Interaction kinetic analysis in drug design, enzymology and protein research

VLADIMIR O TALIBOV
Abstract


The work presented here is focused on the phenomenon of molecular recognition – the mutual ability of biological molecules to recognize each other through their chemical signatures. Here, the kinetic aspects of recognition were evaluated, as interaction kinetics reveal valuable dimensions in the description of molecular events in biological systems. The primary objects studied in this thesis were human proteins and their interaction partners. Proteins serve a fundamental role in living organisms, supporting the biochemical machinery by means of catalysis, signalling and transport; additionally, proteins are the main targets for drugs.

In the first study, carbonic anhydrase (CA) isozymes were employed as a model system to address the problem of drug selectivity. Kinetic signatures preferable for the design of selective sulphonamide-based inhibitors were identified. In a follow-up study, the recognition between CA and sulphonamides was separated into two parts, uncovering intrinsic recognition features that genuinely reflect the interaction mechanism. For the first time, the concept of intrinsic interaction kinetics was applied to a drug-target system.

Another model protein studied in this thesis was calmodulin (CaM), as its interactions with other proteins should have specific kinetic signatures to support the dynamics of calcium-dependent signalling. The study evolved around calcium-dependent CaM interactions with the neuronal protein neurogranin (Ng), and revealed its complex nature. Ng was found to interact with CaM both in presence and absence of calcium, but with different kinetics and affinity. This finding supports development of a mechanistic model of calcium sensitivity regulation.

The last two projects were more applied, exploring the druggability of an emerging class of pharmaceutical targets – epigenetic enzymes. Expertise and methodology for biophysically guided drug discovery towards histone demethylase LSD1 and histone methyltransferase SMYD3 were developed. For LSD1, the project assisted the rational design of active site-targeting macrocyclic peptides, and resulted in the development of competitive inhibitors with a well described mechanism of action. A novel biophysical platform for screening was developed for SMYD3. It proved to be successful, as it identified previously unknown allosteric ligand binding site. Both projects were supported by structural studies, expanding the druggable space of epigenetic targets.

Vladimir O Talibov, Department of Chemistry - BMC, Box 576, Uppsala University, SE-75123 Uppsala, Sweden.

© Vladimir O Talibov 2018

ISSN 1651-6214
urn:nbn:se:uu:diva-363323 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-363323)
Science is made up of so many things that appear obvious after they are explained.
List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the publishers. “***” indicates equal contribution.


Reprints were made with permission from the publishers.
Contribution report

- Paper I. Designed and performed all kinetic experiments, analysed data, wrote the manuscript.
- Paper II. Developed kinetic theory, designed experiments, contributed to the writing of the manuscript.
- Paper III. Performed MST and fluorescent measurements, contributed to the writing of the manuscript.
- Paper IV. Produced proteins, performed biophysical and biochemical experiments, crystallized proteins and solved structures, contributed to the writing of the manuscript.
- Paper V. Produced the protein, designed screening strategy, performed biophysical experiments, crystallized protein and solved the structure, wrote the manuscript.
Contents

1 Introduction .................................................................................................................. 11
  1.1 Molecular recognition ......................................................................................... 11
  1.2 Protein-ligand interactions ................................................................................. 12

2 Early-stage drug discovery ......................................................................................... 15
  2.1 Enzymes and enzyme inhibitors ......................................................................... 15
  2.2 Biophysical methods in medicinal chemistry ..................................................... 17
  2.3 Biosensors
    2.3.1 SPR biosensors: detection principles .............................................................. 19
    2.3.2 SPR biosensor-based interaction kinetic analysis ......................................... 19
    2.3.3 General remarks regarding interaction kinetics ............................................ 21
  2.4 Protein crystallography
    2.4.1 Diffraction experiments, data acquisition and analysis ............................... 22
    2.4.2 Structure-based drug discovery ..................................................................... 23

3 Present investigation .................................................................................................. 25
  3.1 Selectivity and interaction kinetics (Papers I&II) ................................................. 25
    3.1.1 Interaction kinetic analysis of the sulphonamide inhibitors ...................... 26
    3.1.2 Intrinsic interaction kinetics ..................................................................... 30
  3.2 Recognition in calcium signalling (Paper III) .................................................... 32
  3.3 Druggability of epigenetic enzymes (Papers IV&V) ........................................... 35
    3.3.1 Macrocyclic peptides target the active site of LSD1 ............................... 35
    3.3.2 Expanding the druggable space of SMYD3 ................................................. 39

4 Summary and future perspectives ............................................................................. 42

5 Sammanfattning ......................................................................................................... 43

6 Acknowledgments ....................................................................................................... 46
Abbreviations

CA Carbonic Anhydrase
CaBP Ca\(^{2+}\)-binding protein
CaM Calmodulin
CaMKII Calmodulin-dependent Kinase II
FBLD Fragment-Based Lead Design
FTSA Fluorescent Thermal Shift Assay
HDM Histone-demethylase
HMT Histone-methyltransferase
HTS High-throughput Screening
LSD1 Lysin-Specific Demethylase 1
MAO Monoamine oxidase
MST Microscale Thermophoresis
Ng Neurogranin
NMR Nuclear Magnetic Resonance
SA Sulphonamide
SAH S-adenosyl-homocysteine
SAM S-adenosyl-methionine
SMYD3 SET-and MYND-domains containing histone methylase 3
SPR Surface Plasmon Resonance
1. Introduction

Chemical space is infinite, owing to its combinatorial nature. One can select multiple combinations of atoms to create a molecule, with no expected limit of the numbers of unique structures. Surprisingly, the number of different endogenic small molecules, even for the most complex organisms, is not much than a few thousand molecular species [1]. Meanwhile, known molecules, either of natural or synthethic origin, with verified activities towards biological targets outnumber this figure significantly [2]. Even when employing mostly biogenic elements and certain structural restraints [3] to reduce the chemical space, the selection of possible structures for modulation of living systems exceeds $10^{60}$ species [4]. Obviously, chemical space includes a countless collection of compounds with desired properties, useful for both studying the biological systems and being pharmaceuticals.

This thesis is about methods suitable for studying biological interactions and their guidance in the exploration of chemical space. The focus has been on characterizing observable interaction dynamic and molecular recognition phenomena, from kinetic to structural points of view.

1.1 Molecular recognition

Complexation is the most basic chemical reaction that occurs in a living system. The formation of a complex, i.e. the non-covalent association of two or more molecular species, is a required event prior to any further chemical transformation. Depending on structural signatures and environmental conditions, molecular and atomic species pose an ability to mutually recognize each other. This ability is denoted as molecular recognition.

The simplest reaction scheme describing a molecular recognition event is the following:

$$A + B \xrightarrow{k_1} \xleftarrow{k_{-1}} AB$$

*Scheme 1.1*. Reversible interaction between two molecules.

The schematic expression of a generic reaction in Scheme 1.1 represents the reversible association event between molecules A and B. The reaction can be described both thermodynamically and kinetically, distinguished by unique numerical parameters, e.g. by a thermodynamic dissociation constant $K_d$ and kinetic rate constants.

At equilibrium, the ratio of concentrations between reagents and products is expressed as follows:

$$K_d = \frac{[A]_{eq}[B]_{eq}}{[AB]_{eq}}$$  \hspace{1cm} (1.1)
The thermodynamic parameter $K_d$, sometimes referred as the *affinity* of a molecule A for a molecule B (or *vice versa*), is linked to the changes in Gibbs free energy for the system, eq. 1.2.

$$\Delta G = RT \ln \frac{K_d}{\text{C}}$$

(1.2)

$$\Delta G = \Delta H - T\Delta S$$

(1.3)

In eq.s 1.2 and 1.3, $\Delta H$ and $\Delta S$ correspond to the changes in the enthalpy and entropy of the system; $T$ is the reaction temperature, $R$ is the universal gas constant and $\text{C}$ is a standard molar reference concentration.

The progress of the reaction, defined as the change in concentration of the complex AB over time, can be depicted employing the reaction rate constants and the law of mass action:

$$\frac{d[AB]}{dt} = k_1[A][B] - k_{-1}[AB]$$

(1.4)

Eq. 1.4 is a differential form of the second-order reversible reaction rate equation. The reaction rate constants, $k_1$ and $k_{-1}$, are sometimes denoted as $k_a$ and $k_d$, or (more correct, but less often) as $k_2$ and $k_1$, indicating the reaction molecularity.

There is a link between eq.s 1.1 and 1.4. Under equilibrium conditions, for the concentration of any species X:

$$\frac{d[X]}{dt} = 0$$

(1.5)

$$K_d = \frac{k_{-1}}{k_1}$$

(1.6)

The parameters in eq.s 1.1-1.4 are crucial for the definition of an interacting system. This is of great importance for rationalization of the recognition events, linking the structures of interactants to the contribution of various molecular properties to the strength of the interaction on static and dynamic levels.

### 1.2 Protein-ligand interactions

Proteins are one of three major classes of biopolymers. Proteins consist of polypeptide chains, folded into various structures that mediate the functions of proteins: catalysis, signalling and transport. All proteins are able to interact with various molecules, which in turn are either directly involved in their activity or able to modulate it. Protein interaction partners are diverse in their nature, and include ions, small organic molecules, polysaccharides, other proteins and nucleic acids.

The corner stone of a protein-ligand interaction is the protein-ligand interface. The interface consists of various functional groups in both proteins and ligands that create an array of non-covalent interactions of electrostatic nature, classified as ionic bonds, hydrogen bonds, $\pi$-stacking interactions and Van der Waals forces. These forces stabilize the complex in pursuance of a molecular complementarity, as the interactants
functional groups have to be close in space to form the favourable contacts. Additional interaction strength is built up through a hydrophobic effect by excluding water molecules from the interface topological layer through a displacement mechanism, reducing the solvent-accessible surface area and therefore increasing the system entropy. The complex might embody one protein-ligand interface, as it is common for a small molecule binding site of the proteins, or has many (so-called hot-spots) in case of protein-protein interfaces. Consequently, small molecules have to possess a significantly higher degree of molecular complementarity than macromolecular ligands to achieve similar affinities.

Molecular complementarity is not an intrinsic property of an isolated ligand or a protein. Often enough, to achieve more favourable contacts within the interface, biomolecules undergo a conformational change prior or upon the initial interaction phase. Often, the conformational change proceeds relatively rapidly and is not a rate-limiting step for recognition. In other terms, it has no significant impact on the observation of the reaction kinetics. However, in the case of a slow transition from the encounter complex to the ground state, the interaction model convolutes:

\[
A + B \xrightarrow{k_1/k_{-1}} AB' \xrightarrow{k_2/k_{-2}} AB
\]

**Scheme 1.2.** Two-state, or an induced fit model. Complex AB forms through an encounter complex AB'.

\[
\frac{d[AB']}{dt} = k_1[A][B] + k_{-2}[AB] - (k_{-1} + k_2)[AB']
\]

\[
\frac{d[AB]}{dt} = k_2[AB'] - k_{-2}[AB]
\]

In eqs. 1.7-1.8, \( k_2 \) and \( k_{-2} \) are first-order rate constants for a reversible conformational change of the encounter complex. The derivation of the thermodynamic parameters is similar to eq. 1.4:

\[
K_{d}^{AB} = \frac{k_{-1}k_{-2}}{k_1k_2}
\]

Thermodynamic parameter to link the state of equilibrium to both AB' and AB is defined as the following, where \( K_{d} = 1/K_{a} \):

\[
K_{a} = K_{d}^{AB'} + K_{d}^{AB}
\]

\[
K_{a} = \frac{k_1}{k_{-1}} \times (1 + \frac{k_2}{k_{-2}})
\]

Proteins often have multiple binding site. Enzymes are often multimeric, with their activity being under an allosteric regulation on a quaternary structural level, while most ion channels consist of multiple identical (or structurally and functionally similar) domains or subunits. Thus, models that account for multiple binding sites should be appropriately used. In the case of the absence of an allosteric effect between binding
sites, the underlying theory is not different from a standard 1:1 interaction. To depict a cooperative systems, different models were derived.

Historically, the first description of a cooperative binding was developed by Archibald Hill:

$$f = \frac{[L]^h}{K_d^h + [L]^h}$$  \hspace{1cm} (1.12)

Eq. 1.12, known as Hill’s equation, emerged as a purely empirical equation and is phenomenological; $f$ is the fractional occupancy and $h$ is the Hill coefficient - a parameter, whose value indicates the type and degree of a cooperativity. The cooperative effect in the Hill model is fixed and not dependent on the degree of saturation.

Adair-Klotz model breaks down the interaction by defining multiple sites of identical function into simplistic processes, assuming that the binding sites are chain-linked and saturation of one activates another. For $n$ sites, the dependency of fractional occupancy from the ligand concentration is expressed as follows ($K_{a,n}$ is the association constant for site $n$):

$$f = \frac{K_{a,1}[L] + 2K_{a,1}K_{a,2}[L]^2 + \ldots + nK_{a,1}...K_{a,n}[L]^n}{n(1 + K_{a,1}[L] + K_{a,1}K_{a,2}[L]^2 + K_{a,1}...K_{a,n}[L]^n)}$$  \hspace{1cm} (1.13)

Other models exists, including the elegant Monod-Wyman-Changeux (MWC) model which takes into account both multiple binding sites and multiple receptor conformations. Nonetheless, the work presented here has primary dealt with the Hill and Adair-Klotz models.
2. Early-stage drug discovery

Drug development is a multi-step process. Spendings for research and development (R&D) are extremely high, starting from ca. 700 million US Dollars (2017, [5]) for an approved molecule to a magnitude higher prices for the total R&D spendings per approved molecule. Generally, the overall paradigm from the formulating of a new molecule for healthcare use to its further marketing and therapeutical application can be split into four stages: discovery and development, pre-clinical studies, clinical trials and post-market monitoring (U.S. Food & Drug Administration, 2018).

The first stage, also known as early-stage drug design, deals with methodologies and approaches to guide all stages from target validation, selection of compounds with desired properties, identification of a clinical candidate for further development [6]. Disregarding the source of molecular entities, either of synthetic, endogenous or natural products origin, methods to address the effect of a compound on a target are of high concern.

Modern early-stage drug design is based on the concept “phenotype to target”. Although such a reductionistic approach is highly debatable [7, 8], a good understanding of the physical chemistry behind target-ligand interactions is absolutely required for the selection of tractable substances, and assisting their evolution into clinical candidates.

2.1 Enzymes and enzyme inhibitors

The vast majority of FDA-approved drugs and substances in clinical trials are small molecules [9]. Among them, up to 47% of the molecules of pharmaceutical interest by 2002 were enzyme inhibitors [10]. To date, despite the recent progress in biopharmaceutical development and the overall crisis of small molecules [8], enzyme inhibitors represent a huge pool of pharmaceutical substances. Development of expertise in inhibitor design is required for increased success in drug design campaigns.

Catalysis is the most important protein function. The generic scheme of enzyme catalysis, as it was formalized by Victor Henri, involves the formation of a transient enzyme-substrate complex. Assuming irreversibility of the conversion step, the general reaction scheme is written in the following way:

\[
E + S \xrightarrow{k_1 \frac{k}{k_{cat}}} ES \xrightarrow{k_{cat}} E + P
\]

Scheme 2.1. Catalysis scheme used in Michaelis-Menten kinetic model.

Following Scheme 2.1, the reaction rate, defined as \(d[P]/dt\), becomes relatively complicated. However, the rate equation can be simplified to the classic Michaelis-
Menten equation in Briggs-Haldane interpretation [11, 12]. Under certain assumptions, *e.g.* a steady-state concentration of the ES complex at the beginning of the reaction, and insignificant concentrations of both enzyme and product relative to the substrate:

\[ v_0 = \frac{k_{\text{cat}}[E][S]}{K_M + [S]} \]  

(2.1)

The integral form of eq. 2.1 is the Henri equation. The model parameter \( K_M \), denoted as the Michaelis-Menten constant, is a thermodynamical parameter that describes the dynamic equilibrium of the ES complex with the rest of the system:

\[ K_M = \frac{k_{-1} + k_{\text{cat}}}{k_2} \]  

(2.2)

Michaelis-Menten model is valid for a relatively large pool of catalytic systems, although more complex reaction schemes exist [13]. Following this model, four types of reversible enzyme inhibitors, *i.e.* chemical probes that suppress the catalytic activity of the enzyme through direct interaction with the protein in a non-covalent manner, can be classified. They differ in the enzyme form they interact with, as shown in Scheme 2.2.

\[
\begin{align*}
A & \quad E + S \quad \rightleftharpoons \quad ES \quad \rightarrow \quad E + P \\
& \quad \uparrow + I \\
& \quad \quad \quad EI \\
B & \quad E + S \quad \rightleftharpoons \quad ES \quad \rightarrow \quad E + P \\
& \quad \uparrow + I \\
& \quad \quad \quad ESI \\
C & \quad E + S \quad \rightleftharpoons \quad ES \quad \rightarrow \quad E + P \\
& \quad \uparrow + I \\
& \quad \quad \quad EI \\
& \quad \quad \quad \quad + S \rightarrow \quad ESI
\end{align*}
\]

*Scheme 2.2.* Three types of reversible inhibitors: **A** - competitive, **B** - uncompetitive, **C** - non-competitive. Mixed inhibition is a subtype of non-competitive inhibition, characterized by different affinity of the inhibitor to the free enzyme and to the enzyme-substrate complex.

Eq. 2.1 can be modified to account for inhibitor effect on catalysis. To exemplify, applying the steady-state approximation to all enzyme species in Scheme 2.2A and 
defining $K_i$ (inhibition constant) as a dissociation constant for complex EI, Michaelis-Menten equation can be expressed in the following way:

$$v_0 = \frac{k_{cat}[E]_0[S]}{K_M(1 + \frac{[I]}{K_i}) + [S]} \quad (2.3)$$

2.2 Biophysical methods in medicinal chemistry

Historically, in vitro early-stage drug discovery has strongly relied on selecting lead compounds on the basis of activity-based assays. Even nowadays, despite being underestimated, enzyme assays remain the prime method for a high-throughput screening. Unfortunately, catalytic activity-based functional read-outs sometimes require assembly of relatively complicated systems. In addition, functional assays can be misleading for certain targets and compounds. To exemplify, the common cause for a false-positive inhibition effect is mediated by the nature of the synthetic molecules - some of them result in protein denaturation, aggregation or interference with co-factors required for catalysis [14]. These set a limitation on assay design and the application, as it is difficult to work with low-potency compounds due the requirements of high concentrations. Additionally, analytical signals in assays designed for having high sensitivity are often generated by spectro-or fluorometrical means, and are easily affected by the optical properties of the tested molecules. A more problematic issue is the absence of reliable activity assays for certain targets or certain types of enzyme activity modulation.

Alternatively, interactions between compounds and target proteins can be assessed directly. Any numerical parameter defined by describing the chemistry of molecular recognition (eqs 1.1-1.4) can serve as a scoring criterion, useful for the ranking of ligands. To date, a relatively broad array of techniques to quantify physico-chemical parameters of protein-ligand interactions exist. Despite the questionable terminology, these techniques are generalized as biophysical methods.

A general overview of the current state-of-art methods can be found elsewhere [15]. At the first approximation, the methods used nowadays (neglecting computational methods and single-molecule techniques) are divided into three categories: equilibrium-based methods, kinetic and structural methods.

Equilibrium-based methods rely on the study of the target-ligand system at equilibrium, quantifying the fraction of the target in complex with the ligand at particular ligand concentration. Amongst others, Isothermal Titration Calorimetry (ITC) is one of the most reliable approaches to characterize molecular interaction, as besides the dissociation constant $K_d$ it can quantify thermodynamic signatures of the interaction, eq. 1.3 [16]. Unfortunately, the method is not robust enough, being sensitive to artefacts and poor experimental design. Another limitation is that it requires large amounts of matter for a single experiment. Still, calorimetry and NMR [17] are the only genuinely label-free techniques with a significant application share.

Amongst other tools useful for addressing equilibrium interaction parameters, it is worth to mention thermal methods. Fluorescent Thermal Shift Assays (FTSA), or thermal unfolding assays, are simple but powerful tools to assess the stability of protein-ligand complexes. The unfolding of protein upon, for example, increase in a tempera-
ture, can be treated as a pseudo-first-order phase transition (all-or-none) [18, 19], but still occurs through an "molten globule" intermediate [20, 21]. At this stage, before overcoming the entropic barrier to complete unfolding, the hydrophobic core of the partially folded protein is solvent-accessible and able to bind various lipophilic probes, including solvatochromic dyes [22]. Monitoring the dye fluorescence as a function of temperature allows the evaluation of the melting temperature for the protein, $T_m$. As a favourable interaction with a ligand stabilises the structure of the protein, its unfolding temperature is expected to be higher than for the ligand-free protein. Nowadays, approaches to translate changes in the melting temperature into interaction parameters, and to analyze protein melting curves, are well-developed [23]. FTSA is an acknowledged biophysical method often involved in the early-stage drug discovery.

Recent progress in device miniaturization has led to completely new methodologies, based on previously disregarded physical phenomena. Microscale Thermophoresis (MTS) is a relatively new technique with emerging applications. The foundation of the method is the Soret effect - an observation of the non-homogeneous distribution of particles (or molecules, ions and other molecular species) in a temperature gradient. The effect can be rationalized in terms of hydration entropy and local equilibrium which differs across a temperature gradient [24]. The change in thermodiffusion properties of the studied protein upon complex formation is used to report the interaction. In current instruments, capillaries are filled with a solution of the fluorescent target (employing either intrinsic UV fluorescence or a fluorescent label) and a ligand. An infra-red laser creates a thermal gradient in a small spot in the capillary. The change in fluorescence, that arises as a result of re-distribution of the target molecule or its complexes in the heated zone, is recorded [25, 26]. The resulting dose-response curve can be approximated with a quadratic binding equation, leading to quantification of the $K_D$ value.

A completely different approach to characterization of protein-ligand interactions is given by kinetic methods, which monitor the rates of interaction between two molecules of interest. There are multiple ways to set up kinetic experiments, with certain methods fulfilling the requirements of a lead discovery campaign. Most kinetic studies of protein-ligand interactions are performed using Surface Plasmon Resonance-based flow biosensors [15], extensively covered in the following section. Nevertheless, alternatives to biosensor-based methods to assess kinetic parameters of the interaction exist, including kinetic ITC (kinITC) [27], ligand-observed NMR [28], time-resolved Förster Resonance Energy Transfer spectroscopy (TR-FRET) [29] and various nonequilibrium separation-based techniques.

2.3 Biosensors
Flow biosensors have a convenient design for automated instruments, able to satisfy the practical requirements for a drug-discovery campaign. A generic biosensor consists of two units: a recognition element, that senses intermolecular interaction, and an analytical element that generates a response and monitors the interaction in a time-dependent manner. Integration of the biosensor with a flow system consisting of fluidic, autosampling and pump modules allows to execute kinetic experiments with a
panel of molecules in a medium-throughput manner and simplifies data analysis as a result of certain approximations within the rate equations.

Signal detection in flow biosensors relies primarily on optical methods (e.g. interferometry [30] and fluorescence [31]). However, non-spectroscopical detection units exist [32, 33]. Nevertheless, Surface Plasmon Resonance (SPR)-based sensors are the most wide-spread analytical systems [15, 34, 35].

### 2.3.1 SPR biosensors: detection principles

Total internal reflection means that the light beam is not transmitted through the interface and is reflected back into medium with a higher refractive index. Although the reflected beam does not lose its energy, an optical disturbances in media of lower refractivity occurs, in a form of a standing wave with exponentially decaying amplitude - a so-called evanescent field.

To some extent, free electrons in metals and atomic nuclei can be approximated as a neutral, plasma-like gas. Since the nuclei are immobile in a metal crystal lattice, only electrons can move freely in the metal bulk. For an infinitely thin metal film, the mean pathway of the movement lies in the film plane, as their perpendicular movements are rather restricted. Restricted oscillations can be induced via the application of an external field. Such oscillations, driven by the oscillation of an electric field of the evanescent wave, are called Surface Plasmon Resonances.

Upon the induction of plasmon oscillations, the incident light beam is dissipated. The dissipating conditions require the field vector component of the evanescent wave to match a similar component of the surface plasmons. The angle $\theta_r$, at which the plasmon oscillations are induced, is sensitive to the optical properties of the medium and can be a function of refractivity of the surface layer. In the Kreschmann SPR configuration, the plasmon oscillations are induced at the surface of a gold layer at the phase interface penetrating into the low-refractivity solution. The intensity of reflected multi-angle light can be recorded for a set of rays with angles $\theta_1...\theta_n$ to identify the value of $\theta_r$, at which the ray has been dissipated to induce the oscillations. In SPR flow biosensors, it is done in a time-resolved manner to record the dynamics of intermolecular interactions.

### 2.3.2 SPR biosensor-based interaction kinetic analysis

The first commercial SPR-based biosensor was introduced in 1991 [36], and the instrument principles have not changed significantly since, Figure 2.1. The analytical signal is generated through the Kretschmann SPR configuration, with a gold layer being a part of a dockable sensorchip. The recognition element is formed on top of the gold surface, usually through functionalization with linker molecules (e.g. carboxymethyl dextran hydrogel matrix) and a target molecule (usually, a polypeptide or a nucleic acid). A microfluidic system is connected to the pump and autosampling modules, and handles the interaction counterpart (denoted as "analyte" for GE Healthcare biosensors, or "ligand" in this thesis) to address the sample over the recognition element. Changes in the optical properties of the surface upon interactions are recorded and plotted as a sensorgram, or an interaction kinetic curve, Figure 2.2A.
Figure 2.1. Design for a SPR-based biosensor detection unit. Panel A represents the optical interface and the flow system: a - source of p-polarized monochromatic light, b - prism, c - sensing surface, e - microfluidic system, d - CCD array. The resonance excitation angle $\theta_r$ is observed as a dip in intensity of the reflected light, panel B.

Figure 2.2. An example of interaction kinetic curves and their analysis. The set of sensorgrams in panel A consists of three parts: pre-injection baseline (a), association phase b and dissociation phase c. Experimental curves (black) were approximated with eq.s 2.4 and 2.5 (red), supporting the mechanism shown on Scheme 1.1. Shaded area corresponds to signals at equilibrium - panel B shows their values, plotted against the ligand concentration and approximated with the Langmuir sorption isotherm (eq. 2.6).

The sensorgram consists of three parts: pre-injection baseline, association phase (injection of ligand) and dissociation phase (dissociation of complex in a flow), shown in Figure 2.2A. The flow setup allows eq. 1.4 to be significantly simplified, as the ligand is presented in a system at much higher concentration than the immobilized target and its depletion upon interaction with the surface is insignificant. Therefore, the rate equation is transformed from second-order to pseudo-first order, under the Vant’Hoff approximation. The integral form is derived as:

$$R = R'_{max} e^{-(k_1 |B| + k_{-1}) t}$$  \[(2.4)\]

In eq. 2.4, $R$ and $R'_{max}$ are the responses at time $t$ and the maximal response at given concentration of the ligand [B], respectively. They correspond to the concentration of complex [AB] in Scheme 1.1 and [AB]$_{eq}$, multiplied by a correction factor. The dissociation phase starts at the end of injection with the condition of $[B] = 0$, the kinetic curve during dissociation therefore reflects a simple exponential decay:

$$R = R' e^{-k_{-1} t}$$  \[(2.5)\]
Usually, kinetic experiment consists of injections of the ligand at different concentrations, resulting in a set of progress curves, Figure 2.2A. Global-curve analysis using non-linear regression with eq.s 2.4 and 2.5 allows the quantification of kinetic parameters, association-and dissociation rate constants \( k_{1} \) and \( k_{-1} \), and the thermodynamic dissociation constant \( K_d \) (eq. 1.1). Sometimes it is impossible to obtain kinetic data. In such cases, the dissociation constant can still be quantified, employing a steady-state analysis of sensorgrams (Figure 2.2B). As the rates of forward and reverse reactions are equal at equilibrium (eq. 1.5), the Langmuir sorbtion isotherm describing the dependency of the signal at steady-state equilibrium on a ligand concentration is the following:

\[
R = \frac{R_{\text{max}}[B]}{K_d + [B]}
\]  

(2.6)

\( R \) is the observed response at concentration of the ligand \([B]\) and \( R_{\text{max}} \) is the maximum theoretical response upon full occupation of the surface binding sites.

Utilization of the kinetic theory above is legitimate only for interactions that are described by Scheme 1.1. More complex models exists, like the two-state (linked reaction) interaction model as given in Scheme 1.2, models that account for surface heterogenity or for polyvalent nature of ligand molecules [37]. Their validity has to be justified either by re-design of kinetic experiment or by employing orthogonal methods [34, 38], as there are interaction-irrelevant artefacts that may also result in apparent deviations from Scheme 1.1 [39].

### 2.3.3 General remarks regarding interaction kinetics

Since a living organism never exists in an equilibrium state, it is unclear if it is meaningful to translate equilibrium measurements into a phenomenological level. Kinetic characterization of protein-ligand systems is valuable. Besides the generic information regarding the interaction strength and its stability, a well-designed kinetic experiment supports the development of a mechanistic model for the interaction, revealing the fine details of a molecular recognition besides just the energy gain during the transition into the ground state. To exemplify, kinetic experiments can expose long-lived intermediates in the formation of a complex, indicating conformational changes or other important molecular events that occur during association and dissociation (Scheme 1.2).

Interaction kinetic analysis is the correct way to address avidity effects of polyvalent ligands, like antibodies or dendrimeric oligomers.

The interaction kinetic parameters \( k_{1} \) and \( k_{-1} \) serve as efficient parameters for the characterization of ligands and ranking them in a useful order. This is important not only during initial lead selection and optimization, but also for the validation of a molecule before it enters further trials. There is strong evidence that dissociation rate constant \( k_{-1} \) correlates well with the therapeutic efficacy of a drug [40, 41] and its residual potency beyond the elimination phase. The half-life parameter \( \tau_{1/2} \), often denoted as the drug-target residence time, is proposed as a criterion for drug evaluation. It is simply defined from eq. 2.5:

\[
\frac{1}{2}[AB]_0 = [AB]_0 e^{-k_{-1}\tau_{1/2}}
\]  

(2.7)

21
\[ \tau_{1/2} = \frac{\ln(2)}{k_{-1}} \]  

(2.8)

It is still debatable if the off-rate constant \( k_{-1} \) and \( \tau_{1/2} \) are superior in comparison to other ranking quantities, as pharmacokinetics can easily inflate the profits of the low dissociation rate constant [42].

2.4 Protein crystallography

Structural methods are placed beside other biophysical tools, as they rarely give direct answers to questions regarding interaction strength and stability. However, while most biophysical methods deal with molecular recognition phenomena in terms of physical chemistry, structural biology sheds light on molecular complementarity. Evaluating QSAR requires not only interaction parameters, but also information that reveals the impact of different chemical modifications on the interaction, and to adopt a rational ligand evolution strategy.

The modern era in protein chemistry started after the determination of low-resolution structures of two proteins, myoglobin [43] and haemoglobin [44]. Since then, the field of structural biology has developed significantly. The current open database of biopolymer structures, the Protein Data Bank (PDB), was established in 1971 [45] and now contains more than 143000 of deposited entities (PDB, August 2018). Importantly, 90% of structures were obtained employing either X-Ray crystallography or hybrid methods.

2.4.1 Diffraction experiments, data acquisition and analysis

Crystals are solid bodies with a periodical packing of atoms. Crystalline material can be described by a crystal lattice, a geometrical abstraction that illustrates the arrangement of particles (ions, atoms or molecules). The periodic nature of crystal lattice allows the whole structure of a macrophasic object to be reduced in a simple way, as it consists of repeatable patterns – unit cells – the smallest group of crystal constituents, super-imposable using symmetry operators and translational vectors. Importantly, the strict definition of a crystal contains an ability to give a sharp diffraction pattern (International Union of Crystallography, IUCr).

Diffraction is defined as the spreading of a wave around obstacles. Covalent bond distances are small, and are within the range of 1.3-1.5 Å for a polypeptide backbone. Interatomic distances in a folded protein are also short. Therefore, in order to probe a tertiary structure and to treat atom electron clouds as diffractors, waves of high frequency and high energy are required, such as Röntgen, or X-ray radiation.

To observe X-ray scattering in the bulk of crystalline matter, the Wulf-Bragg’s conditions must be met (Figure 2.3A):

\[ 2dsin\theta = n\lambda \]  

(2.9)
In eq. 2.9, $\theta$ is the angle of incident electromagnetic radiation, $\lambda$ is its wavelength and $n$ is a positive integer. The parameter $d$ is the spacing between so-called Bragg’s, or diffracting, planes. Two waves diffracting from different planes either have the same or different phases, resulting in their interference. The interference is usually destructive and the resulting collision of photons cancel their intensities. However, during constructive interference their intensities are combined and amplified (Figure 2.3B), which results in a signal - a reflection of the photons on the detector. The set of reflections for all atoms within the crystal, linked through the symmetry operators, over different angles creates a three-dimensional diffraction pattern. It is an indirect representation of the electron density distribution inside the material.

Proteins are studied by single crystal diffraction. Thus, the nucleation of a well-diffracting monocrystal big enough to sustain radiation damage is the main challenge for the method. Structural experiments are performed as shown in Figure 2.4: the monocrystal (panel A) is irradiated with X-rays to obtain diffraction patterns at different angles (panel B), after indexing and assigning, the reflections are used to calculate the distribution of electronic density (panel C).

2.4.2 Structure-based drug discovery

Protein crystallography is a powerful tool for exploration of the finest atomic details of protein structure and function. An interest in the possible application of X-Ray crystal-
lography in the drug discovery pipeline rapidly arose, resulting in the pioneer studies of renin inhibitors [46]. The concept of structure-based drug discovery (SBDD) appeared as a strategy to rationalize the choice of a lead compound and its further modification. Dorzolamide, a sulphonamide-based carbonic anhydrase inhibitor (Figure 2.5), was the first SBDD-based drug to be introduced (Merck, 1995), and soon enough the methodology faced good success rates [47]. Nowadays, obtaining the structure of a protein-lead compound complex is an important part of early stage drug discovery, as it gives not only insights into the binding mode, but also provides a basis for chemoinformatical and rational ways to improve ligand properties [48].

![Figure 2.5](image)

*Figure 2.5. Stereoview of the structures of carbonic anhydrase II in complex with Dorzolamide (blue, PDB 1CIL) and its analogue (cyan, PDB 1CIM). The ethyl substitution at the primary amine causes the histidine residue of the protein to adopt an alternative conformation, releasing a bound water molecule (red) and decreasing the inhibition constant 5-fold [49].*

Although initial studies focused on the characterization of structures between targets and selected compounds, the method also has an important impact on screening. Modern crystallographic methodologies utilize robust methods for data collection and processing. Even low-energy sources, as in-house electron accelerators, can be employed to collect diffraction patterns for hundreds of high symmetry crystals within a week. Although the method is not suitable for high-throughput screening, it has an outstanding efficiency for a fragment-based lead discovery [50, 51].
3. Present investigation

Biophysical methods used for studies of proteins are often selected to answer two simple questions: if the protein of interest interacts with another molecule, and how strong the interaction is. Regularly, conclusions on protein biology, or on the potency of a certain ligand are derived from an incomplete descriptions of interactions, disregarding kinetic and structural factors of recognition.

In this thesis the spotlight is on the phenomenon of molecular recognition per se. A majority of the studies have a medicinal chemistry focus, illustrating the importance of molecular recognition for drug discovery. The studies represents different approaches to rationalize the chemistry of biomolecular interactions and its driving forces. The thesis covers several types of interactions, where proteins act as the binding partner for small molecules (papers I, II and V), ions (paper III), peptides (paper III and IV) and other proteins (paper III).

In the first study, the problem of multiple homologous off-targets was examined. The link between molecular complementarity and interaction kinetics was evaluated. Carbonic anhydrase isozymes were used as a model system. The follow-up study (paper II) highlighted the importance of using kinetic methods for interaction studies. It revealed the kinetics of intrinsic interactions through a chemodynamic approach and, for the first time, intrinsic rate constants for a drug-target interaction were quantified. The third study uncovered molecular dynamics in calcium-signalling pathways via a mechanistic analysis of complex events responsible for signal transduction. The last two studies aimed to assist the design of molecular probes towards epigenetic enzymes, an emerging target class in contemporary drug design. Kinetic and structural analysis uncovered new ligands with novel binding modes (paper IV) and has expanded the druggable space of these targets (paper IV, V).

3.1 Selectivity and interaction kinetics (Papers I&II)

Reversible hydration of carbon dioxide is a basic homeostasis reaction catalysed by enzymes in the carbonic anhydrase (CA) family. Human CAs are $\alpha$-carbonic anhydrases, metalloenzymes with a prosthetic group bearing a zinc cation with four ligands in its coordination sphere – three histidine residues and a substrate water. The tissue and cellular localization varies for the 15 different human $\alpha$-CAs known. They have high structural similarity, albeit few have no catalytic function.

Isozymes I and II are well-studied enzymes, often used as model proteins for enzymological and biophysical studies. Carbonic anhydrase inhibitors represent a well-developed drug class, predominantly used as diuretics or anti-glaucoma agents. Certain members of the enzyme family are considered to be suitable targets for cancer chemotherapy. Membrane-bound forms of CAs, CA IX and XII, have been shown to
be expressed in malignant cells (Figure 3.1B), where they reduce solid tumour acido-
sis.

CA inhibitors are designed to interact with the relatively large and deep active site
cavity of the protein. Sulphonamides are the main chemotype of CA inhibitors, with
multiple molecules confirmed to be useful in clinical practice. Besides general elec-
 trostatic interactions with the protein, sulphonamides form a coordination bond with
the prosthetic Zn$^{2+}$ group of the enzyme, with the amide of sulphonamide functional
group substituting the water molecule in the zinc coordination sphere, Figure 3.2A.

3.1.1 Interaction kinetic analysis of the sulphonamide inhibitors

In order to obtain insights into the interaction kinetics of carbonic anhydrases with
sulphonamides, a series of kinetic experiments between various CA isozymes and a
selection of benzenesulphonamide-based inhibitors was set up. The small molecular
ligands involved in this pioneering study were well-developed inhibitors, previously
characterized by means of activity- and equilibrium-based biophysical assays. The li-
brary consisted of 18 entries that differed little in their structural patterns: meta-or
para-substituted N-alkylated benzimidazole-and S-alkylated thiopyrimidine deriva-
tives, and fluorine substituted benzenesulphonamides, Figure 3.2B. For the given lig-
ands, the interaction kinetics were analysed using six different human CA isoforms:
full-length CA I, II, VII, XIII, and catalytic domains of CA IX and XII.

Kinetic analysis

Also this enzyme family is structurally very conserved, many amino acids within the
active site cavity of the catalytic domains differ. Still, design of isozyme-selective in-
hibitors is known to be difficult. In a first approximation, the sulphonamide functional
Figure 3.2. Sulphonamide-based inhibitors of carbonic anhydrases. Panel A - neutron diffraction-based model of a complex between human CA II and 5-acetamido-1,3,4-thiadiazole-2-sulphonamide (acetazolamide, AZM); 4G0C [53]. The shaded area represents a cross-section of the active site, grey sphere - tetra-coordinated zinc cation. Panel B - generalized structure of inhibitors, studied in Papers I&II. Top: benzimidazole-and thiopyrimidine derivatives (R₂), R₁ - H or Cl atoms. Bottom: substituted fluorobenzenesulphonamides.

group serves as a "warhead" with a pharmacophoric efficiency high enough to initiate recognition in all studied ligand-enzyme isoforms pairs. Despite this common feature, it was possible to observe differences in the interaction kinetics. All studied interactions were well described by the model shown in Scheme 1.1, and kinetic parameters were quantified for a structure-activity relationship (SAR) analysis.

Kinetic analyses revealed certain common features of all ligands, irrespective of their structure. Association rate constants, $k_1$, were in the range of $10^5$–$10^6$ M$^{-1}$ s$^{-1}$ for the majority of the tested compounds. This was surprising, considering the open topology of the active site and small size of the ligands (ca. 300 Da). In addition, CAs are well known to be one of the most efficient enzymes with the second-order association rate constant for Michaelis complex formation being $10^{10}$ M$^{-1}$ s$^{-1}$ for its inorganic substrate, which in turn approaches the diffusion limit for recognition [54, 55].

**Sulphonamide selectivity**

The potency of inhibitors correlated well with the complex dissociation rate. The interaction kinetic analysis confirmed that fluorinated benzenesulphonamides bearing bulky hydrophobic substituents were able to form extremely tight complexes with CAs, additionally highlighting the high stability of these interactions.

To simplify analysis of multiple kinetic data sets, varying both in targets and ligands, a chemical phylogenetic analysis of the interaction kinetics was designed. It is shown for the 12 selected interactions in Figure 3.3. This type of data analysis provides an overview of general QSAR correlations. For example, although it was evident enough that increasing the size of the hydrophobic substituted, e.g. from methyl to tert-butyl (compounds 2 and 3) considerably decreased the dissociation rate constants, it was not

---

1Compounds 1–3 correspond to compounds 13–15 in papers I&II.
as obvious that the nature of the substituent influenced the ligand discrimination between various isozymes. For example, cycloalkyl moiety (compound I) is preferable for ligands targeting CA IX and XII (Table 3.1), while aliphatic substituents shift the selectivity towards CA II and VII. Following this analysis, several ligands were ranked as semi-selective. Even when it was impossible to unambiguously quantify the interaction parameters by means of SPR spectroscopy, qualitative analysis of the acquired sensorgrams was useful to identify tight-binding ligands. Orthogonal measurements of the affinity employing TSA revealed their $K_d$ to be in a subnanomolar range.

Figure 3.3. An illustration of the concept of chemical phylogenetic analysis. Each set of individual kinetic parameters $k_1$ and $k_{-1}$ is color-bar-coded using an interaction kinetic plot. Employing proteins phylogeny in one dimension and ligand structural similarity in another, the interactions are clustered in a bi-directional hierarchical manner. Interactions (c), (f), (g) and (j) are rendered as essentially irreversible.

Table 3.1. Kinetic parameters and equilibrium dissociation constant for the interactions between compound I and studied human carbonic anhydrases.

<table>
<thead>
<tr>
<th>Isozyme</th>
<th>$k_1$, M$^{-1}$ s$^{-1}$</th>
<th>$k_{-1}$, s$^{-1}$</th>
<th>$K_d$, M</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA I</td>
<td>n/d</td>
<td>n/d</td>
<td>n/d</td>
</tr>
<tr>
<td>CA II</td>
<td>$8.6 \times 10^4$</td>
<td>$8.7 \times 10^{-3}$</td>
<td>$1.0 \times 10^{-7}$</td>
</tr>
<tr>
<td>CA VII</td>
<td>$4.9 \times 10^4$</td>
<td>$2.0 \times 10^{-3}$</td>
<td>$4.0 \times 10^{-8}$</td>
</tr>
<tr>
<td>CA IX$_{cd}$</td>
<td>s/d</td>
<td>s/d</td>
<td>s/d</td>
</tr>
<tr>
<td>CA XII$_{cd}$</td>
<td>s/d</td>
<td>s/d</td>
<td>s/d</td>
</tr>
<tr>
<td>CA XIII</td>
<td>$5.1 \times 10^5$</td>
<td>$7.9 \times 10^{-3}$</td>
<td>$1.5 \times 10^{-8}$</td>
</tr>
</tbody>
</table>

n/d - no data; s/d - dissociation too slow for quantification of parameters

Several structures of complexes between the studied inhibitors and human CAs were solved recently [56–58]. It is interesting to explore the correlation between ligand selectivity, interaction kinetics and structural features of the recognition. For the tightest complexes, the cyclooctylamine group was shown to interact with a hydrophobic wall in the active site – a well-known and conserved topological feature of carbonic anhydrases. Positions of the bound inhibitor are identical for both CA IX and CA XII, as shown in Figure 3.4. The amino acid residues forming the hydrophobic wall are also similar and consisted of aliphatic moieties. However, the corresponding region of the active site in other isozymes contains bulkier residues, as exemplified for CA XIII and CA II in Figure 3.4A and B, respectively, reducing the possible area for lipophilic contacts. Reduced dissociation rates and selectivity are mediated through such interactions and could be rationalised from a structural perspective. In conclusion, even for
highly conserved enzyme families, it is possible to evolve ligand selectivity targeting very small differences in the structures of active sites.

Figure 3.4. Impact of hydrophobic contacts on interaction kinetics between CAs and sulphonamides. Compound 1 was co-crystallized with CA I (5E2M), II (4PYY), IX (6FE1), XII (4Q0L) and XIII (5E2N) after paper I was published. Panel A - superposition of models for complexes between compound 1 and CA IX (green), XIII (orange). In the CA IX complex, the cyclooctyl moiety of the ligand (sticks) points towards a hydrophobic pocket in the active site, shielding a patch formed by aliphatic amino acid residues. The phenylalanine residue is located in the same topological part of CA XIII, preventing the docking of the bulky functional group. As a result, the inhibitor adopts an alternative binding position (wires). Panel B - compound 1 in the active site of CA II (purple). The ligand-free structure of CA II is shown as a reference (dark cyan, 3KS3). The bound ligand was refined in two different conformations, and is coloured according to the isotropic B-factor, blue-to-red gradient. Despite the low-temperature conformation being similar to the one found for the high-affinity interaction shown in panel A, a conservative phenylalanine residue adopts unfavourable rotameric configuration and diminishes the recognition. The binding pose of the ligand in a complex with CA XII is given as a wire model. CA XII has a phenylalanine-to-alanine substitution, rendering the site interaction-competent.
3.1.2 Intrinsic interaction kinetics

In the initial study of CA isozyme-sulphonamide interaction kinetics (paper I), the interaction pH-dependency was also evaluated. The reasoning behind was partly relating to the physiological role of CAs. CA IX and XII are membrane-bound proteins, with their catalytic domain exposed to the cellular matrix. Their basic function is to counter-act tumour acidosis - a consequence of the high metabolic demand of cancer cells, as under hypoxic conditions glycolytic pathway is balanced to lactate. The experiments showed that sulphonamide-carboxylic anhydrase interactions had a clear pH-dependency that was hypothesized to be related to the interaction mechanism. The concept of "intrinsic kinetic" was introduced and developed in a follow-up study.

Interaction model and interaction-linked events

Since the finest atomic details of sulphonamide-CA interactions (Figure 3.2A) has been revealed using neutron diffraction [53], it became evident that the ligand is bound in a deprotonated negatively-charged state. The sulphonamide group is ionizable, with a \( pK_a \) of ca. 10 for a benzenesulphonamide. It is believed that in its deprotonated form it substitutes water from the CA prosthetic group. An interaction scheme can be described using the three following chemical equations:

\[
\begin{align*}
RSO_2\text{NH}_2 & \rightleftharpoons K_{a,SA} RSO_2\text{NH}^- + H^+ \\
[(\text{His})_3\text{H}_2\text{O}]\text{Zn}^{2+} & \rightleftharpoons K_{a,CA} [(\text{His})_3\text{OH}^-]\text{Zn}^{2+} + H^+ \\
[(\text{His})_3\text{H}_2\text{O}]\text{Zn}^{2+} + RSO_2\text{NH}^- & \xrightarrow{k_{1,\text{int}} \quad k_{-1,\text{int}}} [(\text{His})_3\text{RSO}_2\text{NH}^-]\text{Zn}^{2+} + \text{H}_2\text{O}
\end{align*}
\]

Scheme 3.1. Sulphonamide interaction with carbonic anhydrase. Only the deprotonated, negatively charged ligand is competent to associate with the protonated, positively charged protein.

Considering the equilibria in Scheme 3.1, the general reaction rate equation 1.4 can be re-written in a form that keeps the association and dissociation rate constants as pH-independent parameters:

\[
\frac{d[\text{CA} - \text{SA}]}{dt} = k_{1,\text{int}} f_{CA} f_{SA} [\text{CA}][\text{SA}] - k_{-1,\text{int}} [\text{CA} - \text{SA}] \quad (3.1)
\]

Rate constants \( k_{1,\text{int}} \) and \( k_{-1,\text{int}} \) in eq. 3.1 are denoted as intrinsic rate constants, and coefficients \( f \) correspond to the fractions of interaction-competent species, in this case of protonated enzyme and deprotonated sulphonamide ligand. Protonated-or deprotonated fractions can easily be quantified knowing the system pH and \( pK_a \) values of ionizable groups of the interaction partners.

Experimental study of the intrinsic interaction kinetics

In order to confirm the validity of the model, a series of kinetic experiments with selected ligands and three isozymes over pH range 5-9, with a pH increment of 0.5,
was set up, Figure 3.5. After the interaction parameters were quantified, a simple graphic analysis highlighted their agreement with the model described in eq. 3.1.

The observed dependency of the kinetic and thermodynamic association constants was U-shaped, Figure 3.5B. This is consistent with the opposite influence of pH on a concentration of the reactive species. The critical points were found at the near-neutral pH, and the values diminished significantly towards alkaline or acidic conditions. As the observable rate constant $k_1$ can be re-defined from eq. 3.1 as a function of pH, intrinsic interaction parameters could be quantified, Table 3.2.

**Table 3.2.** Selected observed values and quantified intrinsic kinetic parameters for CA II-AZM interaction.

<table>
<thead>
<tr>
<th></th>
<th>pH 5.5</th>
<th>pH 7.5</th>
<th>pH 9.0</th>
<th>Intrinsic$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$, M$^{-1}$ s$^{-1}$</td>
<td>$3.0 \times 10^3$</td>
<td>$2.7 \times 10^6$</td>
<td>$2.0 \times 10^5$</td>
<td>$2.2 \times 10^7$</td>
</tr>
<tr>
<td>$k_{-1}$, s$^{-1}$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>$K_d$, M</td>
<td>$1.7 \times 10^{-7}$</td>
<td>$1.7 \times 10^{-8}$</td>
<td>$2.9 \times 10^{-7}$</td>
<td>$2.7 \times 10^{-9}$</td>
</tr>
</tbody>
</table>

$pK_{a,CA II} = 7.1, pK_{a,AZM} = 7.3$.

$^3$the intrinsic $K_d$ value was estimated as $5.7 \times 10^{-9}$ M by ITC.
The chemodynamic analysis revealed certain important features. Sulphonamides should not be considered as “intrinsic-slow” CA ligands, as their intrinsic association rates are high. The observed moderate association rate constants can be attributed to acid-base equilibrium in the system, but not to an inefficient rate-limiting recognition event. More curious was an absence of a clear pH-dependency for the dissociation rate constant \( k_{-1} \). Such behaviour was expected from the model given in Scheme 3.1, but our studies did not resolve the exact dissociation mechanism. It is unlikely that the ligand dissociates from the protein in a first-order manner, as zinc complexes with coordination number 3 are extremely under coordinated. For the second-order substitution mechanism it remains unclear if the synthetic ligand is displaced by a water molecule or a hydroxide (through, for example, trigonal bi-pyramidal transition state), or if the amide of the ligand undergoes protonation prior dissociation. However, it is clear enough that protonation events do not limit the dissociation kinetics. It is likely that dissociation is initiated with a conformational fluctuation in the bound ligand to start ligand displacement.

**Intrinsic kinetics in the context of drug design**

A low degree of pH influence on \( k_{-1} \) increases the importance of dissociation kinetics for the ranking of sulphonamide-based CA inhibitors. Additionally, a logical approach for increasing the interaction strength of a sulphonamide is to decrease the \( pK_a \) of the amide, for example via introduction of substituents with a strong inductive effect. This is particularly important when the ligand is evolved to probe the target at non-standard physiological conditions, e.g. in acidified extracellular liquid.

The concept of intrinsic kinetics is relevant to QSAR evaluation. To exemplify, the methylsulphonamide is a weak CA inhibitor with a micromolar inhibition potency [60], while for its halogenated analogue, trifluoromethylsulphonamide, the \( K_d \) for its interaction with CA II is 130 nM [61]. An obvious question is therefore if an overall 1000-fold gain in potency is achieved due to bond polarization and improved Van der Waal’s contacts of the trifluoromethyl group in comparison to the methyl, or if it is an apparent gain mediated through a lower interaction-competent concentration of the unhalogenated ligand (\( pK_a \) values are 17.5 and 9.7 in DMSO, respectively). Both factors are involved most likely, but as, for example, a crystal structure typically is a snap-shot of an interaction in its ground state, intrinsic interaction parameters should be used to draw QSAR conclusions when high precision is needed.

### 3.2 Recognition in calcium signalling (Paper III)

Signal transmission in biological systems is mediated exclusively by molecular mechanisms. Signal transduction cascades always include multiple recognition events between messengers and their targets, often multiplying the signal strength or diversifying it to attenuate various components of the cellular biochemical machinery.

Calcium ions play a significant role in signal transduction pathways. Calmodulin is a calcium-binding protein whose conformation is influenced by bivalent metal cations. It exists primarily in two forms, apoCaM and calcium-bound Ca\(^2+/CaM\), differing significantly in their conformations. The calmodulin interactome includes a variety of

\(^4IC_{50}\) value towards *Canis. l. familiaris* crude erythrocyte extract is 300 \( \mu M \).
proteins and is fine-tuned by the degree of calcium saturation. Kinetic studies on calmodulin interactions are important, as the dynamics of the interaction translates into a mechanism of response in CaM-mediated signalling pathways.

**Calcium-calmodulin interaction**

An evaluation of the interaction between calmodulin and $\text{Ca}^{2+}$ ions was performed to show that the protein retains its calcium sensing activity in the SPR-based biosensor setup. The developed SPR assay was sensitive enough to reveal interactions with metal ions, despite the low refractive index increment resulting from micromolar concentrations. The interaction was rapid for both association and dissociation. The macroscopic affinity value, quantified using eq. 2.6, was determined to 2.8 $\mu$M. Experiments with two orthogonal fluorescent in-solution methods (MST and UV fluorescent spectroscopy) showed an excellent agreement between the three different assays.

Calmodulin recognizes up to four $\text{Ca}^{2+}$ cations, with a positive cooperativity between the calcium binding sites. However, there was no strong indication of a cooperative interaction on the biosensor-acquired data, unlike in the data obtained by intrinsic fluorescence spectroscopy. To understand this difference in data output, a closer look was taken on the methods employed.

The C-lobe of calmodulin contains two high-affinity calcium sites. After the conformational change, the only tyrosine residue, located at the C-terminus of the protein (Y138), is exposed to the solvent. In the apo-conformation, the intrinsic fluorescence of the residue is quenched and its emission is employed as a reporter of the calcium saturation of the domain; intrinsic fluorescence can not reflect interactions localized to the N-lobe. To investigate the difference in observed cooperativity between SPR-and fluorescence-acquired data, a simple simulation of dose-response curves using Adair-Klotz cooperativity model (eq. 1.13) was performed. Two experimental setups were considered: two sites at the C-domain (intrinsic fluorescence) and all four sites of the protein (SPR biosensors and MST), Figure 3.6. The values of macroscopic Klotz constants were obtained from experimental measurements, performed by Crouch and Klee [62]. As for the experimental observations, the four-site model was in accordance with eq. 2.6, where phenomenological macroscopic $K_a$ represents a single variable parameter.

---

5 Data was collected from bovine calmodulin. In paper III, both bovine and human calmodulin were used as the sequence identity is 100 ‰.
**Calmodulin interactions with polypeptides**

The prime goal of the present study was to show how calcium saturation influences the dynamics of the interaction between CaM and neurogranin (Ng), a brain-specific protein involved in the regulation of free calmodulin. It is generally believed that neurogranin plays a role of a “calmodulin depo”, increasing the local post-synaptic concentration of CaM, possibly through slowing down the CaM diffusion and promoting its localization proximal to the cytoplasmic membrane [63].

As a control, the interaction kinetics between CaM and a calmodulin-binding peptide derived from the sequence of calmodulin-dependent kinase II (CaMKII) was studied. As expected, the peptide behaved as a tight binder of Ca$^{2+}$/CaM, but did not interact with calmodulin in its apo-form. The association kinetic was extremely rapid and, in the biosensor set up, the rate-limiting step for the association was ligand diffusion.

Initially, neurogranin experiments were also performed using a peptide. Since the Ng CaM binding site is formed by a relatively conservative motif of apoCaM-binding proteins, a so-called IQ motif, peptides consisted of this motif alone have been considered to be a valid model for full-length neurogranin. However, the kinetic experiments did not confirm the hypothesis. The IQ motif-derived Ng peptide (Ng$_{27-50}$) showed no detectable interaction with apoCaM. Moreover, in contradiction with the functional signature, the peptide demonstrated a moderate affinity for the calcium-saturated protein, Table 3.3.

**Table 3.3. Interaction parameters between calmodulin and neurogranin forms. Interactions with Ca$^{2+}$/CaM were too rapid for kinetic analysis.**

<table>
<thead>
<tr>
<th></th>
<th>Ca$^{2+}$/CaM</th>
<th>apoCaM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_d$, M</td>
<td>$k_1$, M$^{-1}$ s$^{-1}$</td>
</tr>
<tr>
<td>Ng$_{27-50}$</td>
<td>$1.4 \times 10^{-5}$</td>
<td>ni</td>
</tr>
<tr>
<td>Ng</td>
<td>$1.9 \times 10^{-5}$</td>
<td>$3.4 \times 10^{5}$</td>
</tr>
</tbody>
</table>

ni - no interaction detected
As a next step, full-length neurogranin was used for kinetic experiments. The interaction with apoCaM had a $K_d$ value in the high nanomolar range, with resolvable kinetics following a simple 1:1 interaction model (Scheme 1.1), as shown in Table 3.3. Ng was still found to interact with calcium-saturated CaM, with a $K_d$ value similar to the IQ-motif containing peptide. It confirmed that the IQ motif alone is not able to efficiently recognize apoCaM.

**General conclusions on CaMBP interaction kinetics**

Fast, diffusion-limited association kinetics between $\text{Ca}^{2+}$/CaM and CaMKII is highly desirable considering the role of this interaction. Calmodulin activates CaMKII, promoting its autophosphorylation. CaMKII remains active after complex dissociation, and the rapid dissociation rate of calmodulin (in this thesis, qualitatively evaluated $k_{-1}$ was found to be around $1 \text{s}^{-1}$) allows multiple re-association events to amplify the following enzymatic cascade. More over, an instant dissolution of a strong, low nanomolar interaction in absence of $\text{Ca}^{2+}$, as shown via injections of chelators over $\text{Ca}^{2+}$/CaM-CaMKII-functionalized surface, is absolutely justified as a strategy to instantly inhibit the entry node of the signalling pathway in the absence of a secondary messenger.

Physiological rationalization of CaM-Ng interactions is more complicated. Ng localization studies showed its non-uniform abundance close to the plasmatic membrane. Likely, the protein promotes apoCaM accumulation, increasing the postsynaptic $\text{Ca}^{2+}$ sensitivity [63]. Calcium-mediated dissociation of $\text{Ca}^{2+}$/CaM, as the high-affinity CaM-Ng complex becomes a low-affinity one, initiates further signal transduction through a modulation of other CaMPB. The residual CaM-binding activity of Ng in the presence of calcium ions can be additionally diminished by intracellular kinases through the phosphorylation of the IQ motif [64], rendering the protein unable to interact with both calmodulin forms. In the short term, this event should increase the intensity of signal transmission even more. Possibly, such two-step mechanism allows the post-synaptic cell to regulate its $\text{Ca}^{2+}$ sensitivity in a selective and tunable manner.

### 3.3 Druggability of epigenetic enzymes (Papers IV & V)

Post-translational modifications (PTM) of proteins extend the chemistry of standard proteinogenic amino acids. Cells utilize the enzymatic machinery to regulate protein function through covalent modifications of amino acid side chains or the polypeptide backbone. The expression of genes is not only regulated by transcriptional factors, but also by the epigenetic state of chromatin: either silenced (heterochromatin) or transcription-competent (euchromatin). One of the factors that define the chromatin state is PTMs on nucleosomal proteins, and there are ample enzymes that alter these modifications.

Epigenetic regulation mediates multiple processes, and is primary responsible for the phenotype of the cell. Malfunction in the epigenetic mechanism often leads to various pathological states, with malignancy being among them. Thus, xenobiotics that can interfere with epigenetic regulation in a selective manner are valuable tools to uncover mechanisms of disease, and paths to novel therapeutical strategies [65, 66].

35
3.3.1 Macro cyclic peptides target the active site of LSD1

Lysine-specific demethylase 1 (LSD1) is a FAD-dependent amine oxidase, which acts on a mono-and dimethylated lysine residue in position 4 of histone H3, predominantly in promoter regions of the genome. The protein controls expression of certain oncogenes and is overexpressed in various cancer types and developmental disorders. LSD1 consists of three defined parts: a monoamine oxidase-like domain (MAO-like) and two reg u latory domains, namely a SWIRM and a tower-domain [67]. While the MAO domain carries a catalytic function, the SWIRM-and tower-domains are responsible for the assembly of macromolecular complexes and more efficient recognition of the substrate nucleosomes.

A majority of developed LSD1 inhibitors are small molecules. Mechanism-based MAO inhibitors, e.g. cyclopropylamine, hydrazine or prop-2-ynamine-based organic molecules, can be evolved to inhibit the demethylase with high selectivity [65]. However, a non-mechanism-based approach is preferable as it reduces the toxicity of a potential pharmaceutical. Molecular scaffolds for evolution into reversible LSD1 inhibitors exist, and tool compounds of high potency have recently been reported [68–70]. An alternative strategy to develop competitive LSD1 inhibitors involves the utilization of its protein binding features, and a variety of linear peptides have been evaluated as active site probes [71, 72]. Here, a structure-based rational approach for the design of peptidomimetic-based LSD1 inhibitors was initiated.

Rational design of active site probes

The macrocyclic scaffolds to be examined in this study were rationally designed on the basis of deposited LSD1-peptide complex structures. The peptides adopt a distinct conformation in their LSD1-bound state. The design strategy evolved around the cyclization of the substrate recognition motif. N-terminal residues of histone H3 were employed as a starting scaffold, and molecular dynamic simulations predicted potential cross-linking nodes to the residues (A1,T3) and (T3,T6).

Active site probes

Interaction kinetic experiments were set up with a series of linear histone H3-based peptides to explore the pharmacology of linear moieties and to choose a minimal sequence for further modifications. Molecular probes were selected to explore the role of N- and C-terminal residues, and to probe individual contacts in the LSD1 active site. SAR observations for the selected linear peptides are shown in Table 3.4.

The affinity diminished significantly with every amino acid removed, and the N-terminal amino acids had the most significant impact on the interaction. The peptide representing the first 11 histone H3 residues with a K4M substitution was chosen for further evolution. Interaction assays revealed critical roles for the arginine residues in positions 2 and 8.

Macrocyclized peptides

In the first stage, intramolecular cross-link was introduced through ring-closing olefin metathesis, probing positions 1-3 and 3-6 for the stapling, varying the length of the alkene bridge and a stereoconfiguration of the α-carbon. Unfortunately, the biophysical analysis of new ligands was severely restricted by their hydrophobic nature, especially for the precursor allyl-and pentinyl glycine bis-substituted linear molecules.
Table 3.4. Selected parameters for the interactions between LSD1 and linear peptides. Overall, 20 different constructs including peptides with non-proteinogenic amino acid substitutions were tested.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length, aar</th>
<th>$K_d$, M</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARTMQTARKSTGGKAPRKQLA</td>
<td>21</td>
<td>$1.4 \times 10^{-7}$</td>
</tr>
<tr>
<td>RTMQTARKSTGGKAPRKQLA</td>
<td>20</td>
<td>$4.5 \times 10^{-5}$</td>
</tr>
<tr>
<td>ARTMQTARKSTGG</td>
<td>13</td>
<td>$3.6 \times 10^{-5}$</td>
</tr>
<tr>
<td>ARTMQTARKST</td>
<td>11</td>
<td>$1.4 \times 10^{-4}$</td>
</tr>
<tr>
<td>ARTKQTARKST</td>
<td>11</td>
<td>$&gt; 2 \times 10^{-4}$</td>
</tr>
<tr>
<td>ARTZQTARKST</td>
<td>11</td>
<td>ni</td>
</tr>
<tr>
<td>AXTMQTARKSM</td>
<td>11</td>
<td>ni</td>
</tr>
</tbody>
</table>

Z - L-norleucine, X - L-citrulline, ni - no considerable interaction detected.

However, macrocyclic ligands consisting of 8 carbon atom bridges between residues 3 and 6 were found to have acceptable interaction profiles, with the lowest degree of unspecific interaction.

To attenuate the problem, the macrocyclic design was revised. The poor interaction properties of the olefin-bridged cycles were attributed to their unsaturated hydrophobic motif. In order to decrease the lipophilicity of the link, an additional series of macrocycles was prepared employing lactamization or copper-assisted 1,3-dipolar cycloaddition. Bridging amino acids (E and K for the amide bridge, propargylglycine and azidolysine for the triazole) were introduced at positions 3 and 6. The activity of novel ligands was assessed by means of an *in vitro* demethylation assay.

The new strategy resulted in ligands with a clear QSAR, with compound potency being strongly dependent on link chemistry. Moreover, the stereoconfiguration of the node amino acid residues shifted ligand potency in a clear manner, indicating a specific recognition. Mechanistic studies showed that the developed macrocycles are competitive inhibitors of LSD1, with lactams found to be the best inhibitors with a significant gain in potency in comparison to the linear parental peptide 2, Table 3.5.

Despite the significant improvement in the inhibition potencies achieved through cyclization, certain SAR features remained unclear. For example, the recognition of parental peptide 2 was highly dependent on arginine residues in positions 2 (Table 3.4) and 8. As shown in Table 3.5, lactam macrocycles were not sensitive to the equivalent substitution (peptides 7, 8), and a methionine-to homomethionine substitution at position 4 of the peptide 4 did not affect its inhibition either. Despite the competitive mode of inhibition of the macrocycles against the mono-methylated histone H3-based substrate, there was an inconsistency in their recognition in comparison to the linear precursor, which indicated a different binding mode.

**Crystallization studies**

Initially, crystallographic experiments employed LSD1 alone. The presence of the active site probes in the crystallization solution severely restricted the nucleation of crystals and their growth. The solved structures did not contain active-site bound ligands. Examination of the protein-protein contacts in the crystal lattice revealed that the entrance to the active site of LSD1 was shielded by the tower-domain of a symmetry-
Table 3.5. Inhibition properties of the lactam macrocycles. Two linear K4M peptides, 21 aar 1 and the parental peptide 2, are shown as references. Inhibition constants for peptides 1, 4, 5, 6, 7 and 8 were quantified using a global curve analysis and eq. 2.3.

<table>
<thead>
<tr>
<th>Code</th>
<th>Sequence</th>
<th>Linkage</th>
<th>$K_i$, M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ARTMQTARTKSTGGKAPRKLQA</td>
<td>linear</td>
<td>$6 \times 10^{-7}$</td>
</tr>
<tr>
<td>2</td>
<td>ARTMQTARKST</td>
<td>linear</td>
<td>$&gt;1 \times 10^{-4}$</td>
</tr>
<tr>
<td>3</td>
<td>ARKMQEARKST</td>
<td>amide</td>
<td>$3.1 \times 10^{-5}$</td>
</tr>
<tr>
<td>4</td>
<td>AREMQKARKST</td>
<td>amide</td>
<td>$4 \times 10^{-6}$</td>
</tr>
<tr>
<td>5</td>
<td>AR<em>KMQ</em>EARKST</td>
<td>amide</td>
<td>$2 \times 10^{-6}$</td>
</tr>
<tr>
<td>6</td>
<td>AR<em>EMQ</em>KARKST</td>
<td>amide</td>
<td>$&gt;1 \times 10^{-4}$</td>
</tr>
<tr>
<td>7</td>
<td>AXEMQKARKST</td>
<td>amide</td>
<td>$4 \times 10^{-6}$</td>
</tr>
<tr>
<td>8</td>
<td>AREMQKAXKST</td>
<td>amide</td>
<td>$8 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

Side chains of the macrocyclic bridge are underlined; *K and *E are D-lysine and D-glutamate, X is L-citrulline. All cyclized peptides are amidated.

related protein chain. Attempts to find conditions resulting in an alternate packing were unsuccessful. Further work was conducted on co-crystallization of LSD1 with its natural interaction partner, nuclear co-repressor REST1 (CoREST1). Crystals were used for soaking experiments.

LSD1-CoREST1 crystal structures with the K4M-substituted 13 amino acid residues linear histone H3 peptide and the most potent macrocycle 5 were solved at low resolution, Figure 3.7. Poor diffraction properties of the LSD1-CoREST1 crystals could be attributed to a high solvent content in the asymmetric unit, approximated to 84%.

![Figure 3.7](image)

Figure 3.7. Cyclization of histone H3-derived peptides changes their binding pose. Panel A - the overall structure of LSD1-CoREST1-macrocycle tertiary complex; LSD1 (grey) and CoREST1 (green) are shown as ribbon models. The macrocyclic moiety was identified in the outer rim of the substrate recognition site, as shown on panel B; black isomesh - OMIT ($F_o - F_c$) difference map for the macrocycle 5, countered at $3\sigma$. Residues 8-11 were not modelled. Panel C shows an overlay of the histone H3 linear peptide and macrocycle 5; side chains are hidden for clarity.

Electron density maps with large circular peaks were observed for crystals, soaked with 5 (Figure 3.7B). The ligand was found in the outer rim of the active site cleft,
distant from the prosthetic FAD. Unfortunately, no significant density was observed for four C-terminal residues, the missing atoms were thus not included in the model. The bound conformation of 13 aar linear peptide was similar to a reported previously for 21 aar K4M histone H3 peptide [71]. The cyclization changed the binding pose of the macrocycles in comparison to its pre-cursor, altering the peptide orientation over partially the same interface.

Although the data quality did not reveal the fine atomic details of the interaction, certain conclusions can be drawn. The amide of the intramolecular cross-link is probably involved in recognition, and its polar contacts have an influence on LSD1 selectivity. The N-terminal amino group was found close to the negatively charged D375/E379 residues. Most important, no polar contacts between the cyclic peptide’s arginine residues were found, rationalizing the observed similar potencies of citrulline-substituted macrocycles (Table 3.5, peptides 7 and 8). Therefore, the designed ligands can potentially be evolved to more drug-like, less polar scaffolds, and potentially to cell-permeable peptidomimetics.

3.3.2 Expanding the druggable space of SMYD3

SET-and MYND domain-containing protein 3 (SMYD3, Figure 3.8) is a lysine methyl transferase (KMT) that was discovered in 2004 as a cancer-associated protein [73]. Elevated expression levels of SMYD3 were identified in colorectal and hepatocellular carcinomas. Some studies showed evidence for a role of SMYD3 in the proliferation of malignant cells [74].

![Figure 3.8. Cross-section of human methyltransferase SMYD3. The methylation tunnel, able to accommodate the lysine residue of the substrate, links the protein substrate and S-adenosyl methionine interaction sites.](image)

Although initially identified as an S-adenosyl methionine (SAM)-dependent histone KMT, SMYD3 also shows good enzymatic activity towards various cytosolic proteins. The most interesting non-histone methylation substrate is a mitogen-activated protein kinase kinase kinase 3 (MAP3K2), an entry node in Ras cancerogenesis-related MAPK signalling pathway. While the nucleosomal activity is replicated in vitro with irregular
consistency, enzymological and structural studies show that MAP3K2 is the substrate for which SMYD3 has the highest specificity [75].

SMYD3 mediates multiple intracellular interactions, with no well-developed theory regarding the roles of SMYD3 in chromatin regulation or other processes it is involved in. Experiments have indicated that, depending on the cellular compartmentalization, the enzyme either forms a complex with heat shock protein 90 (HSP90) or recognizes certain DNA sequences. Additionally, the protein has been shown to be a transcriptional regulator, by being a part of the RNA polymerase II initiation complex.

**Kinetic analysis of SMYD3 interactions**

A biosensor-based SMYD3 interaction assay was designed and employed to characterize its interactions with natural ligands, *i.e.* with the methylation co-factor S-adenosyl methionine and its product, S-adenosyl homocysteine (SAH). The presence of the trialkylthiol moiety in the co-factor translated into a stable high affinity complex: the $K_d$ value for the SMYD3-SAM interaction was found to be 1.7 nM. SMYD3-SAH complex had relatively rapid kinetics and a more than 100-fold lower affinity.

KTM s, including SMYD3, are known as relatively inefficient enzymes with low $k_{cat}$ values. It was therefore interesting that the interaction kinetic analysis revealed that co-factor binding or turnover are not the rate-limiting steps for the methylation reaction.

**Screening assay**

The current project was initiated as an attempt to develop a biophysical platform facilitating SMYD3 ligand design. To date, only two well-developed SMYD3 inhibitors have been disclosed [76, 77], and further research on this protein would benefit discovery of novel, well-characterized ligands.

The developed interaction kinetic assay was modified for screening, allowing multiple referencing and a differential screening set up. The assay was designed to focus on ligands that interact with the substrate binding site of the protein, Figure 3.9A. To achieve it, two analytical surfaces were employed, one having the protein with unmodified active site and the other with active site-blocked protein, employing EPZ031686, a tight-binding inhibitor with a well-described mechanism of action [77]. A set of compounds was screened at a single concentration, and one hit with a good stoichiometric ratio and acceptable interaction profile was identified. This ligand is further denoted as EH05-2$^\text{6}$. The hit compound interacted with both SMYD3 and the SMYD3-EPZ031686 complex. Further experiments involved a study of its interaction in presence of SAM in the running buffer, keeping the protein co-factor site saturated. Competition experiments indicated that the binding of EH05-2 occurs outside the active site.

**Crystallographic evaluation reveals a new binding site**

A good saturation profile and a 1:1 stoichiometry of the discovered ligand, as shown in Figure 3.9B, indicated the presence of a well-defined interaction site. In order to localize it in the structure, crystallization trials were set up. The protein was successfully co-crystallized with EH05-2, and the structure of the complex was solved.

$^\text{6}$EH05-2 corresponds to ligand 2 in Paper V
Figure 3.9. Differential screening strategy utilizes four serial sensing surfaces, panel A: a - denatured SMYD3, b - SMYD3, c - SMYD3-EPZ031686 complex, d - intact matrix. Panel B - sensorgrams for the hit compound EH05-2-SMYD3 interaction and a steady-state analysis of the corresponding data; red line corresponds to a data approximation with eq. 2.6.

The ligand was found bound in a site opposite the methylation site. EH05-2 occupied two hydrophobic cavities on the surface of the C-terminal domain (Figure 3.10) and formed two polar contacts with the side chains of amino acid residues of the protein. Independently, a group of collaborators assisted the project with a computational solvent mapping of SMYD3 surface. Importantly, the interaction hot spots, highlighted through in silico experiments, matched the localization of EH05-2.

Figure 3.10. Stereoimage of the discovered ligand binding site. Black isomesh corresponds to a difference OMIT ($F_0 - F_e$) map of EH05-2, countered at 3σ. Amino acid residues within 6Å from the ligand, crucial for the recognition of HSP90 [78], are shown in yellow.

Potential inhibitors of SMYD3 protein-protein interactions
The discovered compound was found to occupy a hot-spot, proposed to facilitate the recognition between SMYD3 and HSP90 [78], Figure 3.10. This molecular chaperone may possibly serve as an allosteric modulator of the protein, and the discovered site indicates a potential regulatory effect. Additionally, a recent study pointed out an inconsistency in the previously reported involvement of SMYD3 in cancerogenesis [79].
Thus, new model compounds with different mechanisms of action can play important roles as powerful chemical genetic tools to elucidate the problem, clarify protein function and to confirm its validity as a drug target.
4. Summary and future perspectives

A good understanding of a living system can only be developed utilizing molecular theory. The main goal of this thesis was to develop improved expertise in methods that assist research in biochemistry and molecular biology. Methodological progress is extremely important irrespective the aims of the research, being fundamental or applied. Rational approaches for molecular design require a full understanding of the mechanism of action, and biophysical techniques are exclusive tools to assess it.

Biosensor-based kinetic analysis of interactions explore the time-dependent aspect of recognition. Studies presented in this thesis show its application for analysis of binding partners of various nature. Uncharacterised before aspects of recognition were revealed using model systems of carbonic anhydrase isozymes (papers I&II). The concept of intrinsic kinetics is extremely important, as it links compound’s structure, chemistry and interaction kinetics.

Calmodulin interactions, examined in paper III, should have a physiological explanation. *Vice versa* states true – experimentally observed protein-protein interactions and their regulation can explain biological phenomena. Understanding the relationships between signalling proteins under various conditions can contribute to a mechanistical understanding of signal transduction cascades.

In this thesis, it was shown how unusual QSAR properties can be explained employing the biophysical approach, and that an assay-based lead optimization strategy always should be combined with structural studies. This was exemplified by the design of macrocyclic inhibitors of LSD1 (paper IV). The mismatch between *in silico* and *in esse* structures of the bounded macrocycles severely mislead the campaign. Now, that the starting point for further optimization is known, ligands can be evolved to rigid, potent peptidomimetics.

Certain targets do not have sufficient activity or easily detectable reaction signatures, restricting the reliability of *in vitro* functional assays. Interaction assays allow the identification of ligands in a robust manner, especially when the expected potency of probes is low (paper IV&V). Utilizing combined biophysical and structural strategy, a potential allosteric site was found in SMYD3 (paper V). The discovered ligand can be a foundation for a new approach to monitor and to alter SMYD3 function *in vivo*, especially considering an emerging interest in protein-protein interface inhibitors.
5. Sammanfattning


Huvudmålet med denna avhandling var att utveckla kunskap i användningen av molekylära metoder för biokemisk och molekyläri-biologisk forskning, särskilt med inriktning mot läkemedelsutveckling. Metodologiska framsteg är oerhört viktiga oavsett om syftet med forskningen är grundläggande eller applicerad. Rationella metoder för molekylär design kräver en fullständig förståelse av molekylens tänkta verkningsmekanism och biofysiska tekniker är viktiga verktyg för att ta fram denna information. Biosensorbaserade kinetiska analyser av molekylära interaktioner kan användas för att kartlägga hur snabba dessa är i olika steg. De studier som presenteras i denna avhandling visar hur biosensorbaserade analyser kan utformas för att studera hur proteiner interagerar med olika typer av interaktionspartners.

Kalmodulin (CaM) är ett vitalt protein involverad i många regleringsprocesser. Det är känt som ett kalciumbindande protein (CaBP) som åndrar sin förmåga att interagera med andra proteiner efter att det bundit Ca$^{2+}$. Det fungerar därmed som en calciumsensor och spelar en viktig roll vid intracellulär signalering. Dess interaktioner med andra proteiner dock ännu inte är kartlagda på en kinetisk nivå, vilket begränsar vår förståelse av hur CaM fungerar fysiologiskt. Genom studier av protein-proteininteraktioner och deras reglering kan många biologiska fenomen få en förklaring. CaM biologiska roll gör det till ett bra modellsystem för att studera förståelsen av hur relationer mellan signaleringsproteiner under olika förhållanden kan bidra till en mekanistisk förståelse av signaltransduktionskaskader.

I denna avhandling studerades kinetiken för kalciumberoende protein-interaktioner med CaM med biosensorteknik (papper III). Kinetiska parametrar togs fram för interaktionen mellan CaM och Neurogranin (Ng), ett neuronalt, postsynaptiskt protein tidigare känt att binda CaM endast i frånvaro av kalcium. Experimentella resultat visade överraskande en kvarstående förmåga att binna även i närvaro av kalcium. Detta är viktigt med tanke på neurogranins roll i kontroll av tillgängligheten av fritt CaM i synapserna vid Ca$^{2+}$-signalering. Komplexiteten i neurogranins roll i både kalciummättnad och apoCaM kan förklara detaljer i nervcellers signalering och kalciumkänslighet. Studierna avslöjade hur proteiners strukturella dynamik påverkar interaktionerna kinetiskt. En generell slutsats var att den komplexa relationen mellan struktur och biologisk effekt kan förklaras med data från tidsupplösta biofysiska metoder.

Interaktionskinetiska studier tillämpades även för flera studier syftande till framtagning av läkemedel. Modern läkemedelsutveckling utgå upifrån kunskap om proteiner involverade i sjukdomsprocesser och vars funktioner sannolikt kan slås ut med läkemedel som binder till "målproteiner". Målproteinernas biologiska funktioner ofta är svåra att studera in vitro och biofysikaliska metoder behövs för att välja molekyler med önskade egenskaper. Epigenetiska enzymer är exempel på sådana målproteiner, LSD1 och SMYD3 är två viktiga exempel som studerades i artikel IV och V. Båda proteinererna är involverade i utvecklingen av cancer och anses därför vara lämpliga målmolekyler lokalisera i tumörer.


Vissa målproteiner för läkemedel är svåra att studera experimentellt eftersom de inte har tillräcklig funktionell aktivitet eller lätt detekterbara skillnader mellan substrat och produkter. Detta begränsar tillförlitligheten av funktionella biokemiska analyser. Biofysiska analyser av interaktionerna som sådana är mycket fördelaktiga eftersom de tillåter identifieringen av hämmare på ett robust sätt, speciellt när den förväntade potensen av outvecklade substanser är låg (papper IV & V). För SMYD3 screenedes ett

Avhandlingen visar att kinetiska studier är viktiga verktyg i såväl förståelse av proteiners biologiska funktion och reglering, som för identifiering och utveckling av interaktionspartners relevanta för utveckling av läkemedel, men att optimeringen av läkemedelsprototyper bör kompletteras med strukturella studier.
6. Acknowledgments

Certainly, chemistry is the most beautiful science – simply because of the subject, matter and its transformation. I knew that I want to be a chemist since 14 y.o., and after the second year in a chemistry program at Moscow University I chose to focus on the chemistry of proteins. In 2014 I entered a postgraduate school in Uppsala University, and there is a long list of people who contributed significantly to this thesis – an important milestone of my 8 years journey in a world of macromolecules.

I am grateful to Prof. Helena Danielson, my supervisor who taught me how to work independently. Her guidance shaped my mind and had the biggest influence over my skills and expertise to date. I am obliged to mention other members of HD group, present and past, especially Dr. Eldar Abdurakhmanov and Dr. Christian Seeger for sharing their experience, and Edward FitzGerald and Daniela Eriksson, with whom I had a pleasure to work together on certain projects. I am thankful to Prof. Gunnar Johansson for useful discussions regarding the separation techniques, Dr. Doreen Dobritzsch for guidance regarding the protein crystallography, and my co-supervisor Dr. Ylva Ivarsson for keeping me motivated. I must highlight the importance of collaborations I was involved in, and it was a pleasure to work together with organic chemists of the Department of Chemistry - primary, Dr. Jie Yang and Prof. Jan. Kihlberg. Also, I am grateful to my external collaborators - Dr. Vaida Linkuviene, Prof. Daumantas Matulis and Dr. Edoardo Fabini. I am thankful to all other people at Uppsala University I worked with on a daily basis. Separately I want to thank Mohammed Ali, who managed to transfer to me a little of his enormous knowledges of molecular biology techniques.

I had a lot of assistance from my project students: Darius, Henri, Claire, Filip and Martin. Their impact on this work can not be understated.

Fortunately enough I had friends outside the bench. Marcus, Stefan and the rest of the crew - thank you for keeping me social. Two other persons are standing aside - my dearest friends Carla Carvalho and Filip Mihalič, their support was enormous.
Bibliography


A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)