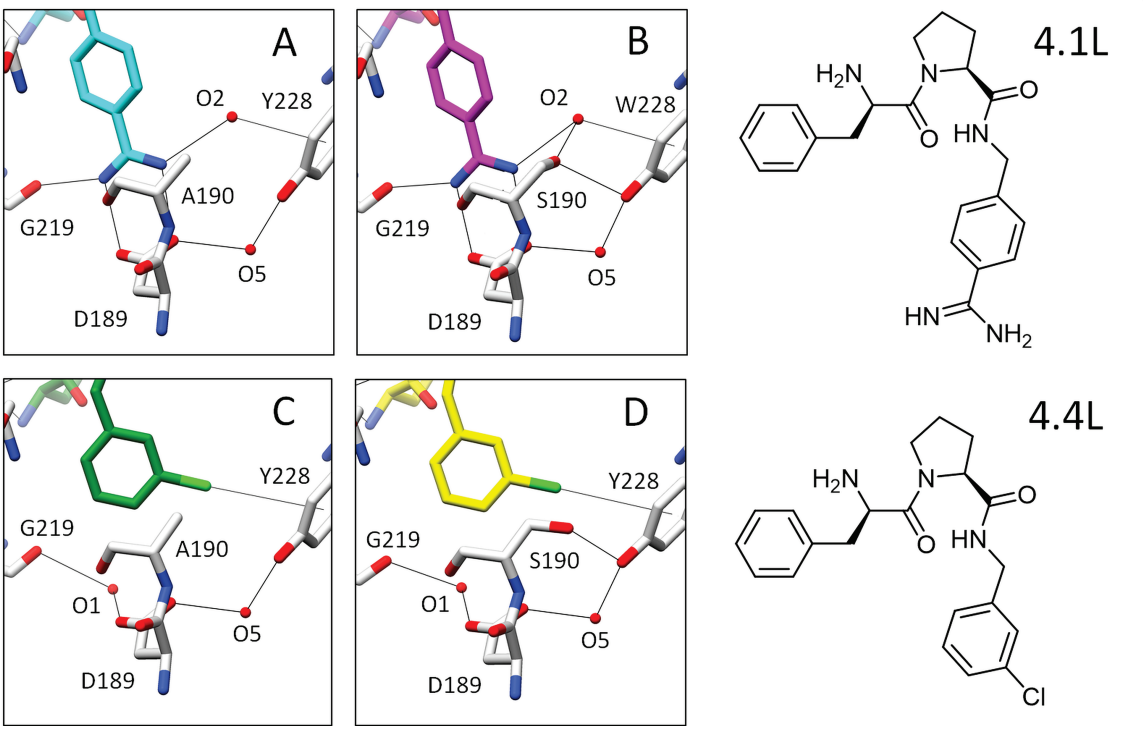


Figure 45: Crystallographically determined binding modes of **4.1L** and **4.4L** with thrombin variants. A) **4.1L** in complex with thrombin wild-type (PDB code: 2ZDA, 1.73 Å[118,125]). B) **4.1L** in complex with mutated thrombin variant A190S (PDB code: 5MM6, 1.29 Å). C) **4.4L** in complex with thrombin wild-type (PDB code: 2ZC9, 1.59 Å[118,125]). D) **4.4L** in complex with mutated thrombin variant A190S (PDB code: 5MLS, 1.62 Å). The structural pairs of the wild-type and mutated variants with the two inhibitors are highly similar. The only apparent difference concerns S190Oγ. Its hydroxyl group stabilized O2 in the complex with **4.1L** through an additional H-bond. In the complex with **4.4L**, the S190 hydroxyl group orients toward Y228 and forms an additional intramolecular H-bond.

To trace possible differences in structural terms, we determined the binding modes of two inhibitors **4.1L** and **4.4L** by crystallography in thrombin and its A190S variant (Figure 45). By

intention, we selected an inhibitor with a basic benzamidino and non-basic *meta*-chloroaromatic benzyl head group for the S_1 pocket. Both inhibitors had already been characterized previously with wild-type thrombin (protein data bank-codes (PDB codes): 2ZDA-**4.1L** and 2ZC9-**4.4L**)[118,125] and showed, according to the enzyme kinetic assay, a negligible affinity difference (**4.1L**-wild-type: 0.55 ± 0.04 nM, **4.1L**-A190S: 0.55 ± 0.06 nM, **4.4L**-wild-type: 590 ± 10 nM, and **4.4L**-A190S: 320 ± 50 nM).

Crystallographically determined binding modes of **4.1L** and **4.4L** with thrombin variants. A) **4.1L** in complex with thrombin wild-type (PDB code: 2ZDA, 1.73 Å[118,125]). B) **4.1L** in complex with mutated thrombin variant A190S (PDB code: 5MM6, 1.29 Å). C) **4.4L** in complex with thrombin wild-type (PDB code: 2ZC9, 1.59 Å[118,125]). D) **4.4L** in complex with mutated thrombin variant A190S (PDB code: 5MLS, 1.62 Å).

All structures are highly similar. In the complex of the mutated variant with **4.1L**, comprising the basic head group, the water molecule O2 connected to the inhibitor experiences an additional hydrogen bond to the side-chain hydroxyl group of S190 (Figure 45). This contact is lacking in the wild type, which possesses instead an alanine at this position. Obviously, the A190S replacement has no effects on the binding mode. We have observed similar differences in other structurally highly related complexes of wild-type thrombin and trypsin[148] and in the binding mode difference of **4.3L**, with both enzymes (Figure 43), which makes us confident that the mutated variant reflects correctly the hardly given differences between both proteases. This is at least valid for the affinity data, the enthalpy/entropy partitioning might be different as suggested by the **4.3L** example above. We refrained from collecting ITC data for the mutated variants as ITC requires large amounts of protein, which are difficult to produce with the recombinant technique. Furthermore, since production in *Escherichia coli* (*E. coli*) does not provide the enzyme in a glycosylated form, the overall solubility of the recombinantly produced thrombin is significantly lower, an aspect that further complicates ITC experiments. In the complex with **4.4L**, which repels O2 for steric reasons from the binding site, S190Oγ in the A190S variant orients away from the hydrophobic P_1 inhibitor portion and forms an H-bond to Y228OH (2.9 Å).

As hardly any differences in the affinity and binding pose are recognized that could be linked to the A190S replacement, we wanted to confirm this observation by selecting a larger set of 54 diverse inhibitors (Figure 70, Supporting Information). The entries of this data set showed distinct binding affinities toward the wild types of thrombin and trypsin Figure 46A, R^2 = 0.51). We therefore tested them in our kinetic enzyme assay for inhibitory potency differences between the wild-type and A190S variant of thrombin. A nearly perfect linear correlation was found (R^2 = 0.98), indicating that no significant difference between both thrombin forms exists (Figure 46B). This assumption is further supported by the fact that the regression line exhibits virtually the same slope as the main diagonal drawn in the diagram (Figure 46B).

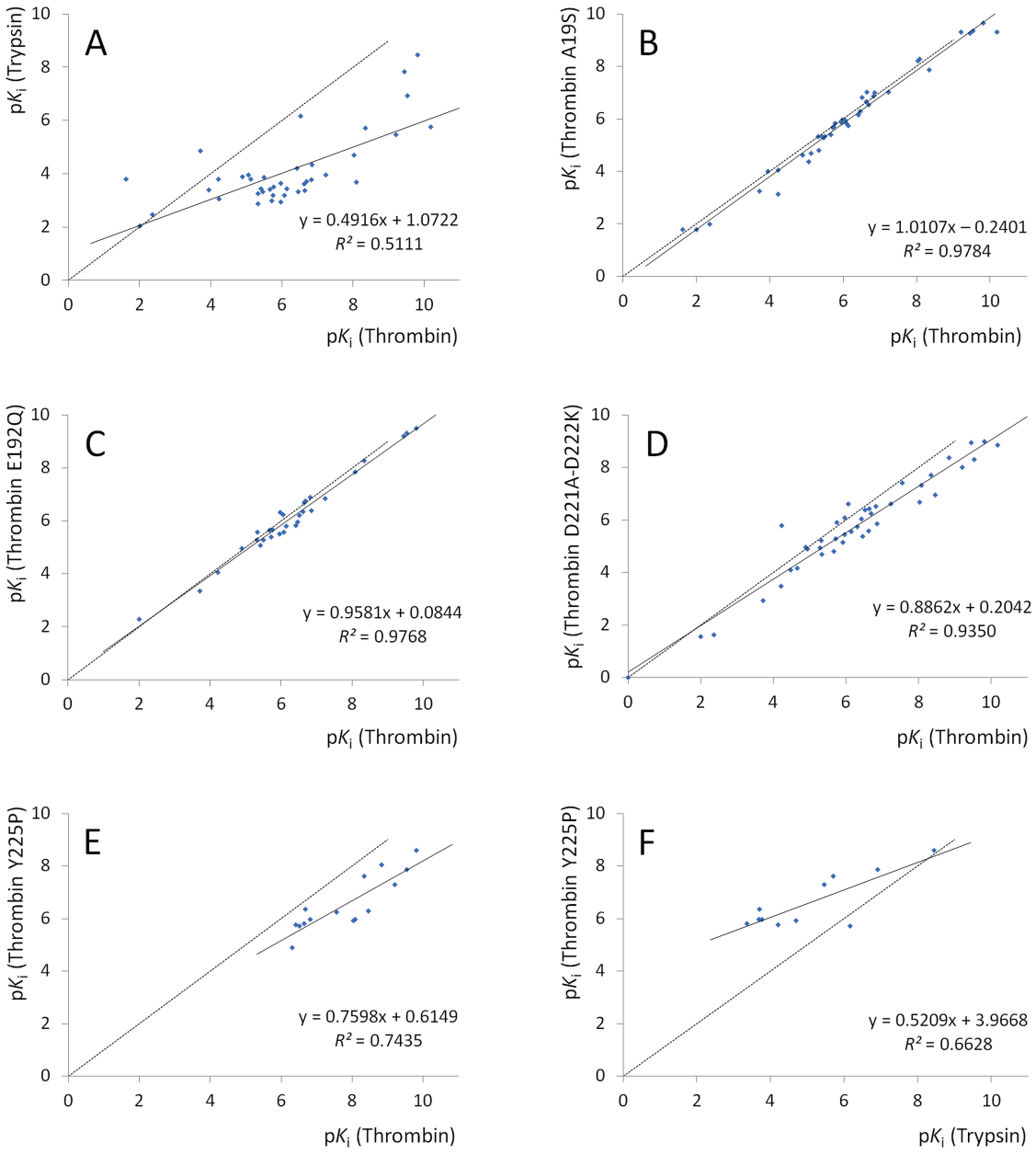


Figure 46: Correlation of assay data for wildtype trypsin and thrombin and four mutated variants of thrombin. The equation for regression line and mean *R* values are listed. A) trypsin vs. thrombin, B) thrombin vs. A190S thrombin, C) thrombin vs. E192Q thrombin, D) thrombin vs. D221A – D222K thrombin, E) thrombin vs. Y225P thrombin, F) trypsin vs. Y225P thrombin, A) – D) 54 data points, E) and F): 15 data points.

4.4.6 Affinity and Structural Data for the Mutated Variant E192Q in Thrombin

A further difference between thrombin and trypsin relates to residue E192, located at the rim of the S_1 pocket. This residue is replaced by an uncharged Q192 in trypsin. E192 shows substantial conformational flexibility, and in the many studied thrombin–inhibitor complexes,[128] it is frequently involved in the binding of particularly polar P_1 head groups.[148] The fragment complex of **4.3F** with thrombin has impressively demonstrated this surprising behavior of E192 resulting in a back-folded conformation of this residue. On first glance, the latter complex suggests that, as the fragment prefers the carboxylate group of the less acidic E192 over that of D189, the concentration of the negative charge on D189 is attenuated in thrombin. This could result from the nearby sodium ion (5.3 Å, PDB code: 6TDT[148]). Obviously, an electrostatic gradient is formed across the S1 pocket spanning from E192 to the Na^+ ion, which might assist in relocating the charges. A similar gradient is not given at the trypsin site. To validate whether this difference in charge distribution has any impact on binding and possibly on selectivity, we generated the mutated E192Q variant of thrombin, which removes the negative charge at the rim of the pocket.

The E192Q variant was tested by the kinetic enzyme assay across the abovementioned reference set of diverse protease inhibitors (Figure 46C). Similar to the A190S exchange, the replacement of charged E192 by uncharged glutamine in thrombin has hardly any impact on the affinity data across the 54 tested compounds (R^2 = 0. 98). A minor difference is perhaps indicated with respect to the A190S variant as the slopes of the regression line deviates slightly from the diagonal.

A crystal structure of the E192Q variant could be determined with D-Phe-Pro-NH-CH_2-4-(2-amino)pyridine (**4.5L**), an inhibitor that was already characterized crystallographically with the wild type of trypsin and thrombin in a previous study (Figure 47B).[148]

Interestingly, all three complexes with **4.5L** are highly similar. However, one important difference is visible. Whereas the binding mode of the inhibitor is ordered in trypsin, in wild-type thrombin, and in the E192Q variant, the inhibitor scatters over two different orientations in the S_1 pocket. While the population of the two orientations in the variant E192Q refines to (I) 47% and (II) 53%, the population of the two orientations in wild-type thrombin converged to (I) 76% and (II) 24%. In our previous study, we could show that the scatter over two orientations has a slight impact on the partitioning of the thermodynamic properties. Disorder over multiple states always increases the entropic contribution on binding and reduces the enthalpic portion. Quite surprisingly, the charge on residue 192 has only minor determining influence in thrombin on the inhibitor's binding mode as we observe the same geometry in the wild-type and mutated variant of E192Q and only the ratio of the two orientations is altered. In trypsin, where only one orientation of **4.5L** is observed, Q192 is directed away from the S_1 pocket. In addition, as seen with the S190A variants, the hydroxyl group of the serine residue

fixes O2, which mediates the contact to the inhibitor. Supposedly, these structural changes have minor influence on the affinity distribution across the series of studied inhibitors (Figure 47C). Nevertheless, the impact of the structural changes is likely more pronounced considering the partitioning of the free energy of binding into its enthalpic and entropic contribution. This is indicated by the values found in Table 6 for inhibitors **4.1L – 4.3L**.

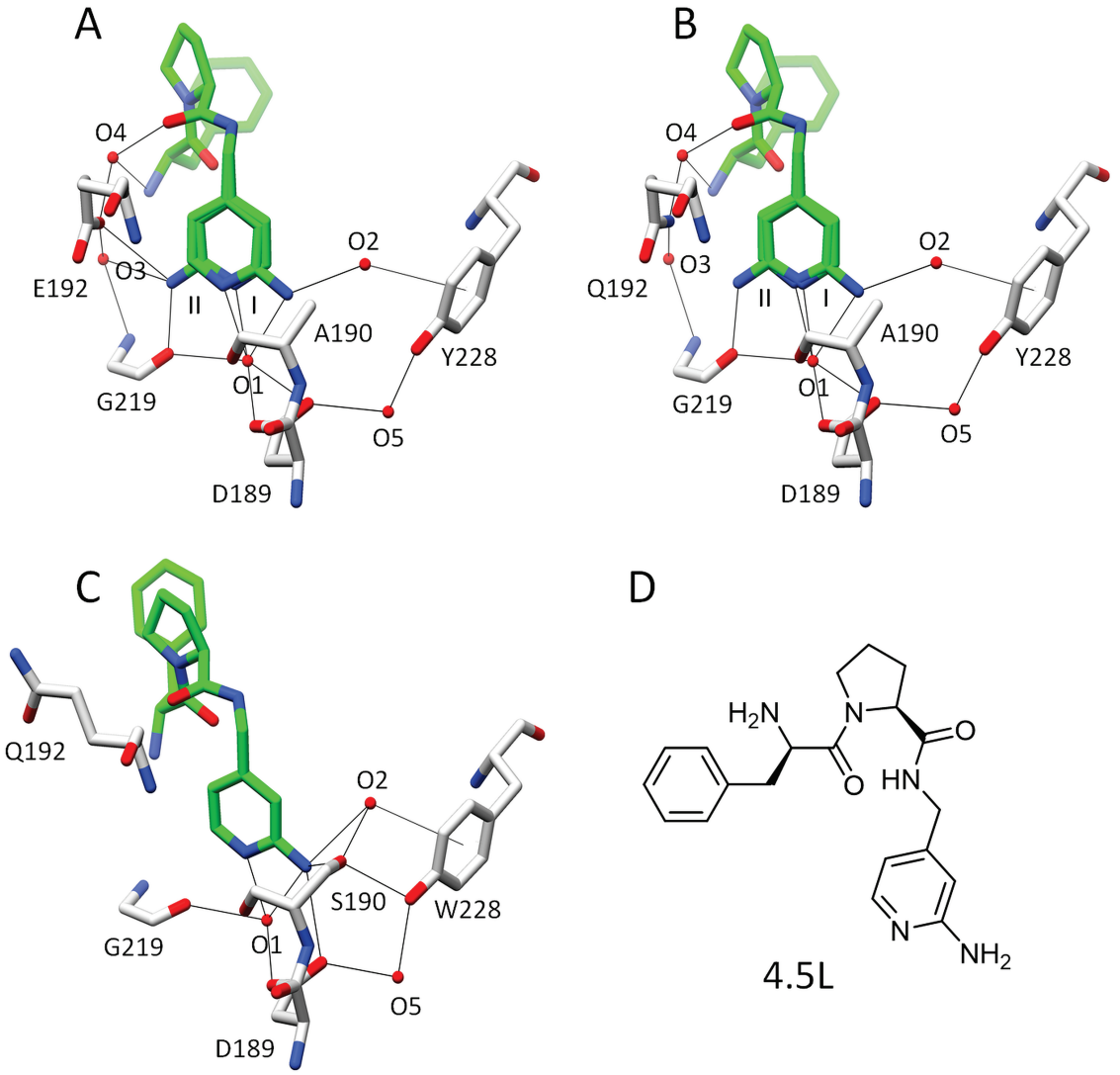


Figure 47: Crystal structures of **4.5L**. A) **4.5L** (light green) in complex with thrombin (PDB code: 6T0P[148]). B) **4.5L** in complex with the E192Q thrombin variant (not deposited as virtually identical with the wild-type structure). C) **4.5L** in complex with trypsin (PDB code: 6T3Q[148]). The binding mode in the S_1 pocket next to D189 is shown. The P_1 residue in both thrombin complexes adopts two orientations. Selected protein residues are depicted as gray sticks and water molecules as red spheres. D) Structural formula of the inhibitor **4.5L**.

4.4.7 Affinity Data for the Mutated Variants D221A – D222K and Y225P in Thrombin

We speculated already above that the sodium ion, exclusively present in thrombin only, has an influence on the charge distribution of D189. We therefore extended our mutational studies on the sodium-binding site in thrombin. However, modifying this site is more challenging as four water molecules and two backbone carbonyl oxygen atoms establish an octahedral coordination sphere around this ion. This geometry is difficult to remove or replace without introducing major changes of the protein architecture. Figure 48A shows a section of the thrombin structure hosting the sodium ion-binding site. In Figure 48B, the related area in trypsin is depicted.

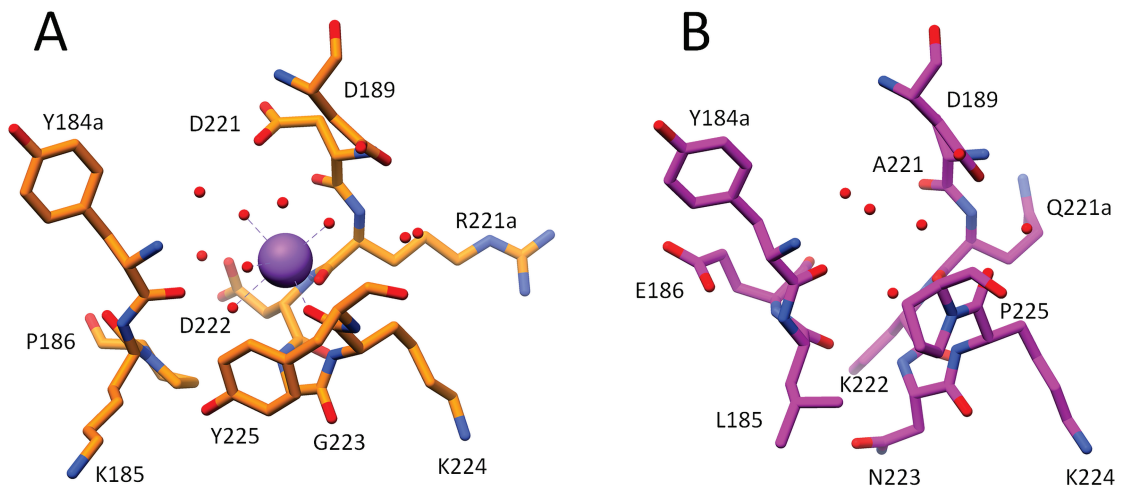


Figure 48: Crystallographic structure of thrombin and trypsin. A) Close-up view of the sodium binding site in thrombin (PDB code: 2UUF[152]) next to the S_1 pocket and D189. B) Corresponding region in trypsin (PDB code: 5MN1[154]), which does not host a similar ion-binding site.

Adjacent to the sodium-ion site, D221 and D222 are found, which likely stabilize the binding of Na^+ in this area. Trypsin, which lacks the sodium-ion site in this region, holds at the corresponding positions alanine and lysine residues. We therefore planned the double mutant D221A – D222K as a promising variant to alter the sodium-ion binding site.[160] The larger and positively charged lysine residue was assumed to interfere unfavorably with the accommodation of Na^+ for steric and electrostatic reasons. As an alternative, we expected that the replacement of tyrosine in thrombin by proline as the corresponding residue in trypsin at position 225 would also make thrombin more like the digestive protease and destabilize the adjacent sodium-binding site.

Both mutated variants of thrombin were generated and tested by our kinetic enzyme assay across the abovementioned reference set of diverse protease inhibitors (Figure 46D – F). The D221A – D222K variant has, at least with respect to the enzyme kinetic analysis, no strong impact on the binding data as suggested by an R^2 = 0.94. However, the slope of the regression

line deviates here more strongly from the main diagonal. This variant has already been described by PINEDA *et al.*[160] and a crystal structure could be determined (PDB code: 1TWX[160]). Overall, the geometrical variations seem to be minor (root-mean-square deviation (rmsd) = 0.5 Å). Small translocations of carbonyl groups coordinating the sodium ion in the wild type and the lack of a density peak at the former sodium-binding site suggest the absence of the ion. However, it remains to be considered that the diffraction power of a sodium ion is nearly identical to that of a water molecule and at a resolution of 2.4 Å, the assignment of water becomes problematic. Nevertheless, it might well be in agreement with our assay results that the D221A − D222K variant is still rather close to the structural properties of wild-type thrombin.

Finally, the kinetic data of the Y225P variant indicate stronger deviations from the wild-type assay results. This is indicated by an R^2 = 0.74 with respect to wild-type thrombin or 0.66 to wild-type trypsin, respectively. Here, the slope of the regression lines deviate significantly from the main diagonal. Due to the strongly reduced activity of the Y225P variant, a large amount of protein material was required to perform the measurements. Since the expression of this variant provided only a very low yield, we had to limit our testing of the inhibitory power to a set of a few selected inhibitors. Unfortunately, we also did not succeed to crystallize the Y225P variant to determine its structure. Nevertheless, if we assume that the mutated Y225P variant alters the stability and in consequence the occupancy of the sodium-binding site in thrombin more strongly than in the other studied variants, the assay data support our hypothesis of the sodium-ion influence. The presence of the positively charged sodium ion seems to have a deviating impact on the binding features of both proteases, supposedly via a significant attenuation of the negative charge concentrated on the carboxylate group of D189 in thrombin. Functionally, the occupancy or absence of a sodium ion in thrombin has been discussed[161] to explain why a kinetically slow and fast form of the enzyme exists. The present comparative study of trypsin and thrombin, investigated in structural and thermodynamic terms under equilibrium conditions, suggests affinity differences but does not allow conclusions on differences in the enzyme kinetic properties.

4.5 Conclusions

In fact, the selectivity-determining features in the S_1 pockets of trypsin and thrombin are difficult to impose to one single dominating factor, such as a single exchange of an amino acid in the S_1 pocket. Hence, selectivity results as a complex interplay of several aspects. The local geometry of both pockets is highly conserved and the sole A190S replacement may be regarded as perfectly conserved, as indicated by the crystal structures determined in parallel for thrombin, trypsin or the A190S and E192Q variants of thrombin. Thus, simple steric

features cannot explain the selectivity difference. The E192Q replacement at the rim of the S_1 pocket has little impact with regard to steric and dynamic properties, both residues require similar space and multiple crystal structures prove the flexibility of the residue in position 192. The affinity data demonstrate the observed high similarity, remarkably a deviating partitioning in enthalpy and entropy contributions points more strongly to a given difference. Likely, this results from the stronger scattering of residue 192 in thrombin compared to trypsin, which also takes impact on the adopted binding pose of the bound inhibitor. In terms of charge distribution, E192, together with the thrombin-specific sodium ion-binding site adjacent to D189, helps in creating an electrostatic gradient across the S_1 pocket, a feature definitely differing in trypsin. The observed induced protonation effects, reported in our previous paper for the basic pyridine head groups (pK_a = 5.0), along with the deviating binding poses of **4.3F** in both proteases, were a first indication for a significant charge attenuation on the carboxylate group of D189 in thrombin compared to trypsin.[135,148] We advocate that this attribute is one of the prime selectivity-determining features between both proteases. In consequence, it influences and controls the other, on a first glance, less obvious selectivity-discriminating features, such as differences in the solvation pattern and ordering of water molecules in both enzymes. The difference in the partitioning of enthalpy and entropy for the trypsin and thrombin complexes with **4.3L** further underscores the importance of the solvation pattern for the properties of the formed complexes (Table 6). Unfortunately, our current understanding of the influence of water solvation features on inhibitor and substrate binding, particularly with respect to solvation binding kinetics, is quite rudimentary. Nevertheless, even though both proteases comprise very similar recognition pockets, their properties strongly differ in the internal architectures and charge distribution, enabling the formation of deviating water inventories important for inhibitor binding and unbinding. Already, in the uncomplexed proteins, the carboxylate group of D189 in thrombin (PDB code: 2UUF) is solvated by a network of three water molecules, whereas in trypsin, the same group is solvated by only two water molecules. Particularly, the neutron diffraction study of *apo* trypsin demonstrates that the immediate solvation pattern of the carboxylate group of D189 is far from ideal.[135] The imperfect and highly perturbed geometry of the first solvation shell around D189 is an important prerequisite for a potent substrate and, in consequence, inhibitor binding. Otherwise, the desolvation of the charged carboxylate in the deep S_1 pocket would be energetically very costly. Furthermore, in trypsin, a water reservoir is found below D189, which is the source of water molecules needed for the association and dissociation process of inhibitors. Remarkably in thrombin, a water channel, which facilities water exchange with a bulk water phase, replaces the water reservoir in trypsin. Elaborate MD simulations have shown that this difference has a decisive influence on the solvation kinetics and, in consequence, on the selectivity of inhibitors binding to both enzymes.[151] As a result, the dissociation of inhibitors from trypsin affords a larger barrier because they must dissociate before the site becomes rehydrated, while in thrombin, re-hydration and inhibitor dissociation proceed simultaneously.

Therefore, the inhibitor-binding mechanism is not only determined by the established protein-inhibitor interactions but also by the differing solvation barriers of both proteins, which adds to the deviating desolvation properties of the different inhibitor molecules, yet another factor in selectivity discrimination.

The series of inhibitors studied with trypsin by our neutron diffraction investigations unravel unexpected differences in the residual dynamics of water molecules in the S_1 pocket. Surprisingly, the rotational and translational behaviors of the water molecules differ not only between uncomplexed and complexed states but also between the complexes hosting different inhibitors. In the trypsin–benzylamine complex studied here, water O1 alters its dynamic properties from a disordered to an ordered state (Figure 39A, B). Water molecule O2 is also disordered in *apo* trypsin and becomes ordered in the complex with **4.3F** by experiencing a weak H-bond with S190Oγ. This residue is replaced by alanine in the S_1 pocket of thrombin and suggests on a first glance that the lacking Oγ and thus the loss of a hydrogen bond have consequences for selectivity. The thermodynamic data suggest no difference in Δ*G* but in the enthalpy/entropy partitioning. Our panel of 54 tested inhibitors also speaks against an impact on affinity; however, the partitioning of enthalpy and entropy cannot be excluded and remains to be shown for a larger number of cases. As our crystallographic analysis shows, the A190S thrombin variant is structurally closer to trypsin at position 190, but with respect to the remaining solvation features of the pocket, it is still a chimera closer to thrombin than trypsin. This makes the interpretation of mutational differences rather inscrutable. Unfortunately, the X-ray structures collected with thrombin do not disclose the required details about the dynamics of the water molecules in the S_1 pocket. Accordingly, more experimental data in terms of neutron diffraction studies complemented by MD simulations are required to further trace the water influence of selectivity between these closely related serine proteases.

Chapter 5

Thrombin Inhibitors Interacting with π-System of Y228

5.1 Crystal Structures of Human Thrombin Inhibitor Complexes

In the following, nine crystal structures of human Thrombin complexes with **5.1** – **5.9** (Figure 49) are presented and analyzed. The crystallographic data collections of each complex range in resolution between 1.33 and 1.79 Å. All structures, except for the Thr • **5.1** and Thr • **5.6** complex, were deposited in the PDB. All inhibitors analyzed in this part of the thesis bind in the S-shaped conformation in the binding pocket of thrombin, in which the P_3 residue is oriented toward the $S_{3/4}$ pocket (Figure 50 – Figure 52). This conformation has already been described in more detail in chapter 3.5. In all complex structures, the P_2 proline of the inhibitor is located in the S_2 pocket, covered by Y60a and W60d of the previously described 60's-loop (Chapter 1.2.2.4). The solvent exposed $S_{3/4}$ pocket above W215 has hydrophobic character and accommodates the phenyl moiety of the P_3 portion of the compounds. In all structures, an H-bond with d = 2.8 – 3.0 Å is formed between the P_1 amido-NH and the S214 carbonyl oxygen, which fixes the inhibitor at the boundary between the S_1 and S_2 pockets. In addition, the structures are similar with respect to the preorganization by intramolecular interactions as already discussed in Chapter 3. The intramolecular water-mediated H-bonding (via O3 with d = 2.7 – 3.0 Å) between the terminal, probably charged P_3 ammonium group. The P_2 backbone carbonyl oxygen forces the inhibitor prior to binding into the conformation that it also adopts in the protein after binding. However, the inhibitors lack the aromatic P_1 group decorated with a hydroxylmethylene or aminomethylene substituent potentially available for intramolecular preorganization, as described in chapter 3.5. Instead, inhibitors **5.1** – **5.9** all have P_1 groups capable of interacting with Y228 either directly or O1-mediated via a halogen-π-interaction. In this way, the inhibitor series differs from other inhibitors that bind instead via D189 in the S_1 pocket of thrombin. An example of this is the benzamidine portion of **3.1**, which establishes a strong salt bridge to the deprotonated D189 at the bottom of the S_1 pocket.

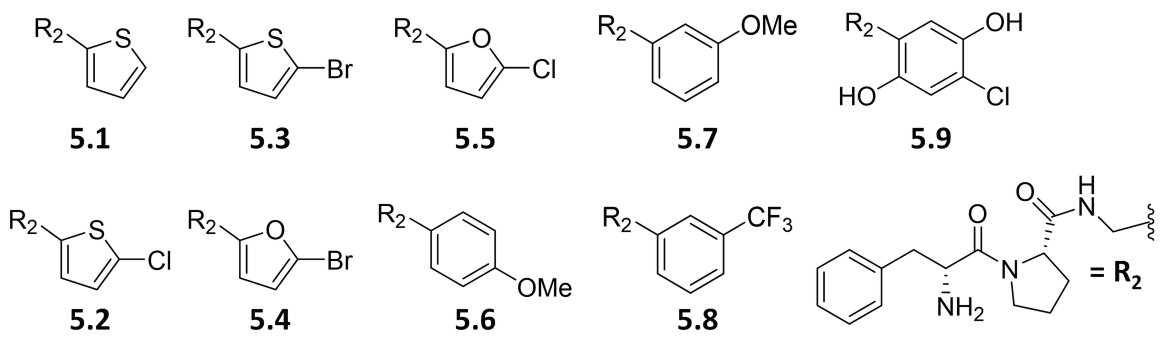


Figure 49: Schematic representation of thrombin inhibitors **5.1** – **5.9** with the parent scaffold R_2.

A closer inspection of the P_1 groups within the series reveals several interesting binding features. Inhibitors **5.1** – **5.3** all contain a P_1 group whose major component is a thiophene, and yet this part of the inhibitor binds in different ways. While the sulfur of **5.3** forms the well-known O1-mediated interaction with the π-System of Y228 (Figure 50C),[162] the sulfur of **5.1** is flipped to the opposite direction and points toward the inhibitors own P_1 and P_3 carbonyl oxygens. In this case, O1 has no direct contact with the inhibitor (Figure 50A). The sulfur atom of the chlorothiophene of **5.2** is also oriented toward Y228, but the water O1 required for a π-interaction with Y228 is not present in this structure (Figure 50B). Although a positive electron density is still evident in the $F_O - F_C$ map at 4 ¢ in the structure of thrombin complex **5.2** at the usual position of O1, it is not significant enough to be assigned in the structure with sufficient confidence. The chlorine is in a reasonable position to interact with one of the oxygen atoms of D189 at d = 2.3 Å. Simultaneously, it forms a bond to O6, which is additionally fixed by D189 and G219. The bromine of **5.3** behaves in a similar way. Again, an interaction is formed to one of the oxygen atoms of D189. However, unlike the complex Thr • **5.2**, O6 is too far away to interact with the bromine; instead, the halogen atom can interact with O1 (d = 3.3 Å) and consequently water-mediated with the π-system of Y228. Remarkably, in the case of **5.1**, the side chain of E192, whose flexibility is already known from previous studies, could not be assigned satisfactorily to the $F_O - F_C$ difference electron density.[125,128] In contrast, once the inhibitor contains a chlorothiophene or bromothiophene in the P_1 position, the difference electron density of that amino acid is well defined. Whereas in the case of **5.2** E192 points to a position away from the inhibitor, in the complex with **5.3** E192 forms an interaction with O3, which is significantly involved in the preorganization of the inhibitor (d = 2.7 Å). A striking feature of the complex Thr • **5.1** was the absence of a visible density of the sodium atom at the usual sodium bonding site. This phenomenon could be a highly interesting discovery, or it could be due to poor crystal quality or problems during data acquisition. In either case, it requires further investigation, which was not done within the scope of this work. For this reason, **5.1** has not been deposited in the PDB.

When the P_1 thiophene scaffold is replaced by a furan, it is remarkable that the halogen, i.e., bromine in **5.4** (Figure 50D) and chlorine in **5.5** (Figure 51A), is in direct π-π-contact with Y228. O1 is not present in these structures and its position is occupied by the halogen atom instead. Additional fixation of the P_1 component via water molecule O6 is no longer possible in either case. While O6 is still present in complex with **5.5**, it can no longer be localized in complex with **5.4**. The side chain of E192 could only be completely assigned in the structure with **5.4**, where, orientated away from the inhibitor, it is not involved in any other strong interactions, similar as in the case of the complex Thr • **5.2**. In contrast, the side chain of E192 in Thr • **5.5** appears to be too flexible to be accurately localized by difference electron density. Comparatively poor resolution of the crystal structure of **5.5** (1.73 Å) also complicates interpretation of the position of this highly flexible amino acid.

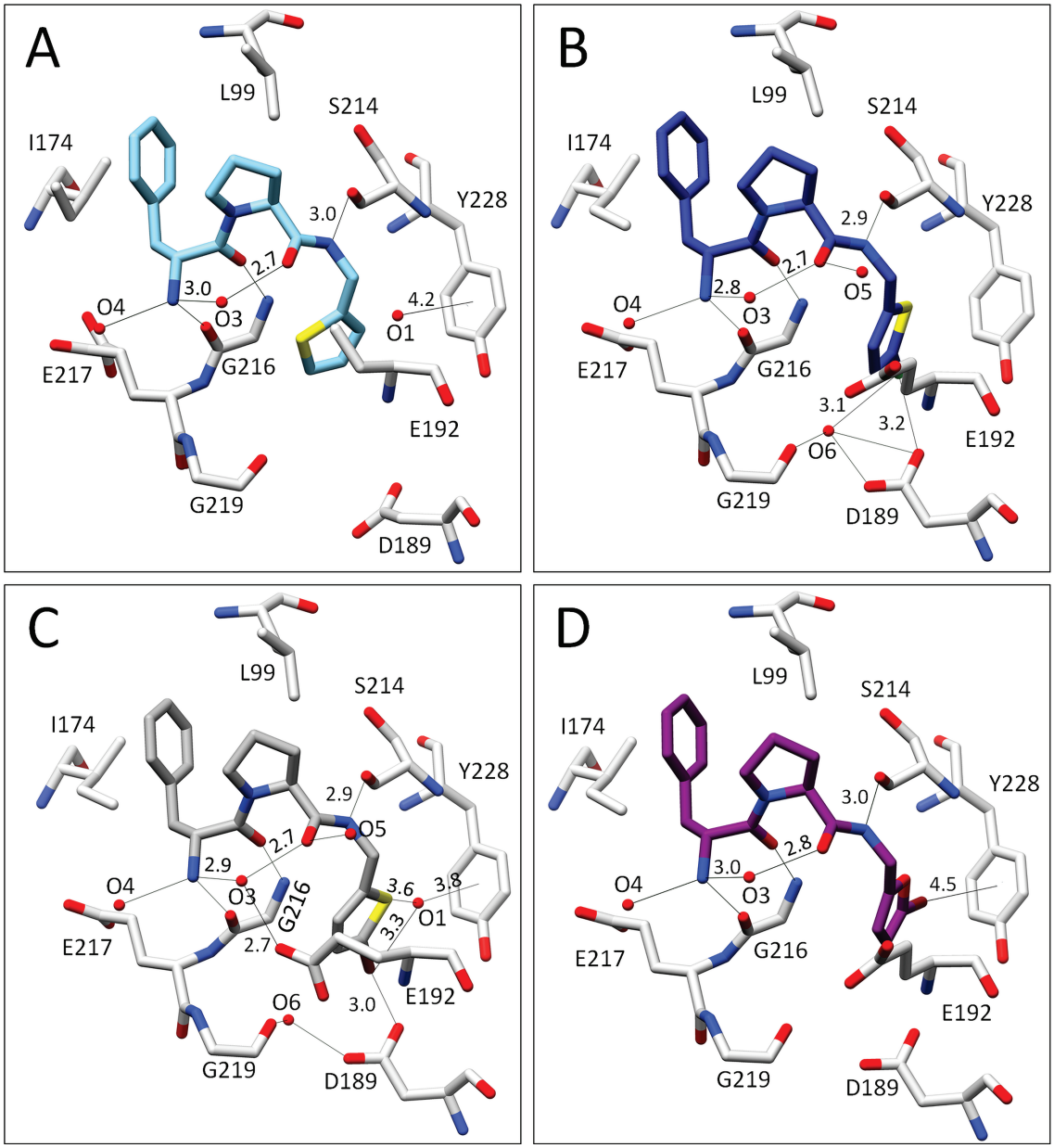


Figure 50: Inhibitors **5.1** – **5.4** bound to the active site of thrombin. Inhibitor **5.1** (not deposited) [(A) light blue], **5.2** (6YHJ) [(B) blue], **5.3** (6T54) [(C) grey] and **5.4** (6Y02) [(D) purple]. Water molecules are shown as red spheres within a radius of 5 Å around the inhibitor. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow, bromine atoms in brown and chlorine atoms in green. The interactions are indicated as black lines.

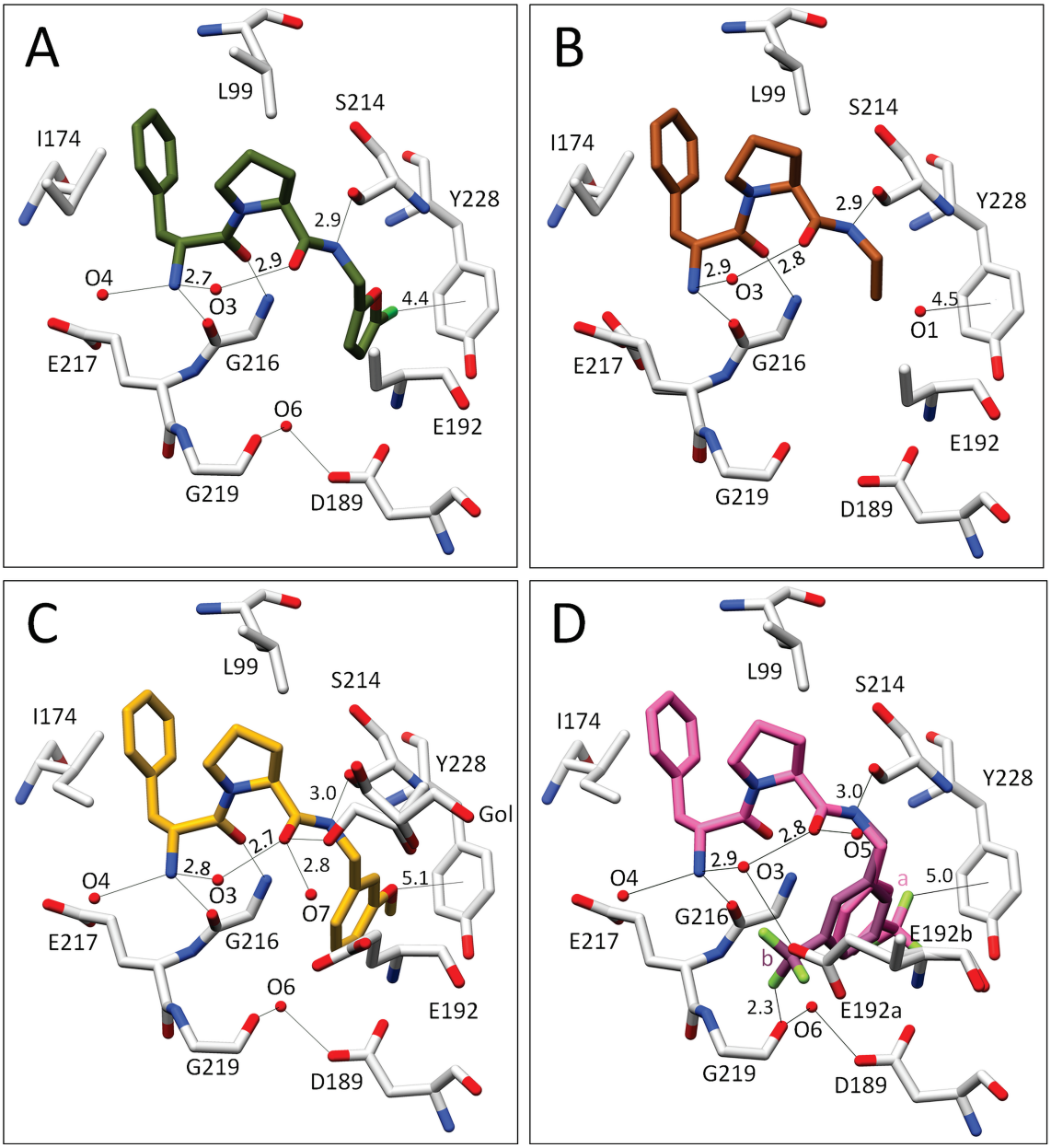


Figure 51: Inhibitors **5.5** – **5.8** bound to the active site of thrombin. **5.5** (6ZGO) [(A) dark green], **5.6** (not deposited) [(B) brown], **5.7** (6YHG) [(C) gold] and **5.8** (6YH9) [(D) pink] **5.8** is shown in conformation a with 56% in pink and b with 44% in a slightly darker color. Water molecules are shown as red spheres within a radius of 5 Å around the inhibitor. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, chlorine atoms in green and fluorine atoms in light green. The interactions are indicated as black lines.

Neither is it possible to determine the position of E192 in this case. Nevertheless, it can be argued that O1 finds its position in the structure and interacts with Y228 in the usual way (d = 4.5 Å). However, it remains unclear whether there is an interaction with the P_1 inhibitor moiety. When the methoxy substituent is in *meta*-position at the P_1 six-membered ring, as in **5.7**, the position of this inhibitor moiety is better fixed, thus the electron density allows its

assignment to the crystal structure (Figure 51C). The methoxy group is definitely oriented toward the aromatic ring of Y228 and performs a π-π-interaction with a rather large distance of d = 5.1 Å. It therefore does not have the possibility to interact with O6. E192 points to a position away from the inhibitor and consequently is not able to participate in its fixation in the protein through interactions.

Another peculiarity in this series is **5.8**. The trifluoromethyl-phenyl substituent in *meta*-position adopts two conformations upon binding to the protein (Figure 51D). With an occupancy of 56%, conformation a points toward Y228, with one of the fluorine atoms interacting with the π-system of the tyrosine (d = 5.0 Å). In contrast, conformation b with an occupancy of 44% binds rotated by 180° and points with the trifluoromethyl in direction of G219. One of the fluorine atoms is located at a distance of d = 2.3 Å toward the carbonyl oxygen of G219. Similar to the Thr • **5.3** complex, E192a (occ.: 56%) can also interact with O3 (d = 2.8 Å). However, if the P_1 substituent of the inhibitor is in conformation b, there is no longer sufficient space for E192a, forcing this highly flexible amino acid to switch to conformation b as well (occ.: 44%). In this case, the side chain points away from the inhibitor but does not adopt a fixed position. For this reason, the side chain of E192a could not be assigned to the final structure. O6 also finds its position in this structure but cannot establish an interaction to the inhibitor.

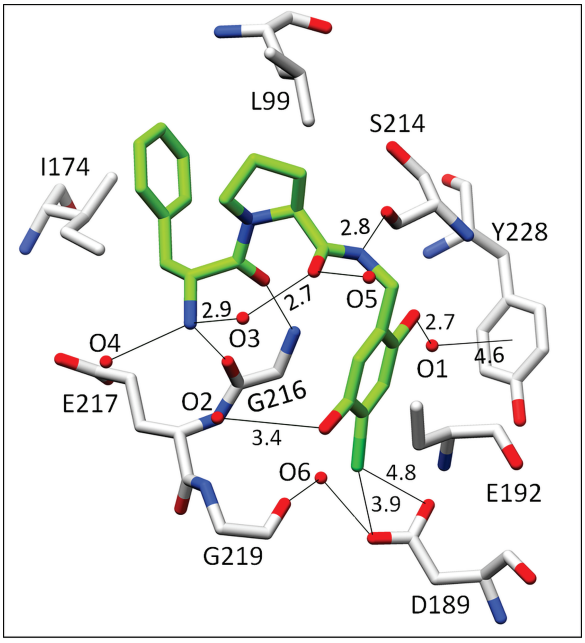


Figure 52: Inhibitor **5.9** bound to the active site of thrombin. Inhibitor **5.9** (6YB6) is shown in green. Water molecules are shown as red spheres within a radius of 5 Å around the inhibitor. Selected residues are displayed for better orientation. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow and chlorine atoms in green. The interactions are indicated as black lines.

Also, the P_1 part of **5.9** is sterically very demanding. However, it binds only in one position in the S_1 pocket of thrombin (Figure 52). The hydroxy group in the *ortho*-position is oriented in a way to form an O1-mediated *π*-*π*-interaction with Y228. The second hydroxy substituent of the P_1 phenyl ring of **5.9** is located in the *meta*-position, being the only one of the inhibitor series that can form an elongated H-bond to the sulfur atom of C220 (d = 3.8 Å). In addition, O2, which is not present in the other crystal structures of the series, is fixed via this hydroxyl group (d = 3.4 Å). The *para*-positioned chlorine binds toward the side chain of D189 (d = 3.9 Å, d = 4.8 Å). However, O6 is not able to interact with the inhibitor. In the position opposite of G219, E192 would not find enough space for the usual orientation in this complex. As described already for other cases, it turns to the outside of the active site and remains in such flexible state that an unique localization of the side chain is not possible.

5.2 Discussion

For the following analysis and discussion of the binding behavior of the inhibitors shown in Figure 49, only the X-ray crystal structures described in Chapter 5.1 are used. Other methods for obtaining results, such as enzyme kinetic assays or ITC measurements, were not considered.

Considering inhibitors **5.1 – 5.3**, it becomes obvious that one and the same P_1 thiophene scaffold can bind in two different positions. Although the inhibitors differ only in the presence or absence of a halogen substituent, respectively, their binding behavior is clearly different. While the sulfur of the thiophene of **5.1** faces the inhibitor's own carbonyl groups P_2 and P_3, the same atom points toward Y228 in case of the halogen-substituted inhibitors **5.2** and **5.3**. For the latter two inhibitors, the position of the P_1 moieties is significantly influenced by the halogen atoms. The chlorine of **5.2** forms an interaction to one of the oxygen atoms of D189. Additionally, it interacts with G219 via an O6-mediated contact. Bromine from compound **5.3** is also able to form an interaction with one of the oxygen atoms of D189. In this case, however, the halogen is too far away from O6 to additionally stabilize the P_1 part of **5.3** via G219. Instead, however, it interacts with O1 (d = 3.3 Å). Comparing both P_1 moieties, the bromothiophene moiety binds slightly twisted in the S_1 pocket of thrombin compared to chlorothiophene one (Figure 53), which increases the distance to O6. Furthermore, it is remarkable that the sulfur of **5.3** can perform the well-known O1-mediated interaction with the *π*-system of Y228. However, a comparable interaction is not observed for **5.2**. Here, it was not possible to detect O1 based on the $F_O - F_C$ difference electron density. Calculating the distance that the sulfur atom of the chlorothiophene would have to the anticipated position of the putatively present O1 (d = 3.3 Å), a reasonable contact appears possible. However, the interaction of

the chlorine with O1 seems to be rather destabilized due to the short distance of d = 2.5 Å. This is probably the reason why O1 does not bind in the case of **5.2**.

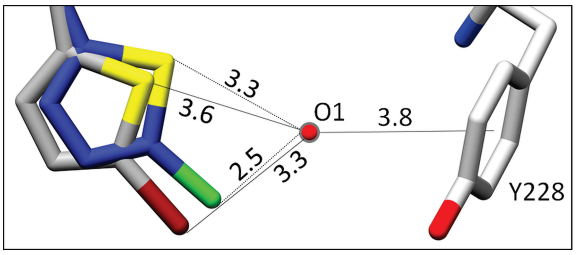


Figure 53: Comparison of the positions of the P_1 inhibitor moieties in the S_1 pocket of thrombin. **5.2** (6YHJ) is shown in blue, **5.3** (6T54) is shown in dark grey. Water molecules O1, only present in the crystal structure of the complex Thr • **5.3** is shown as a red sphere. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow, fluorine atoms in green and bromine atoms in brown. Interactions are indicated as black lines; calculated distances are indicated as dotted black lines.

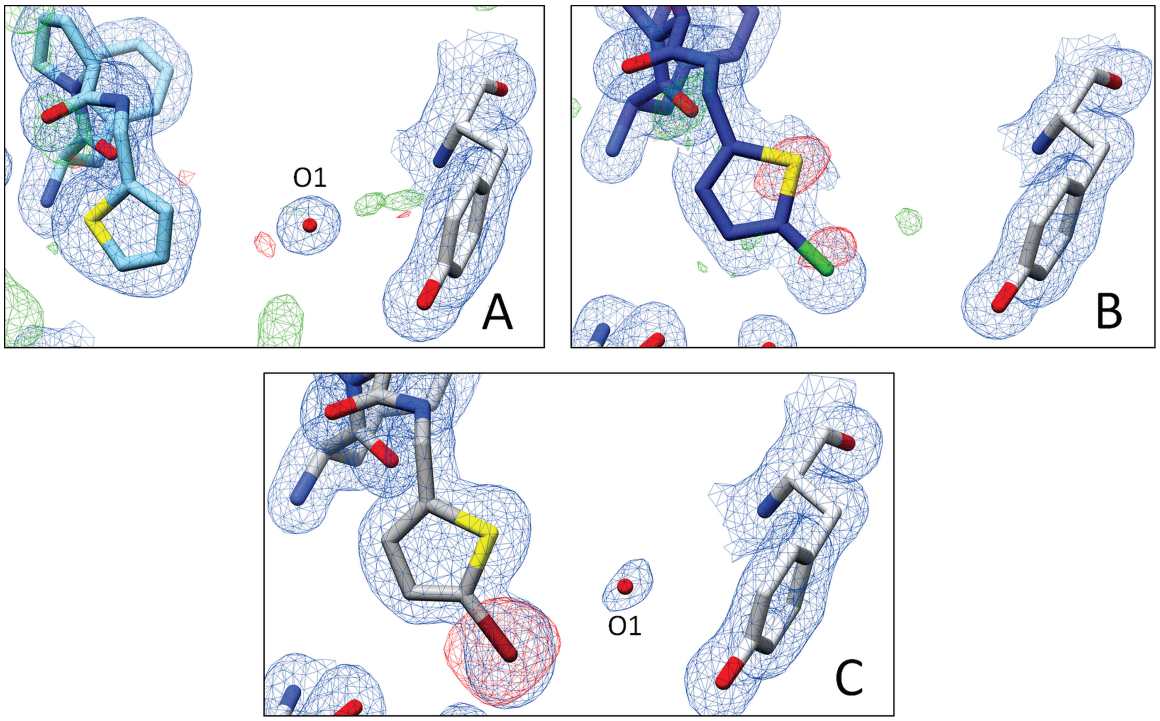


Figure 54: Representation of **5.1** – **5.3** and its electron density. A) Inhibitor **5.1** (not deposited) in light blue, B) Inhibitor **5.2** (6YHJ) in dark blue and C) Inhibitor **5.3** (6T54) in dark grey. Y228 is shown in light grey. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow, fluorine atoms in green and bromine atoms in brown. Electron density difference map (F_O – F_C) is depicted as red, and green meshes at the 3σ level. The 2F_O – F_C density is depicted as blue mesh at the 1σ level.

Why O1 could not be assigned in the structure of **5.2** may have several reasons. At low resolutions of crystal structures, it is quite possible that water molecules are not or only poorly visible in the electron density. However, at a resolution of 1.44 Å, as is the case for the structure of **5.2**, a multiply fixed water at this position should be readily visible. However, the small amount of $F_O - F_C$ difference electron density at the actual position of O1 suggests a low occurrence of this water molecule (Figure 54B). A possible reason for this could be that the chlorine atom instead prefers to interact with O3.

In contrast, the P_1 moiety of **5.1** does not appear to form polar interactions. However, the $F_O - F_C$ difference electron density indicates that the sulfur of the thiophene binds with full occupancy in the direction of the inhibitor's own P_1 and P_3 carbon groups (Figure 54A). The highly electronegative halogen substituents on the thiophene of **5.2** and **5.3** withdraw electrons from the 5-membered ring. Possibly, this reduces electron density on sulfur and might turn it into an acceptor for an H-bond to water molecule O1. Instead, the sulfur of the thiophene ring in **5.1** lacks this electron-withdrawing support. Consequently, it might be less attractive to form an H-bond with O1.

While chlorine and bromine attached to a thiophene form interactions with D189 located at the bottom of the S_1 pocket, the halogen atoms attached to a furan in the same position establish a direct halogen-π-interaction with Y228 (Figure 55). The center of the 5-membered P_1 ring of the inhibitor shifts by 1 Å toward Y228, resulting in a distance d = 4.4 Å of the halogen to the π-system of the tyrosine. However, if thiophene inhibitors **5.2** and **5.3** were placed with their P_1 moiety in the same position as furan inhibitors **5.4** and **5.5**, it is reasonable to assume that the halogen interaction would still be significantly less favorable than for **5.2** and **5.3**. One reason for this could be the altered angle $\alpha_F = 108°$ in furan compared to alpha in thiophene $\alpha_T = 91°$. The geometry of the 5-ring changes in such a way that a halogen-π-interaction becomes possible without major geometric distortions. However, to obtain more precise information about the extent by which the loss of these direct halogen-π-interaction affects the affinity of **5.2** and **5.3**, and whether the polar contact to D189 that occurs instead compensates for this, further thermodynamic and kinetic measurements would need to be performed.

The *para*-methoxy-P_1 moiety of **5.6** could not be assigned to the structure due to a poorly defined difference electron density. In crystallographic analysis, it often occurs that inhibitor structures or side chains of amino acids are difficult to localize. In many cases, this results because of poor resolution of the collected data. However, in the case of the Thr • **5.6** complex, the resolution is 1.5 Å, so the difference electron density can be interpreted accurately, as shown in Figure 56A. Nevertheless, the P_1 part of **5.6** cannot be unambiguously localized. Although a positive difference electron blob is visible at the presumed position of the P_1 part after insertion of the inhibitor, it is not sufficiently defined to determine the exact position of the P_1 moiety of **5.6** (Figure 56B). Thus, it is likely that the *para*-methoxy substituent cannot

form strong bonds to the protein backbone and remains with high residual mobility, allowing this portion to adopt multiple positions in the S_1 pocket of thrombin.

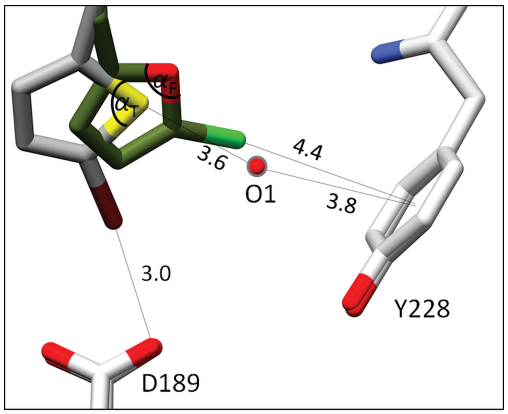


Figure 55: Comparison of the positions of the P_1 inhibitor moieties in the S_1 pocket of thrombin. **5.3** (6T54) is shown in dark grey and **5.5** (6ZGO) is shown in dark green. Water molecule O1, only present in the crystal structure of the complex Thr • **5.3** is shown as a red sphere. Oxygen atoms are displayed in red, nitrogen atoms in blue, sulfur atoms in yellow, chlorine atoms in green and bromine atoms in brown. The angles α_F = 108° and α_T = 91° are indicated in black and interactions are indicated as black lines.

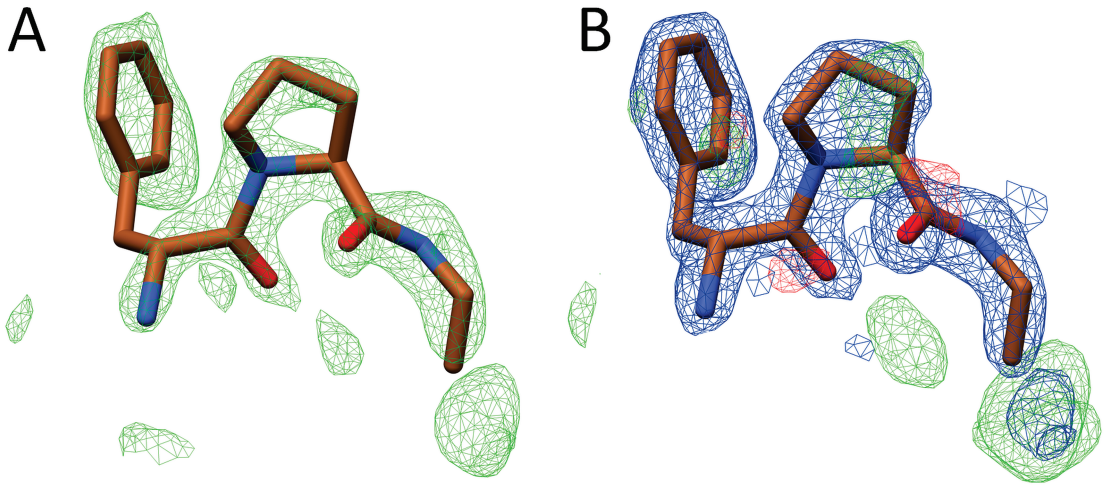


Figure 56: Representation of **5.6** and its electron density in complex with human Thrombin. A) Superposition of the model of **5.6** (not deposited), shown in brown and the difference electron density map ($F_O - F_C$) prior to its insertion into the model. B) Representation of **5.6** and the electron density map ($F_O - F_C$) as well as difference electron density map ($2F_O - F_C$) after its partial insertion into the model. The *para*-methoxy-P_1 part could not be assigned to the structure due to a poorly defined differential electron density. Oxygen atoms are displayed in red and nitrogen atoms in blue. The difference electron density map ($F_O - F_C$) is depicted as red (negative), and green (positive) meshes at the 3σ level. The ($2F_O - F_C$) density is depicted as blue mesh at the 1σ level.

Comparing the positions of the three inhibitors with six-membered rings in the complex with thrombin, it is clear that in the case of **5.7** the ring occupies a slightly shifted position compared to **5.8** and **5.9** (Figure 57). Consequently, the oxygen of the *meta*-methoxy-P_1 substituent can interact with the π-system of Y228 (d = 5.1 Å). The recruitment of the additional water O7, which contacts the inhibitor's own carbonyl group via an H-bond, deprives the highly flexible amino acid E192 of the opportunity to participate in the process. It cannot form an H-bond to O3, found in case of in **5.3** or **5.8**. Consequently, E192 rotates away and does not form any further direct interaction with the inhibitor or the inhibitors directly linked network. Compared to **5.8** and **5.9**, inhibitor **5.7** forms rather few interactions via its P_1 moiety. It can therefore be assumed that this ligand binds enthalpically significantly worse in complex with thrombin than **5.8** and **5.9**. However, **5.7** is likely to have an advantage with respect to the binding process in a protein due to its comparatively less costly desolvation. The CF_3 moiety of **5.8** is highly electronegative and therefore strips off its hydrate shell only with high energy expenditure. The inhibitor **5.9** has several polar groups, including a highly electronegative chlorine atom. Also here, significant larger desolvation costs can be expected, which is likely to be comparatively low for **5.7**. However, the extent to which this affects the affinities of the inhibitors would have to be discussed in more detail in the further course of research by thermodynamic and kinetic measurements.

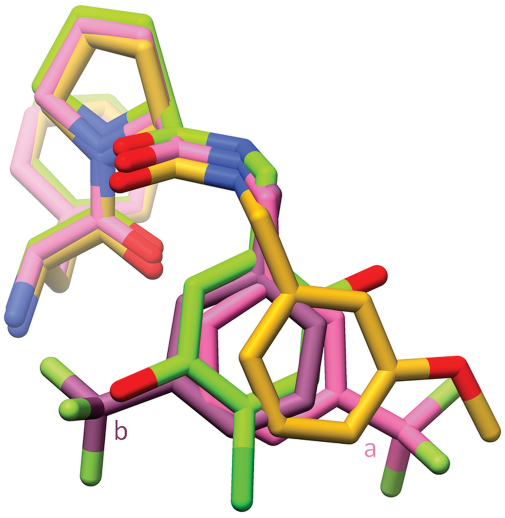


Figure 57: Superposition of ligand structures **5.7** – **5.9** in complex with human thrombin. Inhibitor **5.7** (6YHG) is shown in gold, **5.8** (6YH9) in pink, and **5.9** (6YB6) in green. To distinguish the position of conformations a and b of **2.8**, b is highlighted by a slightly darker color. The oxygen atoms are shown in red, the nitrogen atoms in blue, the chlorine atoms in green, and the fluorine atoms in light green.

With its scatter over two orientations of the P1 moiety in the S_1 pocket, **5.8** represents a special situation. In conformation a, which is with 56% the higher occupied position, the inhibitor can form the halogen-π interaction to Y228 (d = 5.0 Å). Simultaneously, E192 is able to form

an H-bond to O3 (d = 2.8 Å), similar as in the case of **5.3**. In conformation b, however, the P_1 part of the inhibitor is rotated by 180°. In consequence, the halogen-π interaction is no longer possible, and instead the CF_3 part attaches to G219. (d = 2.3 Å). However, as a result, E192 experiences a steric hindrance and has to swerve. It flips away to the outside and remains disordered. Thus, an entropic advantage can be expected for this binding pose. Nevertheless, conformation b is the less preferred conformation. This is possibly due to the loss of the additional H-bond to O3 leading to a likely unfavorable enthalpic contribution. However, since both conformations have a similar population, this suggests that the affinity of both forms is approximately the same and consequently the enthalpic loss of conformation b compared to a is compensated by the entropy gain.

The P_1 portion of **5.9** is the only one of the inhibitors with six-membered rings that is linked to Y228 via O1, which is already known from **5.1** and **5.3**. Even though the inhibitor also has a hydroxy group in the *meta*-position that would be able to interact with the π-system of Y228 without the bypass of O3, the P_1 moiety binds inverted by 180°. The *ortho*-hydroxy group forms an H-bond with O1 which allows the *meta*-hydroxy group to form a prolonged H-bond to O2 (d = 3.4 Å). These two enthalpically favorable bonds are therefore probably the reason why the P_1 moiety binds in this position and Y228 does not face the *meta*-hydroxy group. Compared to **5.7** and **5.8,** O1 and O2 are two additionally water molecules picked up by the complex. Therefore, the entropic contribution to the binding of **5.9** could likely be significantly less favorable compared to the other two complexes. Moreover, the entire P_1 part of the inhibitor is strongly fixed by two additional interactions of the *para*-chlorine with D189. Due to its large space requirement, the P_1 unit of **5.9** does not provide room for interactions with E192. This residue rotates away and remains disordered, similar to the case of **5.8**. It is likely that this provides an entropic advantage for complex formation. In addition, **5.9** forms many enthalpically favorable bonds that supposedly compensate for the entropic deficit that - as mentioned above - presumably results from the two additional water molecules and the strongly fixed P_1 unit.

The fact that only the crystal structures are available for analysis in this project does not allow a clear conclusion about the affinity and selectivity in thrombin. Similar to other chapters in this thesis, it would be desirable to perform additional thermodynamic and kinetic measurements of the binding process. Also, a comparison or estimation of the extent by which inhibitors that contact Y228 rather than D189 affect affinity would require additional data or computer simulations. Due to a lack of time and the late takeover of the data evaluations from a former PhD student in this chapter made these studies in the current thesis impossible.

Chapter 6

Experimental Section

6.1 Site-Directed Mutagenesis of Thrombin Variant E192Q

Mutation of the Thrombin E192Q gene and mutated variants thereof were done as previously described by F. IMMEKUS.[45]

Mutagenesis. The PureYield™ plasmid miniprep system from Promega was used to obtain small amounts of plasmid DNA for the expression of the thrombin variants E192Q. A sample of the DNA obtained via plasmid mini-preparation was sequenced by Eurofins Genomics (Ebersberg, Germany) and translated into a protein sequence by Expasy.[163] The primers were produced by Eurofins Genomics. The complete Thrombin gene was re-sequenced by Eurofins Genomics to confirm both the presence of the desired mutation and the absence of any further, inadvertent mutation. The complete plasmid was amplified by PCR using a thermocycler from MJ Research (St. Bruno, Quebec, Canada). A small remaining amount of matrix DNA was removed by *Dpn I*-digestion to prevent wildtype expression. The mutated free DNA was then transferred into the cloning and expression strains of *E. coli* BL21 (DE3) by heat shock transformation at 315 K. After plating out the centrifuge supernatant on an agar plate with the antibiotic corresponding to the resistance gene, they were incubated overnight at 310 K. A success control by plasmid mini-preparation and subsequent sequencing could be performed using the overnight culture prepared from colonies grown on the agar plate. Clones containing the desired genetic material were stored as glycerol stock at 193 K.

6.2 Expression and Purification of Thrombin Variant E192Q

Overexpression of the thrombin E192Q gene and mutated variants thereof as well as protein purification were done as previously described by F. IMMEKUS.[45] During the entire purification process, the identity and purity of the protein was constantly verified using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The molecular weight of the thrombin mutant is about 35 kDa.

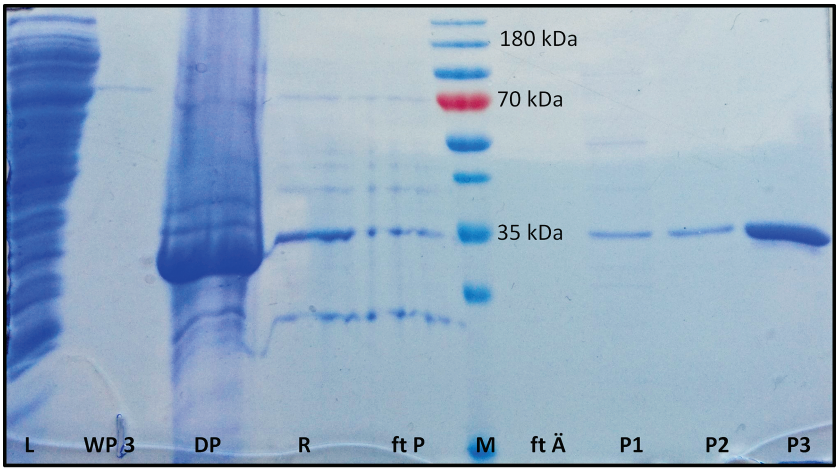


Figure 58: Example of an SDS-PAGE to verify the purity of the expressed and purified thrombin variant E192Q. L: lysate, WP3: washing buffer 3, DP: denatured protein, R: refolded protein, ftP: flow through peristaltic pump, M: marker, ftÄ: flow through Äkta™ FPLC, P1: Peak 1 of the eluate, P2: Peak 2 of the eluate, P3: Peak 3 of the eluate

6.2.1 Expression

First, a preculture of 400 mL LB-medium, 400 µL ampicillin in H_2O (100 mg/mL) and 400 µL chloramphenicol in EtOH (35 mg/mL) was prepared and incubated overnight at 310 K. 50 mL of the preculture were then transferred into 2 L LB-medium containing the same antibiotics. Subsequently, the main culture was also grown at 310 K for 2.5 − 3 h until an optical density at 600 nm (OD_{600}) = 0.8 − 1.0 was achieved. To each of the 6 flasks 2 mL of a 1 M Isopropyl-*β*-D-thiogalactopyranosid (IPTG) solution was added as inductor, resulting in a final concentration of 1 mM in each 2 L flask. After expression at 303 K and 130 rpm for 5 − 6 h in a shaking incubator cell harvest was performed with Avanti J26-S high speed centrifuge, Rotor JA-10, 81,800 x g (Beckman Coulter, Indianapolis, USA) at 10000 rpm. The supernatant was discarded while the resulting cell pellets were frozen in liquid nitrogen and stored at 193 K.

6.2.2 Purification

For cell disruption, the frozen cell pellets were thawed on ice, resuspended in PBS (140 mM NaCl, 10 mM NaH_2PO_4, 2.8 mM KCl, 1.7 mM KH_2PO_4, pH 7.3) buffer and homogenized using a glas douncer. Physical cell disruption was performed with a Branson Ultrasonics™ Sonifier Modell 250 CE (Fisher Scientific, Schwerte, Germany) for three times, 5 sec at a duty cycle of 30%, output control of 7 (microtip limit). After 1.5 − 2.5 h centrifugation, the supernatant was discarded and the remaining cell pellets and containing inclusion bodies were washed in washing buffer 1 (WP1) (20 mM Tris-HCl, 20 mM EDTA, 20 mM DTT, 10% (v/v) Triton X-100,

pH 7.4) followed by a wash step with WP2 (20 mM Tris-HCl, 20 mM EDTA, 20 mM DTT, 1 M NaCl, pH 7.4) and WP3 (20 mM Tris-HCl, 20 mM EDTA , pH 7.4) and centrifuged after each washing step. The incorrectly folded pre-thrombin 2, which was still present in the washed pellets, was dissolved in denaturing buffer (6 M guanidinium-HCl, 23 mM ammonium acetate, 10 mM ethylenediaminetetraacetic acid (EDTA), 30 mM cysteine, pH 9.5) on ice and consequently unfolded. To perform *in-vitro*-folding the concentrated protein solution was dropped into the 150-fold volume of pre-cooled refolding buffer (50 mM Tris, 20 mM $CaCl_2$, 500 mM NaCl, 600 mM arginine, 1 mM EDTA, 1 mM cysteine, 0.1 mM cystine, 5% (v/v) glycerol, 0.2% (m/v) Polyethylene glycol hexadecyl ether (Brij 58), pH 9) using a peristaltic pump with constant stirring at 277 K. To concentrate the highly diluted protein solution in the refolding buffer to 120 mL, a Vivaflow 200 filter membrane with a permeability of 10,000 Da molecular weight (Sartorius, Göttingen, Germany) was used. A large amount of misfolded protein, which was still present in the pellet, was denatured and refolded again. Purification by affinity chromatography at 277 K was done with a HiTrap™ Heparin HP column (GE Healtcare Life Sciences, Freiburg, Germany), using the Äkta™ FPLC instrument (GE Healthcare). An example of the elution profile is shown in Figure 59. The concentrated protein, still present in refolding buffer, was eluted from the column using a gradient of Elution Buffer 1 (EB1) (25 mM Tris, pH 7) and EB2 (25 mM Tris, 1 M NaCl, pH 7) and thus purified. Finally, the protein was centrifuged in VivaSpin 10,000 ultracentrifugation device (Vivascience, Merck, Darmstadt, Germany) at 277 K, rebuffered in crystallization or assay buffer and concentrated. The corresponding protein band of Thrombin variant E192Q is visible in Figure 58 at a band height of 35 kDa (P1 – P3).

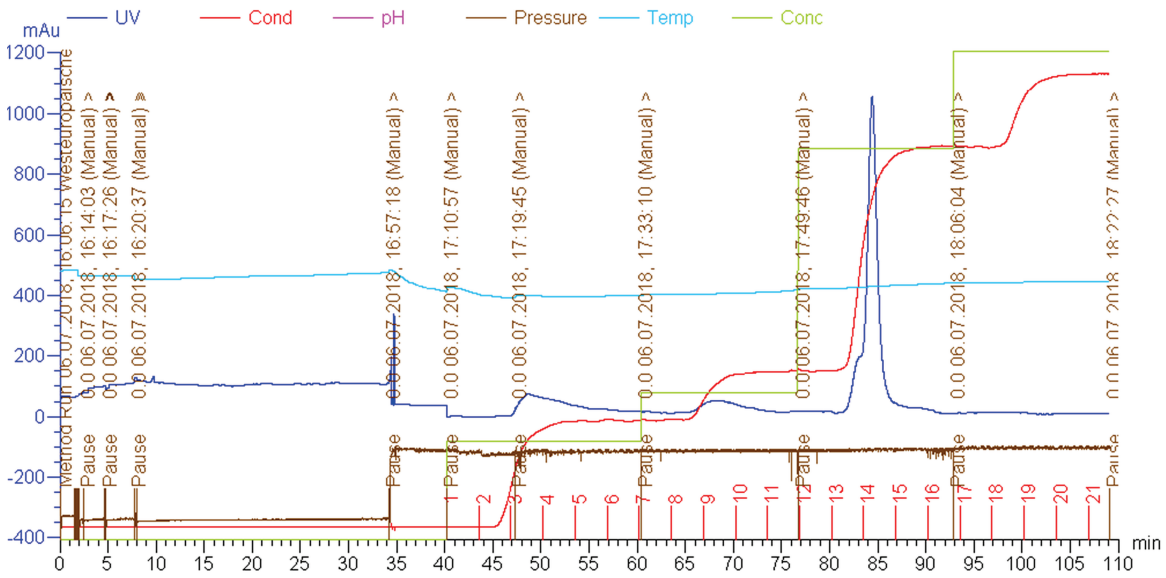


Figure 59: Example of an elution profile of prethrombin E192Q mutant with HiTrap™ Heparin HP column. The UV absorption of the eluate at 280 nm relative to time is shown in blue. Shown in green is the percentage of Elution Buffer 2 in the solvent.

6.2.3 Activation

Concentrated pre-thrombin 2 of the thrombin variant E192Q was activated by cleavage with Ecarin. 0.5 U/µL Ecarin protein activator was dissolved in the respective buffer required for further procedure. A sufficient amount of this solution was added in order to use 4 U activator per 1 mg protein. The mixture was then incubated at 310 K for 3 h under constant shaking. Finally, the protein solution was again stored on ice.

6.3 Crystal Preparation of Thrombin and Thrombin Variant E192Q

6.3.1 Crystallization, Soaking and Freezing

Whereas the preparation of the hanging drop crystallization plates with silicon grease was done at room temperature, all other preparation steps as well as the pipetting of the crystallization samples and the storage of the crystallization plates were performed at 277 K. Human alpha Thrombin obtained from Enzyme Research, South Bend, USA was dissolved in crystallization buffer (20 mM NaH_2PO_4, 350 mM NaCl, 2 mM benzamidine, 0.5 mM sulfated hirudin fragment (54 − 65), acquired from Bachem, Bubendorf, Switzerland, pH 7.5) at a concentration of 10 mg/mL. Each crystallization drop consisted of 1 µL protein solution in crystallization buffer (8 mg/mL thrombin, 2 mM benzamidine, 0.5 mg/mL sulfated hirudin fragment (54 − 65) and 1 µL of reservoir solution (20 mM NaH_2PO_4, 27% (w/v) PEG 8000, pH 7.5). To each reservoir 500 µL of reservoir solution were added. The drop was seeded with former thrombin wildtype crystals using the streak seeding method. The crystals finished growing after 3 − 4 weeks.

For soaking, a 50 mM stock solution of the inhibitors was prepared in DMSO, which was diluted in soaking solution with a ratio of 1:10. The final soaking concentration resulted in 5 mM of the respective compound and 10% DMSO. A crystal without visible imperfections was transferred on a cover slip into a drop of the soaking solution. Similar to the crystallization, the wells contained 500 µL reservoir buffer and were closed with the cover slip and sealed by use of silicone gel. The soaking was stopped after 3 to 24 h. The soaking process was performed identically for all inhibitors and thrombin variants.

After 24 h the soaking was stopped and the crystal was transferred to a cryoprotectant buffer (20% glycerol, 16 mM NaH_2PO_4, 21.6% (w/v) PEG 8000, 5 mM inhibitor) for about one minute. For data collection, the crystal was flash-frozen in liquid nitrogen. The cryoprotection process was performed identically for all inhibitors and thrombin variants.

6.4 Expression and Purification of ALR-2

Expression and Purification of human aldose reductase was performed according to previously described protocols.[5,164,165] During the entire purification process, the identity and purity of the protein was constantly verified using SDS-PAGE. The molecular weight of ALR-2 is about 36 kDa.

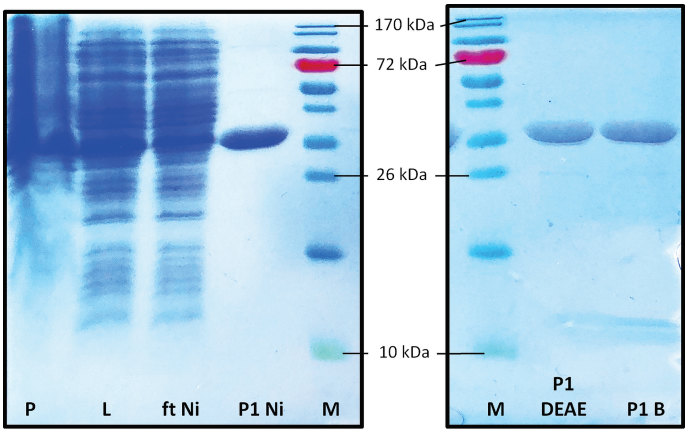


Figure 60: Example of an SDS-PAGE to verify the purity of the expressed and purified protein ALR-2. P: pellet, L: lysate, ft Ni: flow through HiTrap™ Chelating HP with 0.1 M $NiSO_4$ using the peristatic pump, M: marker, P1 DEAE: Peak 1 of the eluate using HiTrap™ DEAE FF Sepharose column, P1 B: Peak 1 of the eluate using HiTrap™ Benzamidine FF column.

6.4.1 Expression

Protein expression was performed with strains of *E. coli* BL21 (DE3). To 300 mL autoclaved LB-medium 300 µL of a 100 mg/mL concentrated ampicillin solution were added. Using an autoclaved pipette tip, either a small amount of a previously prepared glycerol stock was scraped off under the sterile bench or a colony was taken from an agar plate and added to the overnight culture. The preculture covered with aluminium foil was incubated overnight under constant shaking at 310 K. 50 mL of the preculture and 1.6 mL of the 100 mg/mL concentrated ampicillin solution were added to 1.6 L autoclaved SLB-medium and incubated at 310 K under constant shaking. Cell proliferation was controlled by measuring the optical density using NanoDrop™ (Thermo Fisher Scientific, Waltham, Massachusetts, USA) at 600 nm. First, an absorption zero value was set with pure LB medium. Subsequently, samples of the expression culture were measured at certain time intervals. An OD_{600} of about 3 was usually achieved after 4 – 5 h. Protein expression was then induced by the addition of 1.6 mL 1 M IPTG solution. After incubating the culture for another 4 h using the same conditions, it was centrifuged at 10,000 rpm for 10 min to extract a pellet. While the harvested cells were stored at 193 K, the supernatant was discarded.

6.4.2 Purification

For purification, the cells were thawed and evenly resuspended in approx. 40 mL Tris 1 (20 mM Tris, 500 mM NaCl, pH 8). Cell disruption was performed either with a Sonifier or EmulsiFlex-C5 (Ottawa, Ontaro, Canada). The digested cell culture was centrifuged at 277 K for 1 h at 20,000 rpm using an Avanti J26-S high speed centrifuge, Rotor JA-25.5, 75,600 x g (Beckman Coulter, Indianapolis, USA), resulting in the extracted protein in the supernatant. The solution was diluted to 100 − 150 mL with Tris 1 and applied to a chelating nickel column (HiTrap™ Chelating HP, GE Healthcare; prepared with 0.1 M $NiSO_4$). Using an Äkta™ FPLC system with ultraviolet absorption at 280 nm the protein fractions were detected. Initially, the protein-loaded column was flushed with Tris 1 at 2.5 mL/min until no virtual UV-absorption could be detected. A short wash step with low concentrated imidazole in washing buffer 1 (20 mM Tris, 500 mM NaCl, 10 mM imidazole, pH 8) for 30 s at 2 mL/min was performed to eliminate unspecifically bound protein. Subsequently, Tris 1 was reapplied to the column. A gradient of Tris 1 and elution buffer (16 mM Tris, 400 mM NaCl, 200 mM imidazole, pH 8) was used to elute the protein from the nickel column. At a flow rate of 2.5 mL/min for 60 min the concentration of the elution buffer increased to 100%. The protein was eluted at about 25% elution buffer.

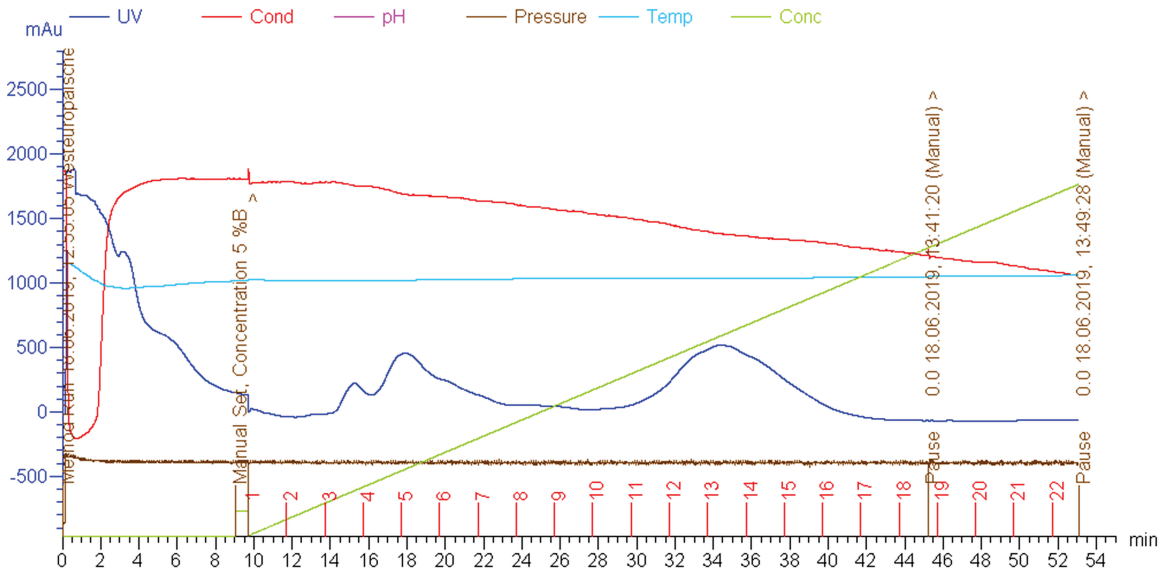


Figure 61: Example of an elution profile of ALR-2 with a HiTrap™ Chelating HP column with 0.1 M $NiSO_4$. The UV absorption of the eluate at 280 nm relative to time is shown in blue. Shown in green is the percentage of Elution Buffer in the solvent.

Fractions collected in 5 mL portions containing the protein were subjected to centrifugal filtration with VivaSpin 10,000 ultracentrifugation device, the buffer was replaced by Tris 2 (10 mM Tris, pH 8) and the protein was concentrated to 15 mL. To separate the histidine tag, the remaining solution with 0.5 U thrombin per mg ALR-2 and 5 mM $CaCl_2$ was filled into a dialysis ZelluTrans tube with a membrane permeability of 10,000 MW (ZelluTrans, Carl Roth + Co. KG, Karlsruhe, Germany) and dialyzed overnight at 277 K in 4.5 L Tris 2. The following day the precipitate was centrifuged at 3,000 rpm and 277 K for 15 min using a Multifuge 3 S-R TTH-750 4.553 x g (Heraeus, Holding, Hanau, Germany) to provide a clear solution for loading the anion exchange column (HiTrap™ DEAE FF Sepharose column, GE Healthcare). The protein solution was diluted with Tris 2 at 1.5 mL/min to 100 − 150 mL and loaded onto the column using a perestaltic pump. After an initial washing on the Äkta™ FPLC system to rinse the loaded column with 100% Tris 2 until the UV absorption remained constant, the protein was eluted through a gradient of Tris 1 and Tris 2. At a flow rate of 1.0 mL/min for 60 min the concentration of Tris 1 increased to 100%. The 5 mL fractions collected on ice and containing the protein were first combined and then rebuffered and concentrated on the respective buffer using a VivaSpin 10,000 ultracentrifugation device. The purity of the protein was verified using an SDS-PAGE. The corresponding protein band of ALR-2 is visible in Figure 60 at a band height of 36 kDa (P1 DEAE).

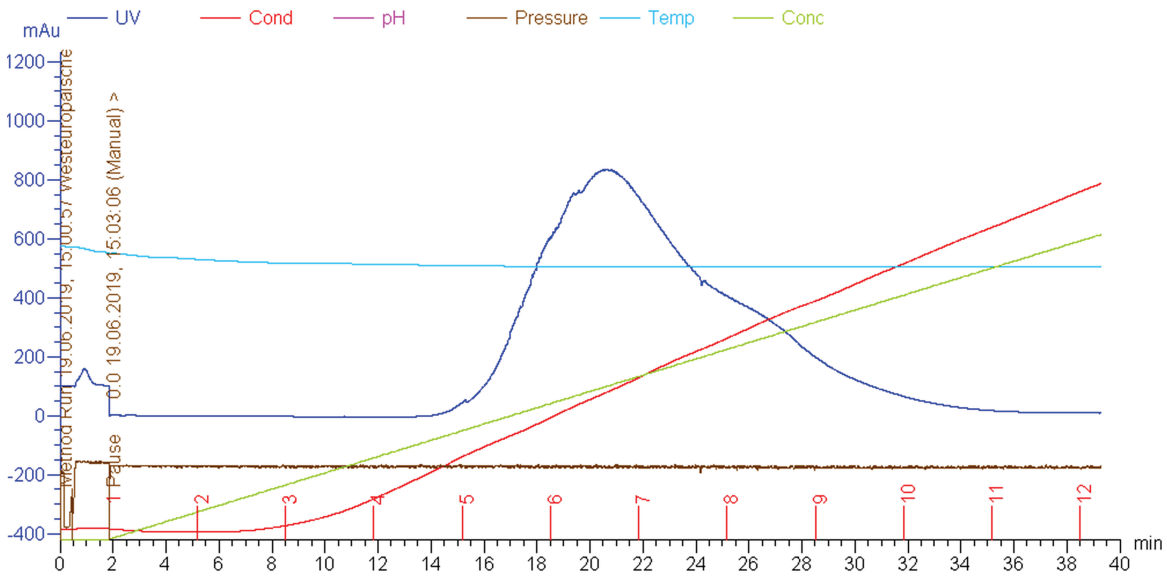


Figure 62: Example of an elution profile of ALR-2 with HiTrap™ DEAE FF Sepharose column. The UV absorption of the eluate at 280 nm relative to time is shown in blue. Shown in green is the percentage of Tris 1 in the solvent.

As an alternative, dialysis was performed in an analogous overnight procedure at 277 K in Tris 1 instead of Tris 2. This offers the advantage that the protein remains dissolved at a constant NaCl concentration of 500 mM and does not bear the risk of precipitation due to the strong reduction of the salinity, as is the case when rebuffering in TRIS 2. Again, the remaining

solution of the previous chelating nickel column with 0.5 U thrombin per mg ALR-2 and 5 mM $CaCl_2$ was filled into a dialysis tube ZelluTrans with a membrane permeability of 10,000 MW to remove the histidine label. Since Tris 1 still has a high salt content, it was not possible to use an anion exchange column afterwards. Instead, a HiTrap™ Benzamidine FF column (GE Healtcare Life Sciences, Freiburg, Germany) was chosen to purify the protein using Tris 1 as a single buffer. The tag-free ALR-2 was previously concentrated to approximately 8 mL using a VivaSpin 10,000 ultracentrifugation device. After rinsing the column with Tris 1 the protein was applied to the column with a 10 ml loop and a syringe at the Äkta™ FPLC system. The protein was eluted at a rate of 2 ml/min. The 5 mL fractions collected on ice and containing the protein were first combined and then rebuffered and concentrated on the respective buffer using a VivaSpin 10,000 ultracentrifugation device. The corresponding protein band of ALR-2 is visible in Figure 60 at a band height of 36 kDa (P1 B)

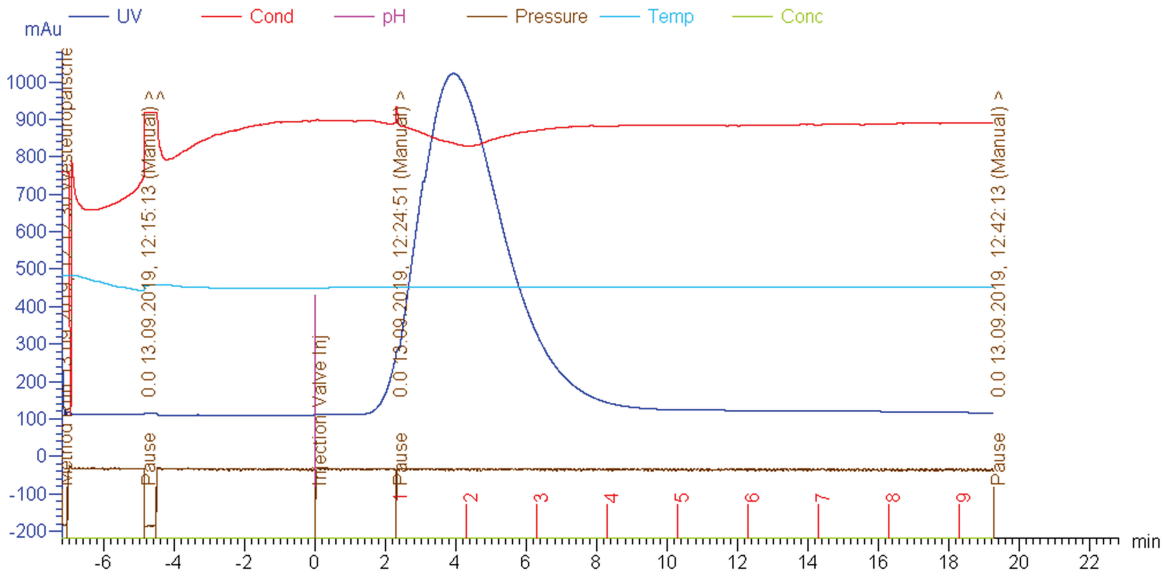


Figure 63: Example of an elution profile of ALR-2 with HiTrap™ Benzamidine FF column. The UV absorption of the eluate at 280 nm relative to time is shown in blue. Shown in green is the per-centage of Tris 1 in the solvent.

6.5 Crystal Preparation of ALR-2

6.5.1 Crystallization, Soaking and Freezing

The crystallization of ALR-2 and mutants was followed known protocols.[37,82,164] After expression, the protein was rebuffered to 50 mM diammonium hydrogen citrate buffer, pH 5 (DAHC) for crystallization and concentrated to about 30 mg/mL. The protein concentration was adjusted photometrically using a NanoDrop™ (Thermo Fisher Scientific, Waltham,

Massachusetts, USA) at 600 nm. To prepare the crystallization solution first Mixture 1 and Mixture 2 were prepared (Table 7) and finally mixed 100 µL of both solutions, resulting in final concentrations of 15 mg/mL ALR-2, 5% (w/v) Polyethylene glycol 6000 (PEG 6000), 5.2 mg/mL Dithiothreitol (DTT) and 0.7 mg/mL NADP disodium salt (Carl Roth + Co. KG, Karlsruhe, Germany). The well solution contained 120 mM DAHC, pH 5.0 and 20% (w/v) PEG 6000. For crystallization, the sitting drop method was used and 9 µL of the crystallization solution was added to each well covered with shark tape (Henkel Adhesives) and equilibrated with 1000 µL reservoir buffer for 3 days at 291 K. After streak seeding with a horsehair, crystals appeared within 1 − 3 days and reached their maximum size after two weeks.

Table 7: Preparation of the ALR-2 crystallization solution in citrate buffer.[37]

Mixture 1		Mixture 2	
50% (w/v) PEG 6000	340 µL	30 mg/mL ALR-2 in 50 mM DAHC	100µL
1 M DTT	34 µL	200 mg/mL $NADP^+$ in 50 mM DAHC	0.66µL
50 mM DAHC	1335 µL		

All crystals were soaked in citrate buffer. For this 2 mg/mL of each inhibitor was dissolved in a soaking solution (120 mM DAHC, pH 5, 25% (w/v) PEG 6000) and a crystal without visible imperfections was transferred into a 9 µL soaking drop and fished after 24 h. To protect crystals 120 mM DAHC mixed with 40% PEG 6000 were used. The crystals were briefly dipped into the cryoprotectant buffer (120 mM DAHC, pH 5, 40% (w/v) PEG 6000, 1 mM inhibitor) and were afterwards flash-frozen in liquid nitrogen.

6.6 Data Collection and Processing

The data collection for thrombin complexes **3.1**, **3.4**, **3.7**, **5.1** − **5.2**, **5.6** and **5.8** as well as for the ALR-2 complexes **2.4** and **2.5** was achieved at 100 K at BESSY II MX beamline 14.1 operated by the Helmholtz-Zentrum Berlin (HZB) in Berlin, Germany. The data of **3.6**, **5.5** and **5.7** in complex with thrombin was collected, at beamline 14.2, respectively. The structures of the ALR-2 complexes **2.3** and **2.4** were also collected at beam line 14.2 using the same conditions. For **3.3** and **5.4** in complex with thrombin data collection was performed at 100 K at European Synchrotron Radiation Facility (ESRF), beamline ID29 in Grenoble, France. Inhibitors **3.2** and **5.9** in complex with thrombin was collected at Elettra beamline 5.2R in Trieste, Italy at 100 K and accordingly **4.5L** in complex with thrombin variant E192Q, at Elettra beamline XRD2. Synchrotron radiation at preselected wavelengths and further data collection statistics are listed in Table 10 − Table 19 in the Appendix. All datasets were indexed, processed and scaled with XDS.[166]

6.7 Structure Determination and Refinement

All structures were determined by molecular replacement with the program PHASER MR[167] implemented in the CCP4 suite[168], using the PDB entry 1H8D[169] as a search model for thrombin and 4PRT[37] for aldose reductase. In the refinement, a subset corresponding to 5% of all reflections were omitted during refinement and used for the calculation of R_{free}. Crystallographic refinement included repeated cycles of conjugate gradient energy minimization and temperature factor refinement performed with the program PHENIX.refine.[170]. Amino acid side chains were fitted to $2F_O - F_C$ and $F_O - F_C$ electron density maps. The program COOT[171] was used to fit the models to the electron density maps. Inhibitors Simplified Molecular Input Line Entry System (SMILES) codes were created with ChemBioDraw Ultra 14.0, (PerkinElmer Informatics, Waltham, Massachusetts, USA) and built with the Grade Web Server[172], which was also used for energy minimization and generation of constraints. As a first refinement step, a Cartesian annealing simulation was used starting with standard parameters. Until the model was completed and provided the best possible explanation of the electron density, the refinement of the XYZ coordinates, occupancies and individual *B*-factors alternated with the structural adaptation in COOT.[171] The temperature factors of complex **4.5L** in E192Q was refined isotropically, or in the case of thrombin complexes **3.2**, **4.1L**, **5.4** and **5.5** anisotropically except water. For the *B*-factors of **3.3** and **5.1** in complex with thrombin, the Translation, Rotation, and Screw-rotation (TLS) refinement was performed with appropriate TLS groups.[173] The definition of the TLS groups was done with PHENIX.refine.[170] For the remaining structures the temperature factors of all atoms, except hydrogen atoms, were refined anisotropically. All structures, except the model of inhibitor **4.5L** in complex with thrombin variant E192Q, were refined with hydrogens automatically added by PHENIX.refine.[170]

The final thrombin and thrombin variant E192Q structures were labeled according to BODE *et al.*[174] Statistics for data collection and refinement are shown in Table 10 − Table 19 in the Appendix..

6.8 Accession Codes

The coordinate files and structure factors of human aldose reductase in complex with inhibitors have been deposited in the RCSB Protein Data Bank (PDB) with the accession codes 6TUF (**2.3**), 6TUC (**2.4**) and 6SYW (**2.6**). 4YS1 (**2.1**) and 4QBX (**2.2**) have already been examined by C. RECHLIN. The electron density for **2.5** was not unambiguous, which prevented accurate interpretation. As a result, the structure of the complex ALR-2 • **2.5** was not deposited in the PDB.

Coordinate files and structure factors of thrombin in complex with the high-affinity inhibitors described in Chapter 2 have been deposited in the PDB with accession codes 5JZY (**3.1**), 5LCE (**3.2**), 5LPD (**3.3**), 6ROT (**3.4**), 6GBW (**3.6**) and 5JFD (**3.7**). **3.5** has already been examined by A. BIELA with the accession code 3RML.[73]

The coordinate files and structural factors of the trypsin complex was deposited in the PDB with the access code 6ZQ2 (**4.1L**).

The structure of complex **4.5L** in E192Q is virtually identical to complex **4.5L** in wild type thrombin (PDB-Code: 6T3Q).[148] Only a significantly poorer resolution differentiates the structures, thus the complex **4.5L** in E192Q was not additionally deposited.

The coordinate files and structural factors of the following inhibitors in complex with trypsin have already been examined by J. SCHIEBEL with the accession codes 5MNN (**4.2F**), 5MNQ (**4.2L**) 5MNH (**4.1F**).[135] K. NGO previously deposited the coordinate files and structural factors of 6T3Q (**4.5L**) in complex with trypsin.[148]

Coordinate files and structure factors of thrombin complexes with the accession codes 2ZDA (**4.1L**), 2ZC9 (**4.4L**) were previously examined by B. BAUM.[125] K. NGO previously deposited the coordinate files and structural factors of 6T0P (**4.5L**) in complex with thrombin.[148] The coordinate files and structural factors of thrombin complexes with the accession codes 4UEH (**4.1F**) and 4UE7 (**4.2F**) have already been examined by E. RÜHMANN.[128]

The coordinate files and structural factors of thrombin in complex with the inhibitors described Chapter 5 were deposited in the PDB with the access codes 6YHJ (**5.2**), 6Y02 (**5.4**), 6ZGO (**5.5**), 6YHG (**5.7**), 6YH9 (**5.8**), 6YB6 (**5.9**). K. NGO previously deposited the coordinate files and structural factors of 6T54 (**5.3**) in complex with thrombin. The electron density for the complexes Thr • **5.1** and Thr • **5.6** were not unambiguous, which prevented accurate interpretation. As a result, the structures were not deposited in the PDB.

6.9 SDS-PAGE

The success of the purification of all proteins was repeatedly checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) during the purification processes. The samples taken were heated to 95°C in a thermal mixer, which denatured the proteins and gave them an even negative charge distribution by SDS. Possible disulfide bridges were broken by the addition of *β*-mercaptoethanol. The polyacrylamide gel, which consisted of a collection and a separation gel, was placed in the chamber, and filled with Laemmli buffer. The composition of the gels is listed in Table 8. The pockets were then filled with samples and markers. The separation is achieved by the migration of the negatively charged proteins to the anode, which are slowed down by the acrylamide mesh according to their size.

Electrophoresis was performed at 130 V (collecting gel) and 180 V (separation gel) for about 1.5 h. The separating gel split from the collecting gel was stained overnight or in the microwave at 180 W for 2 min with Coomassie Blue staining solution. The gel was then decolored using Coomassie decolorizer solution. The protein bands remained stained, which is due to the interaction between basic amino acids and the color tone.

Table 8: Composition of the volumes of the different components of the collecting gel ($V_{Col.}$) and the separation gel ($V_{Sep.}$).

Components	$V_{Col.}$	$V_{Sep.}$
Acrylamide 30%	0.50 mL	1.50 mL
H_2O	1.87 mL	0.74 mL
Collecting- / separating gel buffer	0.60 mL	1.00 mL
SDS-Solution 10%	30.0 µL	30.0 µL
APS	30.0 µL	30.0 µL
TEMED	3.0 µL	3.0 µL

6.10 ITC Measurements

All ITC experiments were performed on a Microcal iTC_{200}™ (Malvern Panalytical, Malvern, U.K.) device at 189 K with a reference power of 5 µcal/s. Data integration and evaluation was done with NITPIC[175] and SEDPHAT.[176]

6.10.1 ITC Measurements of Thrombin

Thrombin samples were prepared by a solution of human α-thrombin (Beriplast, CSL Behring, Marburg, Germany)[177] in a buffer of 100 mM NaCl, 0.1% (w/v) PEG 8000, HEPES, Tricine (N-(tri(hydroxymethyl)methyl)glycine), or Tris at pH 7.8, which was dialyzed at 4°C overnight. The protein treated in this way was subsequently used for further titration experiments. The reported thermodynamic values are given as the average over three measurements.

For each of **3.1** – **3.7**, at least three direct and three displacement titrations have been performed in which the protein was saturated with the weak binder in advance. After a stable baseline had been achieved, all titrations were started. The pre-titration-delay was set to 300 s while the stirring speed was set to 750 rpm. To perform the direct titrations into thrombin, 50 mM inhibitor solution in DMSO was dissolved in the ITC buffer, resulting in an inhibitor concentration of 250 µM. The concentration of thrombin in the cell was set to 25 µM.

Subsequently, the DMSO concentration in all solutions was adjusted to 3%. To prevent artefacts arising from small syringe leakages or air in the syringe, one injection of 0.3 µL (not used in data fitting), was followed by 19 inhibitor injections of 2.0 µL with a duration of 0.6 s and an interval of 180 s between each injection.

The displacement titrations were conducted identically, whereas the protein was saturated with the weak binding inhibitor (WL) **3.8** (Figure 65, Appendix) in advance.[124] **3.8** was added to the cell in a concentration of 83 µM, which corresponds to a saturation D_{sat} [WL] of the protein at approximately 95.5%. The saturation has been calculated, according to Equation 8.[178] The protein concentration, as well as the inhibitor concentration, was doubled. Since a high *c*-value of the isotherms is no obstacle to the correct determination of Δ*H*°, this value was taken from the direct titrations. The association constant of the strong inhibitor (SL) of interest K_a[SL] was calculated using Equation 9,[179] where K_a[obs] corresponds to the observed K_a for displacement titration. [L_{free}] is defined as the amount of weak inhibitor in the cell that has not bound to the protein. Subsequently, the Gibbs free energy Δ*G*°$_{SL}$, the enthalpy Δ*H*°$_{SL}$ and −*T*Δ*S*°$_{SL}$, the entropic term, were determined using Equation 10 and Equation 11 according to KRIMMER and KLEBE.[64]

$$C_{WL} = \frac{D_{sat.}[WL] \cdot C_{SL} - C_{SL} - K_d[WL]}{1 - \frac{1}{D_{sat.}[WL]}} \qquad \text{Equation 8}$$

$$K_a[SL] = K_a[obs] \cdot (K_a[WL] \cdot [L_{free}] + 1) \qquad \text{Equation 9}$$

$$\Delta G_{SL} = -RT \cdot \ln (K_a[SL]) \qquad \text{Equation 10}$$

$$\Delta G_{SL} = \Delta H_{SL} - T\Delta S_{SL} \qquad \text{Equation 11}$$

The ITC experiments for **4.1L** in thrombin were previously published by BAUM *et al*.[125] Inhibitor solutions of 0.5 mM **4.2L** and 0.6 mM **4.3L** were titrated into 50 and 60 µM thrombin, respectively. Inhibitors were dissolved in the same buffer with 3% (v/v) DMSO to assure a complete solubility. The titration protocol consisted of an initial volume of 0.3 µL followed by 15 injections of 1.4 µL for **4.2L** and 12 injections of 1.9 µL for **4.3L** separated by an interval of 180 s.

ITC measurements for **4.1L** and **4.2L** on trypsin were previously described by SCHIEBEL *et al*.[135] **4.3L** was titrated analogous to the protocol given in ref [135] using a direct titration approach. Protein samples were prepared by dialyzing a solution of bovine *β*-trypsin (Sigma Aldrich, product number T8003) against a buffer composed of 80 mM NaCl, 10 mM $CaCl_2$, 0.1% (w/v)

PEG 8000, and 100 mM, HEPES, Tricine, or Tris at a pH of 7.6 and a temperature of 4°C, while **4.3L** was dissolved in buffer contained 3% (v/v) DMSO. The titrations in the three different buffer systems enabled correction for any proton transfer occurring between the buffer and protein upon complex formation. Data analysis followed the same scheme as described in an earlier study.[150] The ITC protocol consisted of an initial 0.3 µL volume injection and 20 – 25 injections with a volume between 1.5 and 1.8 µL separated by a 220 s time interval. The reported thermodynamic values are given as the average over three measurements.

6.10.2 ITC Measurements of ALR-2

After the last purification step, ALR-2 was rebuffered into HEPES buffer for ITC (10 mM HEPES, pH 8) using a VivaSpin 10,000 ultracentrifugation device and concentrated to approximately 15 mg/mL. Before each measurement, a fresh Eppendorf tube was unfrozen, and the concentration was precisely determined photometrically using NanoDrop. At least three direct measurements of **2.3** and **2.4** on ALR-2 were performed in a procedure similar to those applied to thrombin.

The cell was filled with a 100 µM solution containing 67% active ALR-2, an excess of $NADP^+$(0.2 mM, Carl Roth + Co. KG, Karlsruhe, Germany), and 3% (v/v) DMSO in 10 mM HEPES buffer at pH 8.0. For the ITC measurements, the already oxidized co-factor $NADP^+$ was added, which also binds in the protein but can no longer be converted. This prevents falsification of the heat signal of the inhibitor binding by chemical reactions of the cofactor, as demonstrated in previous studies.[87] The syringe was filled with the same concentration of $NADP^+$ and DMSO along with inhibitor (0.75 – 1.0 mM) in 10 mM HEPES at pH 8.0.

One injection of 0.3 µL was followed by 27 inhibitor injections of 1.3 µL with a duration of 0.6 s and an interval of 180 s between each injection. Due to their low affinity, the inhibitors **2.5** and **2.6** were measured by a displacement titration in which the protein was previously saturated with the weak inhibitor (WL) and then titrated with the strong inhibitor (SL) **2.9** (Figure 65, Appendix) which had already been fully characterized previously. The WL was added to the cell corresponding to a saturation D_{sat} [WL] of the protein at approximately 94%, according to Equation 8.[178] Here, one injection of 0.3 µL was followed by 37 injections of 1.0 µL with a duration of 0.6 s and an interval of 180 s between each injection. For the calculation of K_a [WL] of the displacement titrations, Equation 9 was adapted accordingly. Since no exact $\Delta H°$ value can be obtained from a curve with a too small *c*-value, $\Delta H°_{WL}$ was also calculated from the displacement titration according to Equation 12.

$$\Delta H_{WL} = (\Delta H_{SL} - \Delta H_{obs}) \cdot (1 + \frac{1}{K_a[WL] \cdot L_{free}}) \qquad \text{Equation 12}$$

6.11 Kinetic Inhibition Assays

6.11.1 Assay and K_i Determination of Thrombin and Thrombin Variants

The kinetic inhibition constants K_i of a series of inhibitors and fragments (see Figure 70, Appendix), relative to human thrombin, trypsin and thrombin variants A190S, E192Q, D221aA − D222K, and Y225P were determined photometrically by a kinetic fluorescence assay according to STÜRZEBECHER *et al*.[180] For human thrombin and all thrombin variants the tripeptide Tos-Gly-Pro-Arg-AMC • TFA (S1 = 10 µM, S2 = 5 µM, S3 = 2.5 µM) was used as substrate. For Trypsin we used Mes-D-Arg-Gly-Arg-AMC • 2TFA (S1 = 25 µM, S2 = 12.5 µM, S3 = 6.25 µM). Both substrates contained the peptide fluorophoric 7-amido-4-methylcoumarins (AMC). After enzyme-kinetic peptide cleavage, AMC can release the mesomerically stabilized 7-amino-4-methylcoumarine which fluoresces at λ_{ex} = 355 and λ_{em} = 460 nm and can consequently be detected.

The inhibitors were dissolved in DMSO at specific concentrations between 10 pM and 100 mM depending on their expected inhibition strength. With respect to their activity, the proteases were diluted in the assay buffer (50 mM Tris, 154 mM NaCl, 0.01% (w/v) PEG 8000, pH 7.8) to obtain a concentration suitable for measurement. The measurement was performed with a Fluoroskan Ascent™ from Thermo Fisher Scientific, Waltham, Massachusetts, USA at intervals of 20 x 15 s, preceded by an initial mixing time of 40 s. The K_i values were determined as described by Dixon.[181]

6.11.2 Assay and IC_{50} Determination of ALR-2

The IC_{50} values of the respective inhibitors on ALR-2 were performed based on previous prococols.[5,37,182] To more accurately estimate the affinity of the inhibitors, all inhibitors were previously measured in a pretest at concentrations of 800 µM, 80 µM, 8 µM, 800 nM, 80 nM, 8 nM, 800 nM. For the actual measurement the inhibitors were dissolved in DMSO at concentrations ranging from 100 mM to 1 nM, depending on their expected inhibition strength. After expression, ALR-2 was rebuffered to 100 mM phosphate buffer (3.2 mg/mL dipotassium hydrogen phosphate, 11.0 mg/mL potassium dihydrogen phosphate, pH 6.2) using a Vivaspin ultracentrifugation device, concentrated to 12.5 mg/mL and NADPH tetrasodium salt (Carl Roth + Co. KG, Karlsruhe, Germany) 3mg/mL was added. For measurement, the prepared

solutions were pipetted along with the corresponding dilution of the inhibitor and a 50 mg/mL D-xylose solution (Sigma-Aldrich, Merck, Darmstadt, Germany) into 96-well plates (Nuclon™ Delta Surface, Thermo Fisher Scientific, Waltham, Massachusetts, USA) as described in Table 9. Each inhibitor was measured with triplets of 24 different concentrations at a temperature of 298 K and wave lengths of λ_{ex} = 340 and λ_{em} = 496 nm. After shaking for 15 s, a measuring interval of 10 s was followed by a total of 120 repetitions. The assay was evaluated using the program GrapPad Prism version 7.0.0. for Windows (GraphPad Software, San Diego, California, USA). All measurements were performed in triplet; thus, the final IC_{50} values represented a mean of three.

Table 9: Composition of the sample solution in the well for the kinetic fluorescence assay to determine IC_{50} values on ALR-2.

Volume	Solution	c_{final} in Well
20 µL	ALR-2 in phosphatebuffer (12.5 mg/mL)	280 nM
	NADPH (3 mg/mL)	0.28 mM
2 µL	Inhibitor in 100% DMSO	800 µM – 2 pM
20 µL	D-Xylose (50 mg/mL)	26.5 mM
208 µL	Phosphatebuffer	100 mM

Appendix

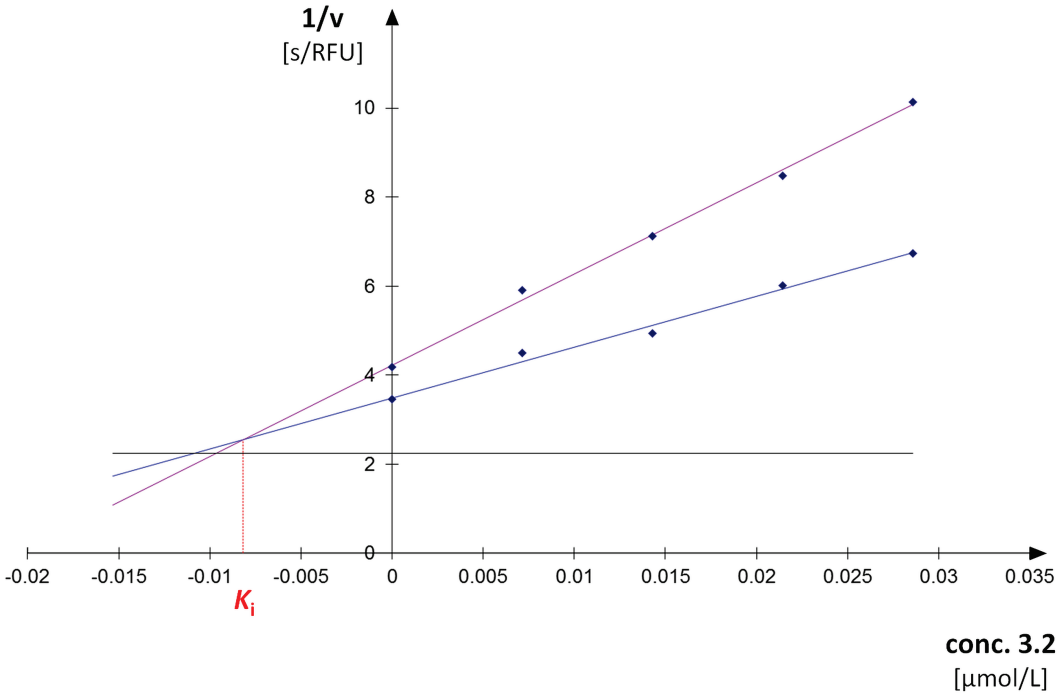
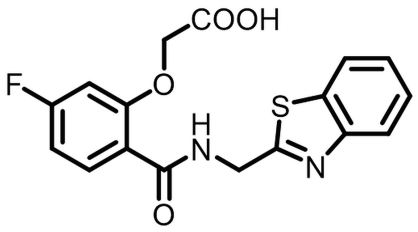
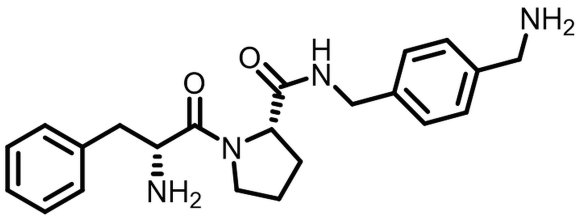


Figure 64: Schematic representation of the determination of the K_i Assay value for thrombin by Dixon-plot using the example of the inhibitor **3.2**. On the x-axis the concentration of inhibitor in µmol/L is plotted. The y-axis indicates the reciprocal of the rate of conversion v. The violet and the blue line indicate the different concentrations of the substrate Tos-Gly-Pro-Arg-AMC·TFA (blue: 10 µmol/L, violet 5 µmol/L). On the x-axis, K_i can be determined by the red dashed line.



2.9



3.8

Figure 65: Reference Inhibitors for ITC. Strong reference inhibitor **2.9** (K_d = 0.42 µM), used for displacement titrations of ALR-2 and weak reference inhibitor **3.8** (K_d = 1.65 µM), used for displacement ITC of thrombin.

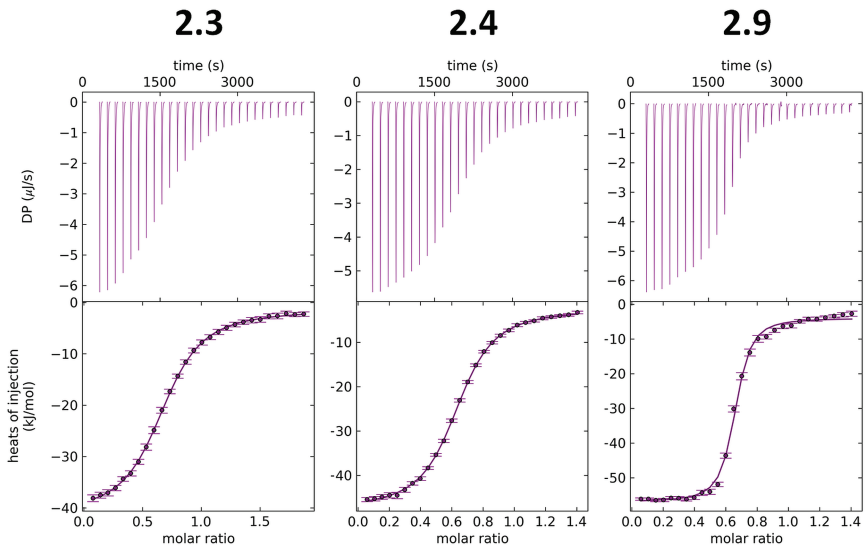


Figure 66: Examples of raw thermograms and integrated heat values for one of at least three direct ITC measurements of the inhibitors **2.1**, **2.4** and **2.9**. For the thermogram on the top the y-axis shows the differential power in µJ/s and the x-axis the measuring time in s. The y-axis of the evaluated data below shows the heats of injections in kJ/mol and the x-axis the molar ratio.

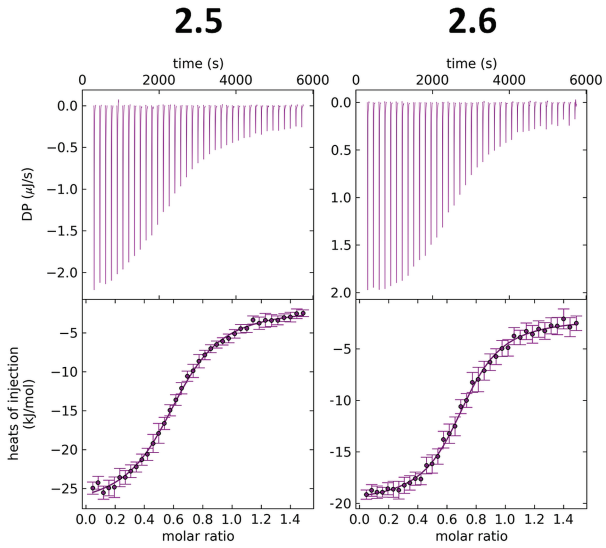


Figure 67: Examples of raw thermograms and integrated heat values for one of at least three displacement ITC measurements of the inhibitors **2.5** – **2.6**. For the thermogram on the top the y-axis shows the differential power in µJ/s and the x-axis the measuring time in s. The y-axis of the evaluated data below shows the heats of injections in kJ/mol and the x-axis the molar ratio.

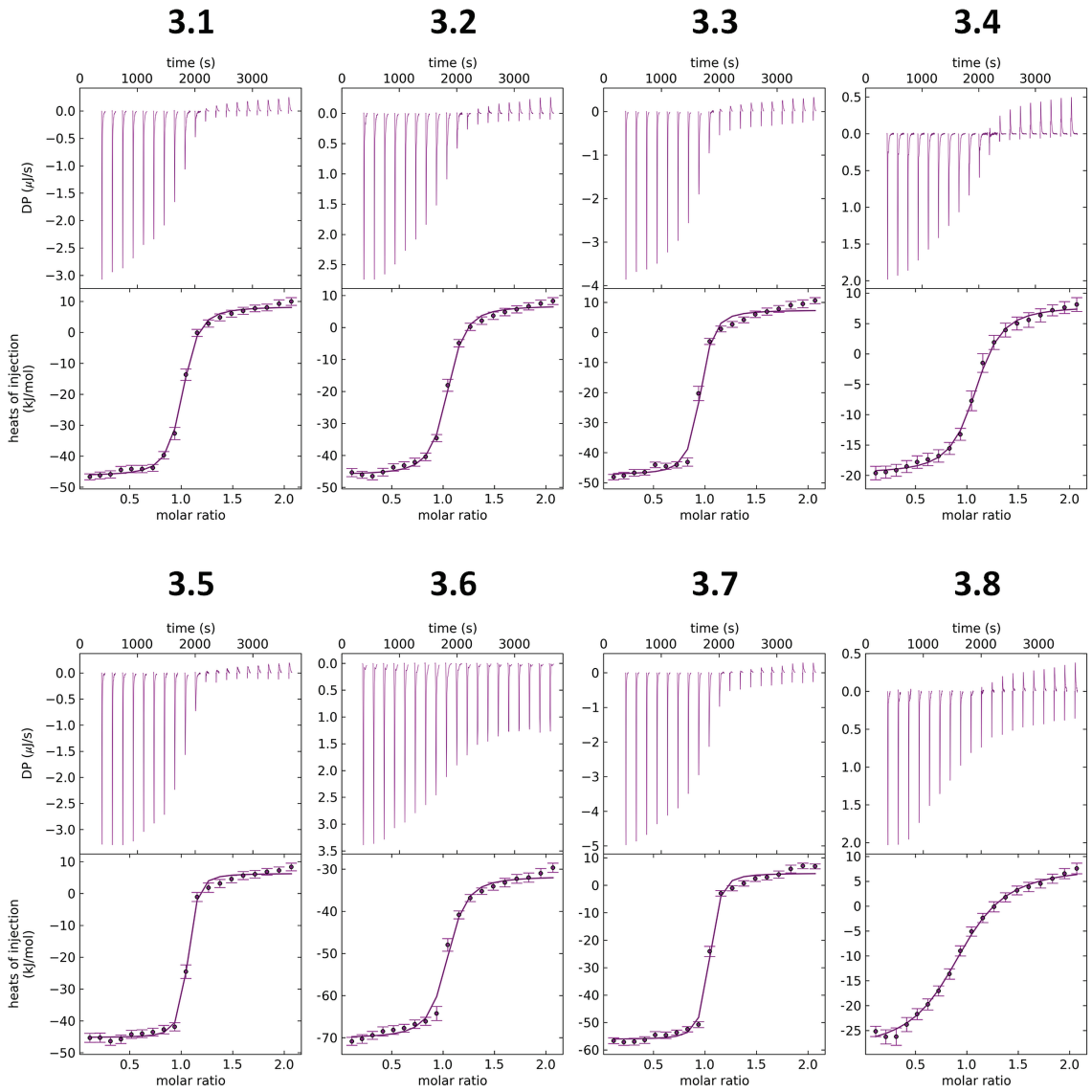


Figure 68: Examples of raw thermograms and integrated heat values for one of three direct ITC measurements of the inhibitors **3.1** – **3.8**. For the thermogram on the top the y-axis shows the differential power in µJ/s and the x-axis the measuring time in s. The y-axis of the evaluated data below shows the heats of injections in kJ/mol and the x-axis the molar ratio.

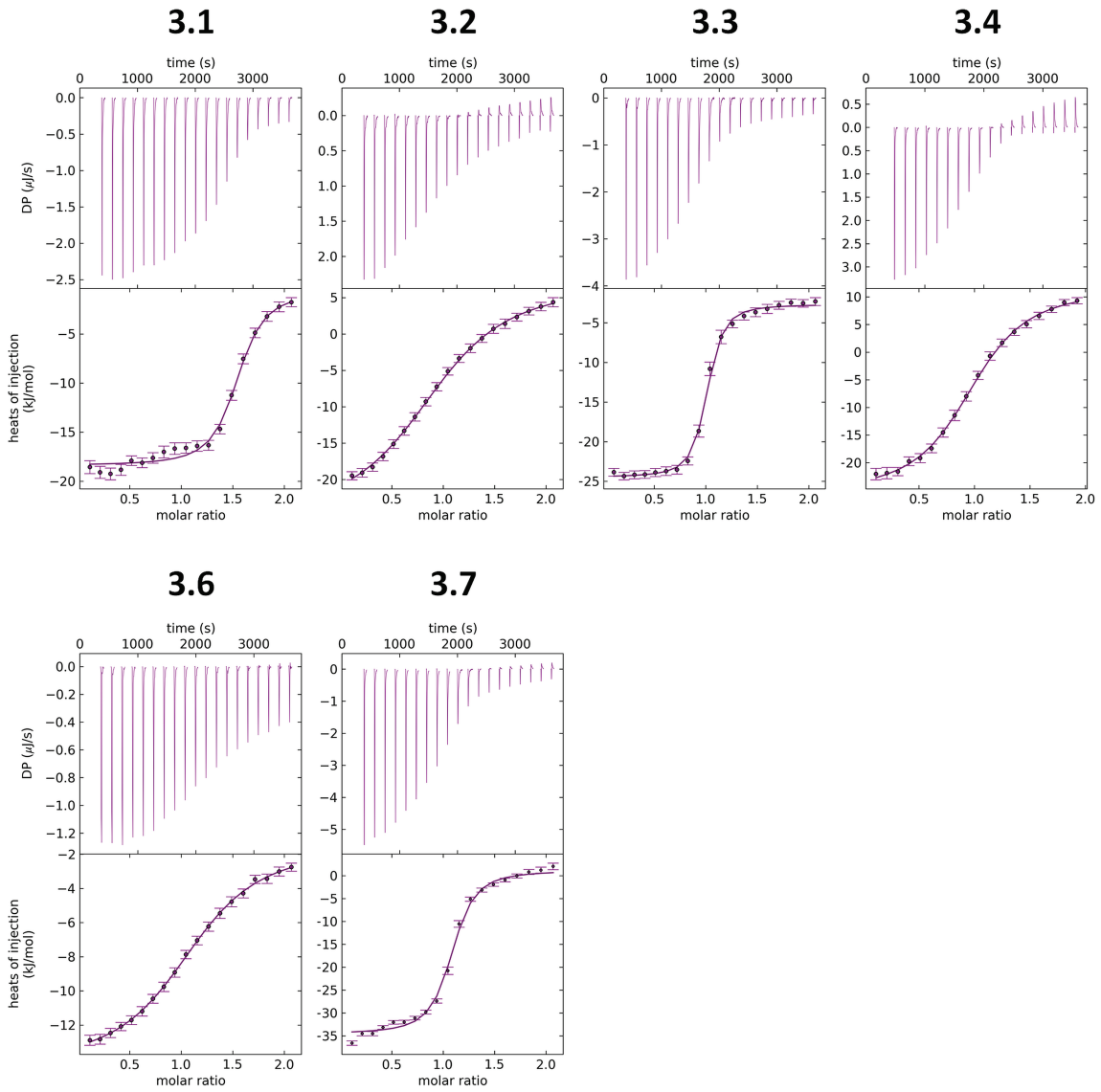


Figure 69: Examples of raw thermograms and integrated heat values for one of three displacement ITC measurements of the inhibitors **3.1** – **3.4** and **3.6** – **3.7**. For the thermogram on the top the y-axis shows the differential power in µJ/s and the x-axis the measuring time in s. The y-axis of the evaluated data below shows the heats of injections in kJ/mol and the x-axis the molar ratio.

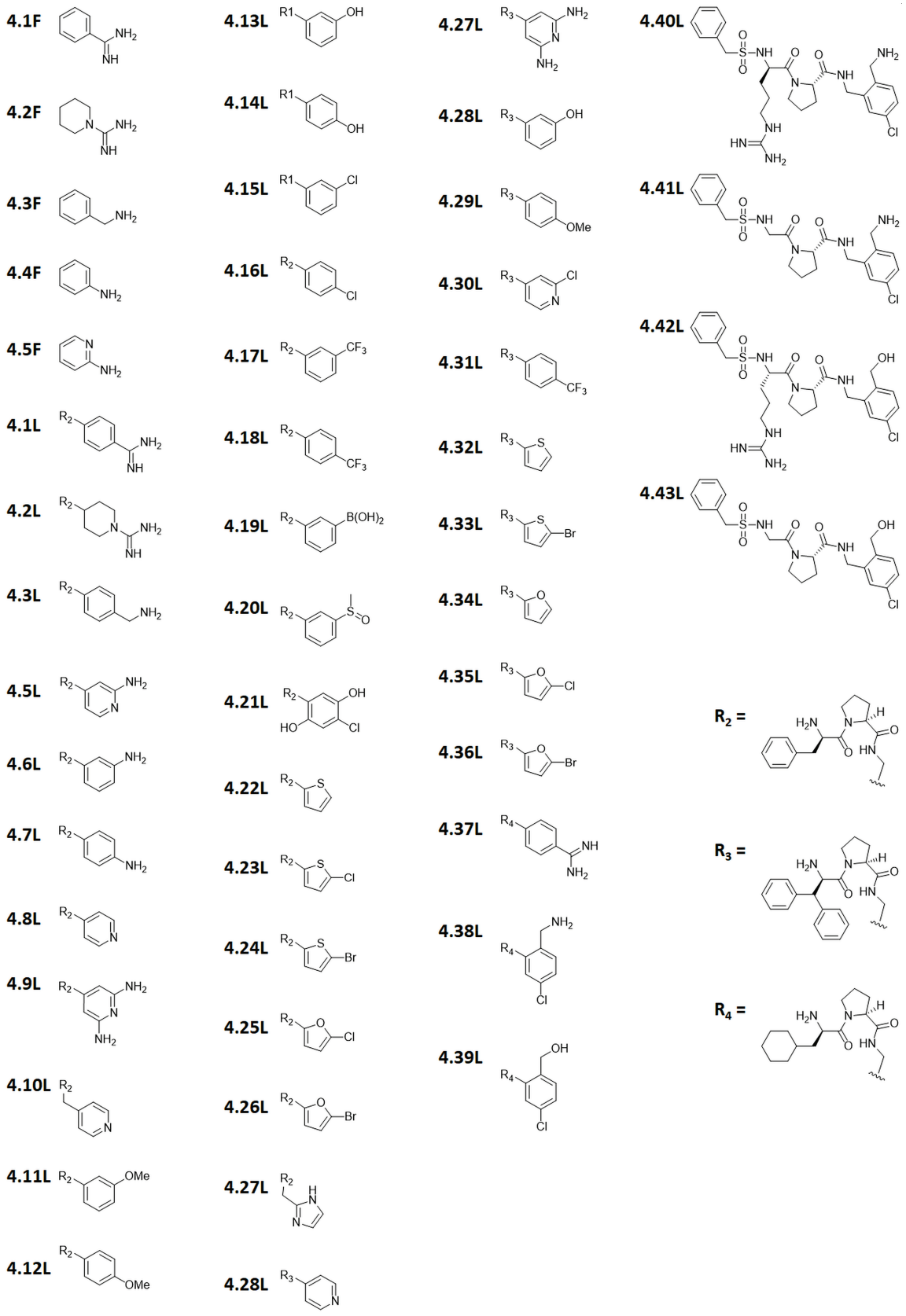


Figure 70: Fragments and inhibitors used in the enzyme kinetic assay on thrombin wildtype and thrombin variants.

Table 10: X-ray data collection and refinement statistics of inhibitors **2.3** – **2.5** in complex with ALR-2 wildtype.

	2.3 (6TUF)	**2.4** (6TUC)	**2.5** (not deposited)
(A) Data collection and processing			
Beamline	Bessy 14.2	Bessy 14.2	Bessy 14.1
Wavelength [Å]	0.9184	0.9184	0.9184
Space group	$P12_11$	$P12_11$	$P12_11$
Unit Cell parameters:			
a, b, c [Å]	47.2, 66.4, 49.2	47.2, 66.5, 49.2	47.3, 66.7, 49.4
α, β, γ [°]	90.0, 92.4, 90.0	90.0, 92.1, 90.0	90.0, 92.8, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.2	2.1	2.2
Solvent content [%]	43	43	43
(B) Diffraction Data[a]			
Resolution range [Å]	47.13 – 1.15 (1.22 – 1.15)	49.19 – 1.06 (1.12 – 1.06)	39.68 – 0.93 (099 – 0.93)
Unique reflections	105125 (16246)	132582 (20046)	198129 (28823)
$R(I)_{sym}$ [%][b]	4.0 (50.7)	7.6 (37.1)	4.2 (30.6)
Completeness [%]	97.3 (93.3)	96.3 (90.3)	96.5 (87.0)
Redundancy	3.7 (3.5)	3.6 (3.4)	3.4 (2.9)
I/σ (*I*)	19.2 (2.3)	14.8 (3.6)	15.0 (2.8)
(C) Refinement			
Resolution range [Å]	34.83 – 1.15	49.24 – 1.06	38.61 – 0.93
Reflections used in refinement			
work	99878	125952	188222
free	5257	6630	9907
Final R values			
work [%][c]	12.9	12.6	10.4
free [%][d]	15.2	14.3	11.7
Number of protein residues	316	313	316
$NADP^+$ atoms	48	48	48
Inhibitor atoms	40	30	17
Water molecules	375	463	454
Other inhibitor atoms	13	13	13
RMSD bonds			
Bond length [Å]	0.006	0.006	0.006
Bond angles [°]	0.99	1.00	1.04
Ramachandran plot[e]			
favored regions [%]	91.0	89.5	91.4
additional allowed reg. [%]	9.0	10.5	8.6
generously allowed reg. [%]	0.0	0.0	0.0
Mean *B*-factor [$Å^2$][f]			
Protein	13.4	8.8	10.1
Inhibitor	18.1	15.0	23.3
Water molecules	25.9	21.7	24.0
$NADP^+$	8.8	4.7	6.5
Other inhibitors	14.8	8.2	10.7

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-\langle I\rangle|}{\Sigma|I|} \cdot 100$ for which *I* = observed intensity and $\langle I\rangle$ = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 11: X-ray data collection and refinement statistics of inhibitor **2.6** in complex with ALR-2 wildtype

	2.6 (6SYW)
(A) Data collection and processing	
Beamline	Bessy 14.1
Wavelength [Å]	0.91841
Space group	$P12_11$
Unit Cell parameters:	
a, *b*, *c* [Å]	47.3, 66.9, 49.3
α, *β*, *γ* [°]	90.0, 92.0, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.2
Solvent content [%]	43
(B) Diffraction Data[a]	
Resolution range [Å]	47.29 – 0.93 (0.99 – 0.93)
Unique reflections	181142 (19356)
$R(I)_{sym}$ [%][b]	5.0 (45.8)
Completeness [%]	88.1 (58.3)
Redundancy	4.5 (3.8)
I/σ (I)	15.7 (2.5)
(C) Refinement	
Resolution range [Å]	39.69 – 0.93
Reflections used in refinement	
work	172084
free	9058
Final R values	
work [%][c]	10.9
free [%][d]	12.3
Number of protein residues	316
$NADP^+$ atoms	48
Inhibitor atoms	38
Water molecules	463
Other inhibitor atoms	13
RMSD bonds	
Bond length [Å]	0.010
Bond angles [°]	1.22
Ramachandran plot[e]	
favored regions [%]	89.9
additional allowed reg. [%]	10.1
generously allowed reg. [%]	0.0
Mean *B*-factor [$Å^2$][f]	
Protein	9.9
Inhibitor	18.4
Water molecules	23.8
$NADP^+$	6.8
Other inhibitors	10.4

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-<I>|}{\Sigma|I|} \cdot 100$ for which I = observed intensity and $<I>$ = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 12: X-ray data collection and refinement statistics of inhibitors **3.1** – **3.3** in complex with thrombin wildtype.

	3.1 (5JZY)	**3.2** (5LCE)	**3.3** (5LPD)
(A) Data collection and processing			
Beamline	Bessy 14.1	Elettra 5.2R	ESRF ID29
Wavelength [Å]	0.91841	1.00000	0.979
Space group	C121	C121	C121
Unit Cell parameters:			
a, b, c [Å]	69.9, 71.5, 72.3	69.9, 71.2, 72.6	70.1, 71.0, 72.8
α, β, γ [°]	90.0, 100.5, 90.0	90.0, 100.4, 90.0	90.0, 100.5, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.5	2.5	2.5
Solvent content [%]	50	50	50
(B) Diffraction Data[a]			
Resolution range [Å]	49.60 – 1.27 (1.35 – 1.27)	43.39 – 1.39 (1.47 – 1.39)	49.46 – 1.50 (1.59 – 1.50)
Unique reflections	91827 (14757)	69244 (10871)	55657 (8805)
$R(I)_{sym}$ [%][b]	3.3 (48.0)	4.0 (47.9)	5.1 (48.5)
Completeness [%]	99.6 (99.2)	98.5 (96.0)	98.8 (97.2)
Redundancy	3.3 (3.2)	2.8 (2.7)	3.8 (3.6)
I/σ (I)	20.2 (2.4)	14.3 (2.1)	15.4 (2.7)
(C) Refinement			
Resolution range [Å]	49.56 – 1.27	43.39 – 1.39	49.46 – 1.50
Reflections used in refinement			
work	87235	65781	52874
free	4592	3463	2783
Final R values for all Reflections			
work [%][c]	12.0	14.0	14.9
free [%][d]	14.2	16.8	17.0
Number of protein residues			
L-chain	29	28	31
H-chain	250	250	251
hirudin	12	11	11
Sodium ions	2	2	2
Inhibitor atoms	29	29	29
Water molecules	332	279	255
Other inhibitor atoms	47	46	31
RMSD bonds			
Bond length [Å]	0.012	0.005	0.010
Bond angles [°]	1.20	0.84	1.03
Ramachandran plot[e]			
favored regions [%]	87.4	87.3	85.5
additional allowed reg. [%]	12.6	12.7	14.1
generously allowed reg. [%]	0.0	0.0	0.4
Mean *B*-factor [$Å^2$][f]			
Protein	17.8	20.7	21.4
Inhibitor	14.3	17.4	16.9
Water molecules	32.7	34.9	31.3
Other inhibitors	31.8	33.8	40.2

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-\langle I\rangle|}{\Sigma|I|} \cdot 100$ for which I = observed intensity and $\langle I\rangle$ = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 13: X-ray data collection and refinement statistics of inhibitors **3.4, 3.6** and **3.7** in complex with thrombin wildtype.

	3.4 (6ROT)	**3.6** (6GBW)	**3.7** (5JFD)
(A) Data collection and processing			
Beamline	Bessy 14.1	Bessy 14.2	Bessy 14.1
Wavelength [Å]	0.91841	0.91841	0.91841
Space group	C121	C121	C121
Unit Cell parameters:			
a, b, c [Å]	69.9, 71.2, 73.0	69.8, 71.1, 72.9	69.4, 71.5, 72.1
α, β, γ [°]	90.0, 100.7, 90.0	90.0, 100.6, 900.0	90.0, 99.8, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.5	2.5	2.5
Solvent content [%]	50	50	51
(B) Diffraction Data [a]			
Resolution range [Å]	43.52 – 1.34 (1.73 – 1.34)	43.43 – 1.45 (1.54 – 1.45)	43.13 – 1.46 (1.55 – 1.46)
Unique reflections	77800 (12438)	61523 (9827)	58841 (9294)
$R(I)_{sym}$ [%][b]	4.4 (58.8)	4.5 (50.4)	3.9 (47.0)
Completeness [%]	98.5 (97.7)	99.0 (97.9)	97.3 (95.3)
Redundancy	3.4 (3.4)	3.4 (3.5)	2.8 (2.8)
I/σ (*I*)	16.6 (2.3)	15.4 (2.3)	15.9 (2.2)
(C) Refinement			
Resolution range [Å]	43.52 – 1.34	43.43 – 1.45	43.13 – 1.46
Reflections used in refinement			
work	73910	58446	55899
free	3890	3077	2942
Final R values for all Reflections			
work [%][c]	13.6	13.2	13.0
free [%][d]	15.7	16.5	16.3
Number of protein residues			
L-chain	28	28	31
H-chain	250	250	252
hirudin	11	11	11
Sodium ions	2	2	2
Inhibitor atoms	32	39	39
Water molecules	247	245	266
Other inhibitor atoms		28	32
RMSD bonds			
Bond length [Å]	0.008	0.008	0.012
Bond angles [°]	1.04	1.00	1.22
Ramachandran plot[e]			
favored regions [%]	86.5	85.7	86.3
additional allowed reg. [%]	13.5	13.9	13.3
generously allowed reg. [%]	0.0	0.4	0.4
Mean *B*-factor [$Å^2$][f]			
Protein	22.3	22.4	22.5
Inhibitor	24.2	24.2	19.5
Water molecules	35.7	34.3	35.8
Other inhibitors	34.8	38.3	41.5

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-\langle I\rangle|}{\Sigma|I|} \cdot 100$ for which *I* = observed intensity and <*I*> = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 14: X-ray data collection and refinement statistics of inhibitor **4.5L** in complex with thrombin variant E192Q.

	4.5L (not deposited)
(A) Data collection and processing	
Beamline	Elettra XRD2
Wavelength [Å]	0.97900
Space group	C121
Unit Cell parameters:	
a, *b*, *c* [Å]	70.0, 71.3, 72.7
α, *β*, *γ* [°]	90.0, 100.7, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.5
Solvent content [%]	50
(B) Diffraction Data [a]	
Resolution range [Å]	35.72 – 2.05 (2.17 – 2.05)
Unique reflections	21507 (3520)
$R(I)_{sym}$ [%][b]	12.2 (50.3)
Completeness [%]	97.0 (97.8)
Redundancy	2.5 (2.5)
I/σ (I)	6.5 (2.0)
(C) Refinement	
Resolution range [Å]	35.72 – 2.05
Reflections used in refinement	
work	20431
free	1076
Final R values for all Reflections	
work [%][c]	20.6
free [%][d]	23.5
Number of protein residues	
L-chain	29
H-chain	251
hirudin	11
Sodium ions	2
Inhibitor atoms	35
Water molecules	151
Other inhibitor atoms	37
RMSD bonds	
Bond length [Å]	0.007
Bond angles [°]	0.91
Ramachandran plot[e]	
favored regions [%]	87.4
additional allowed reg. [%]	12.1
generously allowed reg. [%]	0
Mean *B*-factor [$Å^2$][f]	
Protein	26.5
Inhibitor	23.0
Water molecules	27.4
Other inhibitors	37.1

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-<I>|}{\Sigma|I|} \cdot 100$ for which I = observed intensity and $<I>$ = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 15: X-ray data collection and statistics for inhibitor **4.1L** in complex with trypsin wildtype.

	4.1L (6ZQ2)
(A) Data collection and processing	
Beamline	Bessy 14.2
Wavelength [Å]	0.91841
Space group	$P3_121$
Unit Cell parameters:	
a, *b*, *c* [Å]	54.8, 54.8, 108.5
α, *β*, *γ* [°]	90.0, 90.0, 120.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.4
Solvent content [%]	48
(B) Diffraction Data[a]	
Resolution range [Å]	47.46 – 1.29 (1.37 – 1.29)
Unique reflections	48302 (7699)
$R(I)_{sym}$ [%][b]	4.6 (54.3)
Completeness [%]	99.9 (99.7)
Redundancy	9.7 (9.4)
$I/\sigma(I)$	26.1 (3.7)
(C) Refinement	
Resolution range [Å]	47.49 – 1.29
Reflections used in refinement	
work	45886
free	2416
Final R values for all Reflections	
work [%][c]	12.7
free [%][d]	14.5
Number of protein residues	223
Calcium ions	1
Inhibitor atoms	29
Water molecules	221
RMSD bonds	
Bond length [Å]	0.005
Bond angles [°]	0.90
Ramachandran plot[e]	
favored regions [%]	87.8
additional allowed reg. [%]	12.2
generously allowed reg. [%]	0.0
Mean *B*-factor [$Å^2$][f]	
Protein	14.8
Inhibitor	14.5
Water molecules	24.9

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-<I>|}{\Sigma|I|} \cdot 100$ for which *I* = observed intensity and <*I*> = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 16: X-ray data collection and statistics for inhibitors **4.1L** – **4.4L** in complex with thrombin variant A190S

	4.1L (5MM6)	**4.4L** (5MLS)
(A) Data collection and processing		
Beamline	ESRF ID29	ESRF ID23
Wavelength [Å]	0.978997	0.872899
Space group	C121	C121
Unit Cell parameters:		
a, *b*, *c* [Å]	70.3, 71.7, 72.5	70.4, 71.6, 73.7
α, *β*, *γ* [°]	90.0, 100.5, 90.0	90.0, 100.6, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.5	2.5
Solvent content [%]	51	51
(B) Diffraction Data[a]		
Resolution range [Å]	43.58 – 1.29 (1.37 – 1.29)	43.70 – 1.62 (1.71 – 1.62)
Unique reflections	85703 (13335)	44585 (6991)
$R(I)_{sym}$ [%][b]	3.6 (53.2)	6.8 (50.8)
Completeness [%]	96.3 (93.1)	97.7 (95.9)
Redundancy	3.4 (3.4)	3.8 (3.7)
I/σ (I)	16.3 (2.4)	14.3 (2.8)
(C) Refinement		
Resolution range [Å]	43.60 – 1.29	43.71 – 1.62
Reflections used in refinement		
work	81394	42350
free	4285	2229
Final R values for all Reflections		
work [%][c]	12.3	15.9
free [%][d]	14.2	17.9
Number of protein residues		
L-chain	28	28
H-chain	251	250
hirudin	11	11
Sodium ions	2	2
Inhibitor atoms	29	26
Water molecules	224	189
Other inhibitor atoms	45	46
RMSD bonds		
Bond length [Å]	0.007	0.009
Bond angles [°]	1.1	1.1
Ramachandran plot[e]		
favored regions [%]	85.8	85.7
additional allowed reg. [%]	13.8	13.9
generously allowed reg. [%]	0.4	0.4
Mean *B*-factor [$Å^2$][f]		
Protein	20.8	22.2
Inhibitor	17.1	16.9
Water molecules	32.8	26.4
Other inhibitor atoms	29.8	29.6

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I - \langle I \rangle|}{\Sigma|I|} \cdot 100$ for which I = observed intensity and $\langle I \rangle$ = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O - F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 17: X-ray data collection and statistics for inhibitors **5.1** – **5.2**, **5.4** in complex with thrombin wildtype

	5.1 (not deposited)	**5.2** (6YHJ)	**5.4** (6Y02)
(A) Data collection and processing			
Beamline	Bessy 14.1	Bessy 14.1	ESRF ID29
Wavelength [Å]	0.91841	0.91841	0.979078
Space group	C121	C121	C121
Unit Cell parameters:			
a, *b*, *c* [Å]	70.5, 71.2, 72,9	70.3, 71.4, 72.6	70.1, 71.4, 72.9
α, *β*, *γ* [°]	90.0, 100.8, 90.0	90.0, 100.6, 90.0	90.0, 100.5, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.5	2.5	2.5
Solvent content [%]	51	51	51
(B) Diffraction Data [a]			
Resolution range [Å]	43.6 – 1.70 (1.81 – 1.70)	43.53 – 1.44 (1.53 – 1.44)	43.55 – 1.48 (1.57 – 1.48)
Unique reflections	386551 (6199)	63397 (10134)	58061 (9148)
$R(I)_{sym}$ [%][b]	5.7 (48.4)	6.6 (37.7)	5.1 (50.4)
Completeness [%]	99.7(99.3)	99.2 (98.3)	98.0 (95.8)
Redundancy	3.8 (3.8)	3.4 (3.3)	3.0 (3.0)
I/σ (I)	13.41 (2.66)	10.7 (3.3)	11.7 (2.2)
(C) Refinement			
Resolution range [Å]	43.6 – 1.70	43.55 – 1.44	43.57 – 1.48
Reflections used in refinement			
work	36718	60227	55158
free	1933	3170	2903
Final R values for all Reflections			
work [%][c]	19.6	12.7	13.9
free [%][d]	23.3	15.4	16.4
Number of protein residues			
L-chain	28	33	28
H-chain	258	253	250
hirudin	11	11	11
Sodium ions	0	2	2
Inhibitor atoms	25	52	26
Water molecules	188	292	205
Other inhibitor atoms	34	28	40
RMSD bonds			
Bond length [Å]	0.008	0.009	0.008
Bond angles [°]	1.03	1.15	1.00
Ramachandran plot[e]			
favored regions [%]	84.8	86.8	85.3
additional allowed reg. [%]	14.8	12.8	14.7
generously allowed reg. [%]	0.4	0.4	0.0
Mean *B*-factor [$Å^2$][f]			
Protein	29.4	20.9	23.9
Inhibitor	26.4	23.6	22.1
Water molecules	34.5	33.2	32.3
Other inhibitors	54.2	38.5	39.2

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-<I>|}{\Sigma|I|} \cdot 100$ for which I = observed intensity and $<I>$ = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 18: X-ray data collection and statistics for inhibitors **5.5** – **5.7** in complex with thrombin wildtype

	5.5 (6ZGO)	**5.6** (not deposited)	**5.7** (6YHG)
(A) Data collection and processing			
Beamline	Bessy 14.2	Bessy 14.1	Bessy 14.2
Wavelength [Å]	0.9184	0.9184	0.9184
Space group	C121	C121	C121
Unit Cell parameters:			
a, *b*, *c* [Å]	70.1, 71.9, 72.6	70.2, 71.6, 72.2	70.3, 71.3, 72.7
α, *β*, *γ* [°]	90.0, 100.6, 90.0	90.0, 100.2, 90.0	90.0, 100.6, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.5	2.5	2.5
Solvent content [%]	50	51	51
(B) Diffraction Data [a]			
Resolution range [Å]	49.41 – 1.79 (1.90 – 1.79)	43.44 – 1.50 (1.59 – 1.50)	43.53 – 1.33 (1.39 – 1.33)
Unique reflections	32977 (5253)	53775 (8345)	152846 (23613)
$R(I)_{sym}$ [%][b]	6.9 (49.1)	7.2 (50.2)	4.1 (45.0)
Completeness [%]	99.5 (98.6)	95.5 (91.9)	91.5 (87.1)
Redundancy	3.8 (3.9)	2.8 (2.9)	2.1 (2.1)
I/σ (I)	11.7 (2.4)	9.3 (2.3)	13.8 (2.2)
(C) Refinement			
Resolution range [Å]	33.54 – 1.79	43..44 – 1.50	35.63 – 1.33
Reflections used in refinement			
work	31328	51086	76676
free	1649	2689	4036
Final R values for all Reflections			
work [%][c]	18.5	14.8	12.3
free [%][d]	22.5	18.4	14.5
Number of protein residues			
L-chain	28	27	31
H-chain	259	257	252
hirudin	12	11	11
Sodium ions	2	2	2
Inhibitor atoms	26	21	28
Water molecules	159	262	297
Other inhibitor atoms	27	32	46
RMSD bonds			
Bond length [Å]	0.012	0.014	0.007
Bond angles [°]	1.17	1.28	1.06
Ramachandran plot[e]			
favored regions [%]	85.4	85.7	85.9
additional allowed reg. [%]	14.2	13.9	14.1
generously allowed reg. [%]	0.4	0.4	0.0
Mean *B*-factor [$Å^2$][f]			
Protein	34.7	21.2	15.7
Inhibitor	35.2	30.1	12.1
Water molecules	35.3	32.4	28.7
Other inhibitors	52.0	35.5	32.0

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-<I>|}{\Sigma|I|} \cdot 100$ for which I = observed intensity and $<I>$ = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Table 19: X-ray data collection and statistics for inhibitors **5.8** and **5.9** in complex with thrombin wildtype

	5.8 (6YH9)	**5.9** (6YB6)
(A) Data collection and processing		
Beamline	Bessy 14.1	Elettra 5.2R
Wavelength [Å]	0.91841	1.00000
Space group	C121	C121
Unit Cell parameters:		
a, *b*, *c* [Å]	70.3, 71.3, 72.5	70.4, 71.2, 72.3
α, *β*, *γ* [°]	90.0, 100.5, 90.0	90.0, 100.2, 90.0
Matthews coef. [$Å^3 \cdot Da^{-1}$]	2.5	2.5
Solvent content [%]	51	50
(B) Diffraction Data [a]		
Resolution range [Å]	49.64 – 1.48 (1.57 – 1.48)	35.59 – 1.33 (1.41 – 1.33)
Unique reflections	57330 (9094)	79740 (12752)
$R(I)_{sym}$ [%][b]	7.0 (46.8)	5.9 (58.5)
Completeness [%]	97.5 (95.9)	99.0 (98.3)
Redundancy	3.1 (3.0)	3.4 (3.1)
$I/\sigma(I)$	10.5 (2.3)	11.4 (2.1)
(C) Refinement		
Resolution range [Å]	49.67 – 1.48	35.60 – 1.33
Reflections used in refinement		
work	54463	75753
free	2867	3984
Final R values for all Reflections		
work [%][c]	13.8	13.1
free [%][d]	17.0	15.3
Number of protein residues		
L-chain	28	28
H-chain	250	250
hirudin	11	11
Sodium ions	2	2
Inhibitor atoms	60	29
Water molecules	262	249
Other inhibitor atoms	30	45
RMSD bonds		
Bond length [Å]	0.008	0.010
Bond angles [°]	1.00	1.17
Ramachandran plot[e]		
favored regions [%]	86.1	86.5
additional allowed reg. [%]	13.5	13.5
generously allowed reg. [%]	0.4	0.0
Mean *B*-factor [$Å^2$][f]		
Protein	19.9	20.5
Inhibitor	26.6	24.0
Water molecules	32.3	32.9
Other inhibitors	34.6	37.8

[a] values in parenthesis are statistics for the highest resolution shell. [b] $R(I)_{sym} = \frac{\Sigma|I-<I>|}{\Sigma|I|} \cdot 100$ for which *I* = observed intensity and <*I*> = statistically weighted average intensity of multiple observations. [c] Calculated by MOLEMAN.[183] [d] R_{free} = same definition as for R_{work} for a cross validation set of ≈ 5% of the reflections. [e] Calculated by PROCHECK.[184] [f] $R_{work} = \frac{\Sigma|F_O-F_C|}{\Sigma|F_O|} \cdot 100$ for which F_O = observed structure factor amplitudes and F_C = calculated structure factor amplitudes.

Bibliography

1. Lobanovska, M. & Pilla, G. Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? *Yale J. Biol. Med.* **90**, 135–145, 2017.
2. Klebe, G. *Wirkstoffdesign*. (Springer, Berlin, Heidelberg, Germany, 2009).
3. Bohacek, R. S., McMartin, C. & Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50, 1996.
4. Fink, T. & Reymond, J.-L. 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 Discov. *J. Chem. Inf. Model.* **47**, 342–353, 2007.
5. Steuber, H. Structural and Thermodynamic Characterization of Inhibitor Binding to Aldose Reductase: Insights into Binding Modes, Driving Forces, and Selectivity Determinants. (PhD thesis, Philipps-Universität Marburg, Germany, 2007).
6. Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 4046–4050, 1981.
7. Congreve, M., Chessari, G., Tisi, D. & Woodhead, A. J. Recent Developments in Fragment-Based Drug Discovery. *J. Med. Chem.* **51**, 3661–3680, 2008.
8. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **1**, 187–192, 2009.
9. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discov. Today* **8**, 876–877, 2003.
10. Gleeson, M. P. Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. *J. Med. Chem.* **51**, 817–834, 2008.
11. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333–408, 2013.
12. Harreiter, J. & Roden, M. Diabetes mellitus - Definition, classification, diagnosis, screening and prevention (Update 2019). *Wiener Klin. Wochenschr.* **131**, 6–15, 2019.
13. Scheidt-Nave, C., Du, Y., Paprott, R., Kroll, L., Lampert, T. & Heidemann, C. Soziale Ungleichheit und Diabetes mellitus - zeitliche Entwicklung bei Erwachsenen in Deutschland. *Das Gesundheitswes.* **78**, 12–30, 2016.
14. Kerner, W., Brückel, J. & Böhm, B. O. Definition, Klassifikation und Diagnostik des Diabetes mellitus - Evidenzbasierte Leitlinie DDG, *Diabetologie und Stoffwechsel* 1–11, Berlin, Germany, 2004.

15. Unwin, N., Gan, D., Mbanya, J. C., Ramachandran, A., Roglic, G., Shaw, J., Soltèsz, G., Whiting, D., Zgibor, J., Zhang, P. & Zimmet, P. *IDF Diabetes Atlas*. *International Diabetes Federation* (International Diabetes Federation, Brüssel, Belgium, 2009).

16. Hauner, H., Schwarz, P., Landgraf, R., Köhler, D., Hoffmann, R., Siegel, E., Sturm, D., Kulzer, B., Drobinski, E., Biermann, E., Tschöpe, D., Diehm, C., Lawall, H., Pommer, W., Hammes, H.-P., Ziegler, D., Danne, T. & Holl, R. *Deutschter Gesundheitsbericht Diabetes*. (DiabetesDE, Berlin, Germany, 2010).

17. Krolewski, A. S., Warram, J. H., Rand, L. I., Christlieb, A. R., Busick, E. J. & Kahn, C. R. Risk of proliferative diabetic retinopathy in juvenile-onset type I diabetes: A 40-yr follow-up study. *Diabetes Care* **9**, 443–452, 1986.

18. Pirart, J. Diabetes mellitus and its degenerative complications: A prospective study of 4,400 patients observed between 1947 and 1973. *Diabetes Care* **1**, 168–188, 1978.

19. Deckert, T., Poulsen, J. E. & Larsen, M. Prognosis of diabetics with diabetes onset before the age of thirtyone. *Diabetologia* **14**, 363–370, 1978.

20. Andersen, A. R., Christiansen, J. S., Andersen, J. K., Kreiner, S. & Deckert, T. Diabetic nephropathy in type 1 (insulin-dependent) diabetes: An epidemiological study. *Diabetologia* **25**, 496–501, 1983.

21. Ballard, D. J., Humphrey, L. L., Melton, L. J., Frohnert, P. P., Chu, C.-P., Fallon, W. M. & Palumbo, P. J. Epidemiology of Persistent Proteinuria in Type II Diabetes Mellitus: Population-Based Study in Rochester, Minnesota. *Diabetes* **37**, 405–412, 1988.

22. Gordon, T., Castelli, W. P., Hjortland, M. C., Kannel, W. B. & Dawber, T. R. Diabetes, Blood Lipids, and the Role of Obesity in Coronary Heart Disease Risk for Women. *Ann. Intern. Med.* **87**, 393–397, 1977.

23. Greene, D. A., Lattimer, S. A. & Sima, A. A. F. Sorbitol, Phosphoinositides, and Sodium-Potassium-ATPase in the Pathogenesis of Diabetic Complications. *N. Engl. J. Med.* **316**, 599–606, 1987.

24. Hostetter, T. H., Rennke, H. G. & Brenner, B. M. The case for intrarenal hypertension in the initiation and progression of diabetic and other glomerulopathies. *Am. J. Med.* **72**, 375–380, 1982.

25. Colwell, J. A., Winocour, P. D. & Halushka, P. V. Do Platelets Have Anything To Do With Diabetic Microvascular Disease? *Diabetes* **32**, 14–19, 1983.

26. Merimee, T. J., Zapf, J. & Froesch, E. R. Insulin-like growth factors. Studies in diabetics with and without retinopathy. *N. Engl. J. Med.* **309**, 527–530, 1983.

27. Dills, D. G., Moss, S. E., Klein, R., Klein, B. E. K. & Davis, M. Is Insulinlike Growth Factor I Associated With Diabetic Retinopathy? *Diabetes* **39**, 191–195, 1990.

28. Gabbay, K. H., Merola, L. O. & Field, R. A. Sorbitol pathway: presence in nerve and cord with substrate accumulation in diabetes. *Science* **151**, 209–210, 1966.

29. Brownlee, M. The pathobiology of diabetic complications: A unifying mechanism. *Diabetes* **54**, 1615–1625, 2005.

30. Lorenzi, M. The polyol pathway as a mechanism for diabetic retinopathy: Attractive, elusive, and resilient. *Exp. Diabetes Research* 1–10, 2007.

31. Brownlee, M. Biochemistry and molecular cell biology of diabetic complications. *Nature* **414**, 813–820, 2001.

32. Barski, O. A., Tipparaju, S. M. & Bhatnagar, A. The Aldo-Keto Reductase Superfamily and its Role in Drug Metabolism and Detoxification. *Drug Metab Rev* **40**, 553–624, 2009.

33. Carper, D. A., Wistow, G., Nishimura, C., Graham, C., Watanabe, K., Fujii, Y., Hayashi, H. & Hayaishi, O. A superfamily of NADPH-dependent reductases in eukaryotes and prokaryotes. *Exp. Eye Res.* **49**, 377–388, 1989.

34. Borhani, D. W., Harter, T. M. & Petrash, J. M. The crystal structure of the aldose reductase·NADPH binary complex. *J. Biol. Chem.* **267**, 24841–24847, 1992.

35. Biadene, M., Hazemann, I., Cousido, A., Ginell, S., Joachimiak, A., Sheldrick, G. M., Podjarny, A. & Schneider, T. R. The atomic resolution structure of human aldose reductase reveals that rearrangement of a bound ligand allows the opening of the safety-belt loop. *Acta Crystallogr. Sect. D* **63**, 665–672, 2007.

36. Bohren, K. M., Brownlee, J. M., Milne, A. C., Gabbay, K. H. & Harrison, D. H. T. The structure of Apo R268A human aldose reductase: hinges and latches that control the kinetic mechanism. *Biochim. Biophys. Acta* **1748**, 201–212, 2005.

37. Rechlin, C. Insights into Protein-Ligand Molecular Recognition: Thermodynamic, Kinetic and Structural Characterization of Inhibitor Binding to Aldose Reductase and Carbonic Anhydrase II. (PhD thesis, Philipps-Universität Marburg, Germany, 2015).

38. Bohren, K. M., Grimshaw, C. E., Lai, C. J., Gabbay, K. H., Harrison, D. H., Ringe, D. & Petsko, G. A. Tyrosine-48 Is the Proton Donor and Histidine-110 Directs Substrate Stereochemical Selectivity in the Reduction Reaction of Human Aldose Reductase: Enzyme Kinetics and Crystal Structure of the Y48H Mutant Enzyme. *Biochemistry* **33**, 2021–2032, 1994.

39. Grimshaw, C. E., Bohren, K. M., Lai, C. J. & Gabbay, K. H. Human Aldose Reductase: Pk of Tyrosine 48 Reveals the Preferred Ionization State for Catalysis and Inhibition. *Biochemistry* **34**, 14374–14384, 1995.

40. Petrash, J. M. All in the family: Aldose reductase and closely related aldo-keto reductases. *Cell. Mol. Life Sci.* **61**, 737–749, 2004.

41. Wierenga, R. K. The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett.* **492**, 193–198, 2001.

42. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501–4523, 2002.

43. Jain, M., Yadav, A., Tomar, V., Kannojia, P., Solanki, S. & Tonpay, S. Possible pharmacological basis for antithrombotic effect of Glycyrrhiza glabra in Sprague Dawley rats. *Int. J. Pharm. Sci. Drug Res.* **1**, 113–115, 2009.

44. Di Cera, E., Dang, Q. D. & Ayala, Y. M. Molecular mechanisms of thrombin function. *Cell. Mol. Life Sci.* **53**, 701–730, 1997.

45. Immekus, F. Klonierung , Expression und positionsgerichtete Mutation von humanem Thrombin. (Diploma thesis, Martin-Luther-Universität Halle-Wittenberg, Germany, 2010).

46. Goldsack, N. R., Chambers, R. C., Dabbagh, K. & Laurent, G. J. Molecules in focus Thrombin. *Int. J. Biochem. Cell Biol.* **30**, 641–646, 1989.

47. Sabir, I., Khavandi, K., Brownrigg, J. & Camm, A. J. Oral anticoagulants for Asian patients with atrial fibrillation. *Nat. Rev. Cardiol.* **11**, 290–303, 2014.

48. Mutschler, E. *Arzneimittelwirkungen*. (Wissenschaftliche Verlagsgesellschaft Stuttgart, Germany, 2013).

49. Esmon, C. T. The role of protein C and thrombomodulin in the regulation of blood coagulation. *Blood Vessel* **15**, 607–615, 1984.

50. O'Brien, P. J. & Mureebe, L. Direct Thrombin Inhibitors. *J. Cardiovasc. Pharmacol. Ther.* **17**, 5–11, 2011.

51. Gurm, H. S. & Bhatt, D. L. Thrombin, an ideal target for pharmacological inhibition: A review of direct thrombin inhibitors. *Am. Heart J.* **149**, S43–S53, 2005.

52. Lee, C. J. & Ansell, J. E. Direct thrombin inhibitors. *Br. J. Clin. Pharmacol.* **72**, 581–592, 2011.

53. Alban, S. Pharmakologie der Heparine und der direkten Antikoagulanzien. *Hamostaseologie* **28**, 400–420, 2008.

54. Huntington, J. A. & Baglin, T. P. Targeting thrombin – rational drug design from natural mechanisms. *TRENDS Pharmacol. Sci.* **24**, 589–595, 2003.

55. Clement, B. & Lopian, K. Characterization of in vitro biotransformation of new, orally active, direct thrombin inhibitor ximelagatran, an amidoxime and ester prodrug. *Drug Metab. Dispos.* **31**, 645–651, 2003.

56. Pötzsch, B. & Madlener, K. *Hämostaseologie*. (Springer, Berlin, Heidelberg, Germany, 2010).

57. Dang, Q. D., Sabetta, M. & Cera, E. Di. Selective Loss of Fibrinogen Clotting in a Loop-less Thrombin, *J. Biol. Chem.* **272**, 19649–19652, 1997.

58. Kühne, W. Ueber das Trypsin (Enzym des Pankreas), *Verhandlungen des Naturhistorisch-Medizinischen Vereins* 194–198, 1876.

59. Fink, B., Fischer, J., Pecoraro, C., Ostendorf, N. & Weiß, A. DocCheck Flexikon. https://flexikon.doccheck.com/de/Trypsin, 2020.

60. Stroud, R. M., Kay, L. M. & Dickerson, R. E. The structure of bovine trypsin: Electron density maps of the inhibited enzyme at 5 Å and at 2·7 Å resolution. *J. Mol. Biol.* **83**, 185–208, 1974.

61. Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515, 2019.

62. Rawlings, N. D. & Barrett, A. J. Families of Serine Peptidases. *Methodes Enzymol.* **244**, 19–61, 1994.

63. John E. Ladbury, Klebe, G. & Freire, E. Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat. Rev. Drug Discov.* **9**, 23–27, 2010.

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

65. Tanford, C. The hydrophobic effect and the organization of living matter. *Science* **200**, 1012–1018, 1978.

66. Wang, H. W. Cryo-electron microscopy for structural biology: current status and future perspectives. *Sci. China Life Sci.* **58**, 750–756, 2015.

67. Doerr, A. Single-particle cryo-electron microscopy. *Nat. Methods* **13**, 23, 2016.

68. Massa, W. *Kristallstrukturbestimmung*, 7. Auflage, Vieweg+Teubner Verlag / Springer Fachmedien Wiesbaden GmbH, Wiesbaden, Germany, 2011

69. Glusker, J. P., Lewis, M. & Rossi, M. *Crystal Structure Analysis for Chemists and Biologists*. (VCH Publishers, Inc, 1994).

70. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 899–907, 2002.

71. Bahar, I., Atilgan, A. R. & Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **2**, 173–181, 1997.

72. Drenth, J. *Principles of Protein X-Ray Crystallography*. (Springer-Verlag).

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

74. Laskowski, R. A., Luscombe, N. M., Swindells, M. B. & Thornton, J. M. Protein clefts in molecular recognition and function. *Protein Sci.* **5**, 2438–2452, 1996.

75. Shah, A., Shinde, R., Kare, P., Hymavathi, V., Chavan, S. & Elizabeth Sobhia, M. Induced fit binding of aldose reductase inhibitors to AKR1B10. *Med. Chem. Res.* **21**, 1245–1252, 2012.

76. Terwesten, F. Proteindynamik Flexibilität in Zielproteinen des strukturbasierten Wirkstoffdesigns. (PhD thesis, Philipps-Universität Marburg, Germany, 2017).

77. Rechlin, C., Scheer, F., Terwesten, F., Wulsdorf, T., Pol, E., Fridh, V., Toth, P., Diederich, W. E., Heine, A. & Klebe, G. Price for Opening the Transient Specificity Pocket in Human Aldose Reductase upon Ligand Binding: Structural, Thermodynamic, Kinetic, and Computational Analysis. *ACS Chem. Biol.* **12**, 1397–1415, 2017.

78. Acker, T. M., Gable, J. E., Bohn, M.-F., Jaishankar, P., Thompson, M. C., Fraser, J. S., Renslo, A. R. & Craik, C. S. Allosteric Inhibitors, Crystallography, and Comparative Analysis Reveal Network of Coordinated Movement across Human Herpesvirus Proteases. *J Am Chem Soc* **139**, 11650–11653, 2017.

79. Bowman, G. R. & Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci.* **109**, 11681–11686, 2012.

80. Bohren, K. M., Bullock, B., Wermuth, B. & Gabbay, K. H. The aldo-keto reductase superfamily. cDNAs and deduced amino acid sequences of human aldehyde and aldose reductases. *J. Biol. Chem.* **264**, 9547–9561, 1989.

81. Sangshetti, J. N., Chouthe, R. S., Sakle, N. S., Gonjari, I. & Shinde, D. B. Aldose Reductase: A Multi-Disease Target. *Curr. Enzym. Inhib.* **10**, 2–12, 2014.

82. Steuber, H., Heine, A., Podjarny, A. & Klebe, G. Merging the Binding Sites of Aldose and Aldehyde Reductase for Detection of Inhibitor Selectivity-determining Features. *J. Mol. Biol.* **379**, 991–1016, 2008.

83. Van Zandt, M. C., Sibley, E. O., McCann, E. E., Combs, K. J., Flam, B., Sawicki, D. R., Sabetta, A., Carrington, A., Sredy, J., Howard, E., Mitschler, A. & Podjarny, A. D. Design and synthesis of highly potent and selective (2-arylcarbamoyl- phenoxy)-acetic acid inhibitors of aldose reductase for treatment of chronic diabetic complications. *Bioorganic Med. Chem.* **12**, 5661–5675, 2004.

84. Broecker, J., Vargas, C. & Keller, S. Revisiting the optimal c-value for isothermal titration calorimetry. *Anal. Biochem.* **418**, 307–309, 2011.

85. Tellinghuisen, J. Isothermal titration calorimetry at very low c. *Anal. Biochem.* **373**, 395–397, 2008.

86. Turnbull, W. B. & Daranas, A. H. On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **125**, 14859–14866, 2003.

87. Steuber, H., Czodrowski, P., Sotriffer, C. A. & Klebe, G. Tracing Changes in Protonation: A Prerequisite to Factorize Thermodynamic Data of Inhibitor Binding to Aldose Reductase. *J. Mol. Biol.* **373**, 1305–1320, 2007.

88. Steuber, H., Heine, A. & Klebe, G. Structural and Thermodynamic Study on Aldose Reductase: Nitro-substituted Inhibitors with Strong Enthalpic Binding Contribution. *J. Mol. Biol.* **368**, 618–638, 2007.

89. Krämer, O. Theoretische und experimentelle Untersuchungen zur Inhibition des Enzyms Aldose Reduktase. (PhD thesis, Philipps-Universität Marburg, Germany, 2003).

90. Misuri, L., Cappiello, M., Balestri, F., Moschini, R., Barracco, V., Mura, U. & Del-Corso, A. The use of dimethylsulfoxide as a solvent in enzyme inhibition studies: the case of aldose reductase. *J. Enzyme Inhib. Med. Chem.* **32**, 1152–1158, 2017.

91. Ruiz, F., Mitschler, A., Joachimiak, A., Karplus, M. & Podjarny, A. The crystallographic structure of the aldose reductase - IDD552 complex shows direct proton donation from tyrosine 48 research papers. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D60**, 1347–1354, 2004.

92. Damm, K. L. & Carlson, H. A. Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures. *Biophys. J.* **90**, 4558–4573, 2006.

93. Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol. Biol.* **857**, 231–257, 2012.

94. McLachlan, A. D. Rapid comparison of protein structures. *Acta Crystallogr. Sect. A* **38**, 871–873, 1982.

95. Martin, A. C. ProFit. http://www.bioinf.org.uk/software/profit/, 2020.

96. Kheirjou, S., Abedin, A. & Fattahi, A. Theoretical descriptors response to the calculations of the relative pKa values of some boronic acids in aqueous solution: A DFT study. *Comput. Theor. Chem.* **1000**, 1–5, 2012.

97. Jezuita, A., Ejsmont, K. & Szatylowicz, H. Substituent effects of nitro group in cyclic compounds. *Struct. Chem.* **32**, 179–203, 2021.

98. Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J. & Friesner, R. A. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **113**, 2110–2142, 2013.

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

100. Benfield, A. P., Teresk, M. G., Plake, H. R., Delorbe, J. E., Millspaugh, L. E. & Martin, S. F. Ligand Preorganization May Be Accompanied by Entropic Penalties in Protein–Ligand Interactions. *Angew. Chemie - Int. Ed.* **45**, 6830–6835, 2006.

101. DeLorbe, J. E., Clements, J. H., Teresk, M. G., Benfield, A. P., Plake, H. R., Millspaugh, L. E. & Martin, S. F. Thermodynamic and structural effects of conformational constraints in protein-ligand interactions. Entropic paradoxy associated with ligand preorganization. *J. Am. Chem. Soc.* **131**, 16758–16770, 2009.

102. Morgan, B. P., Bartlett, P. A., Holland, D. R. & Matthews, B. W. Structure-Based Design of an Inhibitor of the Zinc Peptidase Thermolysin. *J. Am. Chem. Soc.* **116**, 3251–3260, 1994.

103. Wienen-Schmidt, B., Jonker, H. R. A., Wulsdorf, T., Gerber, H. D., Saxena, K., Kudlinzki, D., Sreeramulu, S., Parigi, G., Luchinat, C., Heine, A., Schwalbe, H. & Klebe, G. Paradoxically, Most Flexible Ligand Binds Most Entropy-Favored: Intriguing Impact of Ligand Flexibility and Solvation on Drug-Kinase Binding. *J. Med. Chem.* **61**, 5922–5933, 2018.

104. Delorbe, J. E., Clements, J. H., Whiddon, B. B. & Martin, S. F. Thermodynamic and structural effects of macrocyclic constraints in protein-ligand interactions. *ACS Med. Chem. Lett.* **1**, 448–452, 2010.

105. Schwizer, D., Patton, J. T., Cutting, B., Smieško, M., Wagner, B., Kato, A., Weckerle, C., Binder, F. P. C., Rabbani, S., Schwardt, O., Magnani, J. L. & Ernst, B. Pre-organization of the core structure of E-selectin antagonists. *Chem. - A Eur. J.* **18**, 1342–1351, 2012.

106. von Nussbaum, F., Li, V. M-J, Allerheiligen, S., Anlauf, S. Bärfacker, L. Bechem, M., Delbeck, M., Fitzgerald M. F., Gerisch, M. Gielen-Haertwig, H., Haning, H. Karthaus, D., Lang, D., Lustig, K., Meibom, D., Mittendorf, J., Rosentreter, U., Schäfer, M., Schäfer, S., Schamberger, J., Telan, L. A. & Tersteegen, A. Freezing the Bioactive Conformation to Boost Potency: The Identification of BAY85-8501, a Selective and Potent Inhibitor of Human Neutrophil Elastase for Pulmonary Diseases. *ChemMedChem* **10**, 1163–1173, 2015.

107. Riel, A. M. S., Decato, D. A., Sun, J., Massena, C. J., Jessop, M. J. & Berryman, O. B. The intramolecular hydrogen bonded-halogen bond: A new strategy for preorganization and enhanced binding. *Chem. Sci.* **9**, 5828–5836, 2018.

108. McConnell, A. J. & Beer, P. D. Heteroditopic receptors for ion-pair recognition. *Angew. Chemie - Int. Ed.* **51**, 5052–5061, 2012.

109. Mc Kim, S. K. & Sessler, J. L. Ion pair receptors. *Chem. Soc. Rev.* **39**, 3784–3809, 2010.

110. Hua, Y., Ramabhadran, R. O., Karty, J. A., Raghavachari, K. & Flood, A. H. Two levels of conformational pre-organization consolidate strong CH hydrogen bonds in chloride-triazolophane complexes. *Chem. Commun.* **47**, 5979–5981, 2011.

111. Molina, P., Zapata, F. & Caballero, A. Anion Recognition Strategies Based on Combined Noncovalent Interactions. *Chem. Rev.* **117**, 9907–9972, 2017.

112. Clardy, J. & Walsh, C. Lessons from natural forests. *insight Rev. Artic.* **6**, 829–837, 2004.

113. Bode, W. The structure of thrombin, a chameleon-like proteinase. *J. Thromb. Haemost.* **3**, 2379–2388, 2005.

114. Gurm, H. S. & Bhatt, D. L. Thrombin , an ideal target for pharmacological inhibition: A review of direct thrombin inhibitors. *Am. Heart J.* **149**, 43–53, 2005.

115. Steinmetzer, T. & Stürzebecher, J. Progress in the Development of Synthetic Thrombin Inhibitors as New Orally Active Anticoagulants. *Curr. Med. Chem.* **11**, 2297–2321, 2004.

116. Gerlach, C., Münzel, M., Baum, B., Gerber, H. D., Craan, T., Diederich, W. E. & Klebe, G. KNOBLE: A knowledge-based approach for the design and synthesis of readily accessible small-molecule chemical probes to test protein binding. *Angew. Chemie - Int. Ed.* **46**, 9105–9109, 2007.

117. Menear, K. Direct thrombin inhibitors: current status and future prospects. *Expert Opin. Investig. Drugs* **8**, 1373–1384, 1999.

118. Baum, B., Mohamed, M., Zayed, M., Gerlach, C., Heine, A., Hangauer, D. & Klebe, G. 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, 2009.

119. Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162, 1967.

120. Sandner, A., Hüfner-Wulsdorf, T., Heine, A., Steinmetzer, T. & Klebe, G. Strategies for Late-Stage Optimization: Profiling Thermodynamics by Preorganization and Salt Bridge Shielding. *J. Med. Chem.* **62**, 9753–9771, 2019.

121. Stubbs, M. T., Oschkinat, H., Mayr, I., Huber, R., Angliker, H., Stone, S. R. & Bode, W. The interaction of thrombin with fibrinogen-A structural basis for its specificity. *Eur. J. Biochem.* **206**, 187–195, 1992.

122. Lumma, W. C., Witherup, K. M., Tucker, T. J., Brady, S. F., Sisko, J. T., Naylor-Olsen, A. M., Dale Lewis, S., Lucas, B. J. & Vacca, J. P. Design of novel, potent, noncovalent inhibitors of thrombin with nonbasic P1 substructures: Rapid structure-activity studies by solid-phase synthesis. *J. Med. Chem.* **41**, 1011–1013, 1998.

123. Tucker, T. J., Lumma, W. C., Mulichak, A. M., Chen, Z., Naylor-Olsen, A. M., Lewis, S. D., Lucas, R., Freidinger, R. M. & Kuo, L. C. Design of highly potent noncovalent thrombin inhibitors that utilize a novel lipophilic binding pocket in the thrombin active site. *J. Med. Chem.* **40**, 830–832, 1997.

124. Ngo, K. Synthese von D-Phe und D-DiPhe-Pro-basierenden Inhibitoren zur thermodynamischen Untersuchung der Bindungseigenschaften in der S1-Tasche von Thrombin. (PhD thesis, Philipps-Universität Marburg, Germany, 2016).

125. Baum, B., Muley, L., Heine, A., Smolinski, M., Hangauer, D. & Klebe, G. Think Twice: Understanding the High Potency of Bis(phenyl)methane Inhibitors of Thrombin. *J. Mol. Biol.* **391**, 552–564, 2009.

126. Matter, H., Nazaŕ, M., Güssregen, S., Will, D. W., Schreuder, H., Bauer, A., Urmann, M., Ritter, K., Wagner, M. & Wehner, V. Evidence for C-Cl/C-Br···π interactions as an important contribution to protein-ligand binding affinity. *Angew. Chemie - Int. Ed.* **48**, 2911–2916, 2009.

127. Marvin. ChemAxon Package, Budapest, Hungary, 2019.

128. Rühmann, E., Betz, M., Heine, A. & Klebe, G. Fragment Binding Can Be Either More Enthalpy-Driven or Entropy-Driven: Crystal Structures and Residual Hydration Patterns Suggest Why. *J. Med. Chem.* **58**, 6960–6971, 2015.

129. Krimmer, S. G., Cramer, J., Betz, M., Fridh, V., Karlsson, R., Heine, A. & Klebe, G. Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands. *J. Med. Chem.* **59**, 10530–10548, 2016.

130. Witek, J., Mühlbauer, M., Keller, B. G., Blatter, M., Meissner, A., Wagner, T. & Riniker, S. Interconversion Rates between Conformational States as Rationale for the Membrane Permeability of Cyclosporines. *ChemPhysChem* **18**, 3309–3314, 2017.

131. Witek, J., Wang, S., Schroeder, B., Lingwood, R., Dounas, A., Roth, H.-J., Fouché, M., Blatter, M., Lemke, O., Keller, B. & Riniker, S. Rationalization of the Membrane Permeability Differences in a Series of Analogue Cyclic Decapeptides. *J. Chem. Inf. Model.* **59**, 294–308, 2019.

132. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **16**, 1339–1350, 1994.

133. Grossfield, A. WHAM: an implementation of the weighted histogram analysis method Version 2.0.9. http://membrane.urmc.rochester.edu/sites/default/files/wham/doc.html, 2019.

134. Rittle, K. E., Barrow, J. C., Cutrona, K. J., Glass, K. L., Krueger, J. A., Kuo, L. C., Lewis, S. D., Lucas, B. J., McMasters, D. R., Morrissette, M. M., Nantermet, P. G., Newton, C. L., Sanders, W. M., Yan, Y., Vacca, J. P. & Selnick, H. G. Unexpected enhancement of thrombin inhibitor potency with o-aminoalkylbenzylamides in the P1 position. *Bioorganic Med. Chem. Lett.* **13**, 3477–3482, 2003.

135. Schiebel, J., Gaspari, R., Wulsdorf, T., Ngo, K., Sohn, C., Schrader, T. E., Cavalli, A., Ostermann, A., Heine, A. & Klebe, G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat. Commun.* **9**, 1–15, 2018.

136. Fennell, C. J., Wymer, K. L. & Mobley, D. L. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *J. Phys. Chem. B* **118**, 6438–6446, 2014.

137. Mobley, D. L., Dumont, É., Chodera, J. D. & Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **111**, 2242–2254, 2007.

138. Cramer, J., Krimmer, S. G., Heine, A. & Klebe, G. Paying the Price of Desolvation in Solvent-Exposed Protein Pockets: Impact of Distal Solubilizing Groups on A ffi nity and Binding Thermodynamics in a Series of Thermolysin Inhibitors. *J. Med. Chem.* **60**, 5791–5799, 2017.

139. Neeb, M., Hohn, C., Rainer, F., Härtsch, A., Heine, A., Diederich, F. & Klebe, G. Occupying a flat subpocket in a tRNA-modifying enzyme with ordered or disordered side chains : Favorable or unfavorable for binding ? *Bioorg. Med. Chem.* **24**, 4900–4910, 2016.

140. Schmidt, A., Jelsch, C., Østergaard, P., Rypniewski, W. & Lamzin, V. S. Trypsin Revisited: Crystallography at (SUB) atomic resolution and quantum chemistry revealing details of catalysis. *J. Biol. Chem.* **278**, 43357–43362, 2003.

141. Sanderson, P. E. J. Small, noncovalent serine protease inhibitors. *Med. Res. Rev.* **19**, 179–197, 1999.

142. Straub, A., Roehrig, S. & Hillisch, A. Oral, direct thrombin and factor Xa inhibitors: the replacement for warfarin, leeches, and pig intestines? *Angew. Chem. Int. Ed. Engl.* **50**, 4574–4590, 2011.

143. Metheny, N. A., Stewart, B. J., Smith, L., Yan, H., Diebold, M. & Clouse, R. E. pH and Concentrations of Pepsin and Trypsin in Feeding Tube Aspirates as Predictors of Tube Placement. *J. Parenter. Enter. Nutr.* **21**, 279–285, 1997.

144. Gustafsson, D., Antonsson, T., Bylund, R., Eriksson, U., Gyzander, E., Nilsson, I., Elg, M., Mattsson, C., Deinum, J., Pehrsson, S., Karlsson, O., Nilsson, A. & Sörensen, H. Effects of melagatran, a new low-molecular-weight thrombin inhibitor, on thrombin and fibrinolytic enzymes. *Thromb. Haemost.* **79**, 110–118, 1998.

145. Czodrowski, P., Sotriffer, C. A. & Klebe, G. Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pKa Calculations and ITC Experiments. *J. Mol. Biol.* **367**, 1347–1356, 2007.

146. Katz, B. A., Sprengeler, P. A., Luong, C., Verner, E., Elrod, K., Kirtley, M., Janc, J., Spencer, J. R., Breitenbucher, J. G., Hui, H., McGee, D., Allen, D., Martelli, A. & Mackman, R. L. Engineering inhibitors highly selective for the S1 sites of Ser190 trypsin-like serine protease drug targets. *Chem. Biol.* **8**, 1107–1121, 2001.

147. Hilpert, K., Ackermann, J., Banner, D. W., Gast, A., Gubernator, K., Hadvary, P., Labler, L., Mueller, K., Schmid, G., Tschopp, T. B. & Waterbeemd, H. van de. Design and Synthesis of Potent and Highly Selective Thrombin Inhibitors. *J. Med. Chem* **37**, 3889–3901, 1994.

148. Ngo, K., Collins-Kautz, C., Gerstenecker, S., Wagner, B., Heine, A. & Klebe, G. Protein-Induced Change in Ligand Protonation during Trypsin and Thrombin Binding: Hint on Differences in Selectivity Determinants of Both Proteins? *J. Med. Chem.* **63**, 3274–3289, 2020.

149. Castañeda, C. A., Fitch, C. A., Majumdar, A., Khangulov, V., Schlessman, J. L. & García-Moreno, B. E. Molecular determinants of the pK a values of Asp and Glu residues in staphylococcal nuclease. *Proteins Struct. Funct. Bioinforma.* **77**, 570–588, 2009.

150. Hubbard, R. E. & Haider, M. K. Hydrogen Bonds in Proteins: Role and Strength. *ELS,* John Wiley & Sons, Ltd: Chichester, Hoboken, New Jersey, 2010

151. Hüfner-Wulsdorf, T. & Klebe, G. Role of Water Molecules in Protein–Ligand Dissociation and Selectivity Discrimination: Analysis of the Mechanisms and Kinetics of Biomolecular Solvation Using Molecular Dynamics. *J. Chem. Inf. Model.* **60**, 1818–1832, 2020.

152. Ahmed, H. U., Blakeley, M. P., Cianci, M., Cruickshank, D. W. J., Hubbard, J. A. & Helliwell, J. R. The Determination of Protonation States in Proteins. *Acta Crystallogr. Sect. D Struct. Biol.* **63**, 906–922, 2007.

153. Adams, P. D., Mustyakimov, M., Afonine, P. V. & Langan, P. Generalized X-ray and neutron crystallographic analysis: More accurate and complete structures for biological macromolecules. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 567–573, 2009.

154. Schiebel, J., Gaspari, R., Sandner, A., Ngo, K., Gerber, H. D., Cavalli, A., Ostermann, A., Heine, A. & Klebe, G. Charges Shift Protonation: Neutron Diffraction Reveals that Aniline and 2-Aminopyridine Become Protonated Upon Binding to Trypsin. *Angew. Chemie - Int. Ed.* **56**, 4887–4890, 2017.

155. Hall, H. K. Correlation of the Base Strengths of Amines. *J. Am. Chem. Soc.* **79**, 5441–5444, 1957.

156. Nguyen, C. N., Kurtzman Young, T. & Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **137**, 973–980, 2012.

157. Nguyen, C., Gilson, M. K. & Young, T. Structure and Thermodynamics of Molecular Hydration via Grid Inhomogeneous Solvation Theory. *arXiv* 1–16, 2011.

158. Ramsey, S., Nguyen, C., Salomon-Ferrer, R., Walker, R. C., Gilson, M. K. & Kurtzman, T. Solvation Thermodynamic Mapping of Molecular Surfaces in AmberTools: GIST. *J Comput Chem.* **37**, 2029–2037, 2016.

159. Dullweber, F., Stubbs, M. T., Musil, D., Stürzebecher, J. & Klebe, G. Factorising ligand affinity: A Combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J. Mol. Biol.* **313**, 593–614, 2001.

160. Pineda, A. O., Zhang, E., Guinto, E. R., Savvides, S. N., Tulinsky, A. & Di Cera, E. Crystal structure of the thrombin mutant D221A/D222K: the Asp222:Arg187 ion-pair stabilizes the fast form. *Biophys. Chem.* **112**, 253–256, 2004.

161. Wells, C. M. & Di Cera, E. Thrombin is a Na^+-Activated Enzyme. *Biochemistry* **31**, 11721–11730, 1992.

162. Sandner, A., Ngo, K., Schiebel, J., Pizarroso, A. I. M., Schmidt, L., Wenzel, B., Steinmetzer, T., Ostermann, A., Heine, A. & Klebe, G. How a Fragment Draws Attention to Selectivity Discriminating Features between the Related Proteases Trypsin and Thrombin. *J. Med. Chem.* **64**, 1611–1625, 2021.

163. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. The Proteomics Protocols Handbook. *Proteomics Protoc. Handb.* 571–608, 2005.

164. Lamour, V., Barth, P., Rogniaux, H., Poterszman, A., Howard, E., Mitschler, A., Van Dorsselaer, A., Podjarny, A. & Moras, D. Production of Crystals of Human Aldose Reductase With Very High Resolution Diffraction. *Acta Crystallogr. Sect. D Struct. Biol.* **55**, 721-723, 1999.

165. Koch, C. Towards Improved Aldose Reductase. (PhD thesis, Philipps-Universität Marburg, Germany, 2011).

166. Kabsch, W. Scaling , space-group assignment and post-refinement. *Acta Crystallogr. Sect. D* **66**, 133–144, 2010.

167. Mccoy, A. J., Grosse-kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674, 2007.

168. Collaborative Computational Project. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. Sect. D* **50**, 760–763, 1994.

169. Skordalakes, E., Dodson, G. G., St, D., Green, C., Goodwin, C. A., Scully, M. F., Hudson, H. R., Kakkar, V. V & Deadman, J. J. Inhibition of Human alpha-Thrombin by a Phosphonate Tripeptide Proceeds via a Metastable Pentacoordinated Phosphorus Intermediate. *J. Mol. Biol.* **57**, 549–555, 2001.

170. Adams, P. D., Pavel, V., Chen, V. B., Ian, W., Echols, N., Moriarty, N. W., Read, R. J., Richardson, D. C., Jane, S. & Thomas, C. PHENIX: a comprehensive Python-based system for macromolecular structure solution research papers. *Acta Crystallogr. Sect. D* **66**, 213–221, 2010.

171. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics research papers. *Acta Crystallogr. Sect. D* **60**, 2126–2132, 2004.

172. Smart, O. & Womack, T. Grade Web Server. *Global Phasing* http://grade.globalphasing.org, 2012.

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

174. Bode, W., Mayr, I., Huber, R., Stone, S. R. & Hofsteenge, J. α-thrombin : interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467–3475, 1989.

175. Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A. & Schuck, P. High-Precision Isothermal Titration Calorimetry with Automated Peak Shape Analysis. *Anal. Chem.* **84**, 5066–5073, 2013.

176. Zhao, H., Piszczek, G. & Schuck, P. SEDPHAT – A platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods* **76**, 137–148, 2015.

177. Dr. Bonacker, L. & Schröder, M. CSL Behring, Marburg, Germany.

178. Rühmann, E. Strukturelle und thermodynamische Charakterisierung der Fragment und Liganden Bindung an Thrombin. (PhD thesis, Philipps-Universität Marburg, Germany, 2015).

179. Sigurskjold, B. W. Exact Analysis of Competition Ligand Binding by Displacement Isothermal Titration Calorimetry. *Anal. Biochem.* **277**, 260–266, 2000.

180. Stürzebecher, J., Stürzebecher, U., Vieweg, H., Wagner, G., Hauptmann, J. & Markwardt, F. Synthetic Inhibitors of Bovine Factor Xa and Thrombin Comparison of their Anticoagulant Efficiency. *Thromb. Res.* **54**, 245–252, 1989.

181. Dixon, M. The Graphical Determination of of K_m and K_i. *Biochem. J.* **129**, 197–202, 1972.

182. Grimshaw, C. E., Bohren, K. M., Lai, C. J. & Gabbay, K. H. Human Aldose Reductase: Rate Constants for a Mechanism Including Interconversion of Ternary Complexes by Recombinant Wild-Type Enzyme. *Biochemistry* **34**, 14356–14365, 1995.

183. Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. *International Tables for Crystallography Volume F: Crystallography of biological macromolecules,* John Wiley & Sons, Ltd: Chichester, Hoboken, New Jersey, 2001.

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

Acknowledgements

First and foremost, I would like to thank my PhD supervisor PROF. GERHARD KLEBE for the valuable discussions and for giving me the opportunity to write my dissertation as a member of his research group. Especially the scientific exchange and meeting other scientists / people at conferences at home and abroad has broadened my horizon and shaped me a lot. His persistent encouragement and support were very helpful for me during these years.

Special thanks also goes to PROF. ANDREAS HEINE, for his patience and perseverance with all my questions, not only about crystallography. His confidence in me always provided me with new courage and strength.

I would like to thank especially PROF. TORSTEN STEINMETZER, who spontaneously agreed to be my second reviewer.

I would also like to thank my amazing intern, best bike buddy and roommate MARTIN ARNOLD, as well as the master's students ANGELA ILSE MARCA PIZARROSO, YANG LU, LINDA SCHMIDT and LEA-SOPHIE HUBERT whom I mentored and who were very supportive of my projects.

I am very grateful to PROF. TORSTEN STEINMETZER, DR. SANDRINE MARCHAIS-OBERWINKLER, DR. KHANG NGO, CRISTIAN SOHN and WILLIAM SCANLAN for their help in setting up and pipetting all assays.

I would like to thank DR. CHRISTOPH P. SAGER for the help with the ESP calculations, DR. ALEXANDER METZ, DR. FELIX TERWESTEN, and JOHANNA SENST for proofreading parts of this thesis.

Thank you to all my friends and colleagues from the AG Klebe and AG Diederich labs, who helped me with all kinds of problems. The “Pratikum” and corrections of the 5th and 8th semesters were made unique by you with a lot of jokes. You all created a pleasant working atmosphere and made my time in Marburg unforgettable even away from science. As well I would like to mention DR. NICOLE BERTOLETTI, DR. ALEXANDER METZ, DR. CHRISTOPH P. SAGER, DR. LUKAS HEYDER, DR. ANDREAS NGUYEN, DR. FELIX TERWESTEN, DR. STEFAN MERKL, CHRISTINA BEHRENS, and JOHANNA SENST who always provided me with advice and support.

I am grateful to LYDIA HARTLEBEN for her administrative support and to CHRISTIAN SOHN and STEFFANIE DÖRR for their technical help.

A special mention goes to my longtime life partner and best friend JONAS BECKER for helping me out in many ways. Thank you for all the energy you invested in me, your constant

encouragement, and your admiration over several years. I would never have come this far without you!

Most of all I would like to mention my best friend and partner WILLIAM SCANLAN. Especially in the final stage of my PhD, you were not only emotionally my strongest and most important support in all the nervous breakdowns and desperation moments.

My greatest thanks belong to my family, especially my parents, who have always given me the greatest possible support, in good times and in not so good times. I am very grateful for all the crisis phone calls, for their patience and honesty, and the encouraging hugs. Thank you so much!

Curriculum Vitae

Die Seiten xxxiii – xxxiv (Lebenslauf) enthalten persönliche Daten. Sie sind deshalb nicht Bestandteil der Online-Veröffentlichung.

Eidesstattliche Erklärung

(gemäß §10 der Promotionsordnung)

Ich versichere, dass ich meine Dissertation

"*Optimization of the Drug Design Process by Ligand Preorganization and Transient Binding Pockets Based on the Model Proteins Thrombin and Aldose Reductase*."

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen sind Zitate als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 21.02.2022

(Anna Sophie Sandner)