fakultät für chemie
und chemische biologie

technische universität
dortmund

Investigations of Protein-Ligand Interactions with in vitro and in silico Methods

Dissertation

zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

Technische Universität Dortmund

Fakultät für Chemie und Chemische Biologie

vorgelegt von

Helge Vatheuer

1. Gutachter: Prof. Dr. Paul Czodrowski

2. Gutachter: Prof. Dr. Daniel Rauh

Dortmund, Januar 2025

Danksagung

Zuallererst danke ich Prof. Dr. Paul Czodrowski für die Betreuung und die Möglichkeit, in seiner Arbeitsgruppe die Doktorarbeit anfertigen zu können. Seine Begleitung der Arbeit war stets von großer Offenheit in wissenschaftlicher wie in privater Hinsicht geprägt.

Prof. Dr. Daniel Rauh danke ich für die Übernahme des Zweitgutachtens, die Aufnahme als Gastwissenschaftler am Ende der Promotion, sowie die gute Kooperation in mehreren Projekten und die Nutzung seiner Ressourcen, ohne die das Gelingen dieser Arbeit nicht möglich gewesen wäre.

Prof. Dr. Edmund Beck danke ich für die Interimsbetreuung und die hilfreichen Tipps und Diskussionen.

Meinem guten Freund und Kollegen Dr. Marcel Baltruschat danke ich sehr für seine Unterstützung und Verlässlichkeit von Tag 1 an, in allen kritischen, aber auch guten Phasen in Dortmund. Durch dich konnte ich sehr viel hinsichtlich Organisation und der Welt der Nullen und Einsen lernen, und auch außerhalb der Arbeit haben wir uns sehr gut ergänzt und verstanden.

Dr. Mike Bührmann danke ich für die gute Zusammenarbeit und den wissenschaftlichen Austausch im p38- und FRAGTORY-Projekt.

Bei Dr. Lisa Göbel möchte ich mich für die Kooperation am KRas-Projekt bedanken.

Für die Zusammenarbeit am Pepstatin A Projekt bedanke ich mich recht herzlich bei Dr. Oscar Palomino und dem Team von CrystalsFirst.

Des Weiteren danke ich Dr. Matthias Müller für die vielen guten Diskussionen gerade am Anfang der Promotion sowie Andreas Arndt für die Laborunterstützung. Aus dem AK Rauh danke ich noch Kiru und Jörn für ihre Laborhilfe und den einen oder anderen nützlichen Hinweis.

Martina Reibner danke ich für ihre Hilfe im Bestell- und Rechnungswesen, sonstige Tipps und ihr Verständnis für einen TU Dortmund-Neuling wie ich einer war.

Meinen ehemaligen Masterstudenten Luca Kröll und Frederik Götz, die mit ihren Forschungsleistungen auch zu dieser Arbeit beigetragen haben.

Lisa Johannknecht danke ich für das Korrekturlesen einiger Kapitel dieser Arbeit. Den anderen aktuellen und ehemaligen Mitgliedern des CzodrowskiLab danke ich für die gute Atmosphäre, die Mensagespräche, den wissenschaftlichen Austausch, die Konferenzen, Workshops, Kurse uvm.

Meinen Freundinnen und Freunden danke ich ebenso für die Zerstreuung in der Freizeit, wie auch den Support zu diesem Unterfangen.

Meinem Bruder danke ich für die Unterstützung und die motivierenden Gespräche, ohne die diese Arbeit ebenso nicht möglich gewesen wäre.

Danke Lara, du warst immer an meiner Seite und hast mich während der gesamten Zeit mit Zuspruch und Motivation begleitet und damit einen großen Anteil am Gelingen dieser Arbeit. Auch deiner Familie danke ich für jegliche Unterstützung.

Contents

1. Abstracts

1.1 Kurzfassung

Strukturbasiertes Design ermöglicht es Forschenden, gezielt Moleküle zu entwickeln, die spezifisch mit Zielstrukturen, häufig Proteinen, interagieren, und diese weiter zu optimieren, um schließlich zu einem zugelassenen Medikament zu gelangen. Das *Matched Molecular Pair* (MMP) Sorafenib und Regorafenib, zwei zugelassene Medikamente, wurde mit isothermaler Titrationskalorimetrie (ITC) untersucht, da strukturelle Vorarbeiten auf eine große Änderung ausgehend von der H $\rightarrow$ F Substitution an einem aromatischen Ring schließen ließen. Obwohl die Substitution die Affinität um mehrere Zehnerpotenzen verändern kann, wie in einer MMP-Analyse basierend auf Kinom-Aktivitätsdaten aus der ChEMBL-Datenbank gezeigt wurde, hatten sowohl die Affinitäten als auch die thermodynamischen Profile der Bindungsreaktion mit der Mitogen-aktivierten Proteinkinase p38α nur marginale Unterschiede aufweisen können.

Fragmentbasierte Ansätze erwiesen sich sowohl im akademischen Umfeld als auch in der pharmazeutisch-industriellen Forschung bei der Entwicklung von zugelassenen Wirkstoffen erfolgreich. Drei von fünf Hits aus zwei kristallographischen Fragment-Screens konnten kalorimetrisch validiert werden. Für ein weiteres Fragment konnten pK_{a}-Rechnungen einen Protonierungseffekt durch die Komplexierung an der Proteinkinase A aufzeigen.

Diese Ergebnisse waren der Ausgangspunkt für weitere detaillierte Untersuchungen zu Protonierungseffekten, da deren korrekte Behandlung die Erfolgschancen im strukturbasierten Design erhöhen kann. Die Interaktion der Aspartylprotease Endothiapepsin und Pepstatin A wurde mittels verschiedener Labor- und Computermethoden untersucht. ITC-Experimente ergaben eine Aufnahme von über 1 mol Protonen bei der Bindung von Pepstatin A an Endothiapepsin. Da diese Experimente unter physiologischen Bedingungen durchgeführt wurden (und nicht bei pH=4.6, bei dem eine große Anzahl an Kristallstrukturen verfügbar ist), wurde eine neue Kristall-

struktur bei pH=7.6 bestimmt, die nur kleine strukturelle Änderungen durch das Erhöhen des pH-Werts hervorbrachte. Dies war die Basis für computerbasierte Studien, um den genauen Ort des Protonierungsereignisses zu ermitteln. Beide computerbasierten Ansätze zeigten eine signifikante pK_a-Verschiebung, die zu einem nichtstandardmäßigen Protonierungszustand führt, und dass die katalytische Dyade für die Aufnahme von Protonen bei der Komplexierung verantwortlich ist. Diese Studie zeigt, dass die Bewertung der Protonierungszustände für zwei getrennte Systeme (Protein und Ligand) zu einer falschen Zuordnung der Protonierungszustände und damit zu einer falschen Berechnung der Bindungsenergie führen kann.

Weitere Fallstudien an Komplexen zweier Aspartylproteasen mit unserem computerbasierten Ansatz zeigten, dass die Interaktion von Ligand und Protein zu großen pK_a-Verschiebungen von mehr als drei Einheiten führen kann, die es in Folgeexperimenten zu bestätigen gilt.

Der pK_a-Wert der Thiol-/Thiolatgruppe des Cysteins spielt in der Wirkstoffentwicklung eine besondere Rolle, da diese häufig das Ziel kovalenter Inhibitoren sind. Diesen Wert experimentell zu bestimmen ist sehr schwierig. Durch die Bestimmung des pK_a von Cystein13 in der G13C-Mutante der GTPase KRas mit einem computerbasierten Ansatz wurde gezeigt, dass das Schwefelatom von kovalenten Inhibitoren angegriffen werden kann.

Ligandenbasiertes virtuelles Screening ist eine weit verbreitete Methode in der modernen Wirkstoffentwicklung, die es ermöglicht, große Datenbanken von Verbindungen schnell nach ähnlichen Strukturen zu durchsuchen und diese zu identifizieren. In diesem Rahmen wurde ein Open-Source-Kommandozeilentool entwickelt, das ein substruktur-, fingerprint- und shapebasiertes virtuelles Screening umfasst. Die meisten der implementierten Funktionen basieren vollständig auf dem Chemieinformatik-Framework RDKit. VSFlow akzeptiert eine breite Palette von Eingabedateiformaten und ist in hohem Maße konfigurierbar. Darüber hinaus können die Screening-Ergebnisse als pdf- und/oder pymol-Datei visualisiert werden.

1.2 Abstract

Structure-based design enables researchers to develop molecules that specifically interact with target structures, often proteins, and optimize them to ultimately become approved drugs.

The matched molecular pair (MMP) sorafenib and regorafenib, two approved drugs, were investigated using isothermal titration calorimetry (ITC), because structural groundwork suggested a significant change resulting from the H $\rightarrow$ F substitution on an aromatic ring. Although the substitution can alter affinity by several orders of magnitude, as shown in an MMP analysis based on kinase activity data from the ChEMBL database, both affinities and thermodynamic profiles of the binding reaction with mitogen-activated protein kinase p38α (p38α) showed only marginal differences.

Fragment-based approaches have proven successful in both academia and pharmaceutical industry research for developing approved drugs. Three out of five hits from two crystallographic fragment screens could be calorimetrically validated. For another fragment, pK_a calculations revealed a protonation effect upon complexation with protein kinase A.

These results served as the starting point for further detailed investigations of protonation effects, as their correct handling can increase success rates in structure-based design. The interaction of the aspartic protease Endothiapepsin and Pepstatin A was studied using various laboratory and computational methods. ITC experiments revealed an uptake of over 1 mol of protons upon binding of Pepstatin A to Endothiapepsin. Since these experiments were performed under physiological conditions (and not at pH = 4.6, where a large number of crystal structures are available), a new crystal structure was determined at pH = 7.6, which showed only minor structural changes due to the increase in pH value. Based on this, computational studies were performed to determine the exact location of the protonation event. Both computational approaches showed a significant pK_a shift leading to a non-standard protonation state and that the catalytic dyad is responsible for proton uptake upon complexation. This study demonstrates that evaluating protonation states for two separate systems (protein and ligand) can lead to incorrect assignment of protonation states and, consequently, to inaccurate calculation of binding energy.

Further case studies on complexes of two aspartic proteases with our computational approach showed that the interaction of ligand and protein can lead to large pK_a shifts of more than three units, which need to be confirmed in follow-up experiments.

The pK_a value of the thiol/thiolate group of cysteine plays a special role in drug

development, as it is often the target of covalent inhibitors. Experimentally determining this value is very difficult. By determining the pK_a of cysteine13 in the G13C mutant of the GTPase KRas with a computational approach, it was shown that the sulfur atom can be attacked by covalent inhibitors.

Ligand-based virtual screening is a widespread method in modern drug development, allowing rapid searching and identification of similar structures within large compound databases. In this context, an open-source command-line tool has been developed that encompasses substructure-, fingerprint-, and shape-based virtual screening. Most of the implemented functions are entirely based on the cheminformatics framework RDKit. VSFlow can process a wide range of input file formats and is highly configurable. Additionally, screening results can be visualized as pdf and/or pymol files.

2. Introduction

2.1 Rational Drug Design

The origins of medicinal therapy can be found in folk medicine. The effects of meadow saffron on gout and the narcotic properties of poppy juice were already known in ancient times. In the 19th century, due to the advances in science, all types of natural products were isolated from its herbal origin; the spectrum ranged from today's everyday active ingredient caffeine from the coffee plant to the now illegal drug cocaine from the leaves of the coca bush.[1] The first synthetic active substance was also discovered during this time, chloral hydrate. Introduced as a sedative-hypnotic, it is still available in some countries.[2] While serendipity played a major role in the drug discovery process until the mid-20th century, more rational approaches to drug design have gained prominence.[3, 4]

In the drug development process, many parameters must be optimized and taken into account, which requires a lot of money, time and thought.[4] As the name suggests, for structure-based drug design (SBDD) a three-dimensional structure of the target of interest is needed.

2.1.1 Success Stories in SBDD: HIV-1 Protease Inhibitors

The acquired immunodeficiency syndrome (AIDS) is an infectious disease caused by the human immunodeficiency virus (HIV). Firstly recognized in the early 1980s, it has taken the lives of 40 million people, with about 39 million people living with HIV at the end of 2022, according to the World Health Organization (WHO).[5, 6] In addition to the reverse transcriptase and integrase enzymes of the virus, the HIV-1 protease (HIV) in particular has emerged as a target in recent decades.

HIV-1 PR is an enzyme encoded in the Gag-Pol precursor polyprotein in the viral genome that is required for viral replication. Its function is to process polypeptides into functional proteins by cleavage. HIV-1 PR belongs to the aspartic protease (AP) family of enzymes. The mature enzyme is a homodimer of 99 amino acid residues

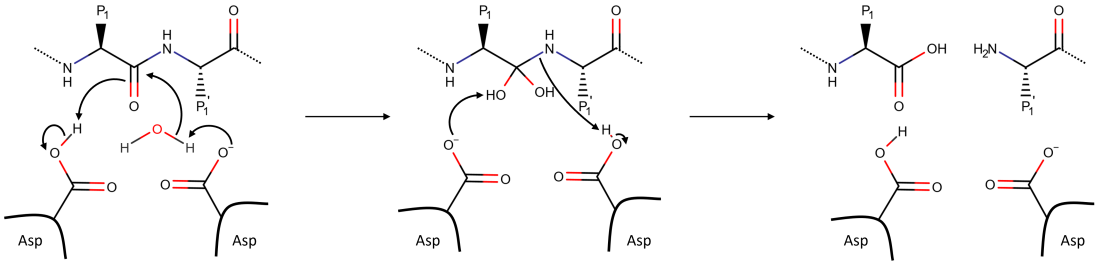


Figure 2.1: Proposed catalytic mechanism for APs

per subunit with the conserved catalytic residues Asp-Thr s^{-1}er-Gly. In the properly folded enzyme, the Asp of both subunits of HIV-1 PR, namely Asp25 and Asp25', are in close proximity and responsible for the hydrolysis of the peptide bond in the polyprotein.[7] The proposed catalytic mechanism of aspartic proteases requires one Asp to be protonated and the other deprotonated (Figure 2.1). Furthermore, a nucleophilic water molecule participates in the cleavage of the peptide bond by attacking the carbonyl carbon atom. It is assumed that the water molecule is closer to the deprotonated Asp residue. The interaction of the water and the deprotonated side chain activates the water. The attack results in the formation of a tetrahedral geminal diol intermediate. Then, the hydroxy group is deprotonated by one Asp of the catalytic dyad and concurrently, the leaving amine is activated by the protonated Asp which leads to the cleavage of the peptide bond.[8]

The first three-dimensional structure of an AP was determined in the 1970s by X-ray crystallography, namely the digestion enzyme pepsin of porcine origin.[9] With 15 members, the AP group is relatively small in the human genome, but it contains several medically relevant targets, including human pepsin, the beta-secretase 1 involved in Alzheimer's disease, the tissue-degrading cathepsin D and the blood pressure-increasing renin.

In addition to HIV-1 PR, renin was the first AP to become a pharmaceutical target. A model compound for finding new AP inhibitors in this early era was pepstatin Ald type (pepA), a naturally occurring small peptide which was obtained from *Streptomyces* filtrates.[10] This hexapeptide, Iva–Val–Val–Sta–Ala–Sta–OH (Figure 2.2), containing the γ-amino acid statine, was originally declared as a specific pepsin inhibitor, but is rather a multi-AP inhibitor, showing also high affinity towards renin and HIV-1 PR.[11, 12]

Based on the findings that the hydroxy group of pepA mimicks the transition state of the enzymatic reaction, peptide analogues have been developed, as pepA A itself does not have sufficient biopharmaceutical properties.[8]

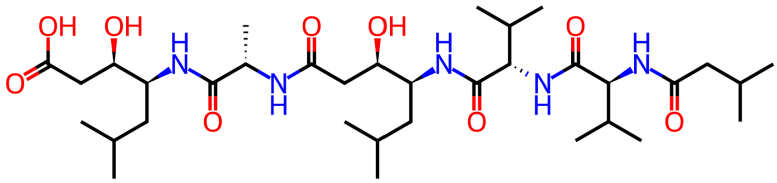


Figure 2.2: 2D molecular depiction of hexapeptide pepstatin A.

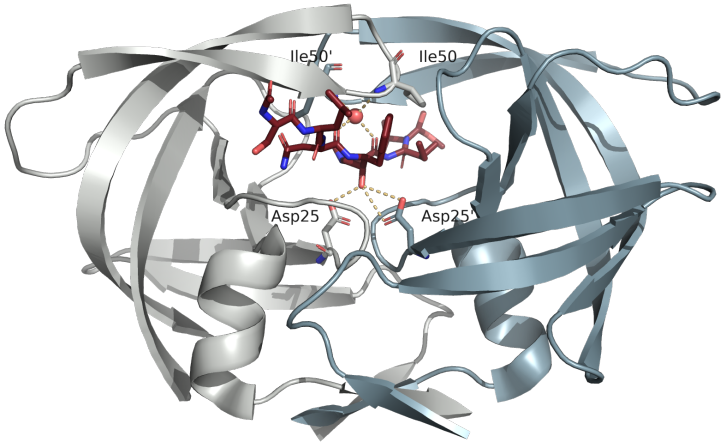


Figure 2.3: One of the first HIV-1 PR structures, co-crystallized with a substrate-based hydroxyethylamine inhibitor (PDB entry: 7HVP). The protease is present as a C_2-symmetric homodimer. The two peptide chains are shown in grey and blue.

The timeline from the identification of HIV-1 PR as a target in the fight against HIV to the first approved inhibitor(s) is truly remarkable: the existence of HIV-1 PR was postulated in 1985, and three years later, experimentally confirmed. In 1989, the three-dimensional structures of both apo and several protein-inhibitor complexes were determined.[13, 14] One of the first published structures of HIV-1 PR co-crystallized with a substrate-based hydroxyethylamine inhibitor, a Gag-Pol polypetide derived heptapeptide (Protein Data Bank (PDB) entry: 7HVP), shows the two subunits forming a dimer, together with some crucial protein-ligand interactions, e.g., with the catalytic dyad and the water-mediated interaction of Ile50 and Ile50' with the backbone carbonyl O of the ligand (Figure 2.3).[15]

The availability of such structures greatly accelerated the development of more effective inhibitors, the first of which was Roche's saquinavir (brand name: Invirase), approved by the U.S. Food and Drug Administration (FDA) in December 1995, making it also one of the first drugs to be developed using structure-based drug design (Figure 2.4).[16] It is a pentapeptide analogue based on one of the natural HIV-1 PR cleavage sites between Phe and Pro. This peptide sequence was not known to be

cleaved by human proteases at the time, and so saquinavir was designed with a hydroxylethylamine transition state analogue in place of the cleavable peptide bond, a bulky decahydroquinoline instead of proline at P1' and a quinoline at P3.[17] Within two years, three other HIV-1 PR inhibitors were approved: ritonavir (Norvir) and indinavir (Crixivan) in 1996 and nelfinavir (Viracept) in 1997.

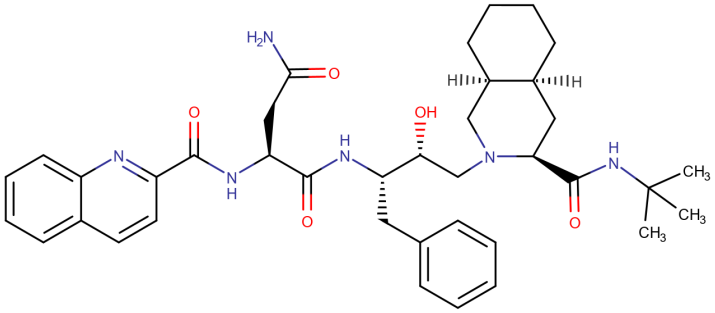


Figure 2.4: Saquinavir - first HIV-1 PR inhibitor on the market.

While this example of a drug designed using an SBDD approach was based on structural data and a lead compound (an oligopeptide) with high affinity for the target, another widely used approach also is fragment-based drug design (FBDD).

2.1.2 Fragment-based Drug Design

As the name suggests, the starting point of FBDD are fragments, small molecules for which the "Rule Of Three" (RO3) applies, i.e. a molecular weight of ≤ 300 Da, a number of hydrogen bond donors and acceptors of ≤ 3, and a cLogP of ≤ 3.[18] Beginning with molecules of this relatively small size, the objective is to identify functional groups or motifs that would otherwise remain undetected due to the hindrance by other parts of the molecule.[19] Screening a target with a fragment library can efficiently sample the chemical space.[20] The subsequent phase following the screening process is the generation of lead compounds, which are drug-sized molecules that serve as the foundation for the advancement of clinical drug candidates.

The development of a lead compound from fragments with generally weak affinity presents a significant challenge. In such cases, information on the fragment's behavior within the binding site is often essential.[19] Having sensitive and robust methods for detecting weak binding interactions is a crucial aspect of FBDD. A range of methodological approaches have been employed to detect the binding of fragments, with nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography being the two most commonly utilized techniques.[20]

Fragment growing is the main approach to optimize the fragment, involving cycles

of X-ray structures and biophysical assays to guide the atom by atom development of the series, an example for this approach emanating in a drug is vemurafenib. Fragment linking/merging strategies have also been successfully applied and turned into the approved drug venetoclax.[20, 21]

Approved drugs typically have an affinity (K_d) in the nano- to picomolar range towards their target as shown for saquinavir and HIV-1 PR.

To achieve this, the focus is on finding molecules that bind to the target of interest in the early stages of drug development - in most cases a protein - and improving these molecules in terms of potency and physicochemical parameters.[1] Protein-ligand binding can be represented by the reaction:

$$\mathrm{P} + \mathrm{L} \underset{\mathrm{K}_d}{\overset{\mathrm{K}_a}{\rightleftharpoons}} \mathrm{PL} \tag{2.1}$$

where a ligand (L) binds to a protein (P) forming a protein-ligand complex (PL). The forward reaction can be characterized by the association constant (K_a), the backward reaction by the dissociation constant (K_d), which is mathematically its reciprocal (equation 2.2). The K_d is the equilibrium concentration of the free ligand at which half of the initial amount of free protein is complexed with the ligand.[22] Since the K_d has the unit of molarity, it is used in preference to K_a.

$$K_a = \frac{[PL]}{[P][L]} = \frac{1}{K_d} \tag{2.2}$$

There are many established methods to determine K_d.[23] However, a more comprehensive view of the interaction is obtained by taking thermodynamics into account.

2.2 Thermodynamics of Protein-Ligand Binding

The thermodynamic relation is applicable to chemical reactions, binding events (such as enzyme-substrate interactions), and is expressed as:

$$\Delta G = \Delta H - T\Delta S \tag{2.3}$$

Equation 2.3 represents the change in Gibbs free energy (ΔG) and is determined by the changes in enthalpy ΔH and entropy ΔS at a given temperature T. In the context of protein-ligand binding, which is the major topic of this thesis, the Gibbs free energy refers to the Gibbs free energy of binding.[24] According to the Equation 2.3, the binding of a ligand to a target protein involves both an enthalpic and an

entropic contribution. Freire emphasized in 2008 that thermodynamic profiles should be considered in drug development and that more enthalpic binders should be given preference.[25] The medicinal chemistry community was very enthusiastic about this approach. More recently, Klebe has pointed out that it is very important to collect systematic data on, for example, congeneric series of ligands to learn more about changes in binding mode, water-mediated interactions or protonation states.[26]

In this context, the entropic binding contribution should not be underestimated as it can negatively affect the overall ΔG balance. In general, the larger the negative amount, the more favorable a binding is. If ΔG<0, a chemical reaction in general, or in this case an interaction, will spontaneously take place.[27] There is a direct thermodynamic relation between K_{d} and ΔG:

$$\Delta G = -RT * \ln K_a = RT * \ln K_d \tag{2.4}$$

which means that thermodynamics provide direct information about the affinity of a new drug, in addition to a detailed description of the enthalpic and entropic binding contribution. However, many other factors such as pharmacokinetics, bioavailability and the physiological responses contribute to the efficacy and are neglected here. Nevertheless, one of the first steps in the search for new drug candidates should be the determination of thermodynamic parameters. Isothermal titration calorimetry (ITC) is a possible method for determining these variables.

2.2.1 Isothermal Titration Calorimetry

For studying molecular interactions in solution, ITC has become the gold standard.[28] The calorimetric measurements carried out in the ITC require a highly sensitive microcalorimeter. This device is able to detect very small amounts of heat, at microjoule level. The interior of such a calorimeter is composed of two cells, the sample cell and the reference cell, both made of a highly efficient thermally conductive material, e.g., Hastelloy (Figure 2.5). Each drop creates a heat signal, resulting in a temperature difference between the sample and reference cell that can be detected by thermopile/thermocouple circuits. A syringe titrates small amounts of solution into the sample cell: in a standard ITC experiment, the syringe contains a ligand solution while the sample cell is equipped with a solution of the macromolecule of interest.[29]

The successful ITC experiment, designed to detect a protein-ligand interaction, gives rise to heat signals during titration, the integration of which produces a sigmoidal

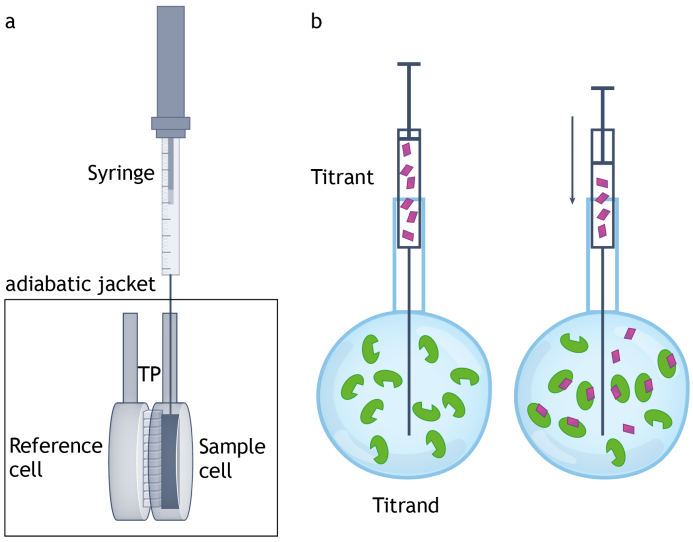


Figure 2.5: Schematic illustration of an ITC instrument. **(a)** The two identical cells in the device used here are coin-shaped and are contained within an adiabatic jacket. A small constant power (e.g., 10 µcal s^{-1}) is applied to the reference cell by the heater. The temperature difference between the reference and sample cell is detected by thermocouple detectors (TP). The cannula of the syringe, which can inject and stir simultaneously, is connected to the sample cell. **(b)** In the experiment, small aliquots of titrant (ligand) are titrated into the titrand (e.g., a protein). During intermolecular interaction, heat is either absorbed or released. Depending on the type of association, the feedback loop increases or decreases the power to the cell to maintain the temperature with the reference. The signal observed in an ITC experiment is the heat per time unit. This is a direct measure of the heat generated by a ligand-macromolecule interaction. Adapted from Bastos et al.[28]

curve from which the parameters K_d, ΔH and the stoichiometry N can be determined (Figure 2.6).

Automated instruments can enhance the throughput and facilitate the handling, since "manual" ITC devices are regarded as low-throughput and loading of the cell requires careful procedure to avoid air in the system.[28]

Of great importance in checking whether an ITC experiment is suitable for the interaction being studied is the c value. It is defined as $c = \frac{[M]_0}{K_d}$ with $[M]_0$ being the concentration of the titrand and should ideally be between 1 and 1000 for reliable determination of the dissociation constant.[30] This means that there is both an upper limit and a lower limit for the determination of the thermodynamic parameters.

There are two ways to determine affinity parameters of weakly binding fragments using ITC: a direct low c value titration and a displacement titration. Low c value titration is feasible and has been shown to give valid K_d values, but it requires a lot

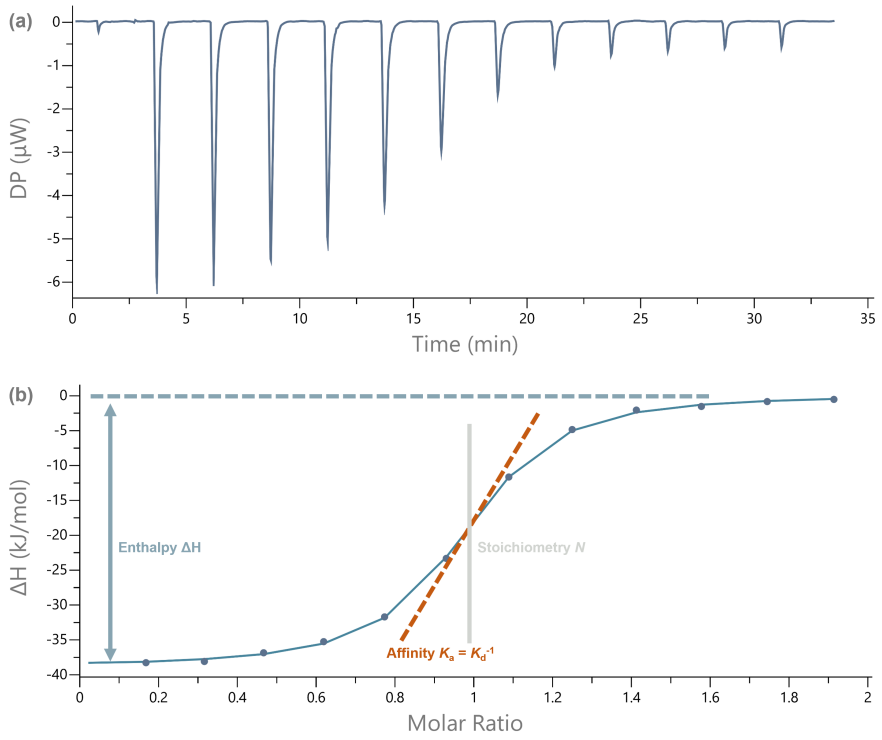


Figure 2.6: (a) Thermogram measured for the interaction of endothiapepsin (EP) and SAP114. (b) Integrated heat values from the thermogram in (a). ΔH is represented by the difference between the two plateaus, N is the value of the molar ratio at the inflection point.

of both fragment and protein and errors regarding ΔH are likely.[31]

Displacement titrations have been carried out for weak binding fragments as well as for ligands with a high affinity in the one-digit nanomolar range. The displacement experiment for compounds showing low affinity towards the target requires the presence of a high affinity ligand that binds to the same binding site as the weak one. Prior to the titration, the protein is incubated with the fragment and in the following ITC experiment, the strong ligand displaces the weak binder from the binding site. The displacement titration as well as the one with the strong ligand yield affinities (K_{observed} for the displacement titration and $K_{\text{reference}}$ for the titration with just the ligand) that can be used to calculate the K_{fragment} of the weakly binding molecule:

$$K_{fragment} = (\frac{K_{reference}}{K_{observed}} - 1) * \frac{1}{c_{fragment}} \tag{2.5}$$

It should be noted that in this case K_{observed}, $K_{\text{reference}}$ and K_{fragment} are association constant. Together with the determined enthalpies $\Delta H_{\text{observed}}$ and $\Delta H_{\text{reference}}$, the enthalpy of the protein-fragment interaction $\Delta H_{\text{fragment}}$ can be calculated[32]:

$$\Delta H_{fragment} = (\Delta H_{reference} * \Delta H_{observed} - 1) * (1 + \frac{1}{K_{fragment} * c_{fragment}}) \quad (2.6)$$

2.3 pK_a Value and Protonation States

According to Brønsted-Lowry theory, an acid is defined as a proton donor capable of donating its proton to a base, a proton acceptor, according to equation 2.7:

$$HA + B \rightleftharpoons A^- + BH^+ \quad (2.7)$$

In the context of medicinal chemistry and biomolecules, we consider an aqueous environment with the following reaction:

$$HA + H_2O \rightleftharpoons A^- + H_3O^+ \quad (2.8)$$

The molar concentration of water is assumed to be constant and can therefore be removed from the equation. We obtain the value of K_a, which is an equilibrium constant, by:

$$K_a = \frac{[A^-][H^+]}{[HA][H_2O]} \quad (2.9)$$

The acid dissociation constant K_a is not to be confused with the association constant in connection with protein-ligand interactions (equation 2.1). The acid dissociation constant is mostly given as the negative decadic logarithm, the pK_a

$$pK_a = -\log_{10} K_a \quad (2.10)$$

Similarly, the *potentia hydrogenii* (pH) is defined as the negative logarithm of the concentration of protons[27]:

$$pH = -\log \frac{c(H^+)}{1M} \quad (2.11)$$

The Henderson-Hasselbalch equation correlates the acid dissociation constant (pK_a) and the pH:

$$pH = pK_a + \log_{10} \frac{[A^-]}{[HA]} \quad (2.12)$$

and can be best visualized if the concentration of deprotonated A^- and protonated species HA is equal, because in this case pH = pK_a.

There are several known methods for obtaining pK_a values of small molecules experimentally.[33] Larger (bio)molecules with multiple titratable groups, such as proteins, are more challenging.

2.3.1 Protonation States in Protein-Ligand Interactions

The titratable side chains of amino acids play a key role in the structural and functional properties of proteins, which are pH dependent.[34] Moreover, protonation states of both ligand and receptor before as well as after binding are very important since the binding process can involve an uptake or release of protons.[35] Out of the 20 proteinogenic amino acids in the standard genetic code, seven have ionizable side chains: Asp, Glu, Tyr, Cys, His, Lys, Arg. Depending on the solvent-accessible surface area, protein residues can be categorized as buried or exposed. Ionizable residues located on the solvent-exposed surface tend to have pK_a values very close to those of the intrinsic pK_a as a result of minimal perturbation by the protein.[36]

The intrinsic pK_a is the pK_a at which the titratable group would be fully exposed to the solvent, with no hydrogen bonds formed and no formal charges nearby.

Model values for this condition have been determined experimentally, for example, with pentapeptides of the sequence Ala-Ala-X-Ala-Ala, where X is the amino acid with the titratable group, with the N- and C-terminus blocked.[37] Usually, ionizable amino acids tend to be at solvent-exposed surface of a protein, with a pK_a similar to the intrinsic one. Conversely, buried ionizable residues embedded in the internal region of a protein are likely to have pK_a values significantly different from those of the model compound due to complex intramolecular interactions.[36]

The Poisson-Boltzmann equation (PBE) 2.13 can be used to calculate the electrostatic potential (ϕ) in such a system:

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) + \epsilon(\mathbf{r})\kappa^2(\mathbf{r}) \sinh \frac{\phi}{k_B T} \tag{2.13}$$

It contains the following parameters: the dielectric permittivity $\epsilon(\mathbf{r})$, the electrostatic potential $\phi(\mathbf{r})$, the permanent change density $\rho(\mathbf{r})$, the Boltzmann constant k_B, the Debye-Hückel parameter κ and the temperature T. For a large macromolecular system such as a protein, it is too complicated to solve the PBE analytically, so it has to be solved numerically.[38]

In a continuum electrostatics based approach to determine pK_a values for protein-ligand complexes, protein and solvent (water) are treated as continuum media with separate and widely distinct dielectric constants. While the dielectric constant assigned to water is ε=80 for water and is usually not changed, the ε for the protein can be less than 10, but it is reasonable to increase it to 20 in some cases considering the conformational flexibility of proteins and their dynamic side chains.[38–40]

For deciphering the electrostatic interaction in a protein-ligand complex, the ligand is first placed in water. It is then moved into the less polarizable protein. This process is associated with a free energy difference $\Delta G_{s\to p}$ (s is for solvent, p for protein), which can be obtained from solving the PBE for the system.

A pK_a shift of an ionizable amino acid side chain or of a ligand with a titratable group in a complex is then detected using the thermodynamic cycle. (Figure 2.7).[41]

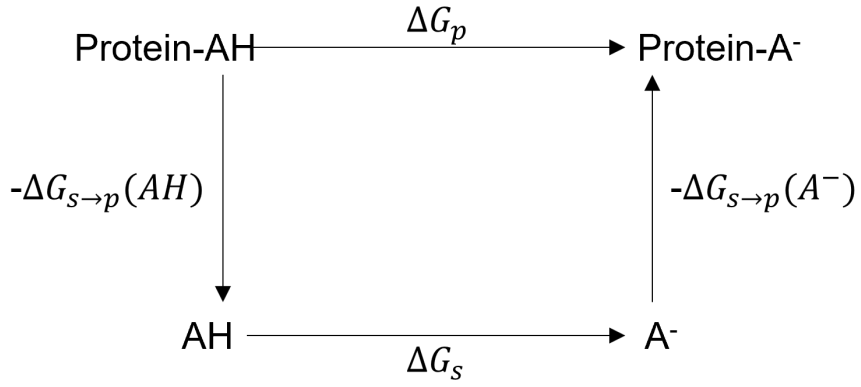


Figure 2.7: The thermodynamic cycle for calculating the pK_a shift. s refers to aqueous solution and p to protein. Protein-AH corresponds to the protonated residue (or ligand) in the protein, while AH corresponds to the protonated compound in solution.

The ΔG_p and ΔG_s free energy differences for the ligand's deprotonation in the solvent and protein, respectively, are included in this cycle. The ΔpK_a, which is the relevant value here, is determined using:

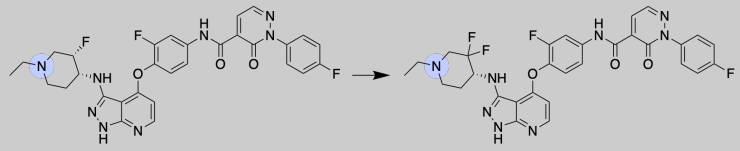
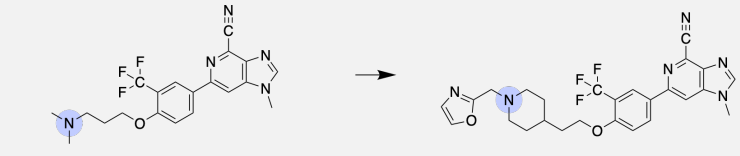
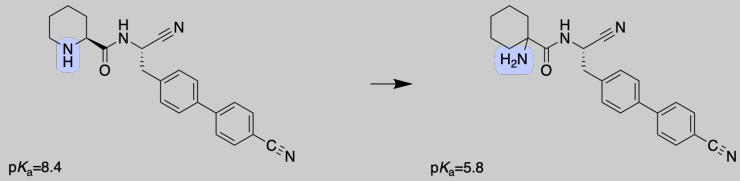
$$\Delta pK_a = \frac{1}{2.303RT}(\Delta G_p - \Delta G_s) = \frac{1}{2.303RT}(\Delta G_{s\to p}(A^-) - (\Delta G_{s\to p}(AH)) \quad (2.14)$$

Depending on the pH value of the environment, a pK_a shift can lead to a change in the protonation state or the degree of protonation, respectively.

The charge state of drugs and drug candidates is an eminent feature in drug discovery. Manallack et al. report in their study that almost 80 % of drugs contain an ionizable group.[42] Such ionizable groups can either occur in the neutral or charged

state. This depends on the pK_a value (and, of course, the present pH value): examples of unexpected large pK_a shifts of small molecules can be found in Table 2.1.

Table 2.1: Examples of small molecules with large pK_a shifts.

Structural similar compounds with their pK_a values	**Effect of pK_a shift**
	Reduced volume of distribution in the mouse with unchanged cellular activity on the kinase Met.[43]
	Reduction of the spleen/plasma ratio by a factor of 13 and constant excellent cellular activity on CathepsinS.[44]
	Minimization of hERG activity and phospholipidosis and increase of selectivity towards other cathepsins.[45]

The charge state can influence all sorts of properties ranging from solubility to distribution coefficient to other absorption, distribution, metabolism, excretion, toxicity (ADMET) properties and even PK/PD properties of the drug inside the human body are influenced by the charge state of the drug. The charge state of these amino acids can also be modulated either by the protein environment or by a bound ligand. This, in turn, can then influence drug binding or structural properties of the protein.

Why is it so important to know the charge state of molecules, whether on the protein or ligand side? In terms of non-covalent interactions between two molecules, salt-bridge interactions contribute significantly to the exothermic binding signature.[46] A salt bridge in a protein-ligand interaction, e.g., between a deprotonated, negatively charged Asp and a protonated basic group of the ligand, such as an amine, or between a positively charged Arg and a deprotonated carboxylic group is often supported by an additional hydrogen bond. While a salt-bridge interaction can be favorable, it is countered by an energetic penalty resulting from the desolvation of charged groups, but it strongly depends on the environment of the protein.[47] Rational design of charge-assisted contacts based on pK_a calculations of the complex

can improve affinity. Moreover, the uncharged form of a drug can cross biological membranes more easily than its corresponding charged form. An approach might be to design a molecule that remains uncharged during transport across membranes, but becomes charged by proton release upon binding to the target through electrostatic interactions.[48]

As mentioned above, the protonation states of a protein and a ligand can change during complex formation, which can be determined experimentally by ITC and computationally by calculating pK_a values of functional groups and their changes during complex formation.[49]

2.4 Targets - from Model Enzymes to relevant Drug Targets

Protein-ligand interactions are being studied on a wide range of targets: some serve as model enzymes to study fundamental biophysical properties, while others are medically relevant, i.e., a target involved in a disease.

2.4.1 Aspartic Proteases

APs are a widely distributed class of enzymes found in vertebrates, plants, fungi, viruses and even archaea.[50, 51] In humans, there are 15 different human APs, involved in processes as diverse as protein-digesting pepsin and blood pressure-regulating renin. One of the best studied APs is the HIV-1 PR (see 2.1.1).[8]

Endothiapepsin

However, the AP EP originates from the fungus *Endothia parasitica* (now called *Cryphonectria parasitica* and belongs to the family of pepsin-like APs that have been identified as molecular targets for the treatment of various diseases.[52] The structural similarity to APs relevant to medicinal chemists, as well as the easy availability and handling of EP, has led to its use as a model enzyme for mechanistic studies and the development of corresponding inhibitors.[53]

2.4.2 Protein Kinases

Since their discovery, protein kinases have been of major importance in medicinal chemistry research in both academia and the pharmaceutical industry. More than 500 of all genes (1.7 %) in the human genome encode protein kinases, making this

family of enzymes a protagonist in cell signalling.[54] Protein kinases are phosphotransferases and thus catalyze the transfer of the γ-phosphate group (PO_4^{3-}) of the co-substrate adenosine 5'-(tetrahydrogen triphosphate) (ATP) to hydroxy groups on the amino acid side chains of their substrates. Due to their substrate specificity, serine-threonine protein kinases are distinguished from tyrosine kinases.

Phosphorylation often induces a conformational change due to the introduced high negative charge density, which in turn influences the activity of the substrate, either by activating or inhibiting effector-mediated regulation.[55]

Abnormal or impaired kinase activity is often associated with diseases resulting in inflammatory and proliferative responses. Cancer, rheumatoid arthritis, cardiovascular and neurological diseases, asthma and psoriasis are the most important examples. These diseases make protein kinases important targets for drug development.[56–58] Although protein kinases are structurally highly conserved, particularly in the ATP-binding site, it is possible to design small molecules that target a specific kinase with high selectivity and also show beneficial pharmaceutical properties.[59] The ATP-binding site is located in the catalytic centre of the catalytic kinase domain between the predominantly α-helical, C-terminal subdomain and the β-sheet-rich, N-terminal subdomain. These two subdomains are connected by the flexible 'hinge' region. Upon binding to the active centre, the adenine moiety of ATP forms hydrogen bonds with the hinge region, while the triphosphate moiety is bound by catalytic amino acid side chains (Glu in α-helix C and Lys in β-sheet 3) and further complexed by two Mg^{2+} ions coordinated via the Asp184 of the DFG (Asp-Phe-Gly) motif in the activation loop. In addition, polar interactions with the glycine-rich loop provide precise positioning of ATP, which is essential for the catalytic reaction. The coordination and the negative charge of the Asp166 side chain towards the hydroxy group of serine or threonine leads to an increase in the nucleophilicity by a proton transfer from serine/threonine to the aspartate, enabling the attack of serine/threonine towards the γ-phosphate group (Figure 2.8).[60]

Two conformational states, DFG-in and DFG-out, are characteristic of kinases: in the ATP-bound state, the kinase adopts an active, so-called DFG-in conformation, in which the DFG-Asp points towards the active centre, while the DFG-Phe is oriented towards the α-helix C, the latter occupying an allosteric pocket adjacent to the ATP binding site. Accordingly, in the inactive DFG-out conformation, there is a reorganization of the two side chains so that the DFG-Asp rotates out of the active centre and the DFG-Phe releases the previously blocked pocket.[61]

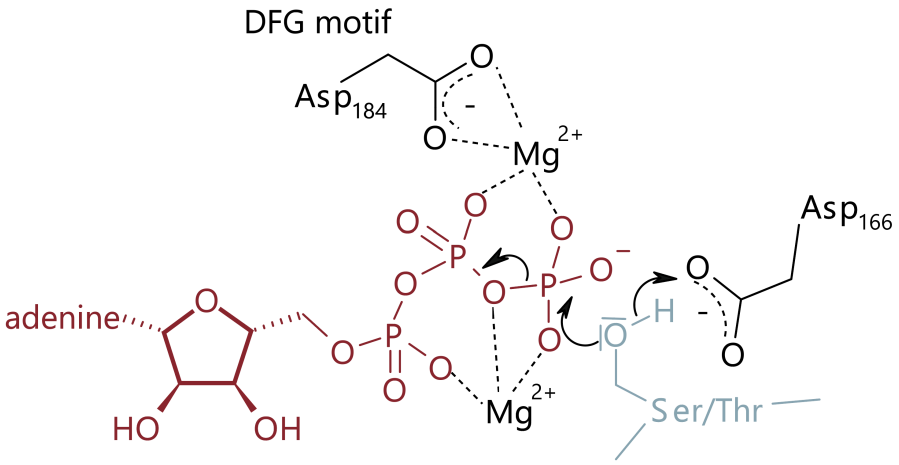


Figure 2.8: Mechanism of a protein kinase phosphorylating its substrate protein at the Ser/Thr-OH. Adapted from [1]

2.4.3 Small GTPases

Small GTPases are a protein superfamily of guanine nucleotide-dependent molecular switches that regulate diverse cellular processes. The superfamily consists of more than 150 family members and can be divided into five smaller subfamilies based on their sequence, structural similarity and functions in the cell: Ras, Rho, Ran, Rab and Arf. Small GTPases cycle between an inactive guanosine diphosphate (GDP)-bound form and an active GDP-bound state, and this cycle is tightly regulated by guanine nucleotide exchange factors (GEFs), GTPase-activating proteins (GAPs) and guanine nucleotide dissociation inhibitors (GDIs). Small GTPases are involved in a wide variety of biological responses and their functional specificity depends on the cellular system, the type of stimulus and their intracellular localization. The rat sarcoma virus (Ras) subfamily is responsible for regulating signalling pathways involved in cell proliferation, morphology, differentiation and survival, while the Rho subfamily regulates actin organization, gene expression and cell cycle progression. The Rab subfamily is involved in vesicle and protein trafficking, while the Arf proteins regulate intracellular transport. The Ran subfamily consists of only one protein involved in nuclear transport.[62, 63]

2.5 Aims of this Work

This work is divided into six projects that aim to find or better understand protein-ligand interactions using different methods in vitro, mainly calorimetric using ITC, and in silico, using an open source screening tool as well as a commercial PBE solver.

This work was preceded by studies of a matched molecular pair, two approved kinase inhibitors, on the medicinal relevant protein kinase mitogen-activated protein kinase p38α (p38α). In this project (Chapter 3), the general influence of fluorine substituents in medicinal chemistry was investigated by an matched molecular pair analysis (MMPA) of protein kinase activity data in the public ChEMBL database, and the interaction of a specific MMP with p38α was then studied calorimetrically.

In another project (Chapter 4), fragment hits for two targets were investigated: first, an unusual binding mode of benzoic acid to the hinge region of the model protein cAMP-dependent protein kinase A (PKA), in particular with respect to the buffer conditions used in a crystallization experiment in which the low-affinity fragment is soaked in the protein crystal, and second, fragment hits from a virtual library screen targeting the model aspartic protease EP again were further investigated thermodynamically.

In addition, the interaction of EP with the multiprotease inhibitor pepA was investigated in another project (Chapter 5), both calorimetrically and computationally, with particular emphasis on the protonation effects that occur as a result of the binding reaction.

More computational case studies focusing on shifts in protonation were performed in another project on the aspartic proteases HIV-1 PR and EP (Chapter 6).

Another small project (Chapter 7) focused on covalent inhibitors targeting the cysteine residue of the oncogenic KRasG13C mutant. This particular mutant has not yet been targeted and the protonation state of Cys13 is important in the design of these inhibitors.

In the absence of structural data, ligand-based virtual screening is used to identify close analogues of known active compounds. The Virtual Screening WorkFlow (VSFlow), an open source command line tool based largely on the RDKit cheminformatics framework, was developed in the last project (Chapter 8).

3. Protein Kinase p38α and Inhibitors Sorafenib/Regorafenib - a Matched Molecular Pair

3.1 Introductory Remarks

This part takes a deeper look into the approved drugs and matched molecular pairs regorafenib and sorafenib and their interaction with the protein kinase p38. Moreover, a general overview of the matched molecular pair (MMP) H-F in medicinal chemistry is given. This work has been made possible thanks to the contributions of Mike Bührmann, Christiane M. Ehrt and Daniel Rauh. The crystal structures and cited affinities are part of Mike Bührmann's PhD thesis whereas the KNIME™ workflow was provided by Christiane M. Ehrt.

3.2 Introduction

One of the protein kinase targets of pharmaceutical research is p38α. Activation of this serine-threonine kinase has been demonstrated to increase the activities of tumor necrosis factor-a and interleukin 1β, both proinflammatory cytokines.[57]

3.2.1 Fluorine and its major Role in Medicinal Chemistry

Fluorine plays an important role in medicinal chemistry: of the 3594 currently approved small molecule drugs (database: ChEMBL32), 405 contain one or more fluorine atoms (11 %).[64] After its isolation in 1886, for which Henri Moissan was awarded the Nobel Prize in 1906, its use in the following decades was limited to military applications and special materials due to the violent reactivity of fluorine gas.

The first pharmaceutical product containing fluorine was fludrocortisone, a hydro-

cortisone in which the 9α-hydrogen is replaced by fluorine. Interestingly, fluorine was not part of the original halogenated series that was synthesized and tested, apparently for reasons of handling the violent reactivity of fluorine gas and the poisonous properties of the fluoroorganic compound occurring in nature.[65] The anti-inflammatory activity of the 9α-halo-17α-hydroxycorticosterone acetates was 1 for the H (cortisone acetate), 0.1 for the I, 0.28 for the Br and 4.0 for the Cl, indicating higher potency for smaller halogen substituents.[66] In a follow-up study, the F analogue fludrocortisone, was investigated and showed a remarkable ten-fold increase in activity compared to the parental hormones.[65, 67] This breakthrough was the starting point for the success of fluorine in medicinal chemistry: not only is fludrocortisone still in use and on the WHO list of essential medicines, but in 2021, seven of the top 20 small molecule drugs – in terms of retail sales - were organofluorines.[68, 69] The incorporation of fluorine into new drugs is a trend still ongoing. This is illustrated by the fact that a total of 33 out of 149 new drugs approved by the FDA between 2019 and 2021 contain at least one fluorine substituent (2019: 11, 2020: 13, 2021: 9).[70–72]

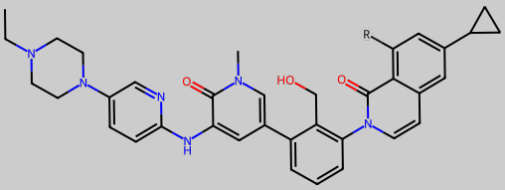
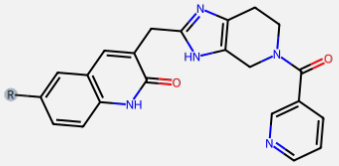
Some of the essential features that give rise to desirable properties in terms of ADMET parameters are: its inductive effects influencing the pK_a of neighbouring titratable groups, the ability to increase metabolic stability and a higher membrane permeability, which conversely can lead to lower solubility. In addition, fluorine atoms are (specifically) incorporated into active compounds to increase potency: a single substitution of a hydrogen atom on an aromatic ring with a fluorine can result in a tenfold increase in affinity – or even more.[73, 74]

Two extreme MMPs of these activity cliffs are presented here as examples (Table 3.1): one of these striking differences, where the ΔpChEMBL is almost three orders of magnitude, involves the target vascular endothelial growth factor receptor (VEGFR-2). The H variant has an IC_{50} of 200 nM while the F variant has an IC_{50} of 3 nM. With no structural data available for either compound in complex with VEGFR-2, a docking approach revealed a potential hydrogen bonding of fluorine to NH of K868. This interaction is not possible in the absence of fluorine.[75]

The other striking pair in this data set showed a 40-fold increase in potency towards Bruton's tyrosine kinase (BTK) upon fluorination. Various MMPs from closely related scaffolds were tested and, in addition to IC_{50} assays, crystallization studies were performed on the respective fluorine variants. The cause of the increase in affinity is explained here by a substituent synergy: the isoquinolone ring system mediates an optimal position for both the cyclopropyl group and the fluorine atom. The fluorination in this series does not require any reduction in the optimal hy-

drophobic interactions of the cyclopropyl group and the full, mutually reinforcing effect of the fluorine atom can be realized.[76]

Table 3.1: Examples of MMPs with activity cliffs.

MMP	Target	Affinity Ratio (H:F)	Structural data
	BTK	666	No structural data
	VEGFR-2	40	Crystal structure of F variant

3.2.2 MMPA – a Means to evaluate the Influence of Fluorine Substitution

A MMP is a pair of structurally related molecules that differ by a specific chemical transformation, such as the replacement of a functional group or the addition/removal of a substituent (3.1).[77]

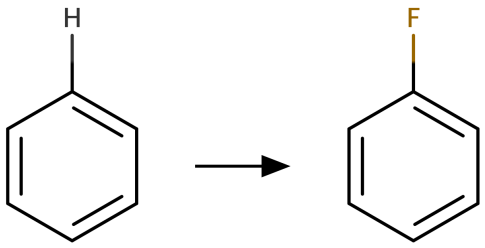


Figure 3.1: Example of a matched molecular pair (MMP).

Relationships between molecular structures can be established by applying the appropriate structural transformations and searching for the products of these transformations in the original database. This is called MMPA and has been found to be widely applicable to structure-property relationships.[77]

There is a plethora of MMPA tools based on different algorithms, both commercial and open-source. In general, all published MMP algorithms can be classified as either supervised or non-supervised methods. Supervised methods are based on a

predefined chemical transformation that generates the MMP. This ensures precise control of the definition of the MMP, which is advantageous for a particular question. Unsupervised methods use an algorithm to try to find all possible pairs in a set of compounds. These are usually maximum common substructure (MCS) or fragmentation approaches. Unlike supervised methods, these methods can lead to the discovery of new and surprising MMPs.[78]

3.2.3 Matched Molecular Pair Drugs - Sorafenib & Regorafenib

Although MMPA is most commonly used in hit-to-lead drug discovery projects to optimize your future drug at a particular site and keep only the one with superior properties, there are several MMPs that have gone through the entire drug discovery process and become approved drugs.[79] One of these is a pair of protein kinase inhibitors: sorafenib (4-4-[3-(4-chloro-3-trifluoromethyl-phenyl)ureido]phenoxy-pyridine-2-carboxylic acid methylamide-4-methylbenzenesulfonate) and regorafenib, which differ in only one atom - sorafenib has a hydrogen substituent on the middle aromatic ring, while regorafenib has a fluorine substituent (Figure 3.2). Sorafenib was developed by Bayer in the 1990s as an inhibitor of the serine-threonine kinase Raf1 ($IC_{50} = 6\,nM$) and has been proven to be effective in reducing tumour growth in xenographic models.[79, 80] It also inhibits the cancer-associated BRAF kinase and p38α. It also inhibits the cancer-associated BRAF kinase and p38α. Regorafenib, also developed by Bayer, has a similar but more advanced efficacy profile.[81] Structures of both compounds obtained by co-crystallization experiments showed significant differences in the binding modes of this specific MMP (Figure 3.3). The part beyond the urea moiety remains largely unchanged. However, in the hinge region, the 4-oxy-N-methylpicolinamide moiety is rotated 180° around the O-C bond of the phenoxy group, resulting in fundamentally different interactions with the enzyme.[82]

Although biochemical assays showed very similar IC_{50} and K_d values, the thermodynamic profiles of both compounds could potentially have different enthalpy-entropy contributions.[82] Therefore, ITC measurements were performed.

3.3 Results & Discussion

To get a more general overview of the effect of an Ar–H → Ar–F substitution on an aromatic ring, the activity data for protein kinases in the ChEMBL database were screened (Figure 3.4). The values have been calculated as the difference between the fluorine and hydrogen variants of the MMP. Negative numbers indicate a

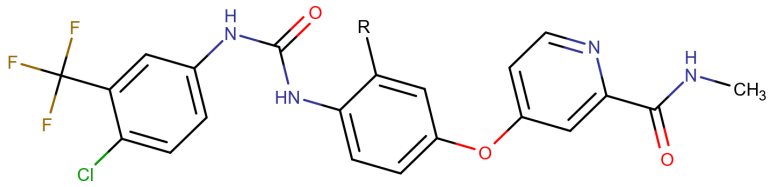


Figure 3.2: Molecular depiction of the matched molecular pair sorafenib (R=H) and regorafenib (R=F).

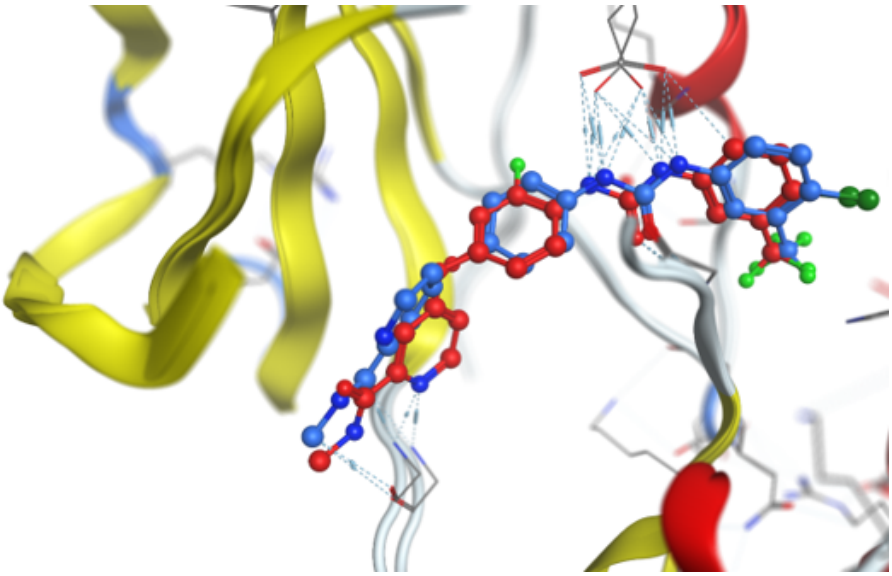


Figure 3.3: Structural alignment of p38α co-crystallized with the matched molecular pair sorafenib (blue) and regorafenib (red).

higher activity of the hydrogen-substituted molecule, while positive numbers reflect a higher activity of the fluorine variant. Over 50 % of the 6133 MMPs (3098) have a ΔpChEMBL between −0.3 and 0.3, corresponding to a doubling/halving in inhibition/affinity (Figure 3.5). Otherwise, there is no general tendency for fluorine to increase IC_{50}, K_d or K_i: the data depicts a normal distribution and both fluorine and hydrogen variants can, in extreme cases, be a hundred times more active than the other.

This is the case for 34 H and 21 F variants. It is certainly desirable to achieve these activity cliffs, because a higher affinity for an active substance could mean that a lower dose would have to be applied as a consequence.

The two extreme MMPs presented earlier were part of the data set (Table 3.1).

There was activity data for eleven different protein kinases for the MMP drugs sorafenib and regorafenib. Regorafenib was more effective against nine of these targets. ChEMBL data for sorafenib-regorafenib were not available for the target of

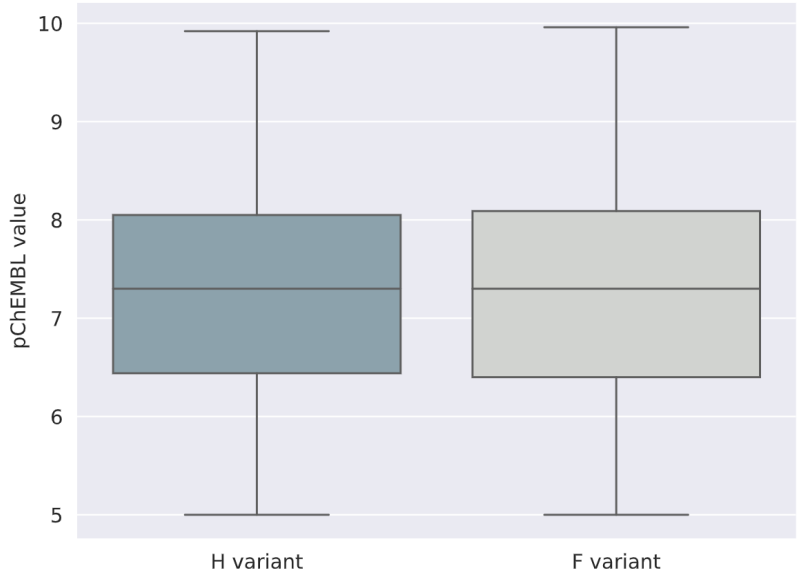


Figure 3.4: Distribution of the pChEMBL of the analysed 6133 H/F MMPs.

interest in this project, p38α.

Next, ITC measurements were performed. Due to the very low solubility of both compounds, the standard ITC protocol, where the protein is in the cell and the ligand ligand in the syringe, had to be reversed. This allowed the concentration of the ligands to be reduced from 200 to 20 µM and later to 10 µM. Even at this concentration, some of the compound precipitated, resulting in an apparent N of > 1.5 when evaluated with the compound concentration added. If the concentration had been reduced further, the detection limit of the instrument would have been reached. Sorafenib tosylate and free regorafenib base were used for the initial measurements (Figure 3.6).

Phan et al. found that the hydrochloride salts of both regorafenib and sorafenib had higher aqueous solubilities than the free bases, both at around 70 µmol l^{-1}.[83]

While these measurements were carried out in distilled water containing 0.2 % sodium lauryl sulfate, one explanation for the higher solubility of the hydrochlorides of the compounds is that the dissolution of the hydrochloric acid molecule decreases the pH of the solution. 70 µmol l^{-1} HCl – assuming a complete dissociation – results in a pH of 4.15. The pK_a of both compounds is not known, but at this pH, the weak basic compounds potentially take up a proton from the solution, and the ionized form has a higher solubility than the neutral form. According to Bhattacher et al. solubility data without reference to pH and pK_a are irrelevant for an ionizable compound. On the other hand, metastable crystalline forms may have a higher apparent solubility.[84]

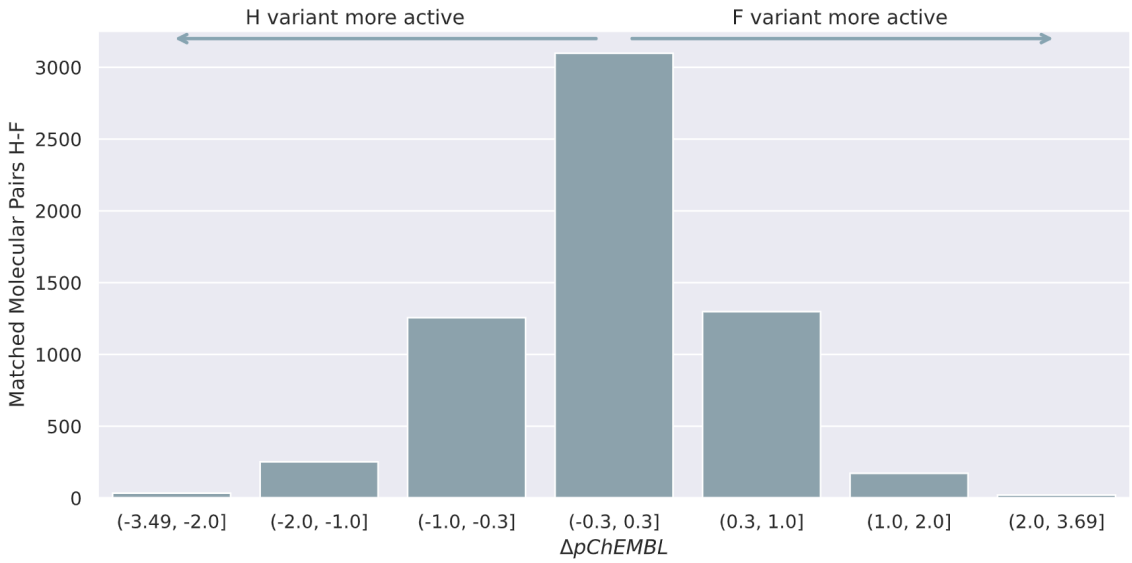


Figure 3.5: Distribution of the ΔpChEMBL of the analysed 6133 MMPs. Negative numbers indicate a higher potency of the Ar–H, positive numbers a higher potency of the Ar–F molecule of the MMP.

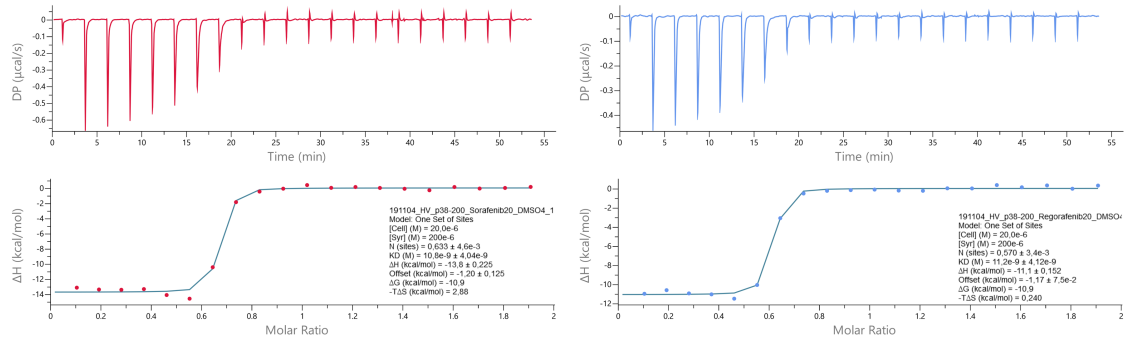


Figure 3.6: Initial thermograms of inverse titration of p38α and sorafenib tosylate (red, left) and regorafenib free base (blue, right).

As a result, the corresponding HCl salts were generated and subsequently tested in the ITC setup. In contrast to the free base of regorafenib and sorafenib tosylate, the hydrochloride salts of the compounds were completely dissolved in the buffer. Control measurements with the better soluble protein kinase inhibitor SB203580 gave affinities in agreement with the literature (Figure A.1).

The stoichiometry N was 1.16 ± 0.06 for sorafenib (Figure 3.7) and 1.17 ± 0.03 for regorafenib (Figure 3.8), respectively.

N values greater than one may result from inaccurately determined concentrations, since available structural data indicate that there is a single binding site, and the ideal N should therefore be 1. Often the value is <1 because not 100 % of the

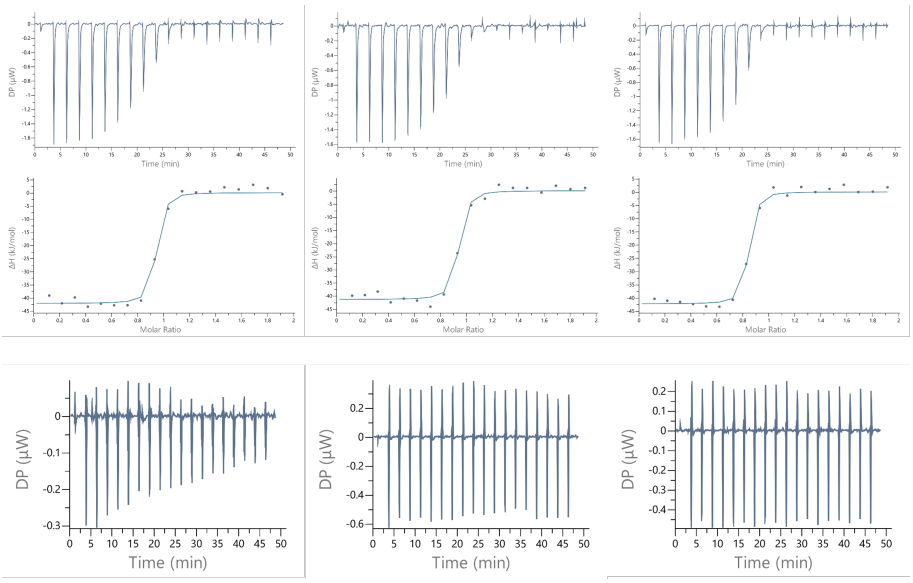


Figure 3.7: Thermograms of inverse titration of p38α and sorafenib hydrochloride (upper row, triplicate measurement) and corresponding control titrations (lower row, left to right: sorafenib (cell) - buffer (syringe); buffer (cell) - p38α (syringe); buffer (cell) - buffer (syringe)).

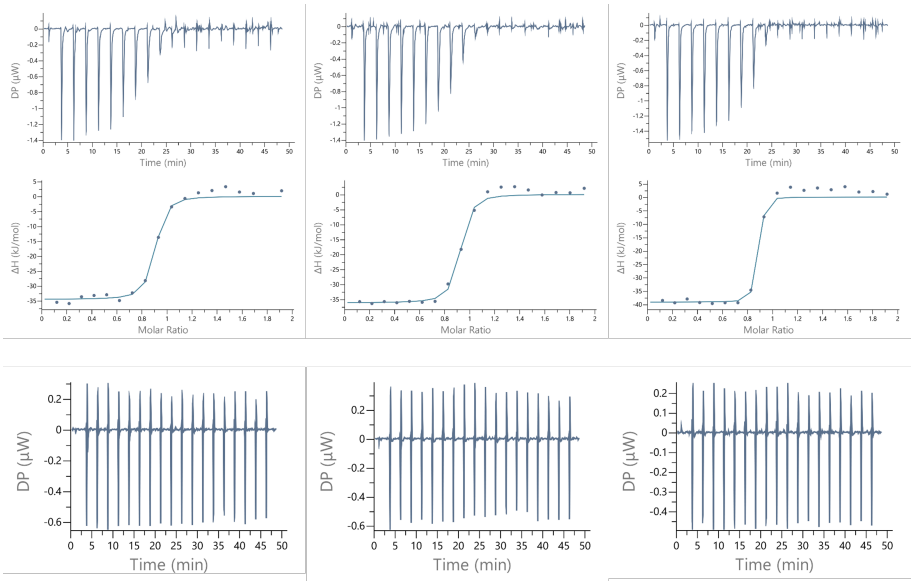


Figure 3.8: Thermograms of inverse titration of p38α and regorafenib hydrochloride (upper row, triplicate measurement) and corresponding control titrations (lower row, left to right: regorafenib (cell) - buffer (syringe); buffer (cell) - p38α (syringe); buffer (cell) - buffer (syringe)).

protein used is active.

While the determined ΔG is essentially the same for both inhibitors, the p38α-sorafenib binding shows a higher enthalpic contribution than the regorafenib binding

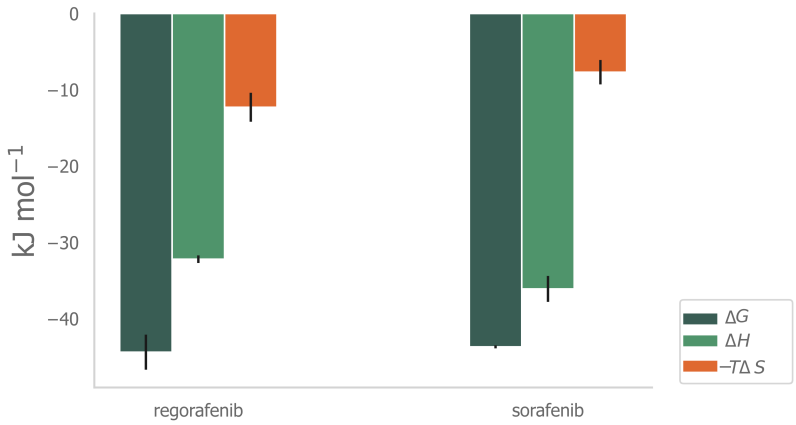


Figure 3.9: Thermodynamic profile of the interaction of p38α and regorafenib/sorafenib, determined by ITC measurements.

(Figure 3.9). The slight differences in enthalpy and entropy are within the error, with –32 ± 1 and –36 ± 2 kJ mol^{-1} for the enthalpy of regorafenib and sorafenib, respectively (Table 3.2). Overall, the two binding reactions are enthalpically and entropically favored, with minimal differences that are within the range of error.

Compared to a previous affinity assay, the calorimetric measurements revealed an affinity that was a power of ten higher in terms of K_d. This is not unusual for different approaches.[82] Nevertheless, three independent measurements taken together showed almost the same results for both regorafenib and sorafenib. Therefore, the shown calorimetric measurements did not provide an explanation for the structural differences between the two crystal structures.

It is already difficult to compare X-ray crystallography results with calorimetric measurements. The conditions in the X-ray are different: in the ITC there should be only as many components in the solution as necessary, whereas the complex crystallization solution contains many more components ($MgCl_2$, DTT, glycerol). As in the initial ITC measurements, regorafenib as free base and sorafenib tosylate were used in the crystallization experiments. That could have had an impact on the structure. It was shown that single crystals of the free bases had different conformations.[83]

3.4 Conclusion & Outlook

In general, the success of fluorine in medicinal chemistry is a good justification for its use. We see that Ar–H → Ar–F exchange is often not very effective and can even cause a loss of affinity. For example, the fluorination of one compound could lead to a higher specificity towards one target. Conversely, this means that the potency

Table 3.2: Thermodynamic parameters of the interaction of p38α and regorafenib/sorafenib, determined by ITC measurements.

	Regorafenib	**Sorafenib**
K_d [nM]	24 ± 2	21 ± 2
N	1.17 ± 0.03	1.16 ± 0.06
ΔG [kJ mol^{-1}]	–44 ± 2	–44 ± 0
ΔH [kJ mol^{-1}]	–32 ± 1	–36 ± 2
–TΔS [kJ mol^{-1}]	–12 ± 2	–8 ± 2

towards other (off-)targets decreases and in the end, the ΔpChEMBL could indicate a higher affinity for the H variant.

The thermodynamic profiles showed no significant differences in the binding of sorafenib/regorafenib to p38α. To improve the accuracy of the measurements, the parameters could be optimized. As the c value was close to 1000 and thus almost at the upper limit, lowering the concentrations, e.g., 100 µM protein and 10 µM ligand, could lead to a more sigmoidal curve and a stoichiometry closer to 1. An optimized ITC protocol would, in principle, be able to provide not only thermodynamic but also kinetic information in a single experiment. This label-free approach where no immobilization is needed could be of great interest.[85] Regarding the crystal structures, it would be interesting to crystallize the different forms (free base, hydrochloride salts) to check if the salt form really has an effect on the conformation in the complex structure, as was seen in the single crystals.[83]

3.5 Experimental Procedure

3.5.1 Expression and Purification of p38α

The expression and purification of p38α was performed as described in the literature.[86] Briefly, the p38α wild type (wt) construct was transformed into chemically competent *Escherichia coli* BL21(DE3), expressed and purified. Overexpression was performed overnight (20 h) at 18°C with shaking at 160 rpm. The target protein was then purified by Ni affinity, anion exchange and finally size exclusion chromatography after proteolytic cleavage of the His_6 tag. The protein was then concentrated to approximately 20 mg ml^{-1}, snap frozen in liquid nitrogen and stored at –80 °C until further use. Prior to ITC measurements, the required amount of p38α was thawed and purified again via a size exclusion chromatography column and concentrated to approximately 10 mg ml^{-1} in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer at pH 7.5 containing 50 mM NaCl and 1 mM tris(2-carboxy-

ethyl)phosphine (TCEP).

3.5.2 ITC Measurements on p38α and Sorafenib/Regorafenib

Due to the low solubility of the hydrochloride salts of sorafenib and regorafenib, a reverse titration was performed, i.e., the protein solution in the syringe was titrated into the ligand solution in the cell. Therefore, 20 µM solutions of both compounds in the same HEPES buffer used for p38α was prepared from 10 mM dimethyl sulfoxide (DMSO) stocks with an intermediate dilution step of a 1 mM compound solution in 100 % DMSO, so that the final DMSO concentration in the protein solution was at 2 % (v/v). The same amount of DMSO was added to the protein solution. Titrations were performed with 19 injections, with a smaller first injection of 0.4 µl followed by 18 injections of 2 µl. The obtained thermogram peaks of all titrations were integrated and fitted with MicroCal PEAQ-ITC Analysis Software 1.41.

3.5.3 Matched Molecular Pair Analysis

The MMPA was carried out with a workflow developed in the Konstanz Information Miner (KNIME) Analytics Platform.[87] The protein kinase targets were selected based on the most recent list provided by the openkinome project containing 552 Uniprot IDs that were used to get activity data from ChEMBL32.[88] For the ChEMBL target IDs obtained from the UniProt IDs, all activity data were pulled from the database, and only IC_{50}, K_i and K_d activities $\leq$10 µM were used for further analysis.

The dataset was then further adjusted by applying the RDKit Salt Stripper node to remove salts and then using MarvinSketch nodes and other RDKit nodes: Substructure Filter to filter out the aromatic F molecules (Figure 3.1, right molecule), Chemical Transformation node to remove the F and then look for matches between the canonical SMILES of the "transformed" molecules and the original H variant for which activity data is available. This approach was similar to that one published by Hajduk and Sauer.[89] Only MMP activity data for the same assay and from the same publication/dataset based on the Assay ChEMBL ID and Document ChEMBL ID were taken into account, respectively. Based on the pChEMBL of the MMP, ΔpChEMBL was calculated as ΔpChEMBL = pChEMBL(F) – pChEMBL(H). A list containing the ΔpChEMBL values were exported to a csv file and for the graphical analysis, the python libraries matplotlib and seaborn were used.

Take Home:

Fluorine can influence substitution of

4. Identification & Characterization of Protein-Fragment Interactions

4.1 Introduction

4.1.1 Fragtory

In order to implement a pharmacophore-based design concept in an academic setting, "Fragtory", an automated KNIME-based workflow has been developed using a modular approach. This workflow allows virtual enumeration of reaction sequences and classifies generated product structures based on filtering criteria and fingerprinting calculations. A comparative PCA helps to identify the most dissimilar product pharmacophores from existing fragments, which serve as potential synthesis candidates. By iteratively expanding the pharmacophore coverage, we aimed to increase the diversity of the library. To evaluate this design concept, we used a well-known model system with an intermediate fragment library of 288 exclusive members, focusing on the identification and characterization of new sp^3-rich protease ligands.

4.1.2 PKA – Benzoic Acid

In another crystallographic fragment screen on PKA, benzoic acid was identified as a binder to the hinge region.[90] The interaction of the carboxylic group with the backbone consists of one oxygen atom of benzoic acid interacting with the backbone nitrogen of Val123 at a distance of 2.9 Å while the other oxygen atom of the carboxylic acid is 2.6 Å away from the backbone oxygen of Glu121 (Figure 4.1). The latter clearly indicates a protonation of the benzoic acid oxygen which is surprising at a crystallization pH of 6.9. The pK_a of benzoic acid in H_2O is 4.2, so it should be completely dissociated at pH 6.9.[91] Therefore, a protonation effect due to the interaction of benzoic acid with the hinge region can be assumed and this effect can be identified with ITC.

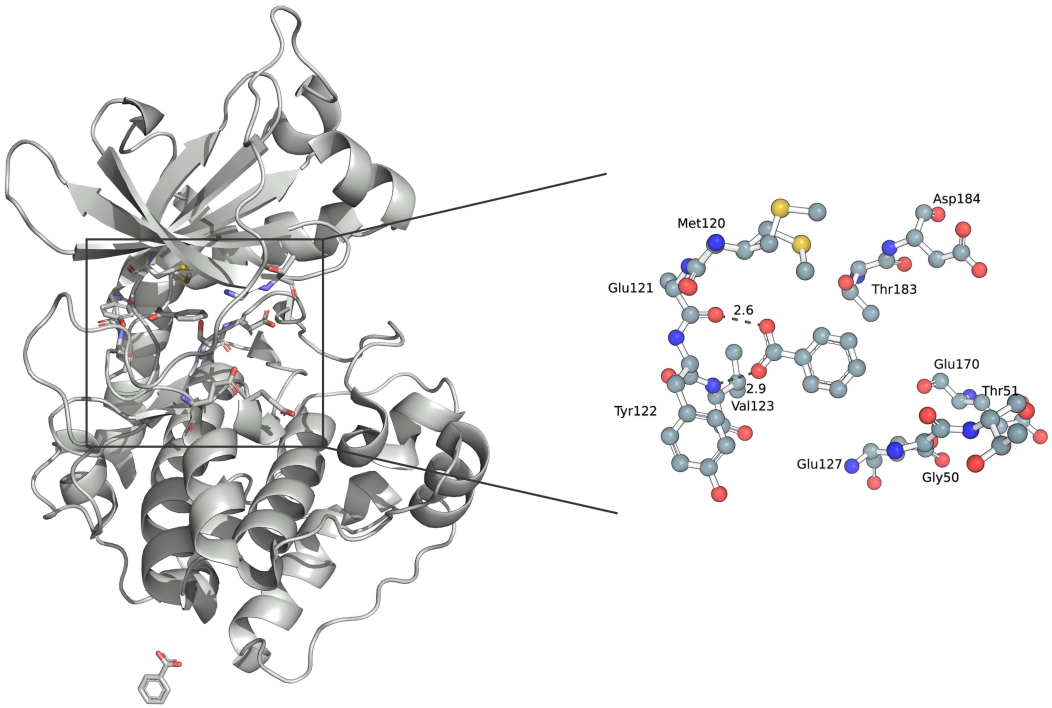


Figure 4.1: PKA-benzoic acid complex (PDB entry: 6SNN) with detailed view of the hinge binding site; distances in Å.

4.2 Results & Discussion

4.2.1 Fragtory

For the initial screening hit, a pyrrolidine-based fragment, no detectable contribution to the ligand binding is made by the bromopyridine residue that faces the solvent (Figure 4.2). There are no interactions between the catalytic dyad Asp35/Asp219 and the fragment, the conserved crystal water is still in place. In contrast, there is a water-mediated contact between the backbone nitrogen of Thr223 and the pyrrolidine nitrogen. In addition, hydrogen bonds are formed between the triazole nitrogen and the backbone NHs of Gly80 and Asp81.

As shown by Rühmann et al., the displacement ligand has an influence of the enthalpy-entropy profile of the fragments. Therefore, only the affinity of the fragments are discussed here, since the ΔG reveals a more robust parameter.[31] The ITC displacement titrations revealed a K_d of 352 µM and a c value of 31.

The next fragment examined 10b had the same triazole linker in the middle, but contained a piperidine instead of a pyrrolidine ring, and the methoxy-bound pyridine with a bromine substituent was replaced by a pyridine (Figure 4.3).

For this interaction, the ITC experiments revealed a K_d of 1.17 mM with a c value

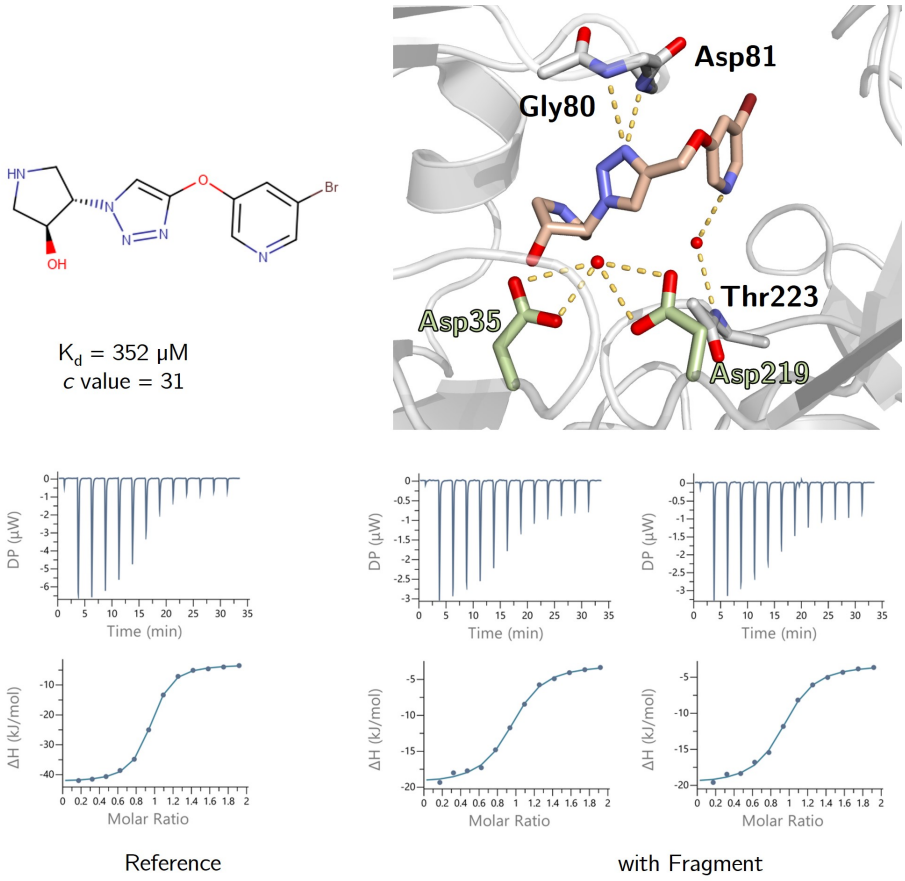


Figure 4.2: Molecular representation of fragment 1 and crystal structure of the EP-fragment complex (PDB: 8C70, top). ITC thermograms of EP (50 µM) and reference compound SAP114 (500 µM) and two displacement titrations where EP was pre-incubated with fragment 1 (500 µM) and titrated against SAP114 (bottom) with corresponding fits.

of 63.

The third investigated fragment 5b had the original pyrrolidine ring of fragment 1, but instead of the triazole, it was bound to a tetrahydroisoquinoline (Figure 4.4).

Here, the nitrogen of the pyrrolidine ring interacted with the catalytic dyad and replaced the crystal water, but remained the only significant interaction. The K_d of fragment 5b was 436 µM with a c value of 37.

When EP was incubated with fragment 10d (Figure 4.5) prior to displacement titration, the solution became cloudy, indicating protein precipitation.

Decreasing the fragment concentration while increasing the DMSO content to 5 % was also unsuccessful, although no precipitation was visible before starting the ITC experiment. Although the compounds were quite pure at >95 %, impurities may be responsible. The only differences between fragment 1 and fragment 10d are the

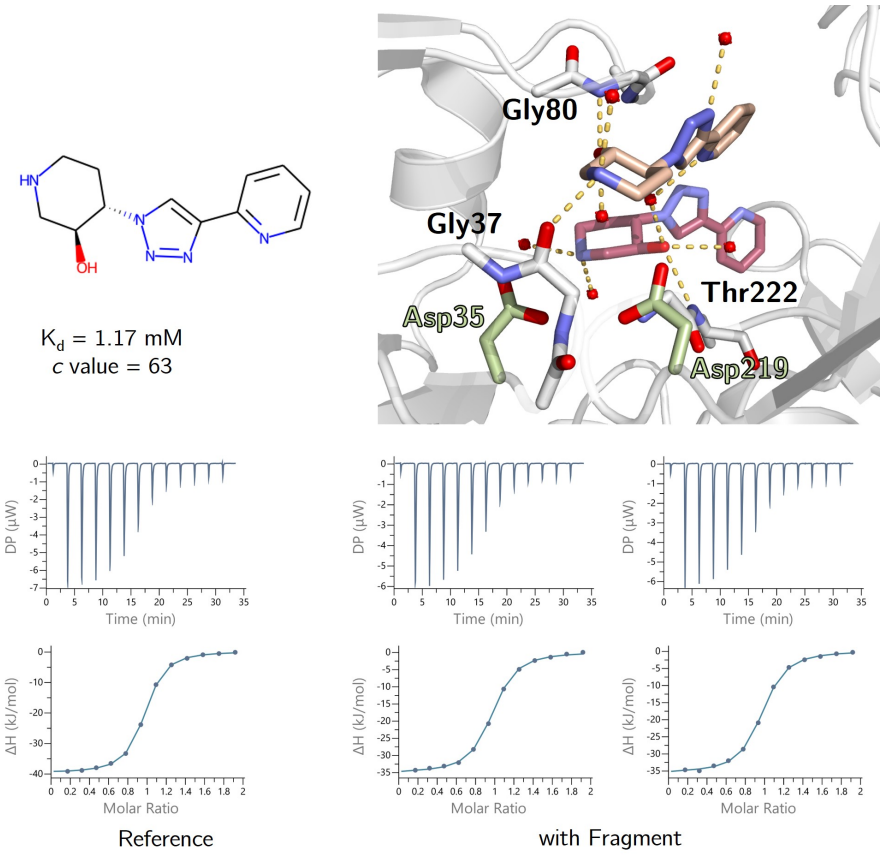


Figure 4.3: Molecular representation of fragment 10b and crystal structure of the EP-fragment complex (PDB: 8C72, top). ITC thermograms of EP (50 µM) and reference compound SAP114 (500 µM) and two displacement titrations where EP was pre-incubated with fragment 10b (500 µM) and titrated against SAP114 (bottom) with corresponding fits.

exchange of the bromopyridine moiety for aniline and that the amino group is in *para* position to the methoxy group while the bromine is in *ortho* position. The poses of the compounds in the crystal structure are almost identical. When predicting the pK_a values of fragment 10d, a small fraction of 11 % of the molecule has a double positive charge on both the amino group and the pyrrolidine-3-ol nitrogen, while 89 % is charged only on the pyrrolidine-3-ol nitrogen. For fragment one, only 3 % of the species is double positively charged, here at the bromopyrodine N and the pyrrolidine-3-ol nitrogen, while 97 % is single positively charged. Thus, fragment 10d increases the ionic strength slightly more than fragment 1, which could be a reason for the precipitation of EP.

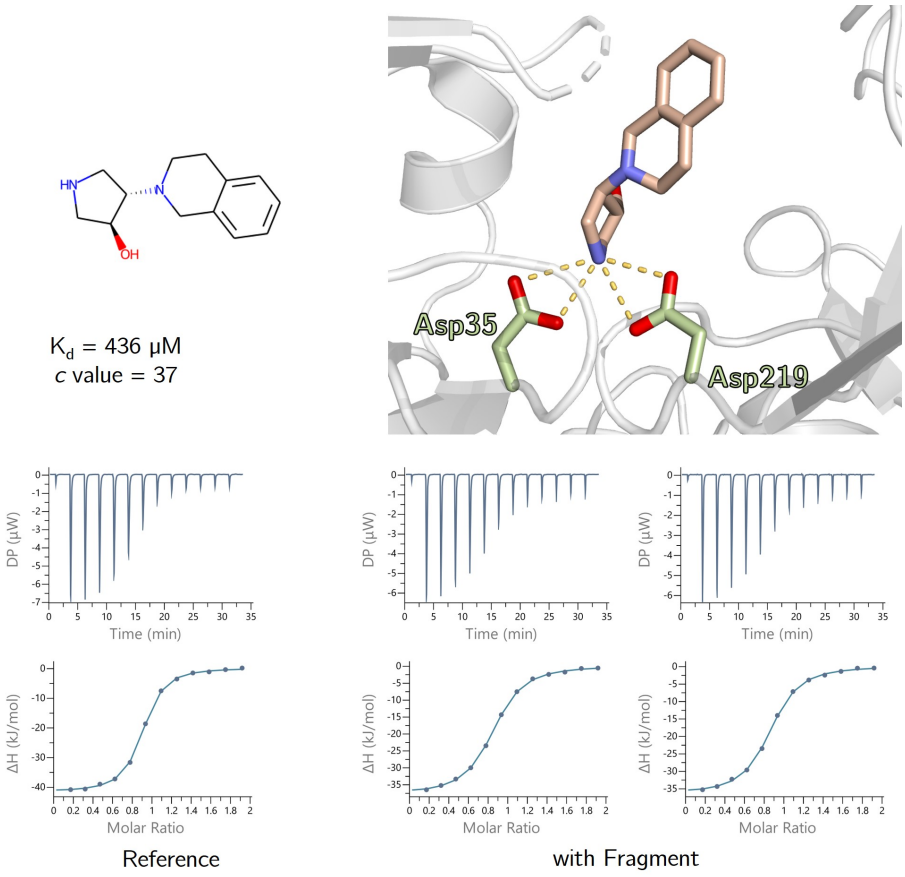


Figure 4.4: Molecular representation of tetrahydroisoquinoline-decorated fragment 5b and crystal structure of the EP-fragment complex (PDB: 8C71, top). ITC thermograms of EP (50 µM) and reference compound SAP114 (500 µM) and two displacement titrations where EP was pre-incubated with fragment 5b (500 µM) and titrated against SAP114 (bottom) with corresponding fits.

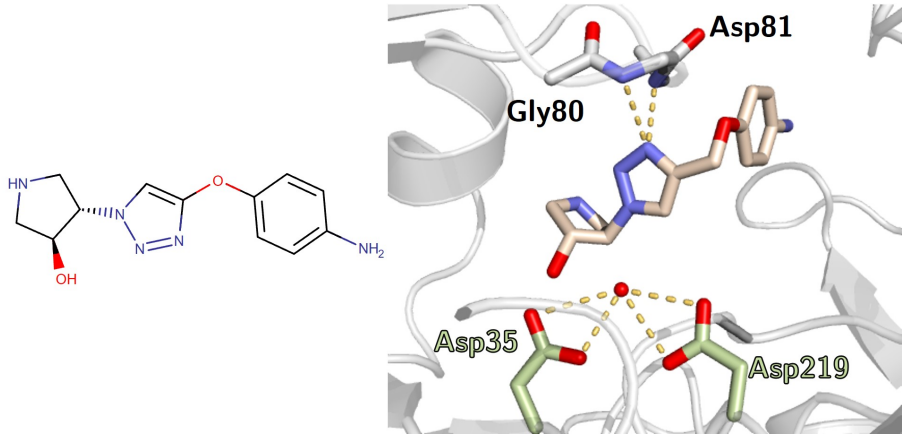


Figure 4.5: Molecular representation of fragment 10d and crystal structure of the EP-fragment complex (PDB: 8C74). This fragment could not be studied thermodynamically.

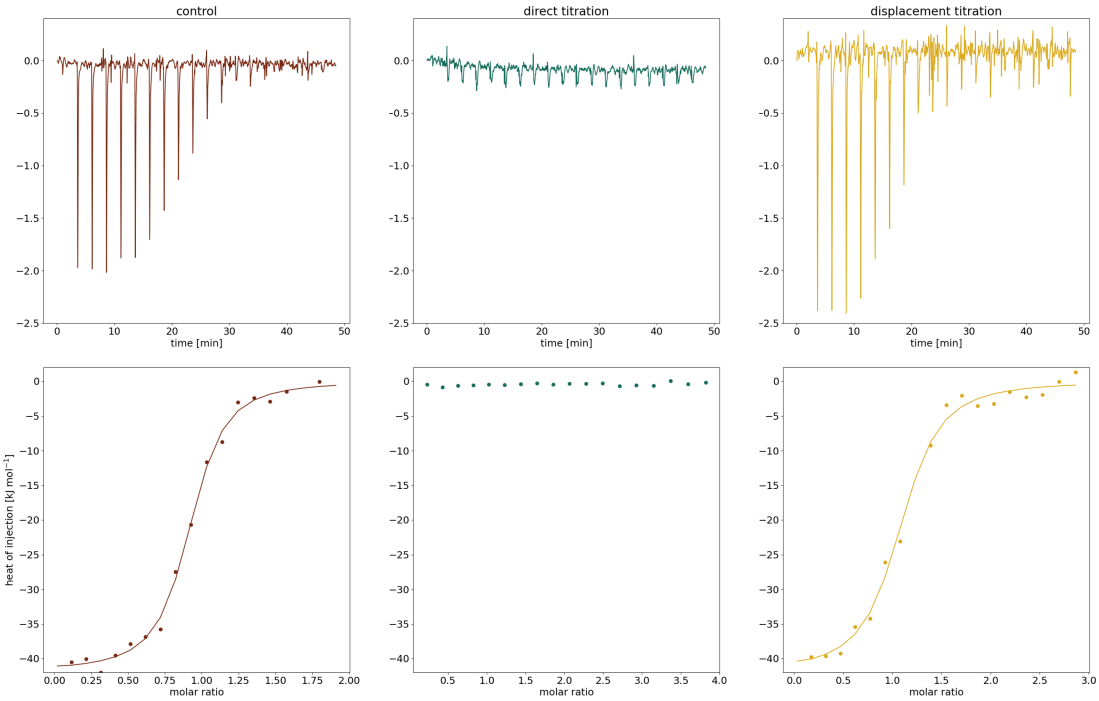


Figure 4.6: Overview of the different ITC experiments to investigate the PKA - benzoic acid interaction. Upper row: raw thermograms. Bottom row: integrated data of the heat signals observed for the measurements and corresponding fitted lines.

4.2.2 PKA – Benzoic Acid

A similar ITC approach was used to study the interaction of PKA with the benzoic acid fragment. However, in addition to the displacement titration with the high-affinity PKA inhibitor fasudil as displacement ligand, a direct low c value titration was performed. As a control, not only for the displacement titration, but also as a quality check as to whether the protein is active, the PKA was titrated with fasudil only. Although no thermodynamic parameters have been published for human PKA, they have been published for the homologous variant from *Cricetulus griseus*, which differs only at six terminal amino acids.[92] The result of this control measurement revealed a stoichiometry of $N = 0.9$, meaning that most of the protein is intact and active (Figure 4.6, left column). The overall ΔG of –37.3 kJ mol^{-1} is also in close proximity of the hamster variant (ΔG = –36.1 kJ mol^{-1}).[92]

In order to test the benzoic acid fragment, we firstly performed a direct titration of PKA and benzoic acid (Figure 4.6, middle column). The used concentrations were 140 µM for the protein and 2.8 mM for the benzoic acid, respectively. Unfortunately, a binding signal could not be detected.

Next, a displacement titration (Figure 4.6, right column) was carried out. For

this purpose, the protein concentrated at 20 µM was incubated with the fragment (4.2 mM) prior to the titration with strong binding fasudil (300 µM), the displacing ligand.

The differences in the thermodynamic profiles of PKA and fasudil and in the displacement titration are so small that they are within the error of the experiments.

There is no experimentally determined affinity for this interaction. However, SeeSAR, a computational tool can estimate the affinity based on the crystal structure and its integrated HYDE algorithm calculates an affinity in the range of 3.8–380 µM. At this range, the displacement titration could probably have confirmed a binding event with the used experimental conditions in the calorimetric measurements.

The equation 4.1 from Rühmann et al. was used to estimate whether the chosen concentrations were reasonable:

$$[fragment]_{cell} = \frac{D_{sat}[Protein] - [Protein] - K_{d\,fragment}}{1 - \frac{1}{D_{sat}}} \tag{4.1}$$

Since the K_d of benzoic acid is not known, we have to assume different affinities. Assuming a K_d of 1 mM corresponds to a saturation (D_{sat}) of about 80 %, while a lower affinity of 10 mM saturates only 30 % of the protein. As a rule of thumb, to achieve sufficient saturation of PKA or any other protein to be tested with the fragment during titration, the final concentration of the fragment added to the sample cuvette must be greater than its estimated K_d value.[31] In the case of benzoic acid, the concentration used would be correct for a hypothetical dissociation constant of 1 mM, whereas for a K_d of 10 mM the fragment concentration should be higher. However, at even higher fragment concentrations, more non-specific binding may occur.

The composition of the soaking solution that led to the structure is given in Table 4.1. This solution contains a significant amount of organic solvents, namely methanol, DMSO and 2-methyl-2,4-pentanediol (MPD), totalling more than 50 % (v/v). In addition, the final concentration of benzoic acid is the same as the buffering capacity of the solution, so a pH shift to a more acidic pH is likely. Therefore, this soaking solution was reconstituted and the pH was measured. And indeed, the pH value decreases 1.6 units, from 6.9 to 5.3. Since water is the most polar solvent in the soaking solution, and more apolar solvents like DMSO, MeOH and MPD tend to prefer an uncharged, in case of benzoic acid protonated form, we can assume that there is some undissociated acid in the solution, which is then soaked into the crystal. In contrast, in an ITC experiment, where the pH is constant at 6.9 due to the higher buffer capacity and much lower fragment and organic solvent concentration,

binding may not take place because the benzoic acid is completely dissociated and the negatively charged benzoate cannot form a hydrogen bond with the backbone oxygen of Glu121.

Table 4.1: Composition of the solution used for soaking benzoic acid into the PKA crystal.

Components	**v/v [%]**
100 mM MES-BIS-TRIS pH 6.9 75 mM LiCl, 0.03 mM Mega8, 1 mM DTT 0.1 mM EDTA, 23 % (v/v) MeOH	60
MPD	30
1 M benzoic acid in DMSO	10

Finally, an additional PB calculation was performed on the complex. The $\Delta \mathrm{p}K_\mathrm{a}$ of the benzoic acid carboxyl group was 1.2, thus shifting from 4.2 (experimental value) to 5.4. This, together with the pH shift due to the composition of the solution, makes a protonated benzoic acid molecule very likely.

4.3 Conclusion & Outlook

As we saw for the three of four fragments in the fragment screen with EP, displacement titrations work well for weak binders. For fragment 10b it could be that the enthalpies have equalized and nothing was seen in the thermogram. On the other hand, the fragment is too similar to fragment 1, as can be seen in the crystal structure, so the enthalpy-entropy profile should be similar. A direct, low c value titration could be performed. This in turn could cause other problems, including possible solubility issues, as a much higher fragment concentration would be required. It remains a challenging undertaking, especially if one wants to rely on entropically or enthalpically more favorable binders/fragments right from the start of drug development.

Due to the sufficiently high solubility the benzoic acid fragment investigated with its target PKA, both approaches were carried out, not revealing any interaction. In this specific case, the protonation state of the carboxy group is essential for the interaction. Structural data obtained from X-ray crystallography, can only provide information on the protonation states in the binding site in high-resolution crystals, while most crystallographic experiments do not resolve hydrogen atoms.[93]

Therefore, other approaches are needed. Protonation effects have been demonstrated in numerous studies using ITC experiments, but this might only be suitable for ligands with an affinity $K_\mathrm{d} < 1\,\mathrm{mM}$ and/or a high enthalpic contribution.[48, 94, 95]

However, for low affinity fragments, high resolution crystallography combined with neutron diffraction structure could be used.[96]

In general, X-ray (fragment) hits often cannot be validated by other experimental methods: for EP, almost half (44 %) out of 71 X-ray hits could not be confirmed by six different biophysical assays.[97] This is not surprising as the experimental conditions between X-ray crystallization and in vitro assays can vary greatly in terms temperature – often cryogenic in crystallization experiments, fragment concentration – mostly higher, and buffer composition – a high amount of organic solvents (e.g., DMSO, MeOH). Moreover, these different conditions can have an impact on the protonation state of a ligand which was recently discussed by Huang et al. who argue that high DMSO concentrations shift the pK_a of of ionizable groups involved in protein–ligand interactions.[98] The influence of organic solvents on the ionization states of small molecules needs to be further investigated. Therefore, the pK_a of benzoic acid in the crystallization solution will be determined experimentally.

4.4 Experimental procedure

4.4.1 Protein Preparation

The isolation of EP from Suparen (kindly provided by DSM Germany GmbH – Food Specialties, Düsseldorf) in 0.1 M NaOAc buffer pH 4.6 was performed as previously described.[99]

For PKA, the expression and purification was performed as follows: the PKA construct was ordered from Addgene, PCR-amplified with an N-terminal His_9-tag and an HRV 3C-cleavage site and cloned into the pET15b plasmid with NcoI/BamHI restriction site. The construct was then transformed into chemically competent *Escherichia coli* Rosetta2 (DE3) and expressed. Purification of the target protein was then performed by Ni affinity chromatography, dialysis with PreScission protease, reverse Ni affinity chromatography, cation exchange and finally size exclusion chromatography. Afterwards, PKA was concentrated to 15 mg ml^{-1}, frozen in liquid nitrogen and stored at –80°C until it was used again. Prior to ITC measurements, the required amount of PKA was thawed and purified again via a size exclusion chromatography column and concentrated to approximately 10 mg ml^{-1} in 30 mM HEPES, 100 mM NaCl, 10 mM $MgCl_2$ at pH 7.4.

4.4.2 Isothermal Titration Calorimetry

For the ITC measurements with EP, experimental set-up was similar as described previously.[97] Based on the reasoning of Krimmer and Klebe, the ITC protocol was adapted to 13 injections.[100] The same buffer used for isolation of the EP batch was used for every other solution needed for the ITC experiments. Affinities of the weakly binding fragments were determined with a strongly enthalpic binder SAP114 used as a displacement ligand.[101] A 500 µM SAP114 solution containing 0.1 M NaOAC (pH 4.6), 3 % (v/v) DMSO and 0.1 % polyoxyethylene-20-sorbitan monolaurate (Tween® 20) was titrated into a 50 µM EP solution of the same buffer that additionally was pre-incubated with 500 µM of the investigated fragments until a final stoichiometry of N=2 (SAP114:EP) was reached. For the reference titration, the same titrant was titrated into the buffer solution without the ligand.

ITC measurements with PKA were carried out in a buffer containing 30 mM HEPES, 100 mM NaCl, 10 mM $MgCl_2$, pH 7.4, 2 % (v/v) DMSO at 25°C. The titration protocol consisted of 19 injections, the first of which was 0.4 µl, followed by 18 titrations of 2 µl each.

All measurements were conducted on a MicroCal PEAQ-ITC Automated ITC (Malvern Panalytical). The obtained thermogram peaks of all titrations were integrated and fitted with MicroCal PEAQ-ITC Analysis Software 1.41. The dissociation constant K_d and the enthalpy ΔH of the fragments was determined as previously described.[32]

4.4.3 PBE Solver

For protein pK_a calculations a program from OpenEye based on the Zap finite difference Poisson-Boltzmann (PB) solver was used. For the partial charges of the protein, Delphi radii and CHARMM36 all-hydrogen partial charges were used. The benzoic acid fragment outside the hinge was removed from the structure. The am1bccsym method was used to assign appropriate charges to the benzoic acid. The pK_a of the ligand's carboxylic group was set to 4.2 according to the experimental value.[91] An inner dielectric of 15, an ionic strength of 0.05 M, and an ionization (i.e., pH) of 6.9 were used.

5. Protonation Effects in Protein-Ligand Complexes - Case Study of Endothiapepsin and Pepstatin A with Computational and Calorimetric Methods

5.1 Introductory Remarks

This chapter has been adapted from the eponymous paper available at bioRxiv.[102] The submission was made possible by contributions from Oscar Palomino-Hernández, Janis Müller, Phillip Galonska, Serghei Glinca and Paul Czodrowski.

5.2 Introduction

The correct determination of protonation states in a protein-ligand complex is crucial, since ionizable amino acids play a key role in the pH-dependent structure, binding interactions and catalysis.[34, 39, 103] From a protein perspective, 7 of the 20 proteinogenic amino acids contain an ionizable group, which is about 1/3 of the amino acid pool found in natural proteins. The charge state of these amino acids can be modulated either by the protein environment or by a bound ligand, or the protein environment can strongly influence ligands with ionizable groups. This in turn can influence the strength of drug binding or the structural properties of the protein.

From a ligand perspective, rational design of charge-assisted contacts between a protein and its ligand, guided by pK_a calculations of the respective complex has been shown to improve affinity, e.g., for a series of *lin*-benzoguanines binding to

tRNA-guanine transglycosylase.[48] Moreover, the uncharged form of a drug can cross biological membranes more easily than its corresponding charged form. One approach could be to design a molecule that remains uncharged during transport across membranes, but becomes charged by proton release upon binding to the target through electrostatic interactions.[48]

One prominent example of modulated protein charge states is HIV-1 protease (PR). HIV-1 PR is an aspartic protease (AP) and its catalytic dyad is made up of two aspartic acids. It has its pH optimum range at 4.0–6.0, at which one of these aspartates is protonated whereas the other aspartate is deprotonated.[104] If the pH value is lowered to more acidic conditions, both aspartates become protonated. Thus, different pH values can alter the protonation states of ionizable residues, which can lead to a conformational change on either the protein or the protein-ligand complex.[105]

To reveal the protonation state in a protein-ligand complex we can employ methods such as high resolution X-ray crystallography, or a combination of high resolution X-ray and neutron diffraction data (Figure 5.1), or isothermal titration calorimetry (ITC). The first two methods are capable of resolving structures with an estimated standard deviation of 0.02 Å in bond length, allowing the distinction between a C-O single bond (1.34 Å) of a carboxyl group with a proton attached and a C=O double bond (1.20 Å) of the carboxyl side chain of an aspartate.[106, 107] Moreover, the addition of neutron diffraction data improves the data-to-parameter ratio, and allows for the hydrogens to also be resolved.[96] The second method, ITC, is regarded as the gold standard for the investigation of intermolecular interactions in solution.[28] In addition to obtaining thermodynamic parameters such as enthalpy and equilibrium constant (from which changes in free energy and entropy can be calculated) this method is able to determine the overall changes in the protonation setup during the binding reaction.[100] This can be achieved by performing measurements in different buffers with different ionization enthalpies, and these protonation events have been demonstrated for a variety of enzyme classes, including aspartic proteases, serine proteases and methyltransferases.[48, 94, 108] However, ITC is not capable to retrieve the exact (at an atomistic level) localization of the protonation events.

In this work, we use EP, an exemplary aspartic protease, as a model enzyme to study protonation effects associated with ligand binding. The present protonation effect in the interaction of EP with the pan-protease inhibitor pepA was demonstrated by thermodynamic investigations using ITC measurements. Our results are compared with a published ITC study [94] and lead to new experiments such as crystal structure determination at physiological pH and novel computational approaches. These new

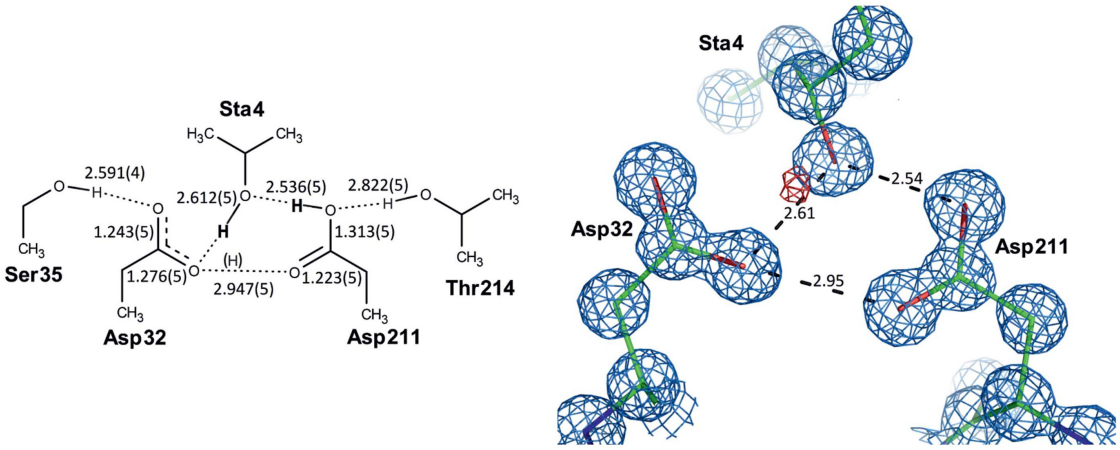


Figure 5.1: Polar interactions in the active site of the aspartic protease Sapp2p structure. *Left*: Schematic diagram with bond distances (values are in Å; estimated standard deviations of the distances are in parentheses). H atoms in bold are unambiguously assigned (the hydroxyl H atom of pepA was visible in the difference electron density; the H atom on the protonated carboxyl of Asp211 was deciphered from the C-O bond distances). The H atom in brackets is hypothetical. *Right*: Detailed structure of the active site in rod representation (PDB code: 4Y9W); hydrogen bonds are shown as dotted lines (numbers represent distances in Å). The $2F_o$-F_c electron density map contoured at the 1.5 level is shown in light blue; the F_o-F_c difference electron density map contoured at the 2 level is shown in red. Figure modified from [107]

experiments underline our experimentally (via ITC) determined protonation event upon pepA/EP complexation.

5.3 Results & Discussion

5.3.1 ITC Measurements reveal Proton Transfer Events

The total proton exchange (Δn_{H^+}) in a binding event can be determined by ITC measurements in different buffers with distinct heats of ionization ($\Delta H_{ionization}$), according to the formula 5.1:

$$\Delta H_{observed} = \Delta H_{corrected} + \Delta n_{H^+}(\Delta H_{ionization}) \tag{5.1}$$

Integrating the raw thermograms yields the experimentally observed enthalpies ($\Delta H_{observed}$), which can be used for a linear regression with the slope corresponding to Δn_{H^+} and the y-axis intercept corresponding to the buffer-corrected enthalpy of the binding reaction ($\Delta H_{corrected}$).

The ITC measurements investigating the interaction of pepA and EP were car-

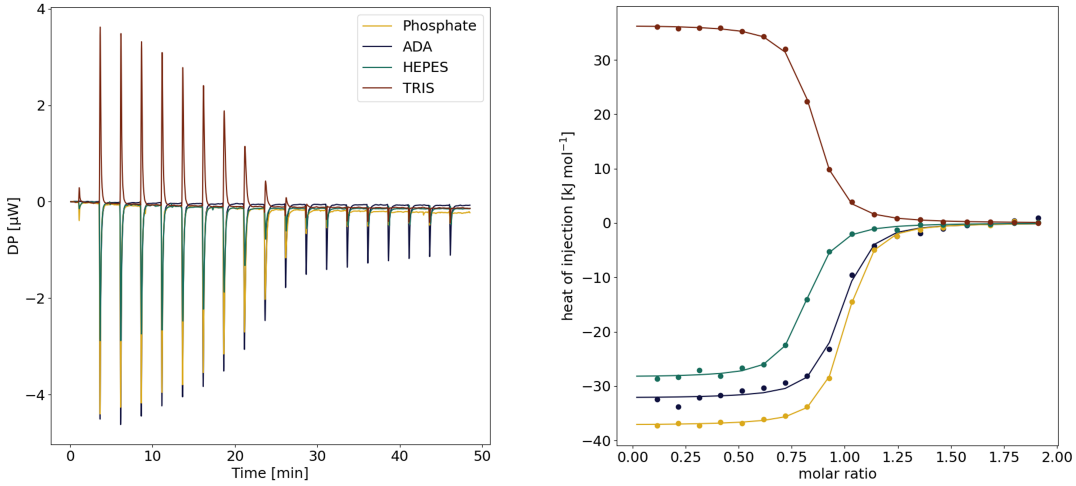


Figure 5.2: Overlay of raw thermograms of the binding reaction between EP and pepstatin A measured in phosphate, ADA, HEPES and TRIS at pH 7.0 (left); integrated data of the heat signals observed for the measurements in the four different buffers (right).

ried out in in phosphate, ADA, HEPES and TRIS buffers at pH 7.0 with ranging $\Delta H_{ionization}$ from 3.60 to 47.45 kJ mol^{-1} (Figure 5.2). A neutral pH was selected since the binding reaction was shown to be pH-dependent and at a lower pH, which would be more suitable for the mykotic enzyme, the dissociation constant is beyond the detection limit of the calorimetric device (K_d<10 nM). Moreover, pepA is not soluble at pH 4.6.[94] From the overlay of the raw thermograms we observe a strong buffer dependency. It is clear that the Δn_{H^+} is not equal to zero, otherwise the raw thermograms would be congruent with each other ($\Delta H_{obs} = \Delta H_{corrected}$).

The ITC measurements indicated an uptake of 1.67 ± 0.12 protons per binding reaction (Figure 5.3). The TRIS "outlier" is not an outlier in the narrower sense since it can be explained by its high positive ionization energy (47.45 kJ mol^{-1}), which leads to a positive enthalpic contribution (30.7 ± 2.9 kJ mol^{-1}) due to the proton uptake of the buffer.

This deviates - at least in terms of the absolute value - with the Δn_{H^+} previously published values of this interaction: 1.1 ± 0.1 protons revealed by ITC by Gómez and Freire.[94] In the following, we will compare our ITC setup with that one of Gómez and Freire, the major differences are given in Table 5.1.

Cacodylate was excluded as a buffering agent due to its toxicity. Measurements in ADA buffer were only performed here, while we did not perform measurements in ACES buffer. PepA was applied from a 20 mM DMSO stock solution, resulting in a final concentration of 2.5 % (v/v). This changed parameter could be a reason for the

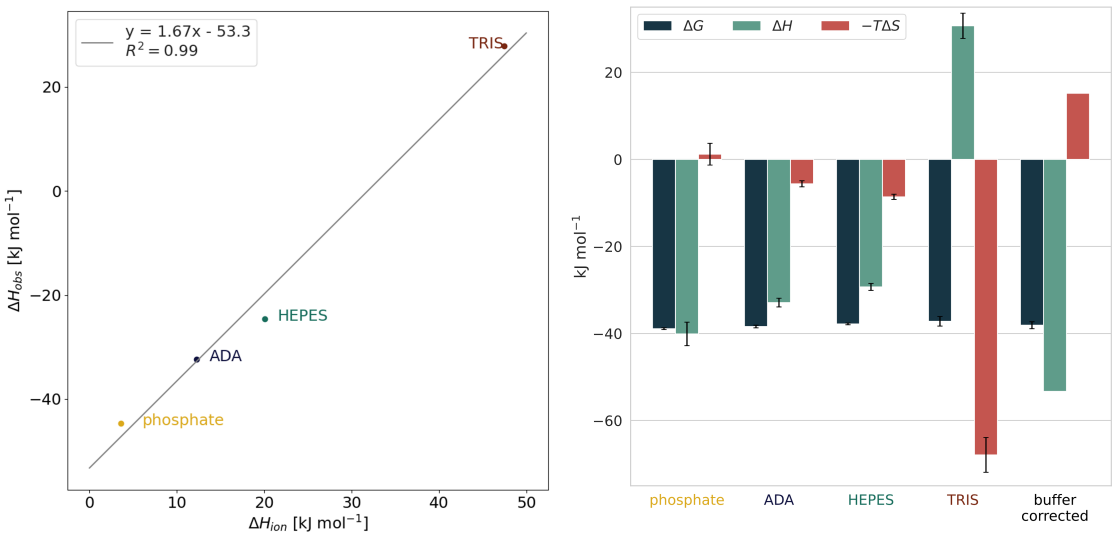


Figure 5.3: Left: Calculation of the heat of ionization. The experimentally observed enthalpies ΔH_{obs} are plotted against the heat of ionization ΔH_{ion} of the respective buffers (phosphate: 3.60 kJ mol^{-1} ADA: 12.23 kJ mol^{-1}, HEPES: 20.04 kJ mol^{-1}, TRIS: 47.45 kJ mol^{-1}[109]). The slope of the line describes the proton uptake during the formation of the protein-ligand complex (on average 1.67 ± 0.12 mol), while its intersection with the ordinate describes the buffer-corrected enthalpy of the binding reaction ($\Delta H_{corrected} = -53.3$ kJ mol^{-1}). Right: Thermodynamic profiles of complex formation in phosphate, ADA, HEPES and TRIS buffers and the buffer-corrected thermodynamic profile. For the buffer-corrected profile, the change in Gibbs free energy ΔG is calculated as the average of ΔG observed in the four buffers, ΔH is obtained as described above, and the entropic contribution –TΔS is calculated from the numerical difference between ΔG and ΔH.

Table 5.1: Experimental conditions used here compared to those used by Gómez and Freire[94]

Parameter	This study	Literature
c[EP]	50 µM	80 µM
c[pepA]	500 µM	1200 µM
c[Buffer]	100 mM	20 mM
Temperature	25°C	16°C
pH	7.0	7.0
DMSO (v/v)	2.5 %	-
Tween20 (v/v)	0.1 %	-
Cell volume	0.2 ml	1.36 ml
Used buffers	ADA, HEPES, Phosphate, TRIS	ACES, Cacodylate, HEPES, Phosphate, TRIS

different results, since it has been shown for several protein-ligand complexes that a higher concentration of DMSO in the samples leads to a decrease in affinity, even at a low concentration of 1 % (v/v).[110] 2.5 % (v/v) DMSO corresponds to a molar concentration of 0.35 M.

Furthermore, different experimental conditions could have affected the proton transfer:

- higher temperature
- concentration of the reagents and buffer/salt
- the presence of DMSO as solvent (already described in the previous paragraph)
- the calorimetric device used

Each parameter alone can have a clear effect on the thermodynamic properties.[100] We used the MicroCal PEAQ-ITC Automated ITC with a cell volume of 200 µl; 25°C is a standard temperature, where most of the experiments are carried out, not only ITC. We followed standard protocols for protein, ligand, and buffer concentrations.[97]

Another parameter influencing the thermodynamic properties with respect to K_d, ΔG, ΔH and $-T\Delta S$ is the protein concentration used. For a ligand binding to thermolysin with an affinity in the low micromolar range, the determined K_d increases with protein concentration by an order of magnitude from 0.86 µM at the lowest protein concentration of 50 µM to over 14 µM at 300 µM thermolysin.[100]

The affinity of pepA is pH dependent and becomes higher at more acidic pH, with a $K_\mathrm{d} < 10$ nM for pH ≤ 5, reaching the upper detection limit of the ITC instrument. This could be another explanation for the deviation between our ITC results and those published by Goméz and Freire[94].

The stoichiometry of the EP-pepA binding reaction is illustrated in Figure 5.4. There is a noticeable difference between the stoichiometry in the respective buffers. Assuming that pepA was taken from the same stock solution, i.e., the ligand concentration is correct, this could indicate that EP is more unstable/less active in HEPES: over 90 % activity ($N \geq 0.9$) in phosphate versus 70% activity ($N \approx 0.9$) in HEPES, since crystallization and previous calorimetric measurements show a 1:1 binding of the ligand to the active site of the protein. At least in TRIS, the results suggest an increasing degradation of the protein: N decreased from above 0.8 to below 0.6 after about 10 hours of exposure at room temperature (RT), according to the batchwise measurements in the automated ITC device. It should be noted that the ITC is able to cool the samples in the tray prior to titration, but there have been no stability problems with EP at RT. We assume that the different stoichiome-

tries should not affect the overall protonation event: it will be constant throughout different measured stoichiometries.

However, EP as fungal enzyme has its optimum at an acidic pH, and most of the published experimental data, whether crystallographic or calorimetric, have been collected almost exclusively in an acetate buffer at pH 4.5–5.0, except for the study by Gómez and Freire.[94] Thus, pH of 7.0 together with temperature (25°C versus 16°C) may have caused the decrease in activity of EP.

The available EP crystal structures in the PDB are exclusively available at an acidic pH of 4.5–5.0 due to the stability of EP. In contrast, the ITC experiments were performed at pH 7.0 which might have affected protein conformational changes. Therefore, we aimed for the determination of EP X-ray at pH values in the neutral range. Additionally, we used two different computational approaches to interpret the ITC data from a structural point of view, because the ITC data itself does not reveal the exact atomistic location of the protonation event.

Figure 5.4: Overview of stoichiometry N of ITC measurements of the binding reaction between EP and pepA in different buffers of 100 mM concentration at a pH of 7.0.

5.3.2 Novel Crystal Structures at higher pH Values reveal no major Conformational Changes

In order to have a structure at the same pH as the calorimetric measurements, EP and pepA were co-crystallized and soaked for 24 hours in soaking conditions with buffers of pH 7.6 and analyzed. The resolution was 1.23 Å, higher than the published 2.0 Å structure (PDB entry: 4ER2), and the all-atom RMSD between the pH 7.6 structure and 4ER2 amounts to 0.89 Å (Figure 5.5). However, the biggest variation within the binding site was located at both termini, where the α-carbon of the the isovaleryl residue at the N-terminus and the C-terminal α-carbon of statine is "flipped" (pepA sequence: Iva-Val-Val-Sta-Ala-Sta). This is because soaking at pH 7.6 results in a modified orientation compared to the structure in the acidic range.

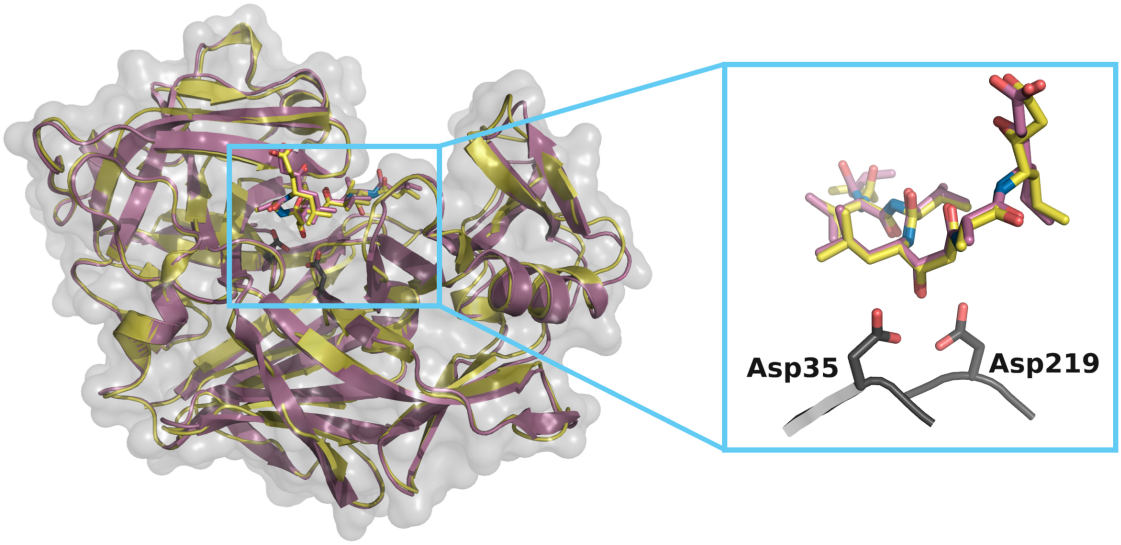


Figure 5.5: Comparison of EP-pepA complexes crystallized at pH 4.6 (purple, PDB entry 4ER2) and soaked at pH 7.6 (yellow, PDB entry 9GFY) with detailed view of pepA in the binding site, with the flip at the N-terminal isovaleryl residue (rear left) and the alternative conformation at the C-terminal statine (upper right).

Although there is a tiny residual density for the "acidic" conformation at pH 7.6 (i.e., a second conformation is observed for the crystal structure at different pH values) and vice versa, these are negligible.

5.3.3 Implicit Solvent pK_a Calculations allow for a structural Interpretation

As a starting point, we used HIV-1 protease (HIV-1 PR) to calibrate our Poisson-Boltzmann (PB) methodology. The pK_a values resulting from the protein pK_a calculations for HIV-1 PR demonstrate that our calculations are in very good agreement to the experimental values: while Asp25 carboxylic side chain is the more acidic one with a pK_a of 3.2, Asp25' carboxylic group has a pK_a of 5.7. The corresponding experimental values are 3.1 and 5.2 (Table 5.2).

Table 5.2: Results of the pK_a calculations on HIV-1 PR (PDB entry: 1XL5).

Method	Asp25	Asp25'
Experiment[111]	3.1	5.2
ZAP-based PBS	3.2	5.7

The structural similarity between HIV-1 PR and EP in the active site is evident, the distances between the aspartates are almost identical (Figure 5.6). This convinced us to apply the HIV-1-PR validation protocol also for EP.

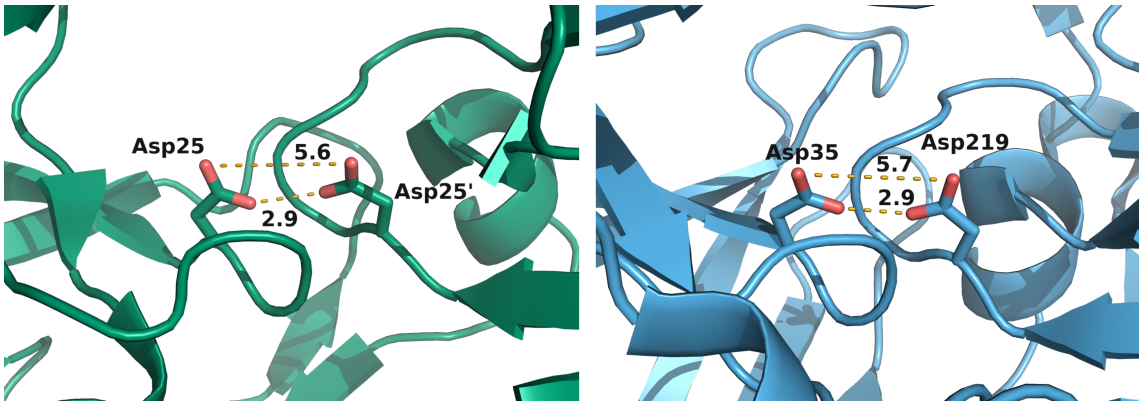


Figure 5.6: Comparison of the active site of HIV-1 PR ligand-deleted (left, PDB entry: 4EJ8) and EP (right, 4Y5L) apo structures. Distance between the relevant atoms in Å.

Next, we carried out pK_a calculations with our PB solver on the three different EP-pepA structures (Table 5.3). The C-terminal carboxy group of pepA was titrated analogous to the titratable amino acid side chains. The only slight pK_a shift of the pepA carboxy group can be explained by the solvent exposure and the associated lack of interactions with amino acid side chains. The overall exchange of protons transferred during the binding reaction (Δn_{H^+}) is 0.79 for the structure 4ER2 and 0.5 for the pH 7.6 structure. In all three structures, the three amino acids Asp33, Asp35 and Asp 219 are mainly responsible for the shift. While Asp81 shows a delta ΔpK_a of 0.5 to 0.6, it does not lead to a protonation shift at an assumed pH of 7.0, since the aspartate is more acidic than the other three. The calculated Δn_{H^+} of the complex at pH 7.6 was 0.5, probably due to the structural changes mentioned above.

Table 5.3: Results of the pK_a calculations on the EP-pepA complexes crystallized and afterwards soaked at different pH and the resulting protonation change Δn_{H^+} (in mols) at pH 7.0 (the pH value of the ITC experiments). The ligand pK_a is the C-terminal carboxy group.

Crystal pH		**Asp35**	**Asp219**	**Asp33**	**Asp81**	**Ligand**	$\sum \Delta n_{H^+}$ **(pH 7.0)**
4.6	Ligand-deleted	5.8	5.5	6.3	4.6	4.32	
	Complex	6.4	7.0	6.8	5.2	4.4	0.79
	Δn_{H^+}	0.14	0.47	0.22	0.02	0	
7.6	Ligand-deleted	6.2	5.5	6.4	4.6	4.32	
	Complex	6.3	7.0	6.7	5.1	4.2	0.5
	Δn_{H^+}	0.03	0.47	0.13	0.01	0	

A closer look was then taken at the binding pocket with pepA in the active site of EP, and in particular at the interacting aspartate residues (Figure 5.7). The structures differ by a maximum of 0.2 Å. The most prominent interacting groups are the carboxylic side chain of both Asp35 and Asp219, the catalytic dyad, with the hydroxyl group of statine 4 of pepA. While the distance between the side chains oxygens of Asp33 and Asp35 is too high to form a hydrogen bond (6.3 Å), one oxygen is in distance of 3.6 Å to the acylic side chain of the pepA ligand. This interaction is probably responsible for the slight pK_a shift of Asp33.

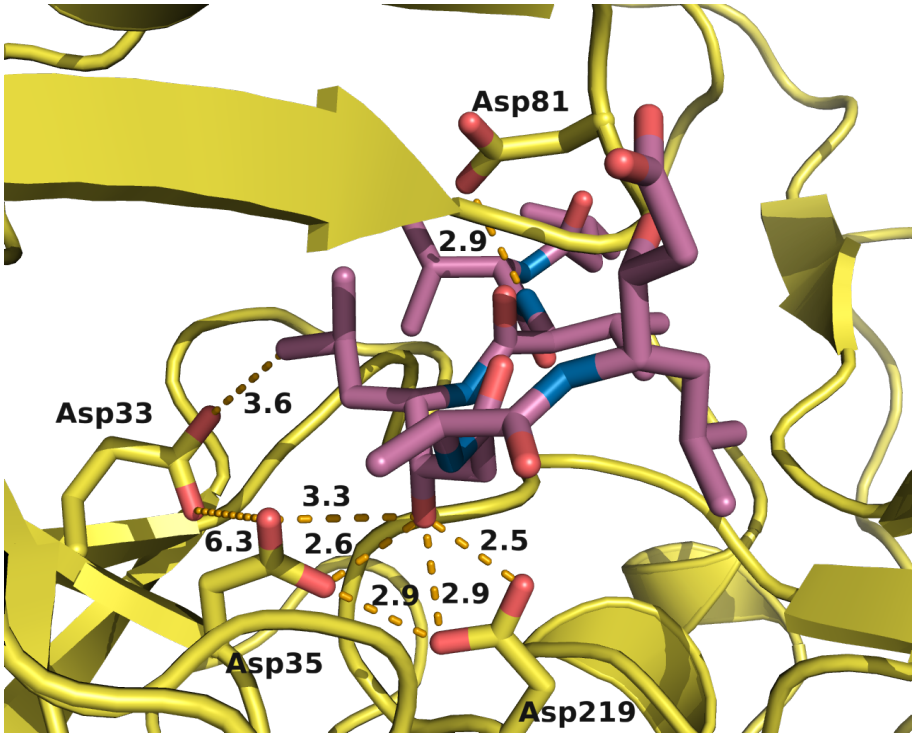


Figure 5.7: pepA (purple) in the active site of EP. The three carboxylic side chains of Asp33, Asp35 and Asp219 are the main contributors to the protonation effect revealed with PB solver. Asp81 is the fourth Asp within the binding site, but not part of the protonation effect according to the calculations. Distance between the relevant atoms in Å (PDB entry: 9GFY - pH 7.6 structure).

Most of the protonation effect based on our calculations occurs at the catalytic dyad. Since Asp35 and Asp219 behave like a dicarboxylic acid due to the given structural conditions and the small distance between them, Asp219 is most likely protonated at pH 7.0, whereas no change for the catalytic dyad occurs at pH 4.6 during complex formation (Figure 5.8).

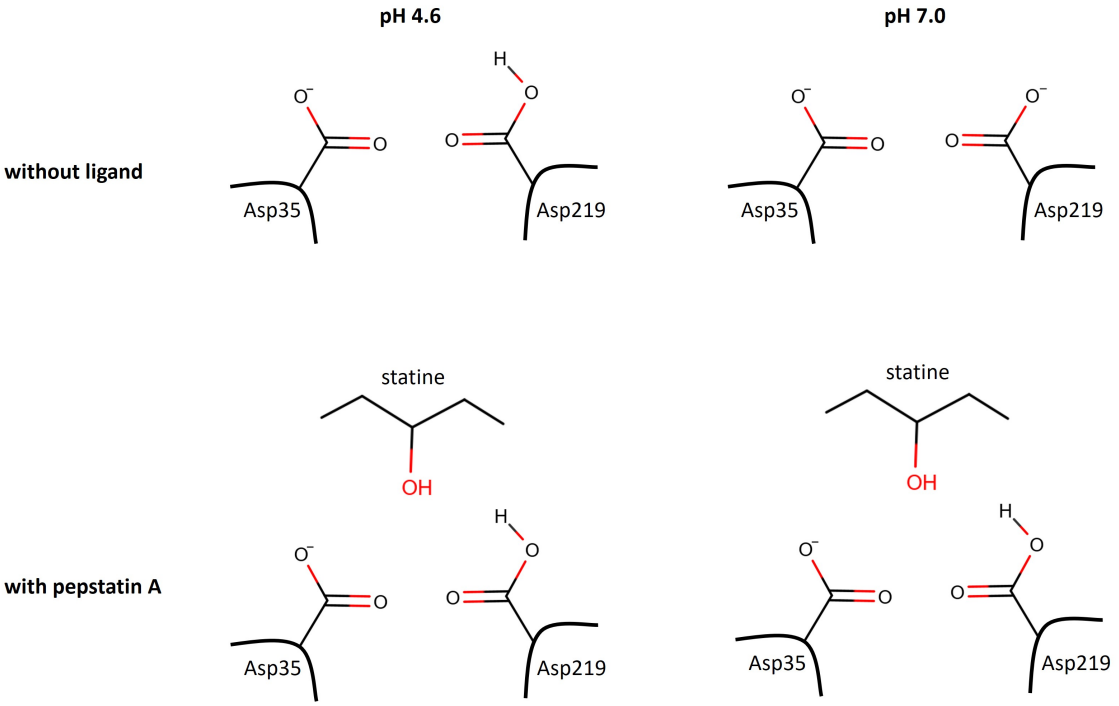


Figure 5.8: Overview of the hypothesized protonation states of the catalytic dyad of EP without ligand at pH 4.6 and 7.0, and with pepA at both pH values. At pH 4.6 the catalytic dyad is monoprotonated without ligand and with pepA, while at pH 7.0 both Asp are deprotonated, but Asp219 takes a proton upon binding of pepA.

5.3.4 Constant-pH MD Simulations further underline the observed Proton Transfer Event

Intrigued by the above results, we set out to understand the behavior of the selected aspartates through constant pH molecular dynamics (CpHMD) simulations. CpHMD is well-suited for investigating molecular systems characterized by pH-dependent dynamics and functions.[112] Thus, we performed simulations in both apo and pepA-bound EP at different pH values.

From a structural point of view, the root mean square deviation (RMSD) of apo and pepstatin-bound EP with respect to the crystal structure of 4ER2 do not show a strong dependence on the pH of the environment (Figure 5.9). Overall, unliganded EP shows higher RMSD values that fluctuate strongly. In comparison, pepstatin-bound EP shows remarkably lower RMSD values, as well as a reduced amount of fluctuations.

Focusing further on C_α fluctuations (RMSF), we observed that the five top regions with the largest contribution to the RMSF correspond to loop-related motions, with no significant change on structured regions. In the pepstatin-bound structure, the motion of the loop covering residues 70–90 (commonly known as the binding flap, [113]) is strongly reduced, likely due to the stabilization with the binding partner. Moreover, we found that there are no large scale conformational motions that show dependence to the pH conditions.

In the previous section, our results suggested that a protonation event should take place upon binding, and that it should be connected to the aspartic acid residues. Thus, we evaluated the protonation propensities at different pH values for the aspartates within the binding site (i.e., Asp33, Asp35, Asp81 and Asp219), as they were the ones closer to pepA. With the protonation propensities, we obtained theoretical deprotonation curves and computed the theoretical pK_a (Figure 5.10).

Indeed, the fraction of deprotonation curves suggest that the binding event shifts up the pK_a of all of the identified aspartates. Moreover, we can observe a high shift in the case of Asp219, from a pK_a of 3.6 to a pK_a of 7.6. This would also support the previous findings: that at pH 4.6 and 7.0, there is a change on total protonation (Δn_{H^+}) close to one. At higher pH values such as pH 7.6, this effects become less prominent.

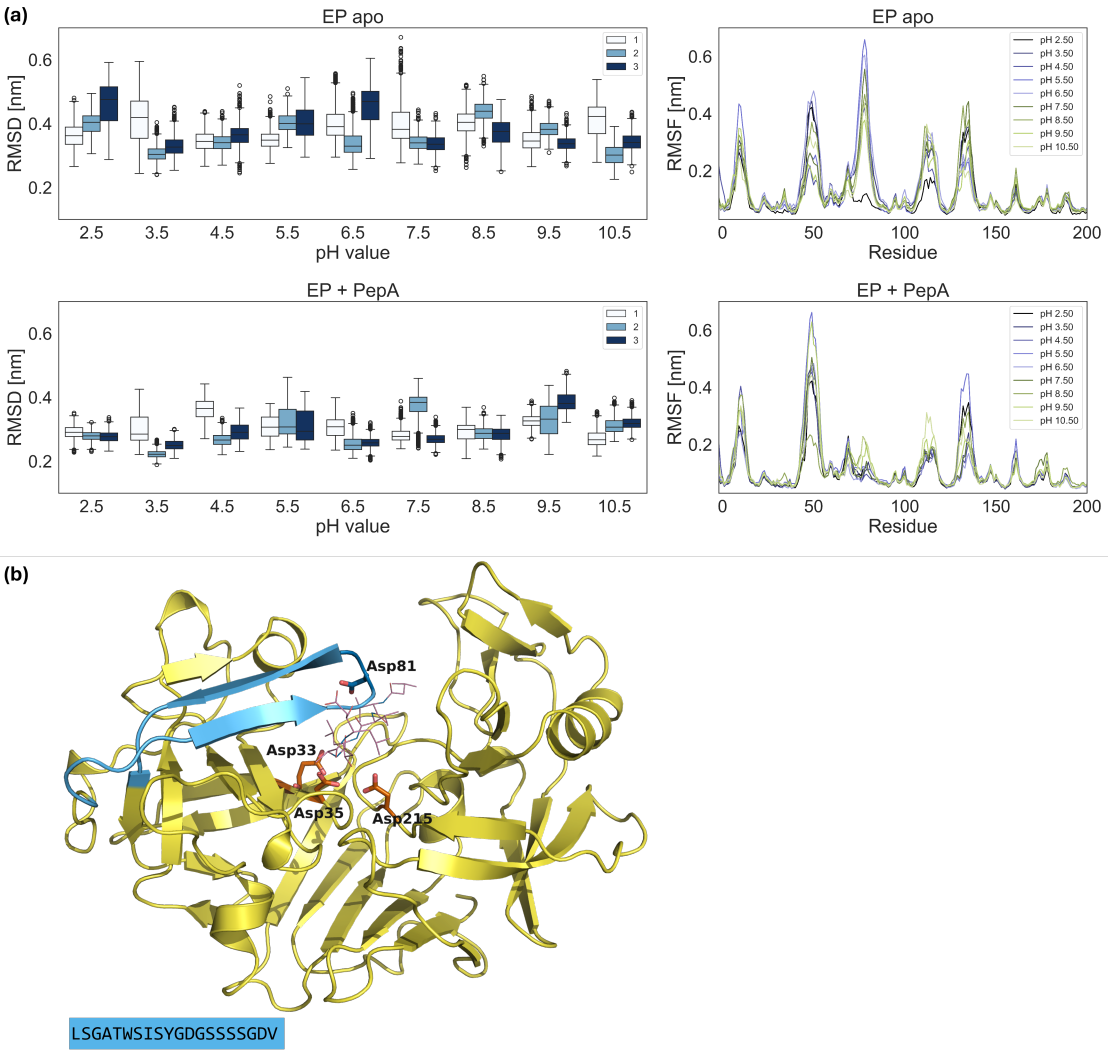


Figure 5.9: (a), *Left*: Root mean square deviations (RMSD) of EP alone ("EP apo") and in presence of pepA ("EP + PepA") at different pH values, with respect to the crystal structure. The boxes represent the interquartile range (from 25th to 75th percentile), while the whiskers extend from the box within 1.5 * IQR, excluding outliers. Labels show the simulation repeat. *Right*: Averaged root mean square fluctuations (RMSF) of the C_α of each residue at different pH values. (b): structure of the EP complex with residues 70-90 and the corresponding sequence highlighted in blue.

5.4 Conclusion & Outlook

We were able to detect a proton transfer event for the complexation of pepA to EP. This was found in ITC experiments which revealed an overall proton uptake of 1.67 ± 0.12 moles of protons. This deviates from a published study [94] and can be explained by differences in the experimental setup. However, both ITC studies (our own and the published one [94]) were run at a pH of 7 which is significantly different from pH of 4.6 at which crystal structures exclusively were determined so far.

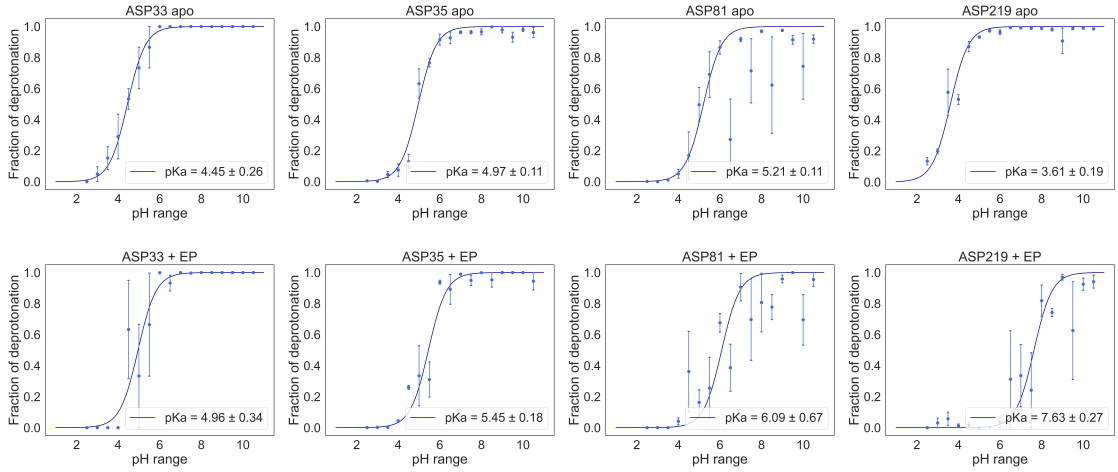


Figure 5.10: Titration curves of EP obtained from constant pH MD simulations for the single protein (top) and with pepA (bottom). For each of the four selected residues in this protein, the dots show the fraction of conformations in which the residue was deprotonated. Errors were estimated from the standard error of the mean for the three different replicas. The lines show the best fits to the Henderson–Hasselbalch equation. The pK_a values for each titratable residue are listed.

This prompted us for follow-up studies such as a novel protein crystal structure determined at higher pH. Its analysis did not reveal any major conformational changes and therefore rules out its impact on the experimentally determined proton transfer event. Furthermore, this outcome justifies the usage of the already existing crystal structure (pH 4.6) as starting point for pK_a calculations.
These calculations were performed in an implicit and explicit solvent manner and both agree in the structural interpretation of the proton transfer event. Additionally, constant pH molecular dynamics simulations show that the different pH conditions do not strongly affect the conformational dynamics of EP. This is in accordance with the novel crystal structures and underlines the power of this simulation technique.
The findings from our study underline the necessity to crucially inspect possible proton transfer events. If those are not correctly accounted for the follow-up structure-based studies might easily evolve in an undesired direction.

5.5 Experimental Procedure

5.5.1 Protein Purification

The isolation of EP from Suparen (kindly provided by DSM Germany GmbH – Food Specialties, Düsseldorf) in 0.1 M NaOAc buffer pH 4.6 was performed as previously described by Köster et al.[99] For the experiments performed at a pH of 7.0, EP

was dialyzed using the Slide-A-lyzer® G2 cassette (ThermoFisher, Waltham, MA, USA) with a 10 kDa molecular weight cutoff. The dialysis was performed in a 600 ml beaker. The membrane of the cassette was soaked in the dialysis buffer for 2 min by holding it (with gloves) in the beaker filled with dialysis buffer at pH 7.0 (TRIS, ADA, HEPES and phosphate, respectively). After pre-soaking, the cassette was carefully dried with a paper towel to prevent buffer from entering the cassette, but without touching the membranes. 3 ml of at least 4 mg ml^{-1} protein solution was added to the cassette using a 200 µl micropipette. When the cassette was closed with the stopper, any remaining air was gently evacuated from the dialysis chamber. The protein was then dialyzed against 500 ml of dialysis buffer for 2 hours at RT with slow stirring. The buffer was then replaced with 500 ml of fresh buffer and stirred again for 2 hours at RT. After another change of buffer, the dialysate was stored in the refrigerator overnight (12-20 hours). The next day, the protein solution was removed using a syringe with a long cannula. The protein was stored on ice before being shock frozen in liquid nitrogen to determine the concentration using NanoDrop™2000c (Thermo Scientific, USA).

5.5.2 Protein Crystallization

The protein was crystallized using an adapted protocol by Köster et al.[99] The protein was concentrated to 5 mg ml^{-1} and stored in 100 mM sodium acetate pH 4.6. Crystals of suitable size were grown in 4 µl drops consisting of 1:1 protein:reservoir solution, equilibrating against 1 ml of 100 mM sodium acetate pH 4.6, 100 mM ammonium acetate, 24 % PEG4000 on Cryschem plates (Hampton Research, USA) at 19°C. More reliable crystal growth could be achieved using streak seeding. Crystals suitable for data collection usually grew in 1-2 weeks. Co-crystallization with pepA was performed similarly, but solid pepA was first immobilized on the crystallization plates by evaporating 2 µl of a saturated pepA DMSO stock (<100 mM) and then crystallization drops were set up on top of the solid. A soaking condition with pH 7.6 was prepared with a saturated concentration of pepA. Then, two crystals grown with pepA were soaked per condition for 24 hours. Crystals were then harvested from the drops with nylon loops and flash frozen with liquid nitrogen.

Data was collected at DESY P11.[114] The data was processed with XDS.[115] The structure was solved via molecular replacement and iterative cycles of model building in Coot[116] and refinement in Phenix[117] were performed.

5.5.3 Isothermal Titration Calorimetry

ITC measurements were performed using the MicroCal PEAQ-ITC Automated ITC (Malvern Panalytical, Worcestershire, UK). A 500 µM pepA solution containing 0.1 M of the different buffers (HEPES, TRIS, ADA, phosphate) of pH 7.0, 2.5 % (v/v) DMSO and 0.1 % (v/v) Tween20 was titrated into a 50 µM EP solution of the same buffer to a final stoichiometry of $N = 2$ (pepA:EP). For the reference titration, the same titrant was titrated into the buffer solution. The obtained thermogram peaks of all titrations were integrated and fitted with MicroCal PEAQ-ITC Analysis Software 1.41.

5.5.4 Implicit Solvent pK_a Calculations

The protein pK_a prediction was performed using a program based on the Zap finite difference PB solver by OpenEye, Cadence Molecular Sciences.[34, 118] As input, the workflow required a PDB file of the complex of interest and starting values for the pH and the pK_a of the ligand. The initial pK_a of pepA was predicted using ChemAxon's Marvin pK_a plugin.[119] The ligand was then extracted from the original PDB file and hydrogens were added. Then, partial charges were assigned to both microspecies using the AM1BCC algorithm using the SZYBKI program by OpenEye, Cadence Molecular Sciences.[120] For the partial charges of the protein, Delphi radii and CHARMM36 all-hydrogen partial charges were used. An internal dielectric of 15, an ionic strength of 0.05 M and a pH of 7.0 were used. Reference pK_a values were as follows: Asp 4.0, His 6.5, Glu 4.5, Tyr 9.8, Lys 10.4 and Arg 12.5. Cys residues were not titrated and modeled in a neutral, protonated state. After protein and ligand were merged into one file and missing hydrogens were added to the protein, the processed microspecies, meaning the protonated and deprotonated form, were saved in a text file, which was used as input for subsequent steps. For the titratable amino acid residues listed above and the ligand, titration reactions were performed and pK_a values were calculated. All hydrogen atoms were explicitly modelled, and except for the orientation of the OH protons, which were sampled in 10° steps, the rest of the structure was static. Optimization of the ionization state and OH orientation was achieved by applying ten million Monte Carlo steps.[118]

For the pK_a calculation of the two aspartates in the catalytic dyad in the HIV-1 PR, the initial results were post-processed due to the electrostatically strongly coupled system: the two residues are in close proximity to each other at about 3Å and therefore behave more like a diprotic acid (e.g. maleic acid). This has already been described by Czodrowski et al.[49]

5.5.5 Constant pH Molecular Dynamics Simulation

The initial structure of bound EP was taken from the PDB data bank (PDB ID: 4ER2 [121]) For the apo structure, the ligand was removed from the binding site manually. Excepting the aspartates, standard protonation states for all side chains were assumed through an in-house script. Both apo and pepstatin-bound structures were solvated in truncated octahedron water boxes with 1 nm between the protein and the boundaries of the box.

Constant pH molecular dynamics (CpHMD) simulations were performed using the GROMACS 2021-dev-beta implementation, where the constant pH approach was implemented by leveraging continuous lambda dynamics.[122] The proteins were represented using the implementation of Buslaev et. al. of CHARMM36m all-atom, due to the improved dihedral sampling in constant-pH simulations.[123] Missing parameters were obtained through the CHARMM Force Field server.[124] Ionic strength was set to 0.15 mM. The systems went through energy minimization, 1 ns heating with the V-rescale thermostat[125] to 300 K , 1 ns pressure coupling with the C-rescale barostat[126] to 1 atm. For the van der Waals interactions, a switching function of 10 to 12 Å was used. A real-space cutoff of 12 Å was applied in the electrostatics calculations with the particle-mesh Ewald method.

In the production run, the lambda dynamics is allowed, effectively exploring the protonation/deprotonation space of the selected residues at a given pH. In the CpHMD simulations, the side chains of the selected aspartates were allowed to titrate. The simulation of both apo EP and EP+pepstatin used 17 pH replicas in the pH range 2.5 to 10.5, with a spacing of 0.5 pH units, and 300 ns per pH-replica. Each system was simulated by triplicate to reduce dependency from the starting seed. The initial 50 ns from each replica were discarded in the analysis.

6. Protein pK_a Prediction on various Protein-Ligand Complexes

6.1 Introductory Remarks

The results of this chapter emerged from Luca Kröll's Master's thesis and are accompanied by figures, tables and, where available, comparisons with experimental data from the literature.

6.2 Introduction

As already mentioned in Chapters 2.3.1 and 5, respectively, the protonation states of proteins and ligands can be crucial in determining how they interact. Here, we focused on protonation effects in protein-ligand complexes of the aforementioned APs EP and HIV-1 PR.

6.3 Results & Discussion

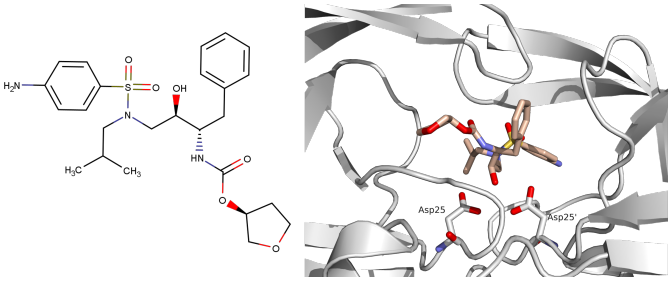
6.3.1 HIV-1 Protease

The wrapper around the protein pK_a prediction tool Zap has been made technically operational. It was to be tested on the basis of published data; a study in which five HIV-1 PR complexes were examined was used for this purpose: all complexes had the typical protonation pattern for HIV-1 PR in its ligand-deleted state, with one protonated and one deprotonated Asp residue in the catalytic dyad.[49]

In all cases, the ΔpK_a between the ligand-deleted Asp25/Asp25' and the Asp25/Asp25' complex increased. In four out of five cases, a hydroxy group of the ligand interacted with the two aspartates of the catalytic dyad; in the other case, a pyrrolidine-N.

For the interaction of HIV-1 PR and the asymmetric peptidomimetic VX-478 (the

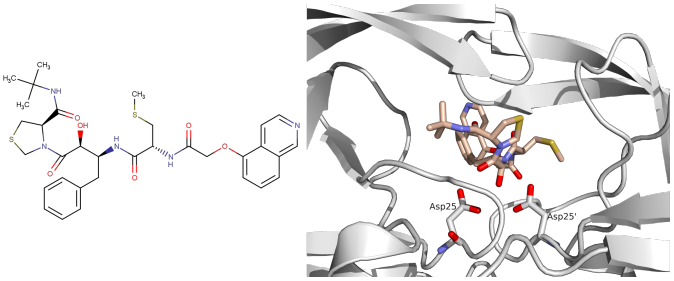
active ingredient in amprenavir, Figure 6.1), the Δn_{H^+} at pH 5.0 from the previous study was –0.01 and here, it was 0.55. ITC measurements in different buffers with distinct ionization enthalpies underline the correctness of the initially published calculations, yielding a Δn_{H^+} of 0.02 ± 0.03.[127]



	pK_a				
	Ligand-deleted		Complex		
	Asp25	Asp25'	Asp25	Asp25'	Δn_{H^+}
Literature	3.6	7.3	2.7	12.5	-0.01
ZAP	4.2	5.1	4.4	6.8	0.55

Figure 6.1: 2D-depiction of peptidomimetic compound VX-478 (left), interaction with the catalytic dyad of HIV-1 PR (PDB entry 1HPV, middle) and comparison of pK_a values in ligand-deleted and complex structure and corresponding Δn_{H^+} at pH 5.0 in literature study[49] and from Zap[128] (right).

In the literature, the asymmetric peptidomimetic KNI-272 showed a similar trend, leading to an increase in the pK_a of one of the aspartates in the complex, whereas with Zap both pK_as increase, causing a shift in the protonation setup (Figure 6.2). ITC measurements in different buffers at different pH values showed alternating proton release/uptake depending on the pH. At pH 5, Δn_{H^+} is about -0.2.[129]

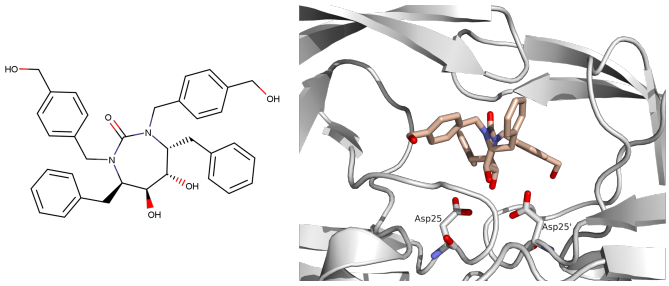


	pK_a				
	Ligand-deleted		Complex		
	Asp25	Asp25'	Asp25	Asp25'	Δn_{H^+}
Literature	3.8	7.2	3.8	13.4	0.01
ZAP	4.3	5.3	5.3	6.5	0.65

Figure 6.2: 2D-depiction of peptidomimetic compound KNI-272 (left), interaction with the catalytic dyad of HIV-1 PR (PDB entry 1HPX, middle) and comparison of pK_a values in ligand-deleted and complex structure and corresponding Δn_{H^+} at pH 5.0 in literature study[49] and from Zap[128] (right).

Binding of the symmetrical inhibitor DMP-323, which contains a cyclic urea core, shows a proton uptake in both approaches, 0.63 in the literature and 1.2 with Zap (Figure 6.3. In both cases, the catalytic dyad, which is assumed to be monoprotonated prior to complexation, takes up 1 mol of protons. NMR and X-ray studies support the diprotonated state and therefore underline the predictions.[130]

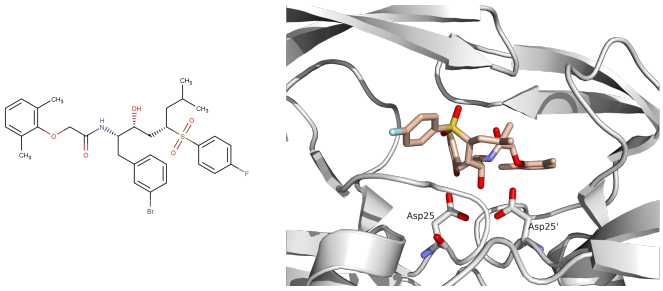
The binding mode of the asymmetric compound S6 (Figure 6.4) containing a hy-



	pK_a				
	Ligand-deleted		Complex		
	Asp25	Asp25'	Asp25	Asp25'	Δn_{H^+}
Literature	3.7	7.2	5.3	10.7	0.63
ZAP	3.9	4.8	5.1	6.2	1.2

Figure 6.3: 2D-depiction of symmetric cyclic urea compound DMP-323 (left), interaction with the catalytic dyad of HIV-1 PR (PDB entry 1QBS, middle) and comparison of pK_a values in ligand-deleted and complex structure and corresponding Δn_{H^+} in literature study[49] and from Zap[128] (right).

droxyethylene sulfone resembles that of KNI-272. While the computation from the literature did not indicate a shift and the catalytic dyad remains monoprotonated at pH 5, the Zap tool revealed a proton uptake from the complex of 0.54.



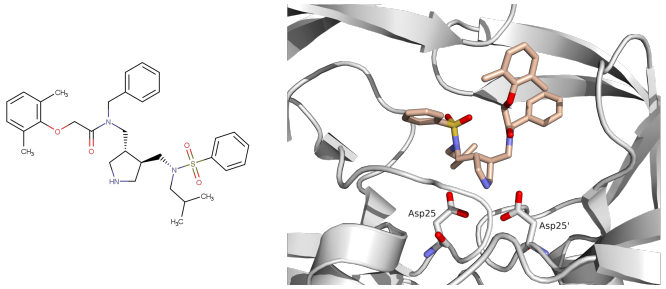
	pK_a				
	Ligand-deleted		Complex		
	Asp25	Asp25'	Asp25	Asp25'	Δn_{H^+}
Literature	2.1	7.2	0.5	15.2	0.01
ZAP	4.4	5.4	4.9	7.2	0.54

Figure 6.4: 2D-depiction of asymmetric compound S6 (left), interaction with the catalytic dyad of HIV-1 PR (PDB entry 1XL5, middle) and comparison of pK_a values in ligand-deleted and complex structure and corresponding Δn_{H^+} at pH 5.0 in literature study[49] and from Zap[128] (right).

Complexation of compound S7 with HIV-1 PR (Figure 6.5) leads to a predominant deprotonation of both aspartates in the complex with both tools, although the calculated pK_a values of the aspartates are much lower in the literature (–9.6 and –1.0) compared to Zap (2.2 and 3.7). Interestingly, the pyrrolidine is expected to be positively charged at pH 5.0, the moiety which is interacting mostly with the catalytic dyad. For both S6 and S7 there are no experimental data from which protonation and pK_a changes can be deduced.

6.3.2 Endothiapepsin

A real strength of the Zap-based tool is that it is able to calculate the pK_a not only for the ionizable amino acid side chains, but also for the titratable functional groups of the bound ligand, resulting in protein-ligand complexes with large protonation



	pK_a				
	Ligand-deleted		Complex		
	Asp25	Asp25'	Asp25	Asp25'	Δn_{H^+}
Literature	3.5	7.2	-9.6	-1.0	-1.02
ZAP	4.2	5.1	2.2	3.7	-0.63

Figure 6.5: 2D-depiction of pyrrolidine-containing compound S7 (left), interaction with the catalytic dyad of HIV-1 PR (PDB entry 1XL2, middle) and comparison of pK_a values in ligand-deleted and complex structure and corresponding Δn_{H^+} at pH 5.0 in literature study[49] and from Zap[128] (right).

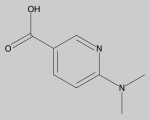
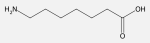
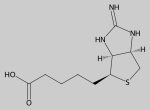
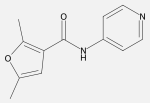
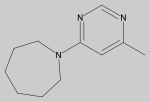
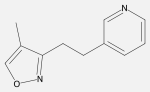
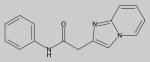
shifts. This was not evident in the EP-pepstatin A interaction, since the C-terminal carboxy group of pepA had a predicted pK_a of 4.3, which was not affected upon complexation (Chapter 5). In the work of Kröll, more than 200 published EP structures from the PDB were analyzed with the Zap-based tool.[128] The top 7 ligands with the largest protonation shifts upon protein-ligand binding are given (Table 6.1): for the EP crystal structure binding 6-(dimethylamino)pyridine-3-carboxylic acid in its active site (PDB entry 4Y3S), they were 1.4 in total. The results of the three complexes with the largest protonation shifts are discussed here.

The Δn_{H^+} of 1.4 results mainly from the shift at the secondary amine of the ligand (0.76) and the carboxylic acid side chain of Asp35 (0.54), suggesting that one mol of protons is taken up by one of them (Figure 6.6). The interaction between the pyridine nitrogen of the ligand and the carboxyl side chain of Asp81 leads to a ΔpK_a of 3.5 for the amine (shifting from 6.5 to 10) and to a shift of –0.6 for the Asp81 side chain (from 4.8 to 4.2), indicating that the amine is protonated and forms an H-bond with the deprotonated carboxyl group of Asp81.

The catalytic Asp residues are located near the carboxyl group of the ligand, which could explain the positive shift of Asp35. This is because the protonation of this residue could potentially create a network of H-bonds. A smaller positive pK_a shift of 0.3 takes place at the Asp33 side chain, likely due to the proximity of the aromatic ring of the ligand.

The other six ligands show a smaller shift in protonation for the entire complex, with the next greatest shift induced by the binding of 7-aminoheptanoic acid (4ZE6), which is predicted to take up 1.0 mol of protons per binding reaction, mostly due to the protonation of Asp35 (0.57) and Asp33 (0.55). Unlike structure 4Y3S, the fragment does not have an impact on the proton setup of the Asp81 carboxyl group (Figure 6.7). It is interesting to note that the ΔpK_a of the amino group of the

Table 6.1: EP-ligand complexes with significant shifts in protonation at pH 7.0. ΔpK_a: changes of the pK_a upon complex formation. Δn_{H^+}: mol of protons taken up or released upon complex formation for the specific residues (ligand, catalytic dyad Asp35/Asp219, Asp33) and the whole complex (Cpl). Modified from [128]

PDB entry	Ligand Structure	Ligand ΔpK_a	Ligand Δn_{H^+}	Asp35 ΔpK_a	Asp35 Δn_{H^+}	Asp219 ΔpK_a	Asp219 Δn_{H^+}	Asp33 ΔpK_a	Asp33 Δn_{H^+}	Cpl Δn_{H^+}
4Y3S		3.5	0.76	1.3	0.54	0.2	0.02	0.3	0.15	1.4
4ZE6		-2.6	-0.20	1.6	0.57	-0.1	0	1.1	0.55	1.0
4ZEA		1.0	0.01	2.6	0.9	0.4	0.05	0.1	0.04	0.9
4Y5C		2.9	0.93	-0.8	-0.06	-0.3	-0.02	0.1	0.03	0.8
4Y3P		2.5	0.88	-0.7	-0.05	-0.6	-0.03	0.1	0.03	0.8
5DQ5		2.5	0.89	-1.3	-0.09	-0.3	-0.02	-0.3	-0.08	0.6
4YD7		3.3	0.93	-1.5	-0.09	-0.2	-0.01	-0.3	-0.09	0.5

ligand is –2.6 (from 10.2 to 7.6), resulting in a release of 0.2 mol of protons.

For the 2-iminobiotin ligand of complex 4ZEA (Figure 6.8), the pK_a shift is 1.0: the carboxyl group shifts from 4.1 to 5.1, resulting in a negligible Δn_{H^+} of 0.01 at pH 7.0. This means the the carboxyl group remains deprotonated. The shift of the Asp35 carboxyl group is much higher (ΔpK_a=2.6), though, and as a consequence, the side chain is likely to be fully protonated at a pH of 7.0.

However, these observations and hypotheses are based on calculations using a static crystal structure and need to be confirmed by experimental methods. Therefore, it would be beneficial to conduct ITC measurements or apply other techniques, such as NMR spectroscopy, to study protonation effects in vitro and further understand

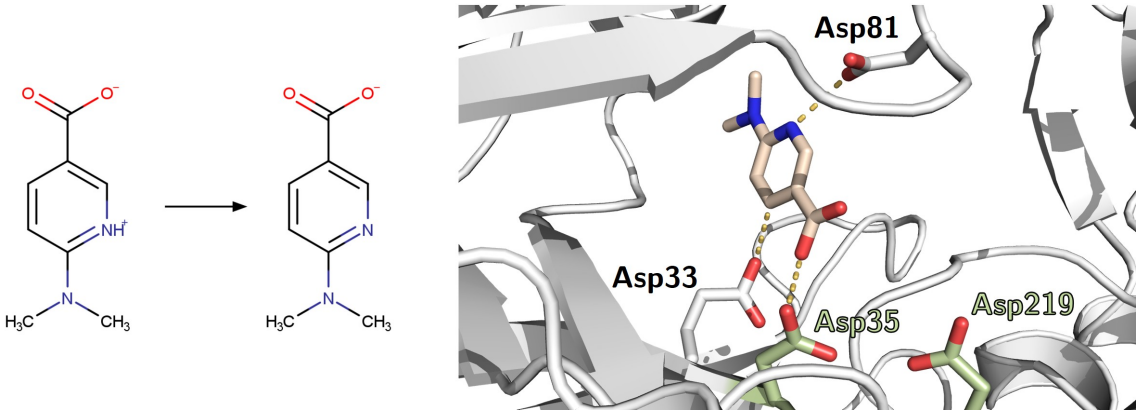


Figure 6.6: Investigated titration reaction of the 6-(dimethylamino)pyridine-3-carboxylic acid fragment in EP structure 4Y3S (left) and three-dimensional depictions of protein-ligand interactions (right). Figure modified from [128]

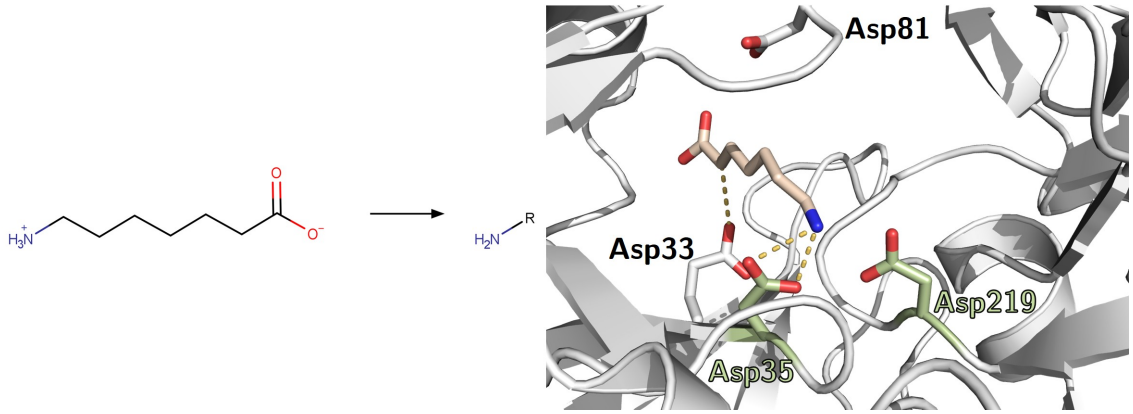


Figure 6.7: Investigated titration reaction of the 7-aminoheptanoic acid fragment in EP structure 4ZE6 (left) and three-dimensional depictions of protein-ligand interactions (right). Figure modified from [128]

the formation of this complex.

6.4 Conclusion & Outlook

Prediction tools for protein pK_a values are important because they influence many factors in biology. A clear advantage of pK_a calculations against experimental methods is the rapid availability of results, if access to a structure is given. What is known about the problems of this method is that the results and thus the accuracy of a PB solver depend on the choice of the dielectric constant, the protein conformation and the charge distribution.[36] Moreover, if the ligands tend to have a low affinity, like the fragments binding to EP discussed here, it is rather difficult to analyze them in an ITC setup.

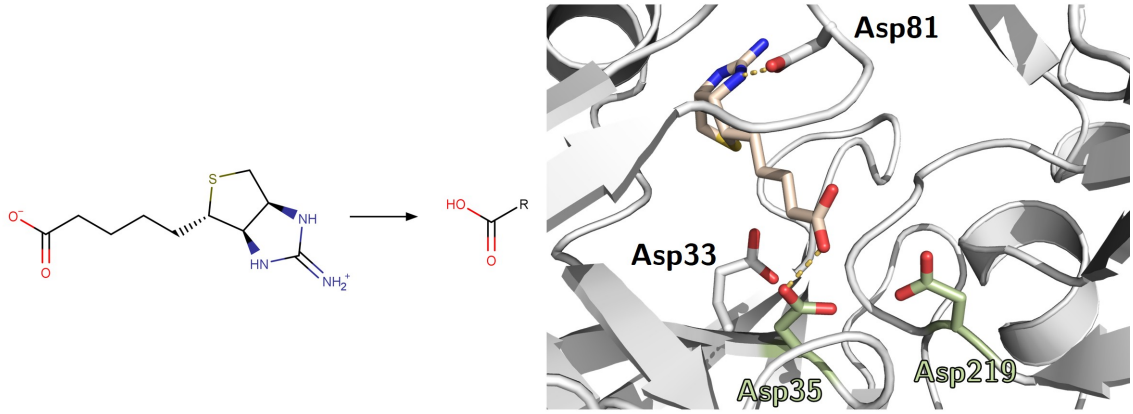


Figure 6.8: Investigated titration reaction of the 2-iminobiotin fragment in EP structure 4ZEA (left) and three-dimensional depictions of protein-ligand interactions (right). Figure modified from [128]

6.5 Experimental Procedure

6.5.1 Implicit Solvent pK_a Calculations

The calculations were carried out analogous to Chapter 5.5.4.

7. Protein pK_a Prediction on KRas Mutant Cysteine Residue

7.1 Introductory Remarks

The calculations addressed in this chapter are part of the publication "Targeting oncogenic KRasG13C with nucleotide-based covalent inhibitors" that was published in eLife.[131] This work has been made possible thanks to the contributions of Lisa Goebel, Tonia Kirschner, Sandra Koska, Amrita Rai, Petra Janning, Stefano Maffini, Paul Czodrowski, Roger S. Goody, Matthias P. Müller and Daniel Rauh. My part was the determination of the cysteine pK_a with the help of our in-house Poisson-Boltzmann solver.

7.2 Introduction

The Ras protein is a contributing factor in many human cancers. [132] The most common features of Ras are that it is a small GTPase that normally switches between a guanosine triphosphate (GTP)-bound active state and a GDP-bound inactive state, enabled in part by the stimulation of GTP hydrolysis by GTPase-activating proteins (GAPs).[133] This switch mechanism can be disrupted by oncogenic Ras mutations, that are found in approximately 25 % of all human cancers, including three of the deadliest forms (pancreatic, colon and lung cancer). KRas, with mutation hotspots at amino acid positions G12, G13 and Q61, is the most frequently mutated Ras isoform (85 %).[133, 134] Glycine at position 12 is the most commonly mutated residue. However, G13 is the second most abundant mutation (14 % of tumours have a mutation at this position) and in 6 % of these cases an acquired cysteine is found.[135, 136]

Of the amino acids, cysteine is one of the least abundant, and yet it plays an important role in a wide range of enzymes such as oxidases, peroxidases, reductases and disulfide isomerases. The pK_a value of a cysteine represents the equilibrium between the thiol-SH and the thiolate-S^-. The protonated thiols are significantly less

reactive than the negatively charged thiolate residue. Since reactivity is important for the functions of most cysteines, in addition to experiments for the determination of the pK_a values of cysteines, it is also important to make progress with the aid of calculations.[137]

7.3 Results & Discussion

In this project, covalent inhibitors targeting the Cys13 of the KRasG13C mutant were designed and the X-ray structures of the covalently bound complexes were presented. In order to calculate the effect of the ligands on the pK_a of Cys13, it was necessary to remove the covalent bond between ethylenediamine guanosine diphosphate (edaGDP)/ butylenediamineguanosine diphosphate (bdaGDP) and Cys13. The PB solver is only able to calculate the pK_a of a free cysteine. The removal of the linker and the resulting KRasG13C-GDP structure is examplarily shown for bdaGDP (Figure 7.1).

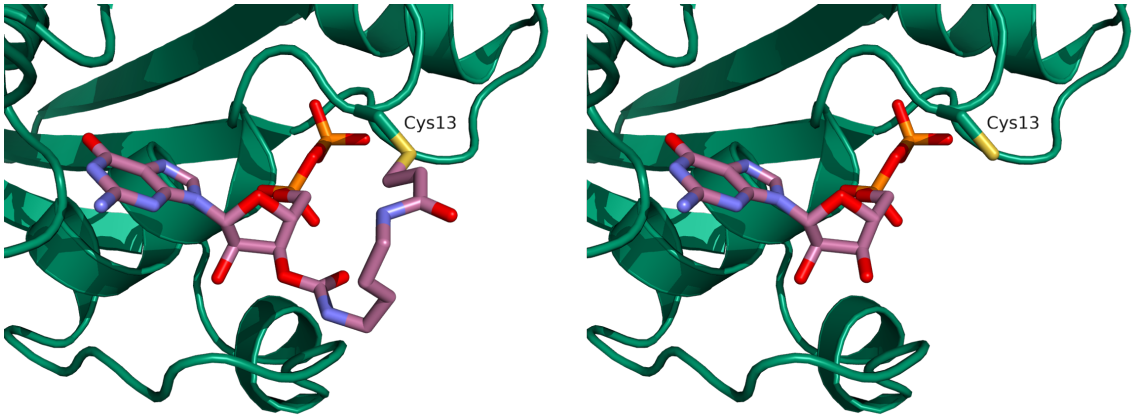


Figure 7.1: The nucleotide-based covalent inhibitor bdaGDP bound to KRasG13C (left, purple; PDB entry: 7OK4) and with the linker to Cys13 removed (right), a step necessary to perform the pK_a calculation.

For the two KRasG13C structures with the modified ligand, the results for the Cys13 pK_a values are the same, whether if the ligand is present (9.85) or not (8.85), the ΔpK_a is therefore 1.0 (Table 7.1). The calculations for Cys12, on the other hand, differ by 0.3 between the structures 4L8G and 4LDJ, but the ΔpK_a between the ligand-deleted structure and the complex is the same for both, namely 1.3. The differences between 4L8G and 4LDJ could be explained by the fact that the X-ray structure of 4L8G is not resolved between residues 60–69, which is a significant part of the switch II region of KRas.

In general, the cysteine pK_a values of both KRasG12C and KRasG13C are thus

Table 7.1: pK_a calculations. Overview of calculated pK_a values of the KRasG12C and G13C mutants in the presence (+) and absence (–) of GDP. The linker in case of the G13C mutants was removed so that all featured structures had GDP and a free cysteine in its active center. The pK_a calculations showed that both cysteines at position 12 and 13 have similar pK_a values indicating that also position 13 should generally be addressable by covalent warheads.

	KRas structure		**pK_a Cys (12/13)**		
KRas Mutant	**PDB**	**Ligand**	+	–	**Δ**
G13C	7OK3	GDP (from edaGDP)	9.85	8.85	1.0
	7OK4	GDP (from bdaGDP)	9.85	8.85	1.0
G12C	4LDJ	GDP	10.65	9.35	1.3
	4L8G	GDP	10.35	9.05	1.3

higher in presence of GDP. This is also logical, because the negative charges of the diphosphate residue lead to the avoidance of another negative charge in the form of the thiolate residue. The KRasG13C structures lack the Mg^{2+} ion, presumably due to precipitation in the buffer used, unlike the KRasG13C structures where Mg^{2+} is present. For the calculations, we did not want any other artifact beyond the pruned ligand in the KRasG13C structures, so the Mg^{2+} ions were removed from both structures.

Conflicting pK_a values for the cysteine of the G12C mutant are found in the literature: Hansen et al. reported a pK_a of 9.0 ± 0.2, which is very close to our calculations, whereas a recent study by Huynh et al. revealed a pK_a of ~ 7.6 with two different methods, by NMR and an independent biochemical assay.[138, 139]

7.4 Conclusion & Outlook

In general, our results show that KRasG13C, analogous to KRasG12C, can be targeted by covalent inhibitors.

In order to obtain an absolute, more realistic picture of the pK_a value of the cysteine, further calculations should be carried out. In this case, however, the Mg^{2+} ion would have to be included in the calculation or, regarding the KRasG13C structures (7OK3 and 7OK4), modelled. It is then expected that the pK_a value will deviate less in the calculations in the presence or absence of GDP, respectively, because the electrostatic environment will be less negative and a negative charge on the sulfur atom will be facilitated as a result of the deprotonation of the thiol.

Analogous experiments, such as those performed by Huynh et al. for KRasG13C,

could provide information on the pK_a value of Cys13.[139] The tendency of the calculations is that Cys13 is slightly less basic, and the binding of GDP has a slightly smaller influence on the pK_a, shifting it less towards basic. If this is confirmed, KRasG13C would be easier to target with covalent warheads than KRasG12C, steric considerations aside.

While the calculations presented lead to plausible pK_a values and can be done in minutes, due to the immense advances in computing power it is now possible to perform constant pH molecular dynamics (cpHMD) simulations on a desktop computer equipped with a single graphics processing unit (GPU) card, as suggested by Harris, Liu, and Shen.[140]

To date, computational approaches cannot replace wet-lab experiments. However, as the determination of pK_a values can be very complex, e.g., using NMR methods, calculations or simulations based on the crystal structure are essential. Depending on the assay being used, it may be necessary to create point mutations in the proteins in order to remove off-target cysteines.[139]

With the rise of covalent inhibitors, determination of cysteine pK_a values has gained importance. Depending on their solvent accessibility and function, the pK_a of cysteine ranges from < 3 to > 9.[141] Since a negatively charged thiolate is much more electrophilic than the thiol residue, the question of whether your protein of interest can be targeted at this or that cysteine stands and falls with the pK_a. In a recent publication by Gao and Günther, a structure- and sequenced based tool employing Machine Learning algorithms called HyperCys was presented.[142]

While machine learning approaches seem promising, they rely heavily on the amount and quality of experimental data. Cysteine residues are particularly difficult to predict because they are less abundant and more buried in proteins than all other titratable residues. This results in an underrepresentation of the residues that show the largest pK_a shifts.[143]

7.5 Experimental Procedure

7.5.1 Implicit Solvent pK_a Calculations

The calculations were carried out analogous to Chapter 5.5.4, with following amendments and differences, respectively: The linker was removed in the bda/edaGDP structures. An inner dielectric of 10, an ionic strength of 0.05 M and ionization (i.e., pH) of 7.5 was applied. For cysteine, a reference pK_a of 8.6 was used. All hydrogen atoms were modeled explicitly, and except for the orientations of the OH and SH

protons, which were sampled in 10° steps, the rest of the structure was static. Optimizing ionization state and SH orientation was achieved by applying ten million Monte Carlo steps.

8. Virtual Screening Command-Line Tool *VSFlow*

8.1 Introductory Remarks

This chapter has been adapted from the eponymous paper published in the Journal of Cheminformatics.[144] The publication was made possible by contributions from Sascha Jung and Paul Czodrowski. My part was the creation of the use cases, together with the corresponding illustrations and the continuous text.

8.2 Introduction

Virtual screening approaches are extensively used computational methods in modern drug discovery projects and they often replace or help to reduce more expensive and time-consuming high-throughput screenings nowadays.[145] There are two major categories of screening approaches: ligand-based and structure-based methods.[146]

Ligand-based methods are typically used if no X-ray structure of the target receptor is available. A single compound or a set of compounds known to bind to a specific target or to be active in a functional assay is typically used as the template to identify similar compounds in a large virtual database. In general, similarity can be evaluated on the basis of 2D and 3D molecular representations.[147] The classical 2D chemical similarity representations is based on molecular fingerprints (e.g., circular fingerprints, topological fingerprints, substructure fingerprints) transforming the molecular representation into a bit vector. The similarity between two vectors is then calculated with various similarity measures, most common is the Tanimoto coefficient. 3D similarity methods mainly consider the shape comparison of two molecules, typically extended by 3D pharmacophoric features, e.g., ROCS is considered the industry-leading commercial program for shape-based screenings.[148]

Structure-based approaches, in most cases classical docking methods, are typically

preferred if the target 3D structure information is available.[149] However, 2D ligand-based methods often require only a fraction of second for a single structure comparison task which allows to perform large screenings within a few hours even on a single, standard CPU. In contrast, docking methods are already considerably more resource demanding and time-consuming, not to mention more elaborated methods such as molecular dynamics simulations.[150] As a consequence, ligand-based methods are very attractive options for initial attempts to identify or filter relevant compounds in large and ultra-large virtual databases.[151] Furthermore, they are valuable tools to identify close analogues of known active compounds in a time efficient manner. In the last couple of years, several methods have been developed to screen non-enumerated chemical spaces up to 10^{15} compounds and beyond in seconds to minutes on standard hardware.[152] The most elaborated technique for large space screening are chemical fragment spaces with corresponding connection rules, e.g., BioSolveIT's fragment spaces in connection with FTrees similarity implemented in their infiniSee software allows the screening of huge chemical spaces (e.g., Enamine REAL space) in seconds on standard hardware.[153, 154]

There are many open-source web servers available for the screening of enumerated compound libraries using a variety of different structure- and ligand-based methods, recently reviewed by Singh et al.[155] For example, many well-known databases such as CHEMBL, PubChem or ZINC include ligand-based similarity search functionalities with molecular fingerprints and/or substructure searches.[64, 156, 157] The web tool SwissSimilarity allows for the 2D fingerprint and 3D shape screening of common public databases and compound libraries of most commercial vendors such as Enamine or ChemDiv.[158, 159] Pharmit additionally offers the possibility to screen large databases based on pharmacophore queries.[160]

Several standalone tools focusing on enumerated 2D ligand-based screening approaches are available, most of which are commercial products.[152] Prominent examples are Schrödinger's GPUSimilarity integrated in their LiveDesign suite using a GPU-powered server in the background, Arthor's NextMove software with a SMARTS-based pattern matcher and Andrew Dalke's chemfp commandline tool.[161–163]

To the best of our knowledge, there is no open-source commandline tool available which is similar to the SwissSimilarity or Pharmit web server and which allows for the comprehensive screening of databases and library files using different 2D and 3D ligand-based screening approaches, all combined in one tool.

In the following, we report an open-source command-line tool called "Virtual Screening WorkFlow" (VSFlow) written in Python and containing three different ligand-based screening modes. It relies on the open-source cheminformatics sofware RDKit.[164]

VSFlow includes a substructure-based and fingerprint-based screening mode (2D) as well as a 3D shape-based screening mode (Figure 8.1). Additionally, it possesses two tools for preparing and managing compound databases for virtual screening.

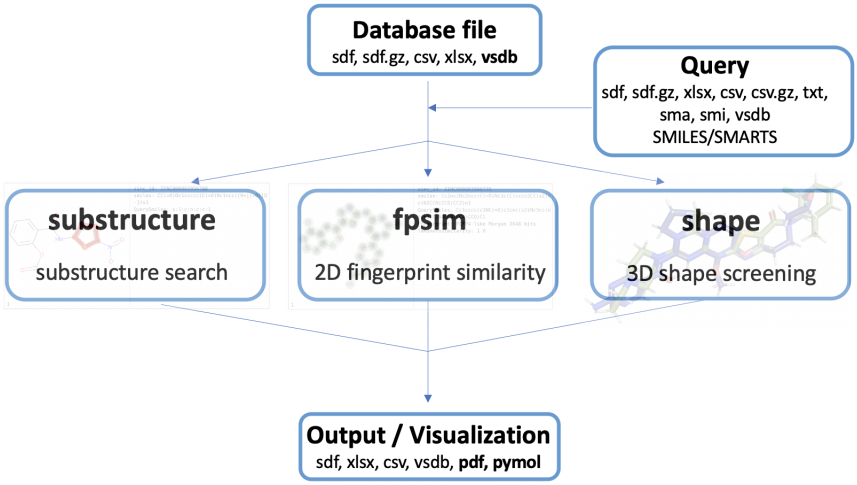


Figure 8.1: Different screening functionalities of VSFlow

8.3 Implementation

VSFlow is written in Python, is open-source and can be downloaded from `https://github.com/czodrowskilab/VSFlow`. It is licensed under the MIT license. As a prerequisite, a working installation of Anaconda or Miniconda is needed.[165] VSFlow including all dependencies can then be installed with the provided yml file as follows:

```
conda env create --quiet --force --file environment.yml
conda activate vsflow
pip install .
```

The Python dependencies are rdkit, xlrd, xlsxwriter, pdfrw, fpdf, pymol-open-source, molvs and matplotlib.[166, 167] VSFlow requires Python version 3.7 or higher.

VSFlow includes 5 separate tools: preparedb, substructure, fpsim, shape and managedb (Figure 8.1). All functionalities of VSFlow can also be run in parallel on multiple cores/threads. Parallelization is implemented via Python's built-in multiprocessing module.

8.3.1 preparedb: Prepare Databases

VSFlow contains a tool to prepare compound libraries for virtual screening (preparedb). It allows for standardization of the molecules, generation of fingerprints

and generation of multiple conformers (Figure 8.2). The output file is a "virtual screening database" (.vsdb) file. The vsdb file is a Python pickle file containing all information in a special Python dictionary format which significantly enhances loading speed compared to SD files, particularly relevant for larger databases. Standardization is done on the basis of the MolVS rules and includes charge neutralization, salt removal and optionally tautomer cannibalization.[166] Fingerprints are generated with the RDKit chemistry framework. Conformers are generated with the RDKit ETKDGv3 method and optimized with the MMFF94 forcefield. [168] The following options are available:

- standardize: standardizes molecules, removes salts and associated charges
- conformers: generates multiple 3D conformers for database molecules
- canonicalize: adds the canonical tautomer to the database
- fingerprint: generates the respective fingerprint for each molecule and stores it in the database

It is also possible to directly download the PDB ligands and the chembl database and store them as .vsdb databases, e.g.::

```
vsflow preparedb -d pdb -s -f ecfp -r 2 -o pdb_ligands.vsdb
```

The above command will download all pdb ligands, standardize the molecules (-s argument), calculate the ECFP2 fingerprint (-f and -r argument) for every molecule and store it along with the molecule in the database (-o argument). You can repeat this for the chembl database, e.g., with a different fingerprint:

```
vsflow preparedb -d chembl -s -f fcfp -r 4 -o chembl_cpds.vsdb
```

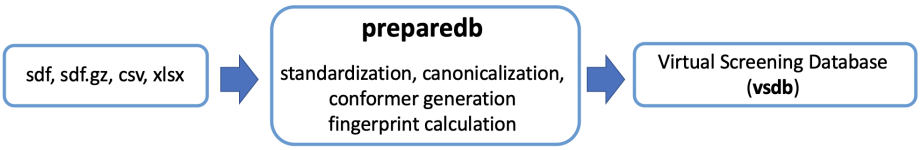


Figure 8.2: Preparedb functionality of VSFlow: prepare compound libraries for virtual screening

8.3.2 substructure: Substructure Search

The substructure search (substructure) is performed based on the GetSubstructMatches() functionality available for RDKit MOL objects.

fpsim: Fingerprint similarity search

The fingerprint generation relies on the RDKit framework. All fingerprints currently implemented in the RDKit (Morgan, RDKit, Topological Torsion and Atom Pairs fingerprint and MACCS keys) are supported and different similarity measures (Tanimoto, Tversky, Cosine, Dice, Sokal, Russel, Kulczynski and McConnaughey similarity) can be used.

shape: Shape-based Screening

Several functionalities of RDKit were combined to perform a screening based on a compounds' molecular shape (Figure 8.3). First, generation of conformers (RDKit ETKDGv3 and MMFF94 forcefield) is done for 2D query structures. Conformers for database compounds can be generated using the preparedb functionality. Then, conformers of each query molecule are aligned to all conformers of each database molecule with the RDKit Open3DAlign functionality, either using MMFF94 force field parameters or Crippen atomic logP contributions (user-defined). In the next step, for every conformer pair the shape similarity is calculated (TanimotoDist, TverskyShape or ProtrudeDist) and the most similar conformer pair for every query/database molecule pair is selected (RDKit rdShapeHelpers). For the selected most similar conformer pair a 3D pharmacophore fingerprint is generated (RDKit Pharm2D) and the fingerprint similarity is calculated. By default, a combined score (combo score), the average of shape similarity and 3D fingerprint similarity, is used to rank the database molecules. The intended use case of the shape screening mode is to screen a database of compounds with multiple conformers (prepared e.g., using the preparedb functionality of VSFlow) and to use a query ligand in a single, bioactive conformation, e.g., from the pdb database.

8.3.3 managedb: Manage Databases

The mode managedb is a convenience tool to update and manage compound databases which are integrated into VSFlow. A detailed description can be found in the VSFlow wiki.[169]

8.4 Results & Discussion

In the following section, the intended usage of VSFlow including some example commands are presented. A detailed description of the multiple possibilities to use VSFlow along with specific examples can be found in the VSFlow GitHub wiki.[169]

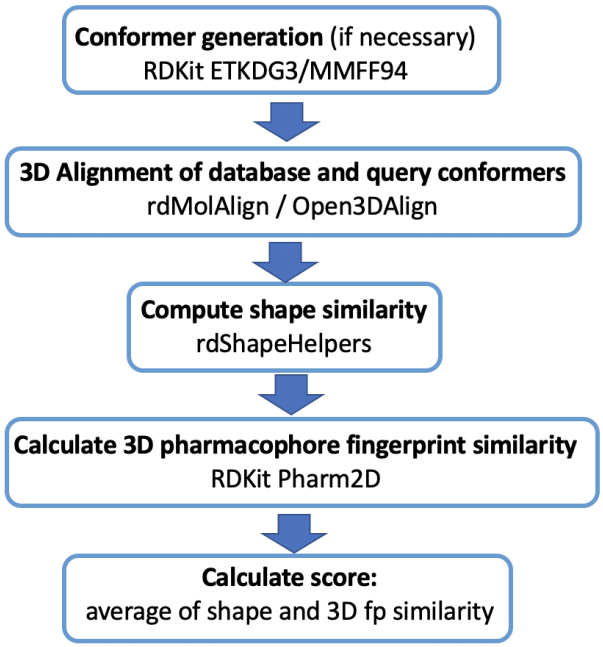


Figure 8.3: Different steps and RDKit functionalities which were combined to perform a screening based on pharmacophore alignment and shape similarity.

In order to demonstrate the three main functionalities of VSFlow together with both its versatile input and output formats, we took the tyrosine-kinase inhibitor dasatinib as query molecule. As database, an SD file of the FDA-approved drugs generated from the ZINC database was used, comprising over 1600 molecules [157]. This database is also available in our GitHub repository.

8.4.1 Substructure Search

For the substructure search, a SMARTS representation of the thiazol function of dasatinib was taken as input to see how many other drugs might have that specific group. Besides the 36 hits (one of them, of course, dasatinib itself) in which the thiazole group was found, three molecules even have two thiazole groups, namely cefditoren, cobicistat and ritonavir. A pdf (supporting information) was generated displaying a table of the found hits with the 2D structures and the found substructure match highlighted in red as well as the information of the hit (e.g., ID, SMILES, Figure 8.4). It should be mentioned that a pdf can only be generated in addition to an sdf, excel or csv file.

```
vsflow substructure -sma "s:1:c:n:c:c:1" -d fda.sdf
-o substructure.sdf --pdf
```

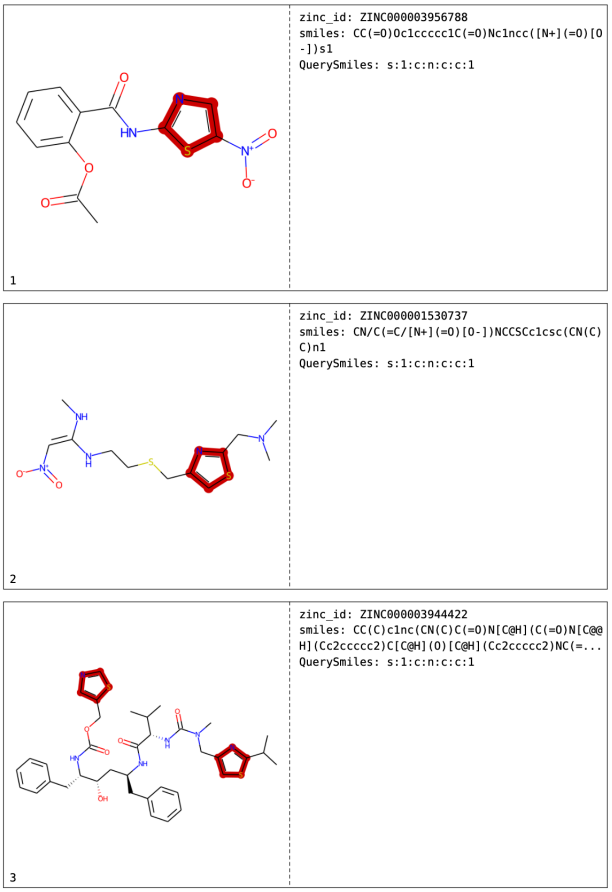


Figure 8.4: Exemplary page of the pdf file generated after substructure search. The left column shows the hits with the substructure matches highlighted in red, the right column the ID of the hits as well as the SMILES and the query SMARTS .

8.4.2 Fingerprint Similarity

For the fingerprint similarity function fpsim, a SMILES input of the molecule was used with default parameters, i.e., an FCFP4-like Morgan 2048 bits of radius 2 for which the Tanimoto coefficient was calculated. A pdf file was selected as output format as well as an Excel file. The simmap parameter will generate a similarity map that visualizes the contribution of the specific atoms to the similarity between the molecules in the database and dasatinib (Figure 8.5). [170]

```
vsflow fpsim -d fda.sdf -o fingerprint.xlsx --pdf --simmap
-smi "Cc1cccc(c1NC(=O)c2cnc(s2)Nc3cc(nc(n3)C)N4CCN(CC4)CCO)Cl"
```

8.4.3 Shape Similarity

In order to perform a shape screening, a new database, containing a maximum of 20 conformers, was generated with the -c argument because the original database only

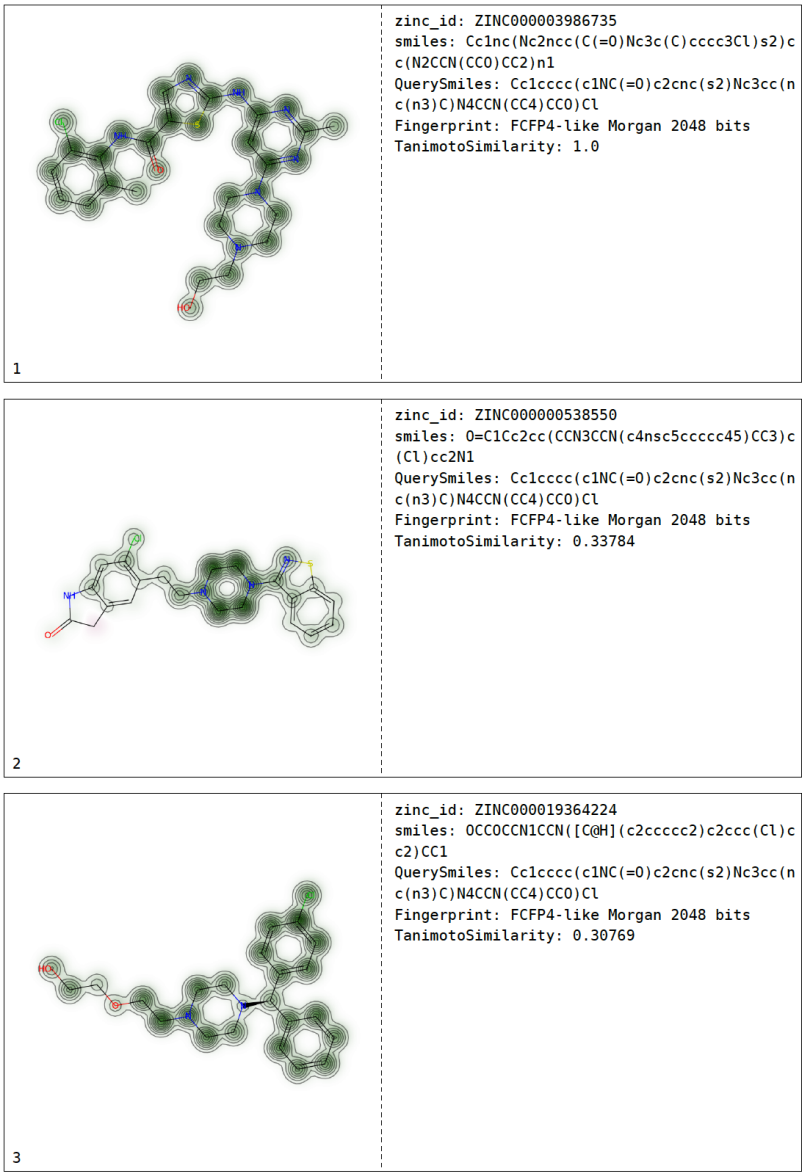


Figure 8.5: Exemplary page of the pdf file generated after fpsim search. The fingerprint similarity (FCFP4-like Morgan 2048 bits) of the molecules with the query molecule dasatinib is visualized in the left column, the right column shows IDs of the molecule as well as the search parameters and the calculated Tanimoto similarity.

had one conformer per compound.

```
vsflow preparedb -i fda.sdf -np 8 -c 20
-o fda_multiple_confs.vsdb
```

Since that is a rather resource-intensive step, multiprocessing was carried out with the help of the -np parameter. The following shape search, also multiprocessed, was then done with the previously prepared vsdb pickle file using the instance coordinates

of dasatinib in complex with tyrosine protein kinase ABL1 (PDB: 2GQG).

```
vsflow shape -i 2gqg_C_1N1.sdf -np 8
-d fda_multiple_confs.vsdb -o shape.sdf --pymol
```

More than half of the top 10 hits were other kinase inhibitors. By default, the shape functionality creates two sd files, one with the query molecule (shape_1_query.sdf) and the found hits as a second file (shape_1.sdf). Additionally, a PyMOL session file was generated (–pymol parameter) so that the aligned structures could be visually inspected directly (Figure 8.6).

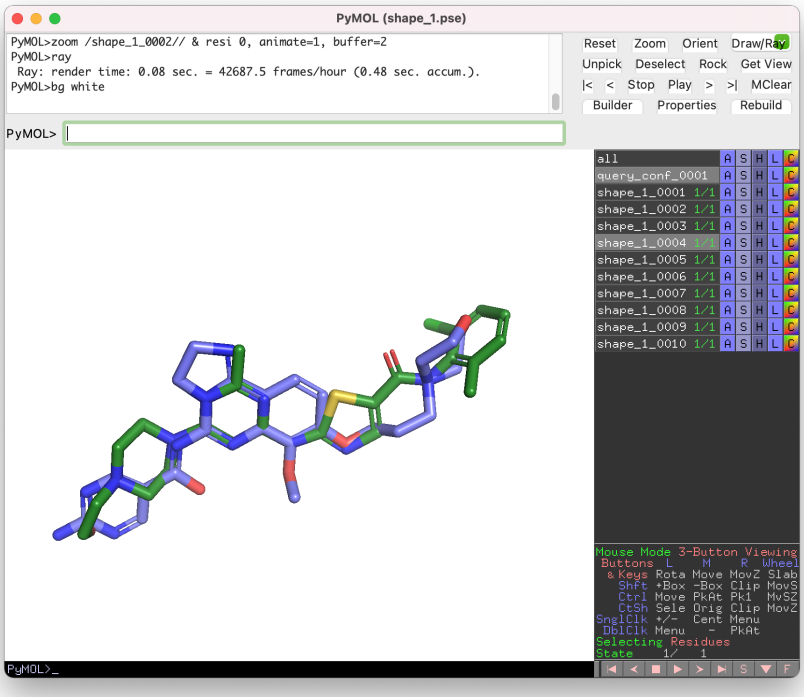


Figure 8.6: Screenshot from the PyMOL session file generated after shape similarity screening. By default, the first ten hits (one of them shown here in blue) are aligned with the query molecule dasatinib (green).

The RMSD spread of the conformer generation process (ETKDG3 followed by MMFF94 minimization) is given in Figure 8.7). It shows a clear upwards trend: the more rotatable bonds, the larger the RMSD.

8.4.4 Runtime Performance

To give the user an idea of the expected runtime performance, we performed a substructure and 2D similarity search in the pdb and chembl28 database.[64, 171] We performed the searches on up-to-date standard notebook hardware, namely a 12th

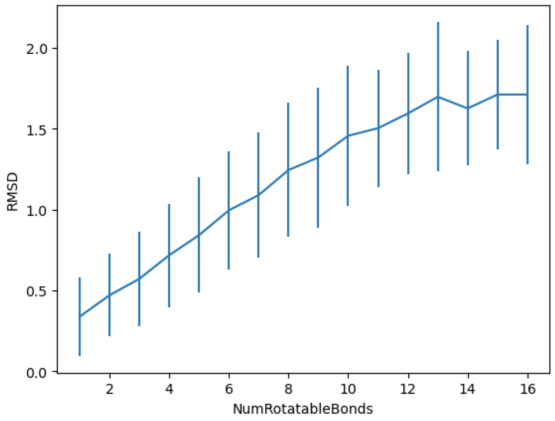


Figure 8.7: RMSD spread of the conformer generation process (ETKDG3 followed by MMFF94 minimization) for the search of the bioactive conformation (Platinum data set)

Gen Intel(R) Core(TM) i7-12700H with 2.70 GHz and 20 cores and 32GB RAM running Windows 11. To get an idea of the performance on your own system, you may execute the following commands accordingly. Both chembl and pdb database can be downloaded and prepared directly within VSFlow:

```
vsflow preparedb -d pdb -o pdb_ecfp4 -f ecfp -np 6
vsflow preparedb -d chembl -o chembl_ecfp4 -f ecfp -np 6
```

With the above calls, the pdb and chembl databases are downloaded into VSFlow and 2048-bit ECFP4 fingerprints are generated for each compound and stored within the output vsdb file. Preparation of the pdb database (containing 36796 unique compounds at 22/05/2022) took 11 seconds on our system, preparation of the chembl28 database (2066377 compounds) took 511 seconds. Now, we performed a substructure and similarity screening using a SMILES as query, once in single-core mode and once on 6 cores:

```
vsflow substructure -smi "C(C1=CC=CC=C1)C1=CC=NC=C1" -d pdb_ecfp4.vsdb
-o pdb_subsearch.sdf
vsflow substructure -smi "C(C1=CC=CC=C1)C1=CC=NC=C1" -d pdb_ecfp4.vsdb
-o pdb_subsearch.sdf -np 6
vsflow fpsim -smi "C(C1=CC=CC=C1)C1=CC=NC=C1" -d pdb_ecfp4.vsdb
-o pdb_fpsim.sdf -f from_db
vsflow fpsim -smi "C(C1=CC=CC=C1)C1=CC=NC=C1" -d pdb_ecfp4.vsdb
-o pdb_fpsim.sdf -f from_db -np 6
vsflow substructure -smi "C(C1=CC=CC=C1)C1=CC=NC=C1"
-d chembl_ecfp4.vsdb -o chembl_subsearch.sdf
vsflow substructure -smi "C(C1=CC=CC=C1)C1=CC=NC=C1"
```

```
-d chembl_ecfp4.vsdb -o chembl_subsearch.sdf -np 6
vsflow fpsim -smi "C(C1=CC=CC=C1)C1=CC=NC=C1" -d chembl_ecfp4.vsdb
-o chembl_fpsim.sdf -f from_db
vsflow fpsim -smi "C(C1=CC=CC=C1)C1=CC=NC=C1" -d chembl_ecfp4.vsdb
-o chembl_fpsim.sdf -f from_db -np 6
```

The following table (Table 8.1) summarizes the overall runtime for each call, e.g., it contains the loading time for the database file, the substructure or similarity search and the generation of the output file.

Table 8.1: Runtime performance of substructure and similarity search on 12th Gen Intel(R) Core(TM) i7-12700H with 2.70 GHz and 20 cores and 32GB RAM running Windows 11

	pdb	chembl	cores
substructure	1 s	162 s	1
	0.8 s	89 s	6
similarity	1 s	157 s	1
	0.75 s	77 s	6

8.4.5 Virtual Screening Performance

To give the user an idea about the performance of the tool in virtual screening practice, i.e., whether it could identify active compounds, we did some basic simulated screenings using the maximum unbiased validation (MUV) dataset.[172] The MUV dataset is based on PubChem bioactivity data and consists of 17 targets, each with 30 actives and 15000 decoys. The choice of actives and decoys is done based on confirmatory and primary screens, which makes the dataset very difficult for virtual screening methods. We performed sample screenings based on 2D fingerprint and 3D shape similarity (mode fpsim and shape). The general performance of 2D fingerprints implemented in RDKit has been studied extensively before, with the MUV dataset being part of a larger evaluation set.[173] We adapted a simplified version of the workflow described before by Riniker and Rohrer.[172, 173] In short, for each of the 17 subsets in the MUV dataset, one of the 30 active compounds was selected as query molecule and the remaining 29 actives were pooled together with the 15000 decoys and used as validation set. This query/validation split was done for all 30 actives. For the resulting 30 query/validation test splits per subset the virtual screening performance was measured by the area under the receiver operating curve (AUC, example curve shown in Figure 8.8) and the mean value was calculated for each subset (mean AUC). The screening consisted of two steps: 1) generation

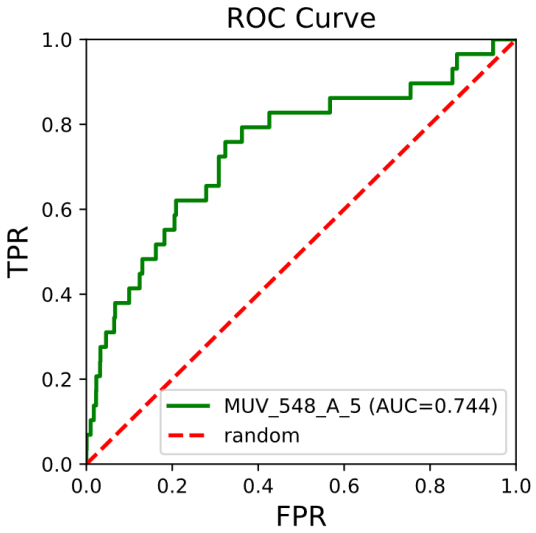


Figure 8.8: Example of a receiver operating curve (ROC) obtained for a query/validation test split after a 2D similarity screening with ECFP2 fingerprint. The MUV subset MUV_548 was the validation set, the query was compound MUV_548_A_5. The area under the curve (AUC) is 0.744. FPR = false positive rate, TPR = true prositive rate.

of a vsdb database with standardized molecules ans pre-computed fingerprints or conformers for the validation set; 2) 2D or 3D similarity screening of the validation set against the query molecule.

The results for 2D similarity screening with various descriptors is summarized in Figure 8.9. They follow, in general, the trend observed by Riniker et al. for 2D fingerprints on the MUV dataset.[173] For some targets, a significant enrichment of actives (e.g., meanAUC = 0.74 for ECFP3 fingerprint for target FactorXIa [MUV_846]) is observed, whereas for other targets no enrichment could be observed based on simple 2D similarity calculations.

Figure 8.10 summarizes the results for the 3D shape-based virtual screenings. Best performance is observed when using the combo score for result ranking for most MUV subsets. However, for MUV_737 (estrogen receptor alpha) and MUV_832 (cathepsin G) scoring with 3D fingerprint yields a better overall enrichment.

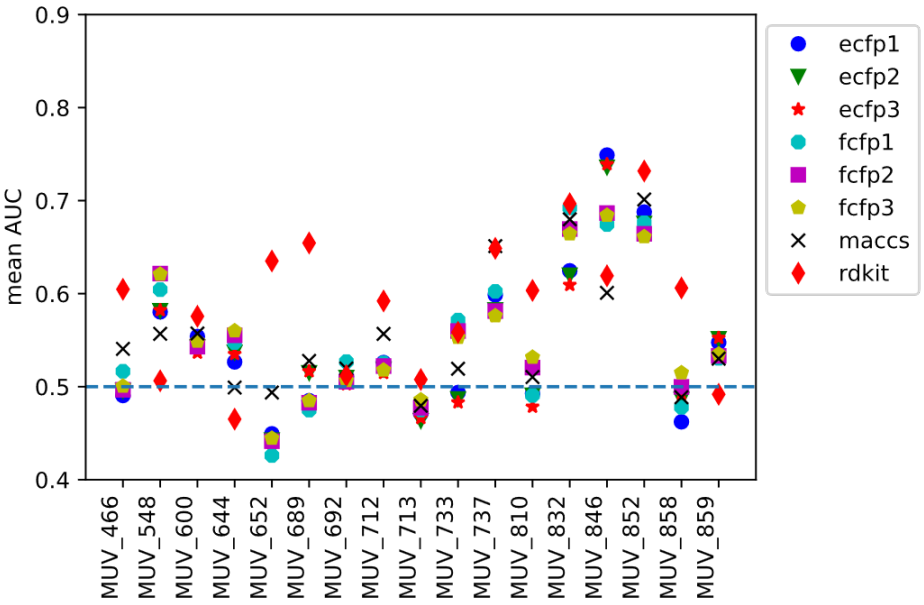


Figure 8.9: Results of virtual screening validation with the MUV dataset for 2D fingerprint similarity. The expectation of mean AUC of 0.5 for random rankings is indicated by the blue dashed line.

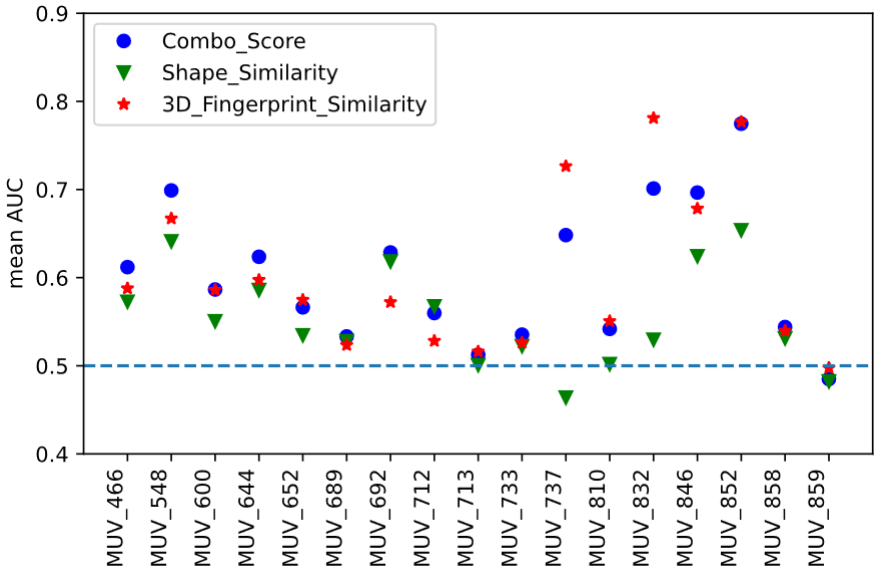


Figure 8.10: Results of virtual screening validation with the MUV dataset for 3D shape-based screenings. The expectation of mean AUC of 0.5 for random rankings is indicated by the blue dashed line.

9. Devices, Consumables and Chemicals

Unless otherwise described, chemicals and reagents from the following companies were used for the preparation of buffer solutions and cultivation media: Carl Roth (Karlsruhe, DE), Sigma-Aldrich (St. Louis, MO, US), Thermo Fisher Scientific (Waltham, MA, US), Merck Millipore (Burlington, MA, US), and Invitrogen (part of Thermo Fisher Scientific).

9.1 Devices

FPLC systems: ÄKTA pure (GE Healthcare, Chicago, IL, US); ÄKTA start (GE Healthcare, Chicago, IL, US)

Incubators: Minitron (Infors HT, Einsbach, DE); Multitron Pro (Infors HT, Einsbach, DE); MaxQ™ 8000 (Thermo Fisher Scientific, Waltham, MA, US); Certomat ® BS-1 (Sartorius, Göttingen, DE)

Microcalorimeter: MicroCal PEAQ-ITC Automated ITC (Malvern Panalytical, Worcestershire, UK)

Microfluidizer: Model M-110L (Microfluidics, Westwood, MA, US)

Pipettes: Research plus (Eppendorf, Hamburg, DE); Xplorer (Eppendorf, Hamburg, DE)

Rotors: Sorvall SA-600 (Thermo Fisher Scientific, Waltham, MA, US); FiberLite™ F20-6x100, FiberLite™ F8-6x1000y (Thermo Fisher Scientific, Waltham, MA, US)

Spectrophotometer: NanoDrop™ 2000 (Thermo Fisher Scientific, Waltham, MA, US)

Thermomixer: ThermoMixer® C (Eppendorf, Hamburg, DE)

Centrifuges: 5424, 5424R, 5804R (Eppendorf, Hamburg, DE); Sorvall Evolution RC (Thermo Fisher Scientific, Waltham, MA, US)

9.2 Consumable Materials

Dialysis tube/cassette: Slide-A-Lyzer® MWCO 10000 (Thermo Fisher Scientific, Waltham, MA, US)

Centrifugal concentrators: Vivaspin MWCO 10 kDa (Sartorius, Göttingen, DE); Vivaspin MWCO 30 kDa (Sartorius, Göttingen, DE)

9.3 Test Substances

Sorafenib (LC Labs, Woburn, MA, US)

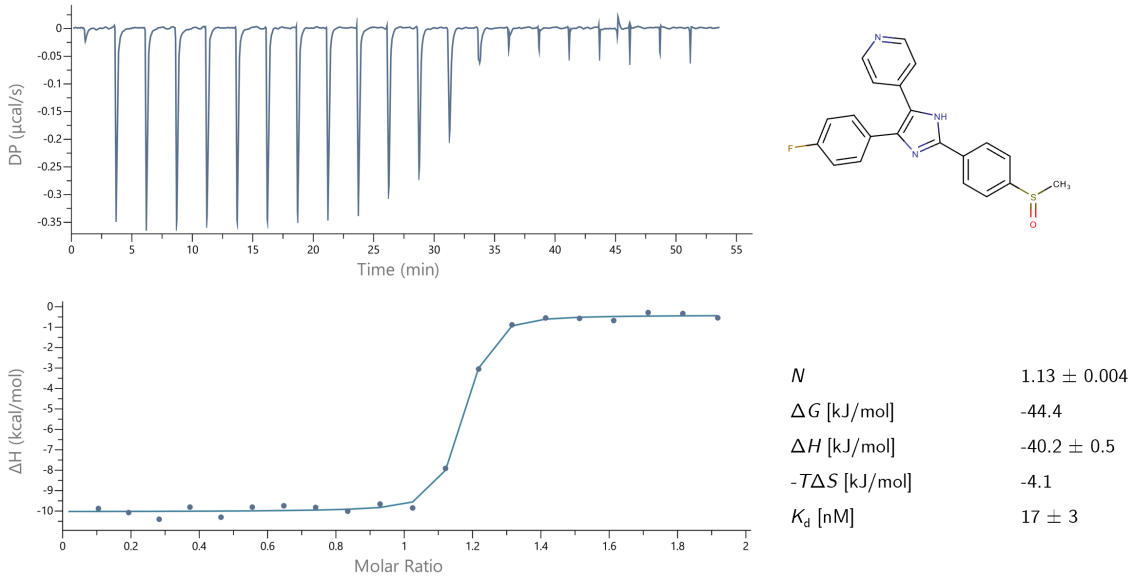
Regorafenib (LC Labs, Woburn, MA, US)

SB203580 (LC Labs, Woburn, MA, US)

SAP114 (Chemspace, Riga, Latvia)

A. Supplementary Material

A.1 Thermodynamic Data



N	1.13 ± 0.004
ΔG [kJ/mol]	-44.4
ΔH [kJ/mol]	-40.2 ± 0.5
$-T\Delta S$ [kJ/mol]	-4.1
K_d [nM]	17 ± 3

Figure A.1: Thermogram of the titration of p38α with reference inhibitor SB203580 (molecular representation top right) and inherent thermodynamic parameters and stoichiometry (bottom right). The K_d for this interaction is 15 nM in the literature[174], but measured with ITC at 30°C and not at 25°C as here.

Abbreviations

pK_a acid dissociation constant

ADMET absorption, distribution, metabolism, excretion, toxicity

AIDS acquired immunodeficiency syndrome

AP aspartic protease

ATP adenosine 5'-(tetrahydrogen triphosphate)

bdaGDP butylenediamineguanosine diphosphate

BTK Bruton's tyrosine kinase

cpHMD constant pH molecular dynamics

DMSO dimethyl sulfoxide

edaGDP ethylenediamine guanosine diphosphate

EP endothiapepsin

FBDD fragment-based drug design

FDA U.S. Food and Drug Administration

GDP guanosine diphosphate

GPU graphics processing unit

GTP guanosine triphosphate

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIV human immunodeficiency virus

HIV-1 PR HIV-1 protease

ITC isothermal titration calorimetry

KNIME Konstanz Information Miner

MMP matched molecular pair

MMPA matched molecular pair analysis

MPD 2-methyl-2,4-pentanediol

NMR nuclear magnetic resonance

p38α mitogen-activated protein kinase p38α

PB Poisson-Boltzmann

PBE Poisson-Boltzmann equation

PDB Protein Data Bank

pepA pepstatin Ald type

pH *potentia hydrogenii*

PKA cAMP-dependent protein kinase A

Ras rat sarcoma virus

SBDD structure-based drug design

TCEP tris(2-carboxyethyl)phosphine

Tween® 20 polyoxyethylene-20-sorbitan monolaurate

VEGFR-2 vascular endothelial growth factor receptor

WHO World Health Organization

wt wild type

List of Figures

List of Tables

References

[1] Gerhard Klebe. *Wirkstoffdesign - Entwurf und Wirkung von Arzneistoffen.* Spektrum, 2009. DOI: 10.1007/978-3-8274-2213-2.

[2] Alan Wayne Jones. "Early drug discovery and the rise of pharmaceutical chemistry". In: *Drug Testing and Analysis* 3.6 (2011), pp. 337–344. DOI: 10.1002/dta.301.

[3] Hugo Kubinyi. "Chance favors the prepared mind - From serendipity to rational drug design". In: *Journal of Receptor and Signal Transduction Research* 19.1-4 (1999), pp. 15–39. DOI: 10.3109/10799899909036635.

[4] Soma Mandal, Mee'nal Moudgil, and Sanat K. Mandal. "Rational drug design". In: *European Journal of Pharmacology* 625.1-3 (2009), pp. 90–100. DOI: 10.1016/j.ejphar.2009.06.065.

[5] Irene T. Weber, Yuan Fang Wang, and Robert W. Harrison. "HIV protease: Historical perspective and current research". In: *Viruses* 13.5 (2021), pp. 1–12. DOI: 10.3390/v13050839.

[6] *HIV and AIDS.* 2023. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids (visited on 01/17/2025).

[7] Ashraf Brik and Chi Huey Wong. "HIV-1 protease: Mechanism and drug discovery". In: *Organic and Biomolecular Chemistry* 1.1 (2003), pp. 5–14. DOI: 10.1039/b208248a.

[8] Jorg Eder, Ulrich Hommel, Frederic Cumin, Bruno Martoglio, and Bernd Gerhartz. "Aspartic Proteases in Drug Discovery". In: *Current Pharmaceutical Design* 13.3 (2006), pp. 271–285. DOI: 10.2174/138161207779313560.

[9] Jordan Tang, ed. *Acid proteases: structure, function, and biology.* 1. 1977, p. 355. ISBN: 9780306326950. DOI: 10.1007/978-1-4757-0719-9.

[10] Hamao Umezawa, Takaaki Aoyagi, Hazime Morishima, Meiki Matsuzaki, Masa Hamada, and Tomio Takeuchi. "Pepstatin, A new pepsin inhibitor produced by actinomycetes". In: *The Journal of Antibiotics* 23.5 (1970), pp. 259–262. DOI: 10.7164/antibiotics.23.259.

[11] Thanh Tam Guyene, Claudine Devaux, Joel Menard, and Pierre Corvol. "Inhibition of human plasma renin activity by pepstatin". In: *Journal of Clinical Endocrinology and Metabolism* 43.6 (1976), pp. 1301–1306. DOI: 10.1210/jcem-43-6-1301.

[12] S. Seelmeier, H. Schmidt, V. Turk, and K. Von Der Helm. "Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A". In: *Proceedings of the National Academy of Sciences of the United States of America* 85.18 (1988), pp. 6612–6616. DOI: 10.1073/pnas.85.18.6612.

[13] Manuel A. Navia, Paula M.D. Fitzgerald, Brian M. McKeever, Chih Tai Leu, Jill C. Heimbach, Wayne K. Herber, Irving S. Sigal, Paul L. Darke, and James P. Springer. "Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1". In: *Nature* 337.6208 (1989), pp. 615–620. DOI: 10.1038/337615a0.

[14] Maria Miller, Jens Schneider, Bangalore K. Sathyanarayana, Mihaly V. Toth, Garland R. Marshall, Leigh Clawson, Linda Selk, Stephen B.H. Kent, and Alexander Wlodawer. "Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution". In: *Science* 246.4934 (1989), pp. 1149–1152. DOI: 10.1126/science.2686029.

[15] Amy L. Swain, Maria M. Miller, Jeremy Green, Daniel H. Rich, Jens Schneider, Stephen B.H. Kent, and Alexander Wlodawer. "X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor". In: *Proceedings of the National Academy of Sciences of the United States of America* 87.22 (1990), pp. 8805–8809. DOI: 10.1073/pnas.87.22.8805.

[16] Boris Turk. "Targeting proteases: Successes, failures and future prospects". In: *Nature Reviews Drug Discovery* 5.9 (2006), pp. 785–799. DOI: 10.1038/nrd2092.

[17] Giovanni Abbenante and David Fairlie. "Protease Inhibitors in the Clinic". In: *Medicinal Chemistry* 1.1 (2005), pp. 71–104. DOI: 10.2174/1573406053402569.

[18] Miles Congreve, Robin Carr, Chris Murray, and Harren Jhoti. "A 'Rule of Three' for fragment-based lead discovery?" In: *Drug Discovery Today* 8.19 (2003), pp. 876–877. DOI: 10.1016/S1359-6446(03)02831-9.

[19] Clara Davoine, Lionel Pochet, and Marianne Fillet. "Trends in Analytical Chemistry Advances in analytical technologies detecting and characterizing noncovalent interactions for fragment-based drug discovery". In: *Trends Anal. Chem.* 166 (2023), p. 117161. DOI: 10.1016/j.trac.2023.117161.

[20] James Osborne, Stanislava Panova, Magdalini Rapti, Tatsuya Urushima, and Harren Jhoti. "Fragments: Where are we now?" In: *Biochemical Society Transactions* 48.1 (2020), pp. 271–280. DOI: 10.1042/BST20190694.

[21] Qingxin Li. "Application of Fragment-Based Drug Discovery to Versatile Targets". In: *Frontiers in Molecular Biosciences* 7.August (2020), pp. 1–13. DOI: 10.3389/fmolb.2020.00180.

[22] Geoffrey A. Holdgate and Walter H.J. Ward. "Measurements of binding thermodynamics in drug discovery". In: *Drug Discovery Today* 10.22 (2005), pp. 1543–1550. DOI: 10.1016/S1359-6446(05)03610-X.

[23] Weina Ma, Liu Yang, and Langchong He. "Overview of the detection methods for equilibrium dissociation constant KD of drug-receptor interaction". In: *Journal of Pharmaceutical Analysis* 8.3 (2018), pp. 147–152. DOI: 10.1016/j.jpha.2018.05.001.

[24] Arne Schön and Ernesto Freire. "Enthalpy screen of drug candidates". In: *Analytical Biochemistry* 513 (2016), pp. 1–6. DOI: 10.1016/j.ab.2016.08.023.

[25] Ernesto Freire. "Do enthalpy and entropy distinguish first in class from best in class?" In: *Drug Discovery Today* 13.19-20 (2008), pp. 869–874. DOI: 10.1016/j.drudis.2008.07.005.

[26] Gerhard Klebe. "Broad-scale analysis of thermodynamic signatures in medicinal chemistry: are enthalpy-favored binders the better development option?" In: *Drug Discovery Today* 24.4 (2019), pp. 943–948. DOI: 10.1016/j.drudis.2019.01.014.

[27] Andreas Hofmann. *Physical chemistry essentials*. 2018, pp. 1–499. DOI: 10.1007/9783319741673.

[28] Margarida Bastos, Olga Abian, Christopher M. Johnson, Frederico Ferreira-da-Silva, Sonia Vega, Ana Jimenez-Alesanco, David Ortega-Alarcon, and Adrian Velazquez-Campoy. "Isothermal titration calorimetry". In: *Nature Reviews Methods Primers* 3.1 (2023). DOI: 10.1038/s43586-023-00199-x.

[29] Michael M. Pierce, C. S. Raman, and Barry T. Nall. "Isothermal titration calorimetry of protein-protein interactions". In: *Methods: A Companion to Methods in Enzymology* 19.2 (1999), pp. 213–221. DOI: 10.1006/meth.1999.0852.

[30] Jana Broecker, Carolyn Vargas, and Sandro Keller. "Revisiting the optimal c value for isothermal titration calorimetry". In: *Analytical Biochemistry* 418.2 (2011), pp. 307–309. DOI: 10.1016/j.ab.2011.07.027.

[31] Eggert Rühmann, Michael Betz, Marie Fricke, Andreas Heine, Martina Schäfer, and Gerhard Klebe. "Thermodynamic signatures of fragment binding: Validation of direct versus displacement ITC titrations". In: *Biochimica et Bio-*

physica Acta - General Subjects 1850.4 (2015), pp. 647–656. DOI: 10.1016/j.bbagen.2014.12.007.

[32] Yan Ling Zhang and Zhong Yin Zhang. "Low-affinity binding determined by titration calorimetry using a high- affinity coupling ligand: A thermodynamic study of ligand binding to protein tyrosine phosphatase 1B". In: *Analytical Biochemistry* 261.2 (1998), pp. 139–148. DOI: 10.1006/abio.1998.2738.

[33] Jetse Reijenga, Arno van Hoof, Antonie van Loon, and Bram Teunissen. "Development of methods for the determination of pK_a values". In: *Analytical Chemistry Insights* 8.1 (2013), pp. 53–71. DOI: 10.4137/ACI.S12304.

[34] Barry Honig and Anthony Nicholls. "Classical electrostatics in biology and chemistry". In: *Science* 268.5214 (1995), pp. 1144–1149. DOI: 10.1126/science.7761829.

[35] Marharyta Petukh, Shannon Stefl, and Emil Alexov. "The Role of Protonation States in Ligand-Receptor Recognition and Binding". In: *Current Pharmaceutical Design* 19.23 (2013), pp. 4182–4190. DOI: 10.2174/13816128113 19230004.

[36] Dilek Coskun. "Computational Approaches for Protein pK_a Calculations". In: *ACS Symposium Series* 1397 (2021), pp. 143–159. DOI: 10.1021/bk-2021-1397.ch006.

[37] Richard L. Thurlkill, Gerald R. Grimsley, J. Martin Scholtz, and C. Nick Pace. "pK values of the ionizable groups of proteins". In: *Protein Science* 15.5 (2006), pp. 1214–1218. DOI: 10.1110/ps.051840806.

[38] Gene Lamm. "The Poisson–Boltzmann Equation". In: *Reviews in Computational Chemistry*. John Wiley & Sons, Ltd, 2003. Chap. 4, pp. 147–365. ISBN: 9780471466635. DOI: 10.1002/0471466638.ch4.

[39] Arieh Warshel. "Calculations of Enzymatic Reactions: Calculations of pK_a, Proton Transfer Reactions, and General Acid Catalysis Reactions in Enzymes". In: *Biochemistry* 20.11 (1981), pp. 3167–3177. DOI: 10.1021/bi00514a028.

[40] Emil Alexov, Ernest L. Mehler, Nathan Baker, António M. Baptista, Yong Huang, Francesca Milletti, Jens Erik Nielsen, Damien Farrell, Tommy Carstensen, Mats H.M. Olsson, Jana K. Shen, Jim Warwicker, Sarah Williams, and J. Michael Word. "Progress in the prediction of pK_a values in proteins". In: *Proteins: Structure, Function and Bioinformatics* 79.12 (2011), pp. 3260–3275. DOI: 10.1002/prot.23189.

[41] Paul Czodrowski, Ingo Dramburg, Christoph A. Sotriffer, and Gerhard Klebe. "Development, validation, and application of adapted PEOE charges to estimate pK_a values of functional groups in protein-ligand complexes". In: *Pro-*

teins: Structure, Function and Genetics 65.2 (2006), pp. 424–437. DOI: 10.1002/prot.21110.

[42] David T. Manallack, Richard J. Prankerd, Elizabeth Yuriev, Tudor I. Oprea, and David K. Chalmers. "The significance of acid/base properties in drug discovery". In: *Chemical Society Reviews* 42.2 (2013), pp. 485–496. DOI: 10.1039/c2cs35348b.

[43] Dolores Diaz et al. "Pharmacokinetic drivers of toxicity for basic molecules: Strategy to lower pK_a results in decreased tissue exposure and toxicity for a small molecule Met inhibitor". In: *Toxicology and Applied Pharmacology* 266.1 (2013), pp. 86–94. DOI: 10.1016/j.taap.2012.10.026.

[44] Wullie Arbuckle et al. "1H-Imidazo[4,5-c]pyridine-4-carbonitrile as cathepsin S inhibitors: Separation of desired cellular activity from undesired tissue accumulation through optimization of basic nitrogen pK_a". In: *Bioorganic and Medicinal Chemistry Letters* 21.3 (2011), pp. 932–935. DOI: 10.1016/j.bmcl.2010.12.065.

[45] Mark Furber et al. "Cathepsin C inhibitors: Property optimization and identification of a clinical candidate". In: *Journal of Medicinal Chemistry* 57.6 (2014), pp. 2357–2367. DOI: 10.1021/jm401705g.

[46] Yasufumi Miyamoto et al. "Discovery of a 3-pyridylacetic acid derivative (TAK-100) as a potent, selective and orally active dipeptidyl peptidase IV (DPP-4) inhibitor". In: *Journal of Medicinal Chemistry* 54.3 (2011), pp. 831–850. DOI: 10.1021/jm101236h.

[47] Renato Ferreira De Freitas and Matthieu Schapira. "A systematic analysis of atomic protein-ligand interactions in the PDB". In: *MedChemComm* 8.10 (2017), pp. 1970–1981. DOI: 10.1039/c7md00381a.

[48] Manuel Neeb, Paul Czodrowski, Andreas Heine, Luzi Jakob Barandun, Christoph Hohn, François Diederich, and Gerhard Klebe. "Chasing protons: How isothermal titration calorimetry, mutagenesis, and pK_a calculations trace the locus of charge in ligand binding to a tRNA-binding enzyme". In: *Journal of Medicinal Chemistry* 57.13 (2014), pp. 5554–5565. DOI: 10.1021/jm500401x.

[49] Paul Czodrowski, Christoph A. Sotriffer, and Gerhard Klebe. "Atypical Protonation States in the Active Site of HIV-1 Protease: A Computational Study". In: *Journal of Chemical Information and Modeling* 47.4 (2007), pp. 1590–1598. DOI: 10.1021/ci600522c.

[50] Milon Mondal, Nedyalka Radeva, Hugo Fanlo-Virgós, Sijbren Otto, Gerhard Klebe, and Anna K.H. Hirsch. "Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment-Based Drug Design Facilitated by Dynamic Combinatorial Chemistry". In: *Angewandte*

Chemie - International Edition 55.32 (2016), pp. 9422–9426. DOI: 10.1002/anie.201603074.

[51] Ina Coburger, Yvonne Schaub, Dirk Roeser, Kornelia Hardes, Patrick Maeder, Nina Klee, Torsten Steinmetzer, Diana Imhof, Wibke E. Diederich, and Manuel E. Than. "Identification of inhibitors of the transmembrane protease FlaK of Methanococcus maripaludis". In: *MicrobiologyOpen* 5.4 (2016), pp. 637–646. DOI: 10.1002/mbo3.358.

[52] L. Coates, P. T. Erskine, S. Mall, P. A. Williams, R. S. Gill, S. P. Wood, and J. B. Cooper. "The structure of endothiapepsin complexed with the gem-diol inhibitor PD-135,040 at 1.37 Å". In: *Acta Crystallographica - Section D Biological Crystallography* 59.6 (2003), pp. 978–981. DOI: 10.1107/S0907444903006267.

[53] Markella Konstantinidou, Francesca Magari, Fandi Sutanto, Jörg Haupenthal, Varsha R. Jumde, M. Yagiz Ünver, Andreas Heine, Carlos Jamie Camacho, Anna K.H. Hirsch, Gerhard Klebe, and Alexander Dömling. "Rapid Discovery of Aspartyl Protease Inhibitors Using an Anchoring Approach". In: *ChemMedChem* 15.8 (2020), pp. 680–684. DOI: 10.1002/cmdc.202000024.

[54] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. "The protein kinase complement of the human genome". In: *Science* 298.5600 (2002), pp. 1912–1934. DOI: 10.1126/science.1075762.

[55] Tony Hunter. "Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling". In: *Cell* 80.2 (1995), pp. 225–236. DOI: 10.1016/0092-8674(95)90405-0.

[56] Irena Melnikova and James Golden. "Targeting protein kinases". In: *Nature Reviews Drug Discovery* 3.12 (2004), pp. 993–994. DOI: 10.1038/nrd1582.

[57] Martin E.M. Noble, Jane A. Endicott, and Louise N. Johnson. "Protein Kinase Inhibitors: Insights into Drug Design from Structure". In: *Science* 303.5665 (2004), pp. 1800–1805. DOI: 10.1126/science.1095920.

[58] Jianming Zhang, Priscilla L. Yang, and Nathanael S. Gray. "Targeting cancer with small molecule kinase inhibitors". In: *Nature Reviews Cancer* 9.1 (2009), pp. 28–39. DOI: 10.1038/nrc2559.

[59] Stephen P. Davies, Helen Reddy, Matilde Caivano, and Philip Cohen. "Specificity and mechanism of action of some commonly used protein kinase inhibitors". In: *Biochemical Journal* 351.1 (2000), pp. 95–105. DOI: 10.1042/bj3510095.

[60] Daniel R. Knighton, Jianhua Zheng, Lynn F. Ten Eyck, Victor A. Ashford, Nguyen Huu Xuong, Susan S. Taylor, and Janusz M. Sowadski. "Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-

dependent protein kinase". In: *Science* 253.5018 (1991), pp. 407–414. DOI: 10.1126/science.1862342.

[61] Daniel K. Treiber and Neil P. Shah. "Ins and Outs of Kinase DFG Motifs". In: *Chemistry and Biology* 20.6 (2013), pp. 745–746. DOI: 10.1016/j.chembiol.2013.06.001.

[62] Janine L Gray, Frank von Delft, and Paul E Brennan. *Targeting the Small GTPase Superfamily through Their Regulatory Proteins*. 2020. DOI: 10.1002/anie.201900585.

[63] Francisco Llavero, Alazne Arrazola Sastre, Miriam Luque Montoro, A Miguel, Joaquín Arenas, Alejandro Lucia, José L Zugaza, Francisco Llavero, Alazne Arrazola Sastre, and Miriam Luque Montoro. "Small GTPases of the Ras superfamily and glycogen phosphorylase regulation in T cells". In: *Small GTPases* 12.2 (2021), pp. 106–113. DOI: 10.1080/21541248.2019.1665968.

[64] Anna Gaulton et al. "The ChEMBL database in 2017". In: *Nucleic Acids Research* 45.D1 (2017), pp. D945–D954. DOI: 10.1093/nar/gkw1074.

[65] Jiang Wang, María Sánchez-Roselló, José Luis Aceña, Carlos Del Pozo, Alexander E. Sorochinsky, Santos Fustero, Vadim A. Soloshonok, and Hong Liu. "Fluorine in pharmaceutical industry: Fluorine-containing drugs introduced to the market in the last decade (2001-2011)". In: *Chemical Reviews* 114.4 (2014), pp. 2432–2506. DOI: 10.1021/cr4002879.

[66] Josef Fried and Emily F. Sabo. "Synthesis of 17α-hydroxycorticosterone and its 9α-halo derivatives from 11-EPI-17α-hydroxycorticosterone". In: *Journal of the American Chemical Society* 75.9 (1953), pp. 2273–2274. DOI: 10.1021/ja01105a527.

[67] Josef Fried and Emily F. Sabo. "9α-Fluoro derivatives of cortisone and hydrocortisone". In: *Journal of the American Chemical Society* 76.5 (1954), pp. 1455–1456. DOI: 10.1021/ja01634a101.

[68] *WHO List of Essential Medicine*. URL: https://list.essentialmeds.org/ (visited on 01/17/2025).

[69] Nicholas A. McGrath, Matthew Brichacek, and Jon T. Njardarson. "A graphical journey of innovative organic architectures that have improved our lives". In: *Journal of Chemical Education* 87.12 (2010), pp. 1348–1349. DOI: 10.1021/ed1003806.

[70] Haibo Mei, Attila Márió Remete, Yupiao Zou, Hiroki Moriwaki, Santos Fustero, Lorand Kiss, Vadim A. Soloshonok, and Jianlin Han. "Fluorine-containing drugs approved by the FDA in 2019". In: *Chinese Chemical Letters* 31.9 (2020), pp. 2401–2413. DOI: 10.1016/j.cclet.2020.03.050.

[71] Yingjie Yu, Aiyao Liu, Gagan Dhawan, Haibo Mei, Wei Zhang, Kunisuke Izawa, Vadim A. Soloshonok, and Jianlin Han. "Fluorine-containing pharma-

ceuticals approved by the FDA in 2020: Synthesis and biological activity". In: *Chinese Chemical Letters* 32.11 (2021), pp. 3342–3354. DOI: 10.1016/j.cclet.2021.05.042.

[72] Jingrui He, Ziyi Li, Gagan Dhawan, Wei Zhang, Alexander E. Sorochinsky, Greg Butler, Vadim A. Soloshonok, and Jianlin Han. "Fluorine-containing drugs approved by the FDA in 2021". In: *Chinese Chemical Letters* 34.1 (2023). DOI: 10.1016/j.cclet.2022.06.001.

[73] Klaus Müller, Christoph Faeh, and François Diederich. "Fluorine in Pharmaceuticals: Looking Beyond Intuition". In: *Science* 317 (5846 2007), pp. 1881–1886. DOI: 10.1126/science.1131943.

[74] Eric P. Gillis, Kyle J. Eastman, Matthew D. Hill, David J. Donnelly, and Nicholas A. Meanwell. "Applications of Fluorine in Medicinal Chemistry". In: *Journal of Medicinal Chemistry* 58.21 (2015), pp. 8315–8359. DOI: 10.1021/acs.jmedchem.5b00258.

[75] Sun Young Han, Jie Won Choi, Jeon Yang, Chong Hack Chae, Jongkook Lee, Heejung Jung, Kwangho Lee, Jae Du Ha, Hyoung Rae Kim, and Sung Yun Cho. "Design and synthesis of 3-(4,5,6,7-tetrahydro-3H-imidazo[4,5-c]pyridin-2- yl)-1H-quinolin-2-ones as VEGFR-2 kinase inhibitors". In: *Bioorganic and Medicinal Chemistry Letters* 22.8 (2012), pp. 2837–2842. DOI: 10.1016/j.bmcl.2012.02.073.

[76] Yan Lou, Zachary K. Sweeney, Andreas Kuglstatter, Dana Davis, David M. Goldstein, Xiaochun Han, Junbae Hong, Buelent Kocer, Rama K. Kondru, Renee Litman, Joel McIntosh, Keshab Sarma, Judy Suh, Joshua Taygerly, and Timothy D. Owens. "Finding the perfect spot for fluorine: Improving potency up to 40-fold during a rational fluorine scan of a Bruton's Tyrosine Kinase (BTK) inhibitor scaffold". In: *Bioorganic and Medicinal Chemistry Letters* 25.2 (2015), pp. 367–371. DOI: 10.1016/j.bmcl.2014.11.030.

[77] Tudor I Oprea. *Cheminformatics in Drug Discovery*. Wiley-VCH, 2005. DOI: 10.1002/3527603743.

[78] Christian Tyrchan and Emma Evertsson. "Matched Molecular Pair Analysis in Short: Algorithms, Applications and Limitations". In: *Computational and Structural Biotechnology Journal* 15 (2017), pp. 86–90. DOI: 10.1016/j.csbj.2016.12.003.

[79] Anne Mai Wassermann, Dilyana Dimova, Preeti Iyer, and Jürgen Bajorath. "Advances in computational medicinal chemistry: Matched molecular pair analysis". In: *Drug Development Research* 73.8 (2012), pp. 518–527. DOI: 10.1002/ddr.21045.

[80] J. F. Lyons, S. Wilhelm, B. Hibner, and G. Bollag. "Discovery of a novel Raf kinase inhibitor". In: *Endocrine-Related Cancer* 8.3 (2001), pp. 219–225. DOI: 10.1677/erc.0.0080219.

[81] Scott M. Wilhelm, Jacques Dumas, Lila Adnane, Mark Lynch, Christopher A. Carter, Gunnar Schütz, Karl Heinz Thierauch, and Dieter Zopf. "Regorafenib (BAY 73-4506): A new oral multikinase inhibitor of angiogenic, stromal and oncogenic receptor tyrosine kinases with potent preclinical antitumor activity". In: *International Journal of Cancer* 129.1 (2011), pp. 245–255. DOI: 10.1002/ijc.25864.

[82] Mike Bührmann. "Strukturbiologische Untersuchungen von Kinaseinhibitoren Entwicklung, Synthese und biophysikalische Charakterisierung von Liganden zur Adressierung einer lipophilen Bindetasche in der Mitogen-aktivierten Proteinkinase p38α". PhD thesis. 2017.

[83] Chi Uyen Phan, Jie Shen, Jiyong Liu, Jianming Mao, Xiurong Hu, and Guping Tang. "Isomorphous crystals formed by the similar supramolecular motifs in sorafenib hydrochloride and regorafenib hydrochloride salts". In: *Crystals* 9.12 (2019). DOI: 10.3390/cryst9120649.

[84] Shobha N. Bhattachar, Laura A. Deschenes, and James A. Wesley. "Solubility: it's not just for physical chemists". In: *Drug Discovery Today* 11.21-22 (2006), pp. 1012–1018. DOI: 10.1016/j.drudis.2006.09.002.

[85] Steffen Glöckner and Gerhard Klebe. "Simultaneous determination of thermodynamic and kinetic data by isothermal titration calorimetry". In: *Biochimica et Biophysica Acta - General Subjects* 1865.2 (2021), p. 129772. DOI: 10.1016/j.bbagen.2020.129772.

[86] Jeffrey R. Simard, Matthäus Getlik, Christian Grütter, Vijaykumar Pawar, Sabine Wulfert, Matthias Rabiller, and Daniel Rauh. "Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors". In: *Journal of the American Chemical Society* 131.37 (2009), pp. 13286–13296. DOI: 10.1021/ja902010p.

[87] Michael R. Berthold, Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinl, Peter Ohl, Christoph Sieb, Kilian Thiel, and Bernd Wiswedel. "KNIME: The Konstanz Information Miner". In: *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer, 2007. ISBN: 978-3-540-78239-1.

[88] *openkinome*. URL: https://github.com/openkinome/kinodata/blob/master/data/human_kinases.aggregated.csv (visited on 01/19/2025).

[89] Philip J. Hajduk and Daryl R. Sauer. "Statistical analysis of the effects of common chemical substituents on ligand potency". In: *Journal of Medicinal Chemistry* 51.3 (2008), pp. 553–564. DOI: 10.1021/jm070838y.

[90] Matthias Oebbeke, Christof Siefker, Björn Wagner, Andreas Heine, and Gerhard Klebe. "Fragment Binding to Kinase Hinge: If Charge Distribution and Local pK_a Shifts Mislead Popular Bioisosterism Concepts". In: *Angewandte Chemie - International Edition* 60.1 (2021), pp. 252–258. DOI: 10.1002/anie.202011295.

[91] Daniel C. Harris. *Quantitative Chemical Analysis*. 7th. 2007. DOI: 10.1016/B978-0-444-40826-6.50009-1.

[92] Barbara Wienen-Schmidt, Hendrik R.A. Jonker, Tobias Wulsdorf, Hans Dieter Gerber, Krishna Saxena, Denis Kudlinzki, Sridhar Sreeramulu, Giacomo Parigi, Claudio Luchinat, Andreas Heine, Harald Schwalbe, and Gerhard Klebe. "Paradoxically, Most Flexible Ligand Binds Most Entropy-Favored: Intriguing Impact of Ligand Flexibility and Solvation on Drug-Kinase Binding". In: *Journal of Medicinal Chemistry* 61.14 (2018), pp. 5922–5933. DOI: 10.1021/acs.jmedchem.8b00105.

[93] *Missing Coordinates and Biological Assemblies*. URL: https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/missing-coordinates (visited on 01/17/2025).

[94] Javier Gómez and Ernesto Freire. "Thermodynamic Mapping of the Inhibitor Site of the Aspartic Protease Endothiapepsin". In: *Journal of Molecular Biology* 252.3 (1995), pp. 337–350. DOI: 10.1006/jmbi.1995.0501.

[95] Khang Ngo, Chelsey Collins-Kautz, Stefan Gerstenecker, Björn Wagner, Andreas Heine, and Gerhard Klebe. "Protein-Induced Change in Ligand Protonation during Trypsin and Thrombin Binding: Hint on Differences in Selectivity Determinants of Both Proteins?" In: *Journal of Medicinal Chemistry* 63.6 (2020), pp. 3274–3289. DOI: 10.1021/acs.jmedchem.9b02061.

[96] Johannes Schiebel, Roberto Gaspari, Anna Sandner, Khang Ngo, Hans Dieter Gerber, Andrea Cavalli, Andreas Ostermann, Andreas Heine, and Gerhard Klebe. "Charges Shift Protonation: Neutron Diffraction Reveals that Aniline and 2-Aminopyridine Become Protonated Upon Binding to Trypsin". In: *Angewandte Chemie - International Edition* 56.17 (2017), pp. 4887–4890. DOI: 10.1002/anie.201701038.

[97] Johannes Schiebel, Nedyalka Radeva, Stefan G. Krimmer, Xiaojie Wang, Martin Stieler, Frederik R. Ehrmann, Kan Fu, Alexander Metz, Franziska U. Huschmann, Manfred S. Weiss, Uwe Mueller, Andreas Heine, and Gerhard Klebe. "Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study". In: *ACS Chemical Biology* 11.6 (2016), pp. 1693–1701. DOI: 10.1021/acschembio.5b01034.

[98] Chia Ying Huang, Sylvain Aumonier, Sylvain Engilberge, Deniz Eris, Kate Mary Louise Smith, Filip Leonarski, Justyna Aleksandra Wojdyla, John H.

Beale, Dominik Buntschu, Anuschka Pauluhn, May Elizabeth Sharpe, Alexander Metz, Vincent Olieric, and Meitian Wang. "Probing ligand binding of endothiapepsin by 'temperature-resolved' macromolecular crystallography". In: *Acta Crystallographica Section D: Structural Biology* 78 (2022), pp. 964–974. DOI: 10.1107/S205979832200612X.

[99] Helene Köster, Tobias Craan, Sascha Brass, Christian Herhaus, Matthias Zentgraf, Lars Neumann, Andreas Heine, and Gerhard Klebe. "A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes". In: *Journal of Medicinal Chemistry* 54.22 (2011), pp. 7784–7796. DOI: 10.1021/jm200642w.

[100] Stefan G. Krimmer and Gerhard Klebe. "Thermodynamics of protein-ligand interactions as a reference for computational analysis: How to assess accuracy, reliability and relevance of experimental data". In: *Journal of Computer-Aided Molecular Design* 29.9 (2015), pp. 867–883. DOI: 10.1007/s10822-015-9867-y.

[101] Maren Kuhnert, Helene Köster, Ruben Bartholomäus, Ah Young Park, Amir Shahim, Andreas Heine, Holger Steuber, Gerhard Klebe, and Wibke E. Diederich. "Tracing binding modes in hit-to-lead optimization: Chameleon-like poses of aspartic protease inhibitors". In: *Angewandte Chemie - International Edition* 54.9 (2015), pp. 2849–2853. DOI: 10.1002/anie.201411206.

[102] Helge Vatheuer, Oscar Palomino-Hernández, Janis Müller, Phillip Galonska, Serghei Glinca, and Paul Czodrowski. "Protonation effects in protein-ligand complexes - a case study of endothiapepsin and pepstatin A with computational and experimental methods". In: *bioRxiv* (2024). DOI: 10.1101/2024.09.12.612797.

[103] Courtney L. Stanton and Kendall N. Houk. "Benchmarking pK prediction methods for residues in proteins". In: *Journal of Chemical Theory and Computation* 4.6 (2008), pp. 951–966. DOI: 10.1021/ct8000014.

[104] Florian Hofer, Johannes Kraml, Ursula Kahler, Anna S. Kamenik, and Klaus R. Liedl. "Catalytic Site pK_a Values of Aspartic, Cysteine, and Serine Proteases: Constant pH MD Simulations". In: *Journal of Chemical Information and Modeling* 60.6 (2020), pp. 3030–3042. DOI: 10.1021/acs.jcim.0c00190.

[105] Natali V. Di Russo, Dario A. Estrin, Marcelo A. Martí, and Adrian E. Roitberg. "pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pK_as: The Case of Nitrophorin 4". In: *PLoS Computational Biology* 8.11 (2012), pp. 88–100. DOI: 10.1371/journal.pcbi.1002761.

[106] A. Wlodawer, M. Li, A. Gustchina, Z. Dauter, K. Uchida, H. Oyama, N. E. Goldfarb, B. M. Dunn, and K. Oda. "Inhibitor complexes of the Pseudomonas

serine-carboxyl proteinase". In: *Biochemistry* 40.51 (2001), pp. 15602–15611. DOI: 10.1021/bi011817n.

[107] Jiří Dostál, Adam Pecina, Olga Hrušková-Heidingsfeldová, Lucie Marečková, Iva Pichová, Pavlina Řezáčová, Martin Lepšík, and Jiří Brynda. "Atomic resolution crystal structure of Sapp2p, a secreted aspartic protease from *Candida parapsilosis*". In: *Acta Crystallographica Section D Biological Crystallography* 71.12 (2015), pp. 2494–2504. DOI: 10.1107/S1399004715019392.

[108] Paul Czodrowski, Christoph A. Sotriffer, and Gerhard Klebe. "Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pK_a Calculations and ITC Experiments". In: *Journal of Molecular Biology* 367.5 (2007), pp. 1347–1356. DOI: 10.1016/j.jmb.2007.01.022.

[109] Robert N. Goldberg, Nand Kishore, and Rebecca M. Lennen. "Thermodynamic quantities for the ionization reactions of buffers". In: *Journal of Physical and Chemical Reference Data* 31.2 (2002), pp. 231–370. DOI: 10.1063/1.1416902.

[110] Dragana Cubrilovic and Renato Zenobi. "Influence of Dimehylsulfoxide on Protein–Ligand Binding Affinities". In: *Analytical Chemistry* 85.5 (2013), pp. 2724–2730. DOI: 10.1021/ac303197p.

[111] Lawrence J. Hyland, Thaddeus A. Tomaszek, and Thomas D. Meek. "Human Immunodeficiency Virus-1 Protease. 2. Use of pH Rate Studies and Solvent Kinetic Isotope Effects To Elucidate Details of Chemical Mechanism". In: *Biochemistry* 30.34 (1991), pp. 8454–8463. DOI: 10.1021/bi00098a024.

[112] Jack A. Henderson, Robert C. Harris, Cheng Chieh Tsai, and Jana Shen. "How Ligand Protonation State Controls Water in Protein-Ligand Binding". In: *Journal of Physical Chemistry Letters* 9.18 (2018), pp. 5440–5444. DOI: 10.1021/acs.jpclett.8b02440.

[113] Lin Hong and Jordan Tang. "Flap Position of Free Memapsin 2 (β-Secretase), a Model for Flap Opening in Aspartic Protease Catalysis". In: *Biochemistry* 43.16 (2004), pp. 4689–4695. DOI: 10.1021/bi0498252.

[114] Alke Meents, Bernd Reime, Nicolas Stuebe, Pontus Fischer, Martin Warmer, Dennis Goeries, Jan Roever, Jan Meyer, Janine Fischer, Anja Burkhardt, Ismo Vartiainen, Petri Karvinen, and Christian David. "Development of an in-vacuum x-ray microscope with cryogenic sample cooling for beamline P11 at PETRA III". In: *X-Ray Nanoimaging: Instruments and Methods* 8851 (2013), 88510K. DOI: 10.1117/12.2027303.

[115] Wolfgang Kabsch. "*XDS*". In: *Acta Crystallographica Section D* 66.2 (2010), pp. 125–132. DOI: 10.1107/S0907444909047337.

[116] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. "Features and development of Coot". In: *Acta Crystallographica Section D: Biological Crystallography* 66.4 (2010), pp. 486–501. DOI: 10.1107/S0907444910007493.

[117] Dorothee Liebschner et al. "Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix". In: *Acta Crystallographica Section D: Structural Biology* 75 (2019), pp. 861–877. DOI: 10.1107/S2059798319011471.

[118] J. Michael Word and Anthony Nicholls. "Application of the Gaussian dielectric boundary in Zap to the prediction of protein pK_a values". In: *Proteins: Structure, Function and Bioinformatics* 79.12 (2011), pp. 3400–3409. DOI: 10.1002/prot.23079.

[119] ChemAxon. *MarvinSketch 19.10*. 2019. URL: www.chemaxon.com.

[120] *SZYBKI 2.7.1.0. OpenEye, Cadence Molecular Sciences, Santa Fe, NM*. 2024. URL: http://www.eyesopen.com (visited on 01/17/2025).

[121] Laurence Pearl and Tom Blundell. "The active site of aspartic proteinases". In: *FEBS Letters* 174.1 (1984), pp. 96–101. DOI: 10.1016/0014-5793(84)81085-6.

[122] Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof, and Berk Hess. "Scalable Constant pH Molecular Dynamics in GROMACS". In: *Journal of Chemical Theory and Computation* 18.10 (2022), pp. 6148–6160. DOI: 10.1021/acs.jctc.2c00516.

[123] Pavel Buslaev, Noora Aho, Anton Jansen, Paul Bauer, Berk Hess, and Gerrit Groenhof. "Best Practices in Constant pH MD Simulations: Accuracy and Sampling". In: *Journal of Chemical Theory and Computation* 18.10 (2022), pp. 6134–6147. DOI: 10.1021/acs.jctc.2c00517.

[124] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields". In: *Journal of Computational Chemistry* 31.4 (2010), pp. 671–690. DOI: 10.1002/jcc.21367.

[125] Giovanni Bussi, Davide Donadio, and Michele Parrinello. "Canonical sampling through velocity rescaling". In: *Journal of Chemical Physics* 126.1 (2007). DOI: 10.1063/1.2408420. arXiv: 0803.4060.

[126] Mattia Bernetti and Giovanni Bussi. "Pressure control using stochastic cell rescaling". In: *Journal of Chemical Physics* 153.11 (2020), pp. 1–11. DOI: 10.1063/5.0020514. arXiv: 2006.09250.

[127] Hiroyasu Ohtaka, Adrian Velázquez-Campoy, Dong Xie, and Ernesto Freire. "Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants

of the HIV-1 protease". In: *Protein Science* 11.8 (2002), pp. 1908–1916. DOI: 10.1110/ps.0206402.

[128] Luca Kröll. "Protonation effects in protein ligand complexes". 2022.

[129] Adrian Velazquez-Campoy, Irene Luque, Matthew J. Todd, Mark Milutinovich, Yoshiaki Kiso, and Ernesto Freire. "Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor". In: *Protein Science* 9.9 (2000), pp. 1801–1809. DOI: 10.1110/ps.9.9.1801.

[130] Toshimasa Yamazaki, Linda K Nicholson, Paul Wingfield, Stephen J Stahl, Joshua D Kaufman, Charles J Eyermann, C Nicholas Hodge, Patrick Y S Lam, and D A Torchia. "NMR and X-ray Evidence That the HIV Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-Based Inhibitor". In: *Journal of the American Chemical Society* 116 (23 1994). doi: 10.1021/ja00102a057, pp. 10791–10792. DOI: 10.1021/ja00102a057.

[131] Lisa Goebel, Tonia Kirschner, Sandra Koska, Amrita Rai, Petra Janning, Stefano Maffini, Helge Vatheuer, Paul Czodrowski, Roger S Goody, and Daniel Rauh. "Targeting oncogenic KRasG13C with nucleotide-based covalent inhibitors". In: *eLife* 61 (2022), pp. 1–19. DOI: 10.7554/eLife.82184.

[132] Sanjana S. Murdande. "Role of rat sarcoma virus mutations in cancer and potential target for cancer therapy". In: *Future Science OA* 6.4 (2020), pp. 1–6. DOI: 10.2144/fsoa-2019-0045.

[133] Adrienne D. Cox, Stephen W. Fesik, Alec C. Kimmelman, Ji Luo, and Channing J. Der. "Drugging the undruggable RAS: Mission Possible?" In: *Nature Reviews Drug Discovery* 13.11 (2014), pp. 828–851. DOI: 10.1038/nrd4389.

[134] G. Aaron Hobbs, Channing J. Der, and Kent L. Rossman. "RAS isoforms and mutations in cancer at a glance". In: *Journal of Cell Science* 129.7 (2016), pp. 1287–1292. DOI: 10.1242/jcs.182873.

[135] Simon A. Forbes et al. "COSMIC: Exploring the world's knowledge of somatic mutations in human cancer". In: *Nucleic Acids Research* 43.D1 (2015), pp. D805–D811. DOI: 10.1093/nar/gku1075.

[136] Marieke Visscher, Michelle R. Arkin, and Tobias B. Dansen. "Covalent targeting of acquired cysteines in cancer". In: *Current Opinion in Chemical Biology* 30 (2016), pp. 61–67. DOI: 10.1016/j.cbpa.2015.11.004.

[137] Goedele Roos, Nicolas Foloppe, and Joris Messens. "Understanding the pK_a of redox cysteines: The key role of hydrogen bonding". In: *Antioxidants and Redox Signaling* 1 (), pp. 94–127. DOI: 10.1089/ars.2012.4521.

[138] Rasmus Hansen, Ulf Peters, Anjali Babbar, Yuching Chen, Jun Feng, Matthew R. Janes, Lian Sheng Li, Pingda Ren, Yi Liu, and Patrick P. Zarrinkar. "The reactivity-driven biochemical mechanism of covalent KRASG12C inhibitors".

In: *Nature Structural and Molecular Biology* 25.6 (2018), pp. 454–462. DOI: 10.1038/s41594-018-0061-5.

[139] Minh V. Huynh, Derek Parsonage, Tom E. Forshaw, Venkat R. Chirasani, G. Aaron Hobbs, Hanzhi Wu, Jingyun Lee, Cristina M. Furdui, Leslie B. Poole, and Sharon L. Campbell. "Oncogenic KRAS G12C: Kinetic and redox characterization of covalent inhibition". In: *Journal of Biological Chemistry* 298.8 (2022), p. 102186. DOI: 10.1016/j.jbc.2022.102186.

[140] Robert C. Harris, Ruibin Liu, and Jana Shen. "Predicting Reactive Cysteines with Implicit-Solvent-Based Continuous Constant pH Molecular Dynamics in Amber". In: *Journal of Chemical Theory and Computation* 16.6 (2020), pp. 3689–3698. DOI: 10.1021/acs.jctc.0c00258.

[141] Ernest Awoonor-Williams and Christopher N. Rowley. "Evaluation of Methods for the Calculation of the p$K_a$of Cysteine Residues in Proteins". In: *Journal of Chemical Theory and Computation* 12.9 (2016), pp. 4662–4673. DOI: 10.1021/acs.jctc.6b00631.

[142] Mingjie Gao and Stefan Günther. "HyperCys: A Structure- and Sequence-Based Predictor of Hyper-Reactive Druggable Cysteines". In: *International journal of molecular sciences* 24.6 (2023). DOI: 10.3390/ijms24065960.

[143] Pedro B.P.S. Reis, Marco Bertolini, Floriane Montanari, Walter Rocchia, Miguel Machuqueiro, and Djork Arné Clevert. "A Fast and Interpretable Deep Learning Approach for Accurate Electrostatics-Driven pK_a Predictions in Proteins". In: *Journal of Chemical Theory and Computation* 18.8 (2022), pp. 5068–5078. DOI: 10.1021/acs.jctc.2c00308.

[144] Sascha Jung, Helge Vatheuer, and Paul Czodrowski. "VSFlow: an open - source ligand - based virtual screening tool". In: *Journal of Cheminformatics* (2023), pp. 1–10. DOI: 10.1186/s13321-023-00703-1.

[145] Eduardo Habib Bechelane Maia, Letícia Cristina Assis, Tiago Alves de Oliveira, Alisson Marques da Silva, and Alex Gutterres Taranto. "Structure-Based Virtual Screening: From Classical to Artificial Intelligence". In: *Frontiers in Chemistry* 8 (2020). DOI: 10.3389/fchem.2020.00343.

[146] Aleix Gimeno, María José Ojeda-Montes, Sarah Tomás-Hernández, Adrià Cereto-Massagué, Raúl Beltrán-Debón, Miquel Mulero, Gerard Pujadas, and Santiago Garcia-Vallvé. "The light and dark sides of virtual screening: What is there to know?" In: *International Journal of Molecular Sciences* 20 (6 2019). DOI: 10.3390/ijms20061375.

[147] Gerald Maggiora, Martin Vogt, Dagmar Stumpfe, and Jürgen Bajorath. "Molecular Similarity in Medicinal Chemistry". In: *Journal of Medicinal Chemistry* 57.8 (2014). PMID: 24151987, pp. 3186–3204. DOI: 10.1021/jm401411z.

[148] *ROCS 3.4.3.0: OpenEye Scientific Software, Santa Fe, NM.* 2022. URL: https://www.eyesopen.com (visited on 01/17/2025).

[149] Pedro H.M. Torres, Ana C.R. Sodero, Paula Jofily, and Floriano P. Silva-Jr. "Key topics in molecular docking for drug design". In: *International Journal of Molecular Sciences* 20 (18 2019). DOI: 10.3390/ijms20184574.

[150] Luca Pinzi and Giulio Rastelli. "Molecular docking: Shifting paradigms in drug discovery". In: *International Journal of Molecular Sciences* 20 (18 2019). DOI: 10.3390/ijms20184331.

[151] Francesco Gentile, Jean Charle Yaacoub, James Gleave, Michael Fernandez, Anh-Tien Ton, Fuqiang Ban, Abraham Stern, and Artem Cherkasov. "Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking". In: *Nature Protocols* 17 (3 2022), pp. 672–697. DOI: 10.1038/s41596-021-00659-2.

[152] Wendy A. Warr, Marc C. Nicklaus, Christos A. Nicolaou, and Matthias Rarey. "Exploration of Ultralarge Compound Collections for Drug Discovery". In: *Journal of Chemical Information and Modeling* 62 (9 2022), pp. 2021–2034. DOI: 10.1021/acs.jcim.2c00224.

[153] Uta Lessel, Bernd Wellenzohn, Markus Lilienthal, and Holger Claussen. "Searching fragment spaces with feature trees". In: *Journal of Chemical Information and Modeling* 49 (2 2009), pp. 270–279. DOI: 10.1021/ci800272a.

[154] *infiniSee version 4.0.0; BioSolveIT GmbH, Sankt Augustin, Germany.* 2022. URL: www.biosolveit.de/infiniSee (visited on 01/17/2025).

[155] Natesh Singh, Ludovic Chaput, and Bruno O. Villoutreix. "Virtual screening web servers: Designing chemical probes and drug candidates in the cyberspace". In: *Briefings in Bioinformatics* 22 (2 2021), pp. 1790–1818. DOI: 10.1093/bib/bbaa034.

[156] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A. Shoemaker, Paul A. Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E. Bolton. "PubChem in 2021: New data content and improved web interfaces". In: *Nucleic Acids Research* 49 (D1 2021), pp. D1388–D1395. DOI: 10.1093/nar/gkaa971.

[157] Teague Sterling and John J. Irwin. "ZINC 15 – Ligand Discovery for Everyone". In: *Journal of Chemical Information and Modeling* 55.11 (2015), pp. 2324–2337. DOI: 10.1021/acs.jcim.5b00559.

[158] Maiia E. Bragina, Antoine Daina, Marta A.S. Perez, Olivier Michielin, and Vincent Zoete. "The SwissSimilarity 2021 Web Tool: Novel Chemical Libraries and Additional Methods for an Enhanced Ligand-Based Virtual Screening Experience". In: *International Journal of Molecular Sciences* 23 (2 2022). DOI: 10.3390/ijms23020811.

[159] Vincent Zoete, Antoine Daina, Christophe Bovigny, and Olivier Michielin. "SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening". In: *Journal of Chemical Information and Modeling* 56 (8 2016), pp. 1399–1404. DOI: 10.1021/acs.jcim.6b00174.

[160] Jocelyn Sunseri and David Ryan Koes. "Pharmit: interactive exploration of chemical space". In: *Nucleic acids research* 44 (W1 2016), W442–W448. DOI: 10.1093/nar/gkw287.

[161] Andrew Dalke. "The chemfp project". In: *Journal of Cheminformatics* 11 (1 2019). DOI: 10.1186/s13321-019-0398-8.

[162] URL: https://github.com/schrodinger/gpusimilarity (visited on 01/17/2025).

[163] URL: https://www.nextmovesoftware.com/arthor.html (visited on 01/17/2025).

[164] *RDKit: Open-Source Cheminformatics Software*. 2022. URL: https://www.rdkit.org (visited on 01/17/2025).

[165] *Anaconda Software Distribution*. Version Vers. 2-2.4.0. 2020. URL: https://docs.anaconda.com/.

[166] URL: https://github.com/mcs07/MolVS (visited on 01/17/2025).

[167] *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.* 2022. URL: https://pymol.org.

[168] Shuzhe Wang, Jagna Witek, Gregory A. Landrum, and Sereina Riniker. "Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences". In: *Journal of Chemical Information and Modeling* 60.4 (2020), pp. 2044–2058. DOI: 10.1021/acs.jcim.0c00025.

[169] URL: https://github.com/czodrowskilab/VSFlow/wiki (visited on 01/17/2025).

[170] Sereina Riniker and Gregory A. Landrum. "Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods". In: *Journal of Cheminformatics* 5.1 (2013), p. 43. DOI: 10.1186/1758-2946-5-43.

[171] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, T N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. "The Protein Data Bank". In: *Nucleic Acids Research* 28 (1 2000). URL: https://www.rcsb.org.

[172] Sebastian G. Rohrer and Knut Baumann. "Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data". In: *Journal of Chemical Information and Modeling* 49 (2 2009), pp. 169–184. DOI: 10.1021/ci8002649.

[173] Sereina Riniker and Gregory A Landrum. *Open-source platform to benchmark fingerprints for ligand-based virtual screening*. 2013, p. 26. DOI: 10.1186/1758-2946-5-26.

[174] Peter R. Young et al. "Pyridinyl imidazole inhibitors of p38 mitogen-activated protein kinase bind in the ATP site". In: *Journal of Biological Chemistry* 272.18 (1997), pp. 12116–12121. DOI: 10.1074/jbc.272.18.12116.

Eidesstattliche Versicherung (Affidavit)

_____ _____

Name, Vorname
(Surname, first name)

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

_____ _____

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

_____

_____

_____

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

_____ _____

Ort, Datum
(Place, date)

Unterschrift
(Signature)