Beyond Affinity: Drug-Kinase Interaction Put under the Microscope

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Pharmazie der

Philipps-Universität Marburg

vorgelegt von Dipl.-Humanbiologin

Barbara Wienen-Schmidt geb. Wienen

geboren in Neuss

Marburg/Lahn 2016

Erstgutachter: **Prof. Dr. Gerhard Klebe**

Zweitgutachter: **Prof. Dr. Andreas Heine**

Eingereicht am:

Tag der mündlichen Prüfung am:

Hochschulkennziffer: 1180

Die Untersuchungen zu den vorliegenden Thema wurden auf Anregung von Prof. Dr. G. Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von März 2012 bis Dezember 2016 durchgeführt.

Meiner Familie

Table of contents

1 Introduction

1.1 Drug design

The origin of many diseases or their symptoms is the dysregulation of particular proteins in certain organs and cells of the human body. The basic concept in drug design is that the activity of these proteins can be artificially modulated through the binding of a drug to these assigned target proteins.

However, the development of such an approved drug is an expensive process that on average takes 13.5 years (example from 2007) and capitalized costs of around $ 1,778 million.[1] An acceleration and cost reduction could be achieved if the prediction of successful drug candidates in the early stage of development was improved. Hence, the major challenge we are facing nowadays, is to reliably predict a drug's *in vivo* efficacy, its potential for resistance development and success in clinical trials.

With respect to this topic, a paradigm shift could be witnessed over the last couple of years. The classical approach in drug design is to solely optimize the affinity between a drug candidate and its target protein in the process of drug development. Nonetheless, many recent studies suggest that affinity between the target protein and a drug alone is not the best descriptor for a drug's *in vivo* effectiveness and success.

Additional biophysical descriptors such as thermodynamics and kinetics of ligand binding to the target protein are taken into account with increasing popularity over the last years (**Figure 1**). These additional parameters allow further differentiation of drug candidates early on in the development

process.[2] Nonetheless, it is still under discussion in what direction for a particular application these biophysical parameters should be optimized. Furthermore, the detailed understanding on how rational chemical modifications of potential drugs really influence kinetics or thermodynamic signatures of protein binding is still insufficient. In order to enable these predictions it is crucial to use and characterize model systems as detailed as possible. Consequently there is a strong requirement for extensive and systematic investigations of drug-target interaction in order to gain understanding and hence improve predictive power for the process of drug design.

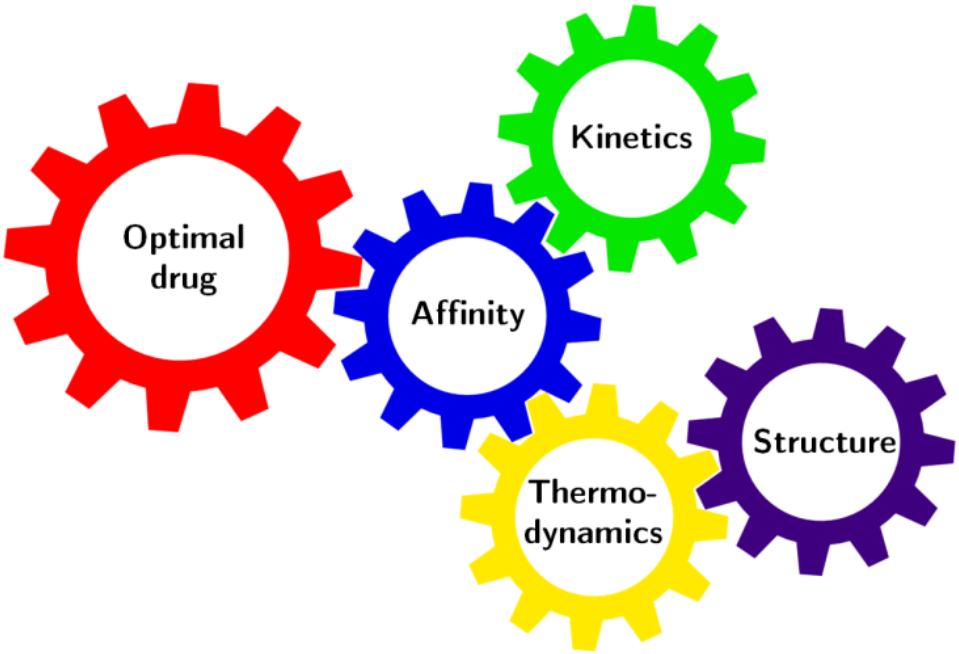


Figure 1: Systematical investigation of various protein-ligand interactions is required to understand the interplay between structure, thermodynamics, kinetics and affinity and how they have to interact in order to produce an optimal drug.

The essence of drug design is the induction of an optimal drug binding process to the target protein leading to an ideal drug-protein interaction. This binding is a highly complex process that is influenced by factors like shape complementarity, hydrogen bonding, van der Waals forces, and changes in protein, drug and protein-drug complex hydration, flexibility and degrees of freedom. In total, the binding affinity reflects the strength of the protein-drug interaction. In the following small molecule drugs, potential drug in the development process, as well as fragments thereof, are referred to as “ligands” or “compounds”.

1.2 Thermodynamics & isothermal titration calorimetry

As described above, thermodynamic analysis of protein-ligand complex formation is used with increasing popularity as an additional descriptor in the process of drug design. To measure thermodynamic profiles the method of isothermal titration calorimetry (ITC) is frequently applied and also employed in this thesis to measure the thermodynamic signature of different protein-ligand complexes. The schematic set-up of an ITC experiment is depicted in **Figure 2**.

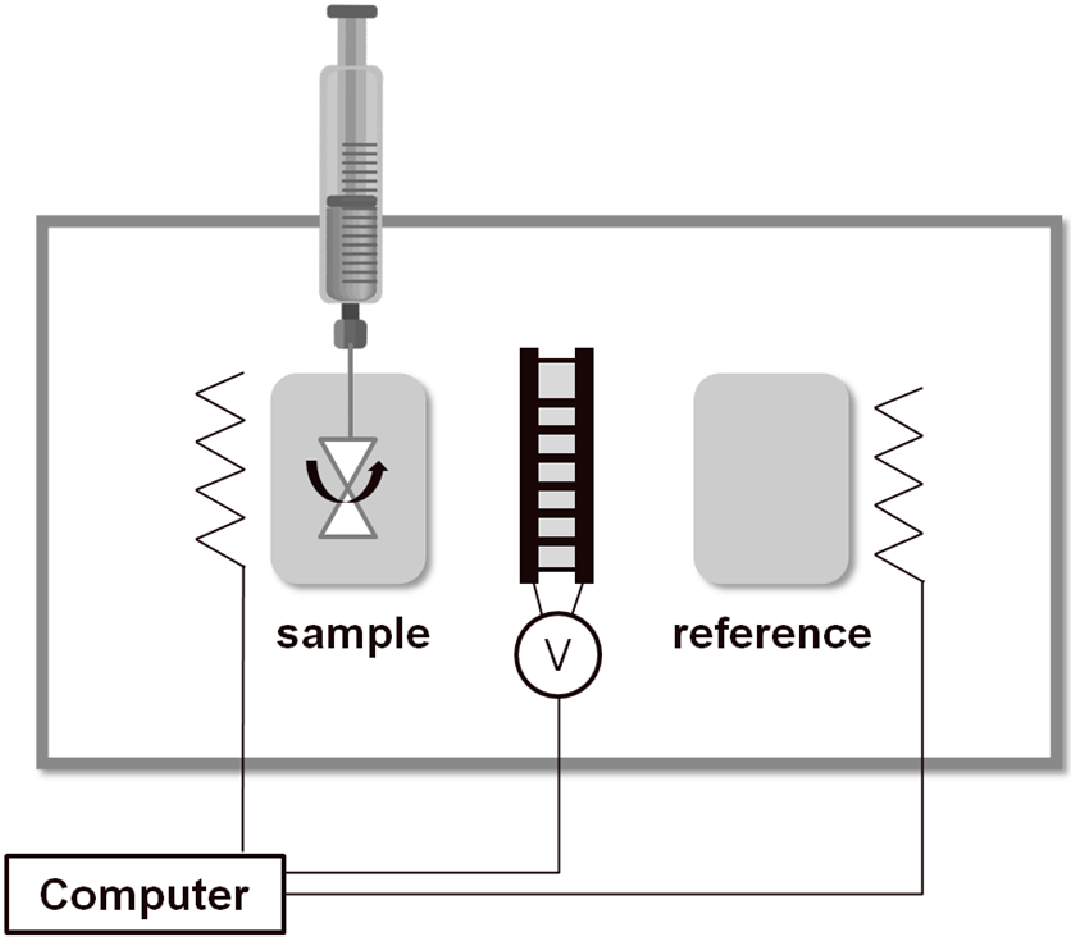


Figure 2. Schematic set-up of an ITC experiment. Under continuous stirring of the syringe paddle, the ligand is stepwise titrated into the sample cell which is filled with protein. The protein-ligand binding leads to a release or absorption of heat. The reference cell is filled with water and kept at a constant temperature.

The ITC device features two cells located within an adiabatic jacket, one cell contains the protein sample (sample cell) the other contains water as a reference (reference cell). Both cells are then heated to an equal, user-defined temperature. A constant power is applied to the reference cell to maintain the defined temperature. During the measurement, a small, defined volume of a ligand-solution is titrated into the sample cell. The binding reaction then

either releases or absorbs heat. Hence the power applied to the sample cell needs to be modified in order to achieve the same temperature as the reference cell (feedback power). For exothermic ligand binding, less power is required, for endothermic ligand binding the power applied needs to be increased. The differential power is proportional to the difference in temperature between the two cells and is recorded.[3]

Multiple aliquots of ligand solution are stepwise titrated into the sample cell. Each injection causes a peak shaped heat signal (**Figure 3**). Peak integration after the measurement yields a sigmoidal curve. From this curve association constant K_a, stochiometry n and enthalpy of binding ΔH° can be read. ΔG° can then be calculated from K_A using the definition of Gibbs free energy of binding for a system at chemical equilibrium, where R is the ideal gas constant (8.314 J mol^{-1} K^{-1}) and T the absolute temperature (K)[4]:

$$\Delta G^\circ = -RTln(K_a)$$

The entropic contribution $-T\Delta S^\circ$ consisting of the change in entropy ΔS° (kJ mol^{-1}), which is weighted by the temperature, is then calculated from ΔG° (kJ mol^{-1}) and ΔH° (kJ mol^{-1}) using[5]:

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

Therefore, only ΔG° and ΔH° are directly determined. Entropy is merely calculated and therefore the most error-prone parameter.

The symbol "°" indicates a binding free energy value in its standard state and is often neglected for convenience. The reference of the standard state is essential for scaling and comparing different measurements. The standard state is defined for 1 mole ligand and 1 mole protein forming 1 mole protein-

ligand complex. The reaction takes place in a hypothetical ideal solution having a unit activity coefficient at a constant pressure of $p^\circ = 10^5$ Pa.[6]

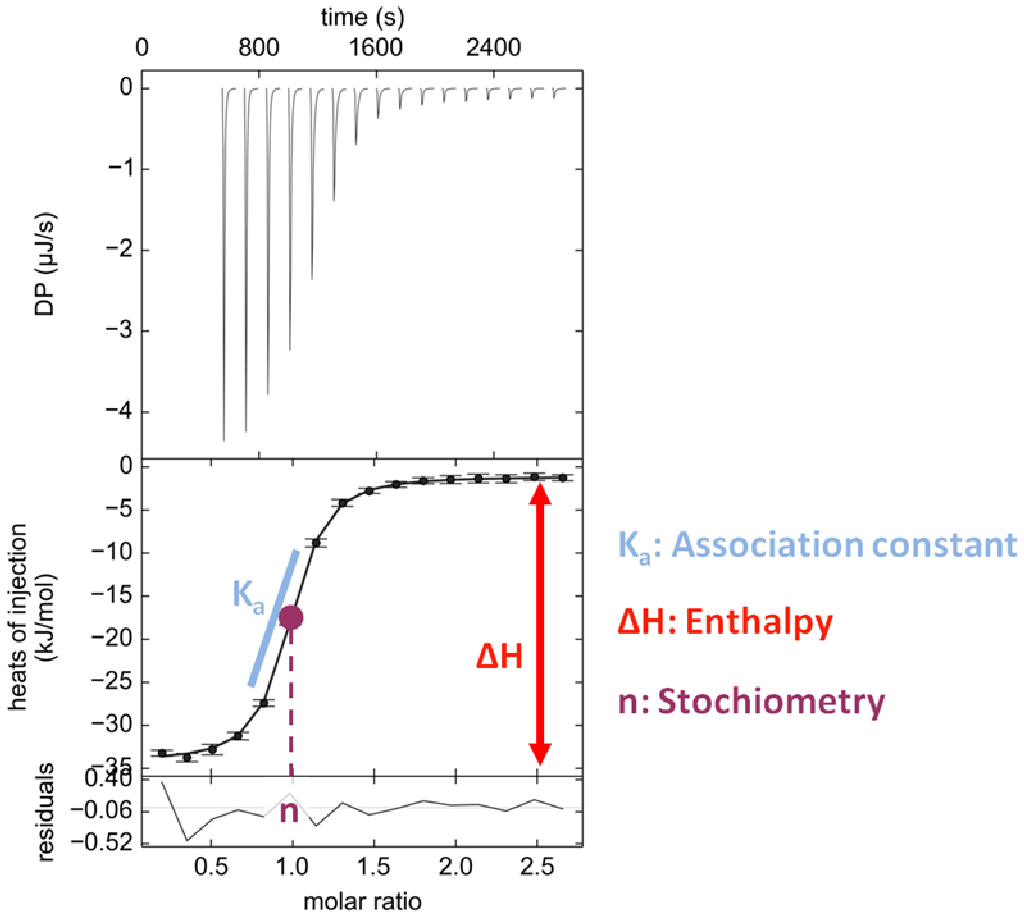


Figure 3. ITC-thermogram. The top graph depicts the different injection peaks during the titration (exothermic signal). The lower graph shows the sigmoidal curve resulting from peak integration. In this way, the experimental values K_a, ΔH and the stochiometry of the reaction are obtained.

1.3 Enthalpy vs. entropy

As described above, the Gibbs free energy of binding ΔG partitions into enthalpy ΔH and entropy $-T\Delta S$. Thereby, enthalpy is a measure for the

amount of heat energy required to reach a certain state, while entropy reflects how easily this energy can be dispensed throughout various molecular energy levels.[4]

A favorable enthalpy for a ligand-protein complex formation can result from favorable hydrogen bonds, the quality of which is defined by their distances and angles. Also van der Waals interactions can have a beneficial impact on enthalpy. They mostly result from hydrophobic interactions and desolvation. The desolvation of polar groups on the other hand has an unfavorable effect on the entropic contribution.[4]

Concerning entropy, binding of two partners to form one common complex inevitably results in an entropic penalty that has to be overcome. Upon binding, the system looses half of its translational and rotational degrees of freedom. Moreover, the ligand and partially also the protein looses conformational degrees of freedom. However, an increase in buried hydrophobic surface through ligand binding, resulting in a release of water from the active site, is considered entropically favorable (hydrophobic effect).[4]

Due to enthalpy-entropy compensation optimizing both terms unfortunately remains challenging. A system with improved specific interaction and hence improved enthalpy is usually less dynamic, resulting in a, partially compensating, unfavorable entropy.[7]

1.4 The physiological relevance of PKA

The cAMP-dependent protein kinase (PKA) is used as a model protein in this thesis. Its ample implication in physiologically important signalling

pathways makes PKA unsuitable as a drug target. Potential side effects could be severe. However, this kinase is well investigated, can be expressed in large quantities, crystallizes easily and can be used for protein NMR as well as SPR measurements. Hence it is the perfect model protein for systematic kinase studies.

PKA is physiologically activated downstream of membrane integrated G-protein coupled receptors (GPCR). Through receptor stimulation outside the membrane, GPCRs can bind and thereby activate the stimulatory G_s-protein. The G_s-protein is a heterotrimer consisting of an α_s-, β- and γ-subunit. The α_s-subunit can bind Guanosindiphosphate (GDP) as well as Guanosintriphosphate (GTP). Upon activation, GDP is exchanged for cytosolic GTP and the $\beta\gamma$-complex as well as the GPCR is separated from the α_s-subunit. The remaining α_s-GTP complex then activates adenylyl cyclase at the inside of the membrane, leading to an increased conversion of ATP into cyclic adenosine monophosphate (cAMP). This rise in the cAMP level can then activate PKA.[8] In its inactive state, PKA is a heterotetramer and consists of two regulatory and two catalytic subunits. The regulatory subunits can cooperatively bind two molecules of cAMP. Binding of cAMP leads to a conformational change resulting in a release of the two separate catalytic subunits. Upon release, the activated catalytic subunits can then phosphorylate serine and threonine residues of various substrate proteins. These phosphorylations regulate a wide range of signaling pathways further downstream of PKA. [9]

For instance, PKA is involved in glycogen metabolism and it stimulates gene expression by phosphorylating the transcriptional activator cAMP-response

element binding (CREB) protein. Furthermore, PKA can phosphorylate potassium channels in neurons, thereby facilitate their closing, leading to an increasing excitability of the cell.[10] Generally, the effects of PKA signaling strongly depend on the cell type.

1.5 The structure of PKA

Overall, kinase structures are widely conserved. Kinases consist of a small N-terminal lobe with 5 β-sheets. The small lobe is connected through the hinge region with the larger C-terminal lobe which is predominately build from α-helices. The active site is located in-between the two lobes.[11]

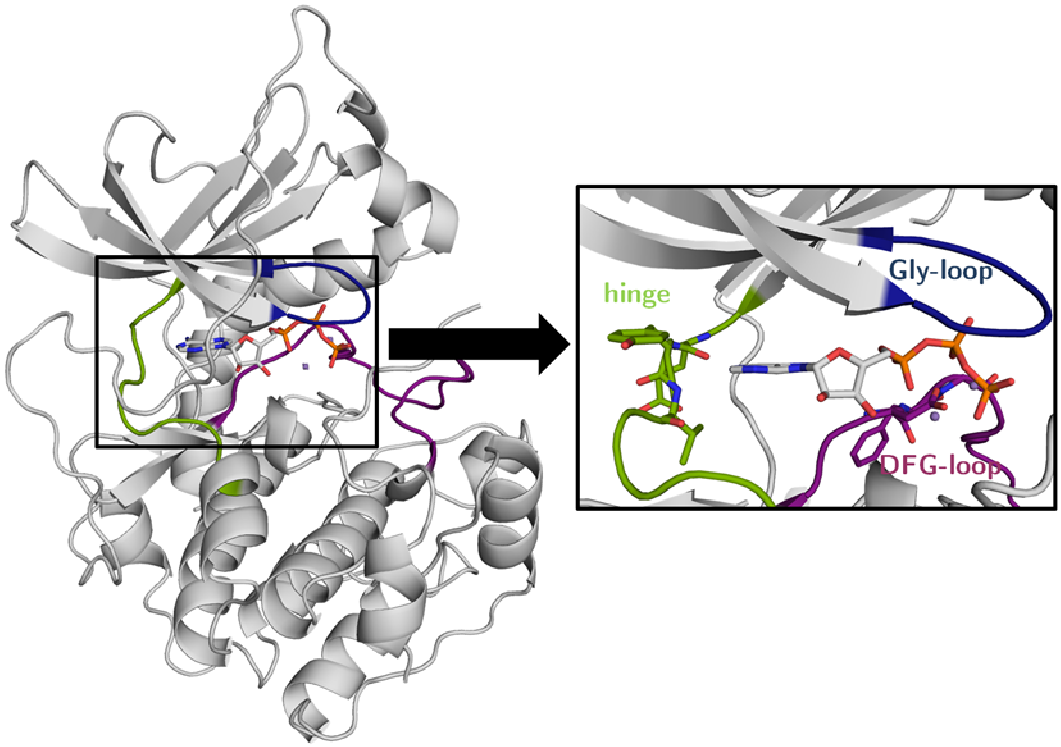


Figure 4. Crystal structure of PKA from *mus musculus* in complex with ATP. The hinge region is colored in green, Gly-loop in blue and the DFG-loop in purple. In the active site ATP is bound together with two Mn^{2+} ions (physiologically Mg^{2+} ions). The structure is taken from the protein data bank (PDB), code 1ATP[12].

The entire kinome can be divided into nine families. PKA itself belongs to the AGC-kinase family of serine/threonine kinases.[13] All 63 kinases included in this group respond to changes in local concentrations of cytoplasmic second messengers such as cAMP or lipids.[14]

In **Figure 4**, the general fold of PKA is visualized; the most important elements for ATP- and ligand-protein interaction are colored. Physiologically, the adenine ring of ATP is positioned in a deep hydrophobic pocket between the small and big lobe and forms hydrogen bonds to the hinge region (green). Two magnesium ions, essential for ATP binding, enable the positioning of the negatively charged ribose-triphosphate tail. Thereby the γ-phosphate that is transferred in the catalytic process is located at the edge of the catalytic cleft. It is stabilized by the glycine-rich loop (Gly-loop) (blue) and the DFG-loop (purple). Another conserved sequence involved in substrate recognition is the APE-motif which located in the C-terminal substrate recognition pocket.

Hinge, Gly-loop and DFG-loop are also crucial interaction sites for ATP-competitive kinase inhibitors and will be referred to in the following chapters of this thesis. The hinge region is the most essential drug interaction partner, possessing two donor and one acceptor function in nearly linear arrangement.[15] The Gly-loop is highly flexible and can adopt closed and opened conformations. Crucial for the activity and for so-called DFG-out inhibitors is the DFG-loop. If the phenylalanine from the DFG-motif is rotated out of the binding pocket, a kinase is in its inactive conformation. Some ligands are able to trigger this conformation. However, for PKA so far

no DFG-out conformation has been observed, hence it remains unclear whether PKA can adopt such a conformation.

1.6 Aim of the thesis

This thesis is subdivided into six parts. The first four chapters focus on systematic thermodynamic and structural examination of drug-kinase interactions.

Protein-drug interaction is analyzed using cAMP-dependent protein kinase (PKA) as a model protein. The use of well-studied model proteins allows capturing elaborate information from a wide range of experimental techniques. Therefore, PKA is a prototype of the wide and clinically relevant protein kinase family. The human genome might encode for up to 518 protein kinases.[13] Many of these kinases might be of pathological relevance. Anomalous kinase activity is particularly important in diseases that originate from inflammatory and proliferatory responses. Examples are cancer, rheumatoid arthritis, cardiovascular and neurological disorders, asthma and psorias. These implications make protein kinases indeed a major target in drug design.[16-18]

Here, the aim was to investigate the interaction of PKA with different ligand subsets. Even though all ligands derive from the approved drug fasudil, they each focus on different structural aspects relevant in drug design. Through minor structural changes in the ligand compositions, major changes in protein-ligand complex structure and thermodynamics of binding can be caused. With this approach current theories can be tested on their

applicability in the field of protein kinase inhibition. Aspects analyzed in the first four chapters are the impact of ligand degrees of freedom, changes resulting from ligand growth, the possible influence of methyl groups decorating the ligands in various positions, and differences in the hinge binding modes of the ligands. The hinge region, being in the active site of the protein, is the crucial area of attack for ATP-competitive kinase inhibitors.

The fifth chapter focuses on the issue of kinase selectivity. A kinase screen was performed and thermodynamic data were checked for correlation with selectivity profiles. Current hypotheses on the topic should be experimentally tested.

The sixth chapter is a methodology analysis. It aims at comparing the two major crystallization methods: Soaking and co-crystallization. Using PKA as an example, pronounced structural differences were observed.

1.7 References

1 Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. *Nat Rev Drug Discov* **2010**, *9*, 203–214.

2 Renaud, J.-P.; Delsuc, M.-A. *Curr Opin Pharmacol* **2009**, *9*, 622–628.

3 Holdgate, G. A.; Ward, W. H. J. *Drug Discov Today* **2005**, *10*, 1543–1550.

4 Chaires, J. B. *Annu Rev Biophys* **2008**, *37*, 135–151.

5 Klebe, G. *Nat Rev Drug Discov* **2015**, *14*, 95–110.

6 Krimmer, S. G.; Klebe, G. *J Comput Aided Mol Des* **2015**, *29*, 867–883.

7 Freire, E. *Drug Discov Today* **2008**, *13*, 869–874.

8 Silbernagl, S. *Taschenatlas Physiologie*; DeL; Thieme, 2012.

9 Smith, C. M.; Radzio-Andzelm, E.; Madhusudan,; Akamine, P.; Taylor, S. S. *Prog Biophys Mol Biol* **1999**, *71*, 313–341.

10 Berg, J.; Tymoczko, J.; Stryer, L. *Biochemistry, Fifth Edition*; Biochemistry; W. H. Freeman, 2002.

11 Taylor, S. S.; Zhang, P.; Steichen, J. M.; Keshwani, M. M.; Kornev, A. P. *Biochim Biophys Acta* **2013**, *1834*, 1271–1278.

12 Zheng, J.; Trafny, E. A.; Knighton, D. R.; Xuong, N.-H.; Taylor, S. S.; Ten Eyck, L. F.; Sowadski, J. M. *Acta Crystallogr D Biol Crystallogr* **1993**, *49*, 362–365.

13 Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912–1934.

14 Turnham, R. E.; Scott, J. D. *Gene* **2016**, *577*, 101–108.

15 Stroud, R., Finer-Moore, J., Eds. *Computational and Structural Approaches to Drug Discovery*; RSC Biomolecular Sciences; The Royal Society of Chemistry, 2007; 001–382.

16 Melnikova, I.; Golden, J. *Nat Rev Drug Discov* **2004**, *3*, 993–994.

17 Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. *Science* **2004**, *303*, 1800–1805.

18 Zhang, J.; Yang, P. L.; Gray, N. S. *Nat Rev Cancer* **2009**, *9*, 28–39.

2 The Surprising Impact of Flexibility on Drug-Kinase Interaction

2.1 Annotations

The following chapter will be published. It is a cooperation of the groups of Prof. Gerhard Klebe and Prof. Harald Schwalbe. The following people contributed to the subject and will therefore be listed as authors on the publication:

Barbara Wienen-Schmidt[‡], Hendrik R. A. Jonker[†], Tobias Wulsdorf[‡], Hans-Dieter Gerber[‡], Krishna Saxena[†], Andreas Heine[‡], Harald Schwalbe[†]*, Gerhard Klebe[‡]*

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

[†]Institut für Organische Chemie und Chemische Biologie, Johann Wolfgang Goethe-Universität Frankfurt, Max-von-Laue-Straße 7 , N160-3.14 , 60438 Frankfurt am Main, Germany

* Corresponding authors

Protein expression for ITC and crystallization was performed by Barbara Wienen-Schmidt as well as ITC-measurements, crystallization and crystal structure determintion. Ligand synthesis and qNMR measurements were performed by Hans-Dieter Gerber. Amide chemical shift perturbations as well as amide ^{15}N-T2-relaxation measurements and related protein expression were performed by Henry Jonker.

2.2 Abstract

Considering biophysical parameters in drug design can accelerate and direct the development of clinically successful drugs. Currently, the selection of drug candidates with particular thermodynamic and kinetic profiles upon binding is under discussion and even prevalently applied. One such concept is the design of rigid ligands in order to reduce the entropic penalty upon binding to the target protein. Nonetheless, systematic studies analyzing to what extent ligand rigidity influences the overall thermodynamic profile are deficient. In this study, the effect of ligand flexibility on the thermodynamic profile of ligand binding to a protein was investigated. For this purpose, the wide and clinically relevant family of protein kinases is represented by cAMP-dependent protein kinase (PKA). All ligands derive from the approved drug fasudil and have an equal number and scope of atom types of heteroatoms while varying in their internal degrees of freedom. Protein-ligand interactions were scrutinized using X-ray crystallography, isothermal titration calorimetry (ITC) and nuclear magnetic resonance spectroscopy (NMR). Three different aspects that influence the resulting thermodynamic profile were checked: 1. Protein-ligand complex flexibility, 2. Ligand flexibility and 3. Protein-ligand complex hydration. From our results we conclude, that protein-ligand complex flexibility does not differ significantly for our series of ligands even though a strong induced fit is observed. Interestingly, ligand flexibility and hence the loss of degrees of freedom upon binding does in no way dominate binding thermodynamics. Water however, more detailed the change in protein-ligand complex hydration, as well as differences of ligand hydration in aqueous solution prior to binding seem to be the crucial factors and are essential to consider for the general understanding of thermodynamic

profiles of ligand binding. Altogether, our results demonstrate that protein-ligand interactions for flexible proteins are even more complex than for rigid proteins. Our data demonstrate that global understanding and potential future prediction of thermodynamic profiles are highly dependent on further systematic studies.

2.3 Introduction

Finding selective, effective and clinically successful drugs is a long and expensive enterprise. Hence, it would be beneficial if this process could be accelerated by improving criteria to predict successful drug candidates and to identify unsuitable ones early on. Therefore, the analysis of parameters beyond affinity is required. In this context thermodynamic analysis of protein-ligand interaction is used with increasing popularity. Accordingly, the design of drugs with certain thermodynamic properties has been described as a promising approach in many articles and reviews.[1-6] Moreover, design guidelines to accomplish purposefully tailored profiles have been intensely discussed, frequently advising the development of rigid, correctly preorganized ligands in order to influence the entropic penalty upon binding of the ligand to its target.[7-10, 2]

However, there remains a lack of systematic studies that explore correlations between ligand structure, target protein characteristics and thermodynamic signature. From our perspective, it is important to distinguish between different types of proteins. An utter generalization cannot lead to satisfactory answers. Particular factors that have to be taken into account are: 1. Size; 2. Flexibility of the target and 3. Water structures. In this study we thermodynamically investigated the wide and clinically relevant family of

protein kinases. Protein kinases are highly flexible proteins [11-13] as indicated by the presence of flexible loops in the active site and helix movement upon ligand binding. They have been rarely investigated by means of comprehensive thermodynamic characterization. Only ten unique kinases are listed in databases that are specialized for thermodynamic annotations, such as Scorpio[14] and BindingDB[15] (CDK2, ERK1/2, JNK2, Pim1, Aurora-A, Thymidine Kinase, Nucleosid Diphosphate Kinase, vSrc, cSrc). Indeed for none of these targets has a systematic investigation of ligand series been conducted. Considering that according to Manning the human genome might encode for up to 518 protein kinases[16] and with respect to their importance as a drug target class, there is a clear need for thermodynamic data to characterize protein kinases. The aim of this study is to start filling this gap, in order to broaden our understanding of kinase dynamics and thermodynamics of ligand binding.

Here, the use of well-studied model proteins allows for the capture of extensive information from a wide range of experimental techniques. Using the cAMP-dependent protein kinase (PKA) as a model system, elaborate information on protein-ligand interaction was obtained using X-ray crystallography, isothermal titration calorimetry (ITC) and nuclear magnetic resonance spectroscopy (NMR). The selected ligands are derived from the approved drug **fasudil**[17], which has been developed as a rho-kinase inhibitor[18] but, nonetheless, displays high affinity toward PKA[19]. All of these ligands have an equal number and scope of atom types of heteroatoms but vary in their internal degrees of freedom (**Figure 1**). In this way, the effect of ligand flexibility on the thermodynamic protein binding profile can be investigated. Thorough analysis of the underlying structural factors

contributing to this thermodynamic profile has been made using X-ray crystallography and $^{1}H^{15}N$ best-TROSY (transverse relaxation optimized spectroscopy) spectra. Furthermore, protein-ligand complex dynamics were analyzed using ^{15}N-T_2-relaxation measurements. Here, we present a unique study where the static data from high resolution crystal structures is faced and combined with the information on the dynamics from detailed NMR measurements.

For the thermodynamic data, ITC was used to measure ΔG and ΔH. Subsequently, –TΔS was calculated using the expression for Gibbs free energy of binding:

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

The K_D was determined from ΔG using the definition of Gibbs free energy for a system at chemical equilibrium:

$$\Delta G = -RTln\left(\frac{1}{K_D}\right)$$

The variables have the subsequent definitions: ΔG: Gibbs free energy of binding; ΔH: Enthalpy of binding; ΔS: Entropy of binding; R: Ideal gas constant; T: Absolute temperature; K_D: Dissociation constant.

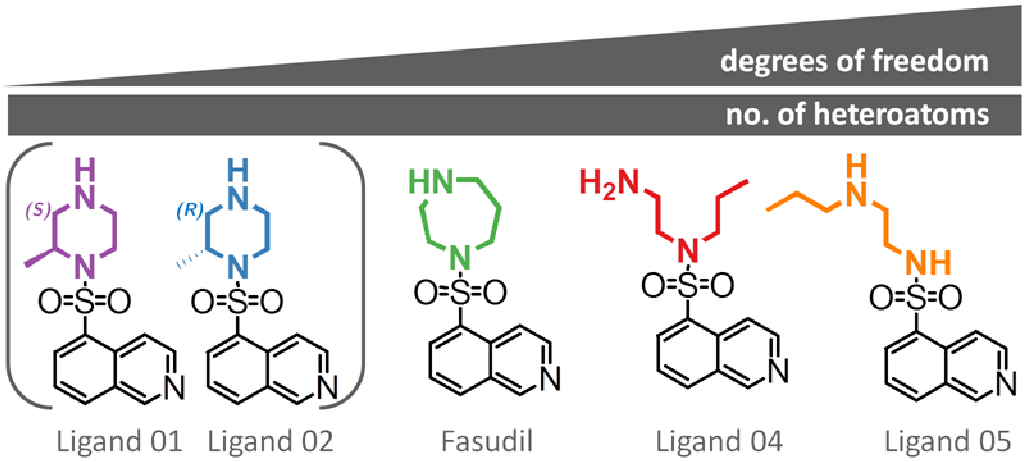


Figure 1. Chemical structure of the five ligands used in this study. All ligands have an equal number of heteroatoms but different conformational degrees of freedom. From the left to the right the internal degrees of freedom increase and will, in any case, require an entropic price to be paid upon binding, as an increasing number of torsional degrees of freedom will be lost. In purple: S-methyl-piperazine substituted fasudil-derivative (Ligand **01**); in blue: R -methyl-piperazine substituted fasudil-derivative (Ligand **02**); in green: **Fasudil**; in red: open-chained fasudil-derivative (Ligand **04**); in orange: long-chained fasudil-derivative (Ligand **05**)

Our study displays striking results that could hardly have been predicted, as counterintuitively the most flexible ligand binds entropically most favorably to the protein. Consequently, the hypothesis stating that a significant loss of a ligand's degrees of freedom is detrimental with respect to the entropic contribution to binding, needs to be questioned and analyzed in the context of the entire binding event. We conclude that a generalization of simple design guidelines is not constructive and will not lead to satisfactory results.

2.4 Experimental section

Protein expression and purification

The catalytic subunit of cAMP-dependent protein kinase from Chinese hamster ovary cells was expressed with a His-tag in a modified pET16b-Vector with an introduced TEV-cleavage site between the protein N-terminus and His-tag. This plasmid was transformed into *E. coli* strain *Bl21 (DE3)/pLysS* (Novagen).[20]

Cell disruption was performed using a high-pressure homogenizer for multiple cycles. After centrifugation (1h at 30.000g) cell lysate supernatant was purified in a first step using a Ni-NTA column that binds the His-tag of the protein and was eluted by an imidazole gradient. The His-tag was then cleaved off by TEV-protease. Afterwards, an inverse Ni-NTA column was employed collecting PKA in the flow-through. Finally, ion exchange chromatography was performed using a MonoS column separating three-fold phosphorylated PKA from the four-fold phosphorylated form using a HEPES buffer with a sodium chloride gradient.[20]

Crystallization

Co-crystallisation was performed using the hanging drop method at 4 °C. The crystallization drops contained the following ingredients: 10 mg/mL PKA (240 µM), 30 mM MBT (MES/Bis-Tris Puffer pH 6.9), 1 mM DTT, 0.1 mM EDTA, 75 mM LiCl, 0.03 mM Mega 8, 0.07 mM PKI (Sigma: P7739), 1.2 mM ligand dissolved in DMSO from a 50-100 mM stock. The well contained a mixture of methanol in water with varying methanol concentrations (v/v) for the different ligands (**Fasudil**: 18% methanol; Ligand **05**: 18% methanol;

Ligand **04**: 14% methanol; Ligand **01**: 16% methanol; Ligand **02**: 19% methanol). In the crystallization setup streak-seeding was performed with apo-crystals as seeds using a horse hair in order to initialize crystal growth. For crystal mounting, crystals were cryo protected in 5 mM MBT (MES/Bis-Tris Puffer pH 6.9), 1 mM DTT, 0.1 mM LiCl, 1.2 mM ligand dissolved in DMSO from a 50-100 mM stock, 16 % (v/v) methanol, 30% (v/v) MPD and flash frozen in liquid nitrogen.

Crystallography

All structures were collected at the storage ring Bessy II Helmholtz-Zentrum Berlin, Germany at Beamline 14.1 on a Pilatus 6M pixel detector. The datasets were processed using XDS[21] and molecular replacement was performed using CCP4 Phaser[22] and PDB-structure of PKA from *bos taurus* 1Q8W as a model. This was followed by simulated annealing, multiple refinement cycles of maximum likelihood energy minimization and B-factor refinement using Phenix[23]. Coot[24] was used to fit amino-acid side chains into σ-weighted 2Fo – Fc and Fo – Fc electron density maps. If appropriate electron density was observed, multiple side chain conformations were built into the model and maintained during the refinement if the minor populated side chain displayed at least 20 % occupancy. Hydrogen atoms were included using a riding model. Ramachandran plots for structure validation were calculated using PROCHECK[25]. Data collection, unit cell parameters and refinement statistics are given in the supplementary information. Analysis of temperature factors was performed with Moleman[26]. Protein and PKI B-factors were anisotropically refined, water B-factors were isotropically refined for all structures. Decision for anisotropic or TLS refinement was based on comparison of R_{free}. Anisotropic refinement was chosen over TLS if the

achieved R_{free} values were at least 0.5% lower for anisotropic than for TLS refinement. R_{free} was calculated using 5% of all reflections which were randomly chosen and not used for the refinement. The required ligand restraint files were created using the Grade webserver[27, 28]. For figure preparation Pymol was used.

Isothermal titration calorimetry

The buffer used for the ITC experiments contained: 30 mM sodium phosphate buffer pH 7.2, 10 mM $MgCl_2$, 100 mM NaCl, 3% (v/v) DMSO. All measurements were repeated 3-5 times. Further buffers were used in order to check for protonation linkage. In these buffers 30 mM sodium phosphate buffer was replaced by 30 mM HEPES and 30 mM triethanolamine (TEA), respectively (both at pH 7.2). Buffer dependency was tested for all measurements in the absence of PKI. For the measurements expressed, purified and dialyzed PKA was used in the ITC-measuring cell. A 15-20 fold higher concentrated ligand solution, diluted in dialysis buffer, was then stepwise injected to the protein solution during the measurement. All measurements were performed at 25 °C. ITC data were analyzed using NITPIC and Sedphat[29, 30]. Raw data and exact values and standard deviations for ΔG, ΔH and $-T\Delta S$ can be found in the supplementary information.

Compound purity was analyzed using quantitative nuclear magnetic resonance spectroscopy (qNMR) and in case of deviation, ligand concentration was corrected accordingly.

NMR

All protein NMR measurements were performed by the group of Prof. Harald Schwalbe in Frankfurt/Germany.

Ligands

Ligands **01**, **05** and **04** were purchased from Uorsy (Ukraine). The *R* and *S* isomers of the 2-methylpiperazine inhibitors **01** and **02** were synthesized starting from 5-(chlorosulfonyl) isoquinoline-hydrochloride prior to this freshly prepared from isoquinoline-5-sulfonic acid via a known literature procedure[31, 32], which was reacted with the respective, commercially available, enantiomerically pure R or S-configured mono-N-Boc-protected 3-methylpiperazine thus rendering the corresponding inhibitor precursors, respectively. Final *N*-Boc deprotection with 4 M HCl in dioxane gave rise to the corresponding inhibitors **01** and **02** as their hydrochloride salts.

2.5 Results

Strongest induced fit is triggered by the most flexible ligand. In order to verify comparable binding modes of all five ligands discussed, co-crystal structures were obtained. For all structures, resolutions between 1.4 and 1.6 Å could be achieved. In all five cases, difference electron densities are well defined and indicate every heteroatom of the bound ligands (supplementary information) with 100% occupancy in the binding site. All structures have been deposited in Protein Data Bank (PDB). The respective codes are listed in **Table 1**.

Importantly, the crystallographic data confirmed that indeed all five ligands display a congruent hinge binding position of their respective isoquinoline

moiety (**Figure 2B**). Moreover, their adjacent sulfonamides occupy a common orientation in all structures, where one oxygen points toward the Gylcine-rich loop (Gly-loop) and the second toward the hinge region. Hence, protein-ligand interactions are highly similar for the isoquinoline-5-sulfonamide portions of all five ligands. In all structures, a hydrogen bond is formed between the isoquinoline-nitrogen of the ligands and the backbone nitrogen of Val 123 of the protein. Besides these interactions, the compounds' sulfonyl-groups do not directly interact with the protein.

Table 1. List of PDB-codes for the different ligand co-crystal structures

Ligand	PDB code
Ligand 01	5LCU
Ligand 02	5LCT
Fasudil	5LCP
Ligand 04	5LCR
Ligand 05	5LCQ

In contrast to the identical ligand core binding, significant changes were noted amongst the five different protein-ligand-complexes considering the attached sulfonamide substituents.

In particular, the αG helix, the APE motif and the position of the protein-kinase-inhibitor-peptide (PKI) are visibly shifted (**Figure 2A**). Yet, the most prominent difference was revealed in the active site. Here, the Gly-loop adopts three distinct positions ranging from a wide-open to a closed conformation when compared to the apo-protein (**Figure 2B**).

The first, most open position of the Gly-loop exhibits the structure with the open-chained ligand **04** (red). Here, the loop is pushed out of the ligand

binding site as result of steric hindrance. The binding of ligand **04** simply requires more space in the area of the Gly-loop than any of the other ligands.

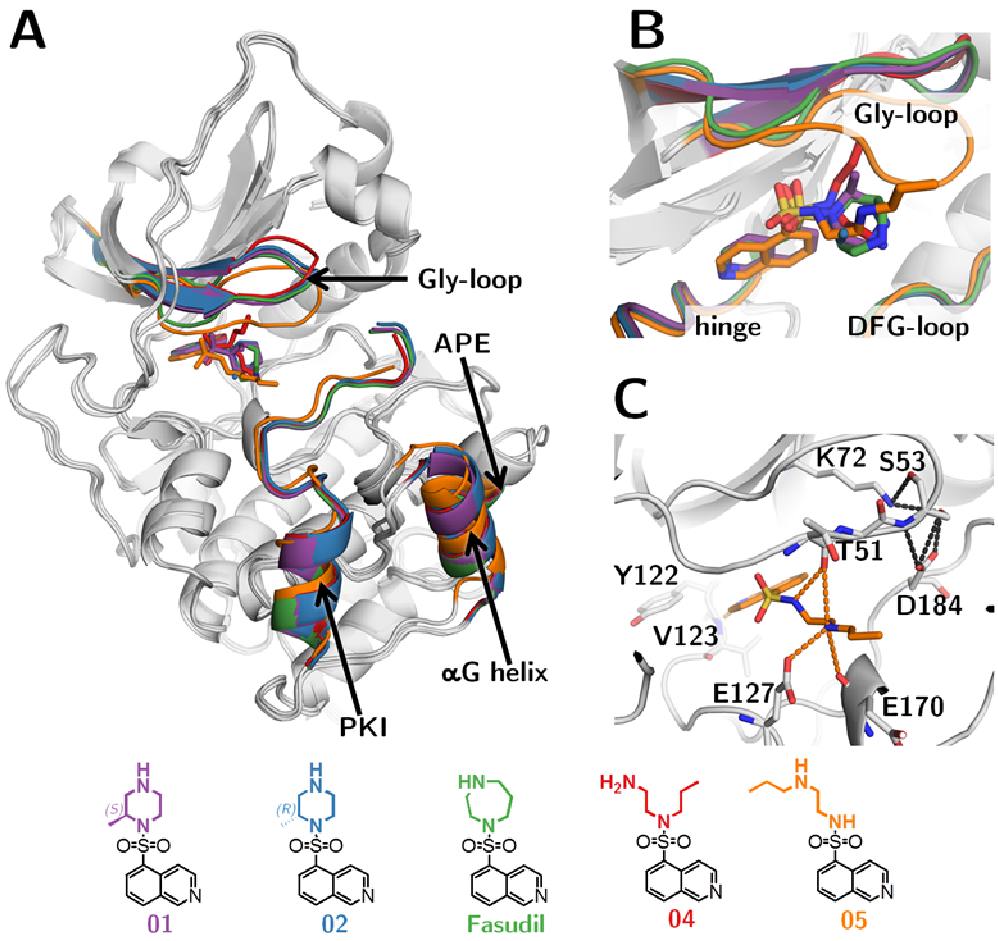


Figure 2. Superimposition of the co-crystal structures of all five ligands. A: Overall view of the protein. Ligand **04** is displayed in red; fasudil in green; **02** in blue; **01** in purple and **05** in orange. **B:** Blow up of the active site. The long-chain ligand **05** induces the strongest conformational change dragging the Gly-loop toward the ligand. **C:** Interactions between **05** and the protein are displayed as orange dotted lines. The key interaction responsible for the downward movement of the Gly-loop is a hydrogen bond to Thr51. This transition is facilitated by the new interactions formed by the Gly-loop to the remaining part of the protein. The latter contacts are displayed as black dotted lines.

The second, half-open conformation of the Gly-loop is found in the structures of S-methyl-piperazine substituted ligand **01** (purple), the R-methyl-piperazine substituted ligand **02** (blue) and **fasudil** (green). Interestingly, all three ligands share a common position of the Gly-loop even though the interaction pattern of the homopiperazine moiety and 2-methyl-piperazine moieties of the ligands with the protein differ significantly as will be described later.

Finally, the third, closed position is depicted in the structure of the long-chain ligand **05** (orange). A strong induced fit resulting in a pulling-down of the Gly-loop can be observed. Responsible for this rearrangement are two hydrogen bonds formed between the backbone oxygen of Thr 51 and the sulfonamide-nitrogen as well as the secondary amine in the long chain of the ligand as shown in **Figure 2C**. Only in ligand **05** is a secondary amide present in the sulfonamide position and it can hence act as a hydrogen-bond donor. All other ligands possess a tertiary amide as an equivalent, which does not have the ability to act as a hydrogen-bond donor. Consequently, the formation of the key interaction to Thr51 is impossible for all other compounds. In addition, steric hindrance would prevent the closed position of the Gly-loop for all ligands other than ligand **05**.

In the case of ligand **05**, the closed conformation of the Gly-loop is facilitated by additional interactions of the loop involving the following amino acids of the protein: Asp184, Ser53 and Lys72. In total, five new interactions are formed (**Figure 2C**, black dotted lines).

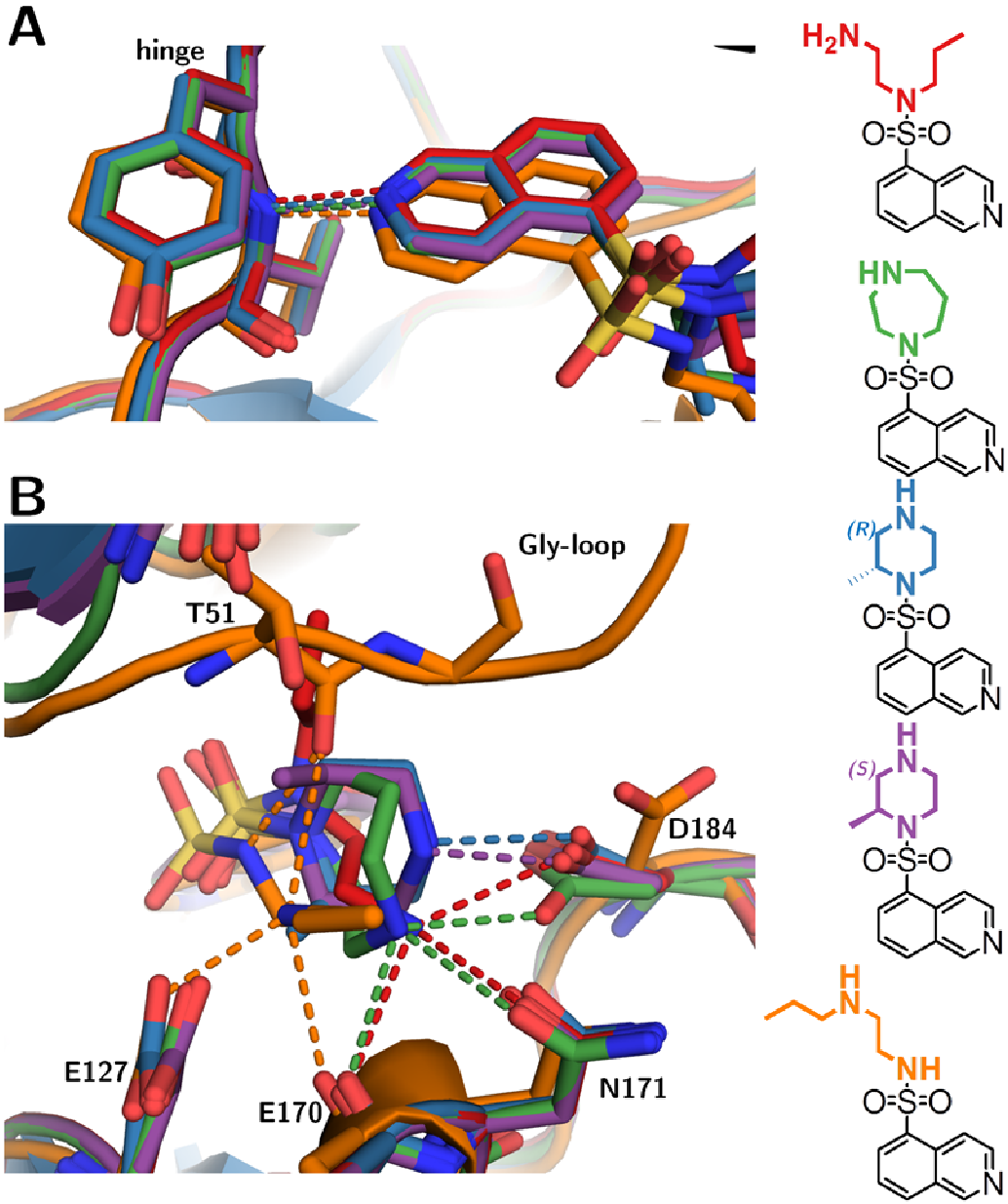


Figure 3: Polar interactions of all five ligands to the corresponding protein structure. Ligand **04** is displayed in red; **fasudil** in green; ligand **02** in blue; ligand **01** in purple and ligand **05** in orange. Polar interactions are represented as dotted lines and colored in conformity with the corresponding ligand**. A:** Interaction of the isoquinoline with the hinge region of the protein. **B:** The different polar interactions of the five ligands additional to those shown in **A.**

From further analysis of the ligand binding mode, differences can be discerned with respect to the deviating substituents (**Figure 3**). The interaction patterns established by the terminal amines differ. The highest number of polar interactions to the protein is recognized by the homopiperazine portion of **fasudil** (green) and the terminal aminoethyl moiety of the open-chain ligand **04** (red). Interestingly, both ligands form comparable interaction patterns. In either case, the terminal amino group forms hydrogen bonds with the backbone carbonyl oxygen of Glu 170 as well as the terminal carboxamide or carboxylate group of the side chains of Asn171 and Asp184. In contrast to **fasudil** (green) and **04** (red), the interaction pattern of the terminal aminoethyl nitrogen of the long-chain ligand **05** (orange) differs. The only common interaction occurs with the backbone carbonyl oxygen of Glu 170. Furthermore, **05** interacts with the side chain of Glu127.

It is also notable that the S-methyl-piperazine substituted **01** (purple) and the R-analog **02** (blue) each establish one hydrogen-bond to Asp 184 of the DFG-loop via their terminal NH group.

A difference in the position of water molecules of the residual solution pattern can be observed for the different complexes. As a matter of fact, the protein flexibility takes impact on the observability of the adjacent water positions. By optimizing the diffraction quality of the studied crystals, the resolution of the collected datasets was maximized. Nonetheless, many putative water molecule positions remained unresolved due to an ambiguous density distribution next to the region showing enhanced residual mobility in the crystal structure. A distinct analysis and quantification of the water

pattern and hence a comparison across the active sites of all complexes is limited due to the flexible nature of the protein, which effects the electron density in this region.

The most flexible ligand binds entropically most favored to the protein. The thermodynamic signature of ligand binding to the protein was determined using ITC. All profiles were assessed for putative buffer dependence. No significant protonation effects were observed in these experiments. The selected buffers differ in their ionization enthalpy by approximately 30 kJ/mol.[33] A slope between -0.07 and -0.15 could be observed for **01**, **05** and **04**, across the three considered buffers, a value considered as insignificant. For **01** and **02**, a slope of +0.15 was revealed. The calculated pKa value for the piparazine nitrogen is 7.3, which is slightly below the applied buffer pH (calculated using http://www.chemicalize.org/). Hence, there might be a slight protonation effect, owing to a partial proton uptake, in particular if compared to the slopes of the other compounds which have calculated pKa values between 8.0 and 10.1. However, also this amount of proton uptake is hardly beyond the significance threshold. Nonetheless, the profiles used for the evaluation of the thermodynamic signatures were determined in phosphate buffer to profit from the very low ionization enthalpy of this buffer. In literature values between 0.93 and 5.12 kJ/mol were reported as heat of ionization for the phosphate group[33].

To reveal a closer match with the crystallographic data we applied two distinct scenarios during the ITC titrations, both potentially important under physiological conditions. In the first set of titrations, the ligand was directly titrated to the protein in the sample cell. In the second set of titrations an

alternative strategy was used: A peptide sequence of the substrate protein to be phosphorylated was added to protein in the sample cell. In our experiment, a 20 residue peptide named PKI was applied. It is assumed that this peptide binds at the same position in the active site, adjacent to the ATP-binding site, as the substrate proteins. PKI mimics the amino acid sequence motif of the substrate proteins exhibiting as only difference the replacement of the Ser/Thr residue that becomes phosphorylated by an Ala.

Accordingly, measurements were performed in the presence (squared symbols) and absence (circular symbols) of the substrate-mimicking inhibitor peptide PKI (**Figure 4**).

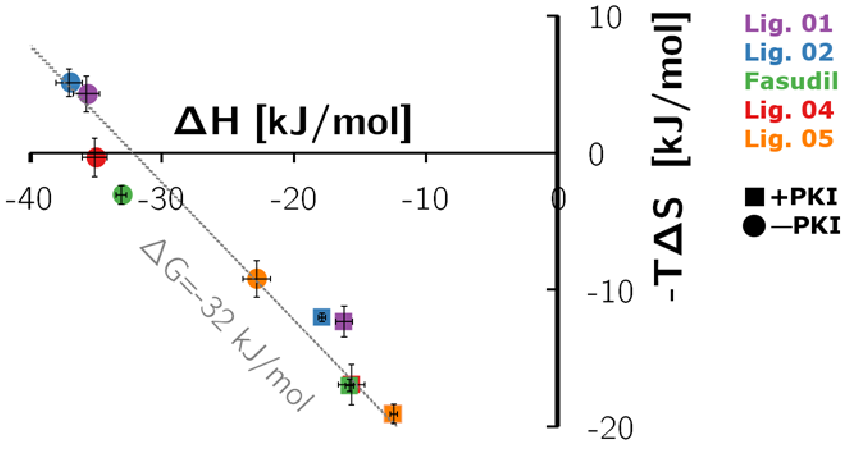


Figure 4. Thermodynamic profiles for all five ligands in absence (circles) and presence of PKI (squares). Entropy is displayed on the Y-axis and enthalpy on the X-axis. The dotted line represents isoaffinity for the mean ΔG values of 32 kJ/mol. Titrations with the different ligands are colored according to the color-code presented in **Figure 1**. The squared symbol for ligand **04** (red) is virtually hidden as the same profile is found for **fasudil** (green).

Upon comparison of the relative differences of the two sets of thermodynamic profiles it is apparent that on an absolute scale an offset between the sets of profiles is given leading to a shift toward more favorable entropy and less beneficial enthalpy in the presence of PKI (squared symbols). Thus, the relative differences between the ligands correlate in both cases and the ligands bind with equal potency.

Moreover, it is striking that the binding signatures of **fasudil** (green), the two methyl-piperazine substituted **01** (purple) and **02** (blue) as well as the open-chain ligand **04** (red) are similar and scatter maximally in $\Delta\Delta H$=4.2 kJ/mol and $-T\Delta\Delta S$=7.8 kJ/mol. In comparison, the long-chain ligand **05** (orange) displays a unique thermodynamic profile significantly reduced in its enthalpic and simultaneously enhanced in its entropic contribution to binding. In consequence, **05** was the most entropically and least enthalpically favorable binding ligand. In case of the presence of PKI, the effect concerning ligand **05** is similar, however, somewhat less pronounced.

Correlation of amide chemical shift perturbations in presence and absence of PKI. To further analyze the structural influence of the peptidic inhibitor PKI, which is present in all of the studied crystal structures, $^{1}H^{15}N$ best-TROSY spectra were measured and the amide chemical shift perturbation (CSP) of 82 amino acids was analyzed in the presence and absence of PKI. **Figure 5** shows the chemical shift perturbations from the $^{1}H^{15}N$-best-TROSYs in the absence and presence of PKI for the five different ligands. For clearity, the graph only depicts the largest shifts. These data indicate unchanged structural properties and demonstrate the relevance of the crystal structures determined in presence of PKI.

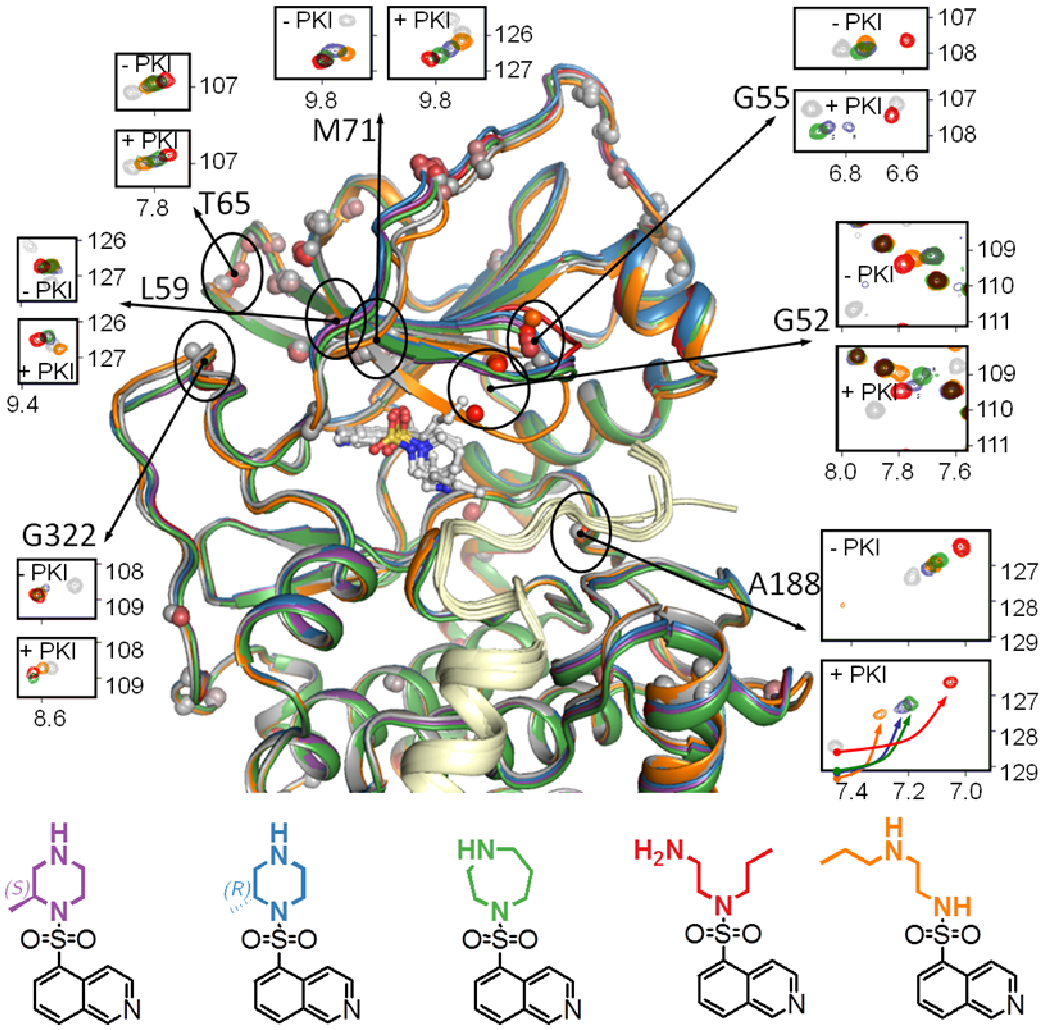


Figure 5. Amide chemical shift perturbations for the residues with the strongest shift; X-axis: ω_2 – ^{1}H (ppm); Y-axis: ω_1 – ^{15} N (ppm)

The largest spatial difference can be observed for residues Gly 55 (not assignable for the long-chain ligand **05**), Ala 188 and Arg 256. Gly 55 is located in the Gly-loop, which is involved in the strongest induced-fit adaptations among the five crystal structures. Ala 188 resides two amino acids C-terminal from the DFG-motif and is in close proximity to the PKI-peptide's C-terminus. Arg 256 is remote from the active site at the bottom of the large subunit located in close proximity to the N-terminus of the PKI

peptide. These NMR data fully support the structural information from crystallographic analysis. The residues with the smallest amide CSPs are: Arg 45, Gly 66, Val 219, Gly 225, Val 266, Glu 245, Arg 280, Asp 301, Val 310, Glu 311, Asp 328, Glu 341, Cys 343, Gly 344. The structural alignment of the crystal structure in **Figure 2** was performed using the backbone atoms of these 14 residues, thereby ensuring an authentic alignment.

The most flexible ligand does not form a flexible protein-ligand complex. In order to analyze the residual flexibility of the resulting protein-ligand complexes ^{15}N-T_2-relaxation NMR measurements were performed. Overall 75 amino acids could be assigned and are averaged in **Figure 6**. The ^{15}N-T_2-relaxation time can be used as a direct measure for backbone dynamics, as the lower the ^{15}N-T_2-relaxation time is determined, the more rigid the protein backbone will be. We assume that the ^{15}N-T_2-relaxation times are representative for the dynamic properties of the entire amino acid residues. On average, the dynamics of all five complexes are very similar. It is striking that ligand **05** with the largest amount of internal degrees of freedom does not form a significantly more flexible protein-ligand complex compared to the other more constraint members of the series.

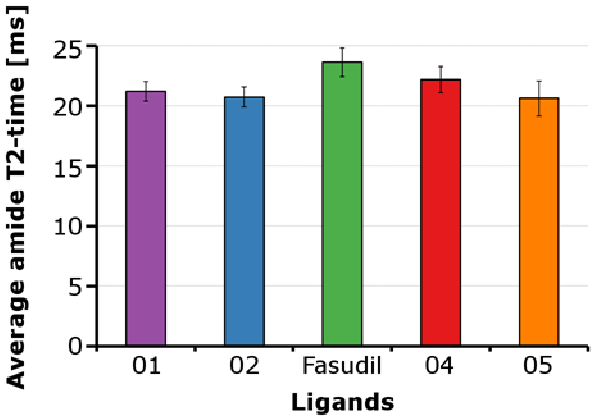


Figure 6. Amide ^{15}N-T2-relaxation time in ms averaged over 75 amino acids for the different ligands, respectively. Overall differences are small. Error bars present the averaged standard deviation of the different amino acids.

2.6 Discussion

For a series of five congeneric ligands studied in this contribution we thoroughly analyzed the interaction with PKA. High-resolution crystallographic data revealed a comparable hinge binding mode of all ligands. As special feature a strong induced-fit mechanism for ligand **05** (orange) is observed. Thermodynamic signatures were measured either in the absence or in the presence of the substrate mimicking peptide PKI. They uncover that ligand **05** was the most entropically favored binder. NMR data unraveled the structural influence of PKI on the protein and allowed analysis of the backbone dynamics.

On first sight, it appears counterintuitive that the long-chain ligand **05** (orange) exhibiting the largest amount of internal degrees of freedom, was the most entropically favored binder in the presence as well as absence of PKI. At first glance, this appears contradictory as the binding of more rigid

and conformationally preorganized ligands reveals an entropic advantage to binding.[34, 35] In the present case, this would be expected for the ligands **fasudil**, **01** and **02** possessing the conformationally restricted cyclic substituents. The in contrast observed entropically more favored binding of ligand **05** to the protein can therefore result from multiple factors: Either from a protein-ligand complex with strongly enhanced residual flexibility, from the displacement of a larger number of water molecules from the active site upon ligand binding to the uncomplexed protein, or a difference in the solvation properties among the five ligands in aqueous solution prior to binding. This contribution even overcompensates the entropic losses to be paid by restricting the flexible ligand **05** to the bound state. The results from ^{15}N-T_2-relaxation NMR measurements suggest however, that **05** forms the most rigid protein-ligand complex (**Figure 6**). This is in accordance with the crystallographic data indicating a strong induced-fit adaptation pulling the Gly-loop toward the bound state. Furthermore, this ligand shows the lowest overall B-factors in the crystal structure when compared to the other ligands (supporting information). Clearly this does not indicate an entropic advantage of **05** upon binding. One possible explanation for the favorable entropy contribution of ligand **05** can originate from displacement of previously well-ordered water molecules from the formed protein-ligand complex. Unfortunately, the amount of water molecules in the active site cannot be reliably evaluated due to a less-well defined electron density in that area of the crystal structures in consequence of enhanced loop flexibility. It is remarkable that the relative differences in the thermodynamic profiles between the five ligands remain similar with and without the substrate-like PKI. However, the overall shift on absolute scale appears reasonable

assuming that the protein is pre-stabilized and structurally organized in a way to better recognize either the co-substrate ATP or, in our case, the inhibitors of the fasudil-type upon binding of the substrate peptide. The thermodynamic profiles of ligand binding are all shifted toward an entropically more favorable but enthalpically less beneficial signature in the presence of PKI. This profile is in accordance with a better preorganization of one of the binding partner, here of the recipient protein.

An alternative or additional contribution influencing the thermodynamic signature of ligand **05** compared to the other members of the series can arise from differences of the ligands in aqueous solution prior to binding. Once the ligands are released from the bulk water phase and accommodate the protein, they have to shed their solvation shells. If these shells show structural differences in the local solvation pattern, deviating thermodynamic signatures will result. In the present case, **05** would assemble a higher ordered local water structure which upon collapse will produce an entropic advantage.

In conclusion, the experimental data demonstrate that the entropically more favored binding of **05** to the protein can only result from an overcompensating displacement of a larger amount of ordered water molecules from the active site upon ligand binding or an entropic benefit of the ligand while leaving the solvation shell in the bulk water phase. Our observation that a ligand with a larger amount of internal degrees of freedom does not necessarily lead to entropically less-favored binding was also observed in other cases, however, without providing a conclusive explanation.[36] Therefore, the hypothesis stating a significant impact of the ligand's degrees of freedom on the entropic binding component, needs to be

questioned and clearly requires a more detailed analysis of the binding event as a whole. It demonstrates that simple design guidelines cannot be generalized and deserve a more careful consideration.

2.7 Conclusion

General and hypothetical design guidelines need to be systematically investigated for different protein classes. Rules that apply for small and rigid proteins do not necessarily apply for large or flexible proteins. From our systematic study on a protein kinase we conclude that the thermodynamic profile is not dominated by the loss of rotational degrees of freedom of the ligand. The hypothesis, that ligand rigidification necessarily results in significantly more entropic ligand binding profiles could not be confirmed. Strikingly, we even recorded the opposite trend in that the most flexible ligand was the most favorable entropic binder. Therefore, the loss of degrees of freedom of the ligand, which, as a matter of fact, has to be paid, does not necessarily dominate the thermodynamic profile of a flexible protein like a kinase. Furthermore, the effect of residual flexibility of the formed protein-ligand complex appears to be minor in the studied case as all complexes under consideration show very similar NMR data. In fact, the dominant effect seems to be the displacement of ordered water molecules from the active site particularly in the complex where the Gly-loop folds upon the bound ligand or from structural differences in the solvation pattern of the ligands in the bulk water phase. These appear to be the crucial factors for the resulting entropically favored binding. In summary, our results demonstrate that protein-ligand interactions formed with flexible proteins are even more complex than with rigid proteins. Our data signalize that global

understanding and reliable predictions of thermodynamic profiles need a view on the entire binding event and require many more systematic studies of congeneric ligand series.

2.8 Abbreviations

Ala: alanine

Asp: aspartate

Arg: arginine

ATP: adenosine triphosphate

CHO: Chinese hamster ovary

CSP: chemical shift perturbation

DFG: aspartat-phenylalanin-glycin motif

DMSO: dimethyl sulfoxide

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

Gly: glycine

Gly-loop: glycine-rich loop

Glu: glutamate

His-tag: histidine-tag

ITC: isothermal titration calorimetry

MBT: MES/Bis-Tris

MPD: 2-methyl-2,4-pentanediol

Ni-NTA: nickel-nitrilotriacetic acid

NMR: nuclear magnetic resonance spectroscopy

PDB: protein data bank

PKA: cAMP-dependent protein kinase

PKI: protein kinase inhibitor

qNMR: quantitative nuclear magnetic resonance spectroscopy

Ser: serine

TEA: triethanolamine

TEV: tobacco etch virus

Thr: threonine

TROSY: transverse relaxation optimized spectroscopy

Val: valine

2.9 References

1 Freire, E. *Drug Discov Today* **2008**, *13*, 869–874.

2 Ladbury, J. E.; Klebe, G.; Freire, E. *Nat Rev Drug Discov* **2010**, *9*, 23–27.

3 Kawasaki, Y.; Freire, E. *Drug Discov Today* **2011**, *16*, 985–990.

4 Núñez, S.; Venhorst, J.; Kruse, C. G. *Drug Discov Today* **2012**, *17*, 10–22.

5 Klebe, G. *Nat Rev Drug Discov* **2015**, *14*, 95–110.

6 Ákos Tarcsay,; Keserü, G. M. *Drug Discov Today* **2015**, *20*, 86–94.

7 Mann, A. In *The Practice of Medicinal Chemistry*, second edition ed.; Wermuth, C. G., Ed.; Academic Press: London, 2003; 233 – 250.

8 Loughlin, W. A.; Tyndall, J. D. A.; Glenn, M. P.; Fairlie, D. P. *Chem Rev* **2004**, *104*, 6085–6117.

9 Chang, C. A.; Chen, W.; Gilson, M. K. *Proc Natl Acad Sci U S A* **2007**, *104*, 1534–1539.

10 Chaires, J. B. *Annu Rev Biophys* **2008**, *37*, 135–151.

11 Wong, C. F. *Biochim Biophys Acta* **2008**, *1784*, 244–251.

12 Huang, Z.; Wong, C. F. *J Comput Chem* **2009**, *30*, 631–644.

13 Kornev, A. P.; Taylor, S. S. *Trends Biochem Sci* **2015**, *40*, 628–647.

14 Olsson, T. S. G.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. *J Mol Biol* **2008**, *384*, 1002–1017.

15 Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. *Nucleic Acids Res* **2016**, *44*, D1045–D1053.

16 Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912–1934.

17 Shibuya, M.; Asano, T.; Sasaki, Y. In *Cerebral Vasospasm*; Seiler, R. W., Steiger, H.-J., Eds.; Springer Vienna: Vienna, 2001; 201–204.

18 Sasaki, Y.; Suzuki, M.; Hidaka, H. *Pharmacol Ther* **2002**, *93*, 225–232.

19 Breitenlechner, C.; Gassel, M.; Hidaka, H.; Kinzel, V.; Huber, R.; Engh, R. A.; Bossemeyer, D. *Structure* **2003**, *11*, 1595–1607.

20 Kudlinzki, D.; Linhard, V. L.; Saxena, K.; Sreeramulu, S.; Gande, S.; Schieborr, U.; Dreyer, M.; Schwalbe, H. *Acta Crystallogr F Struct Biol Commun* **2015**, *71*, 1088–1093.

21 Kabsch, W. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132.

22 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658–674.

23 Adams, P. D. et al. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–221.

24 Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486–501.

25 Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J Appl Crystallogr* **1993**, *26*, 283–291.

26 Kleywegt, G. J.; Zou, J.-Y.; Kjeldgaard, M.; Jones, T. A. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; Springer Netherlands: Dordrecht, 2001; 353–356.

27 Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. http://www.globalphasing.com.

28 Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. *J Chem Inf Comput Sci* **2004**, *44*, 2133–2144.

29 Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. *Anal Chem* **2012**, *84*, 5066–5073.

30 Houtman, J. C. D.; Brown, P. H.; Bowden, B.; Yamaguchi, H.; Appella, E.; Samelson, L. E.; Schuck, P. *Protein Sci* **2007**, *16*, 30–42.

31 Morikawa, A.; Sone, T.; Asano, T. *J Med Chem* **1989**, *32*, 46–50.

32 Makhija, M. T.; Kasliwal, R. T.; Kulkarni, V. M.; Neamati, N. *Bioorg Med Chem* **2004**, *12*, 2317–2333.

33 Goldberg, R. N.; Kishore, N.; Lennen, R. M. *J Phys Chem Ref Data* **2002**, *31*, 231.

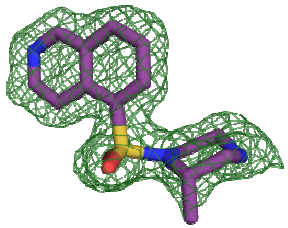
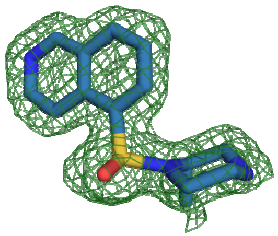
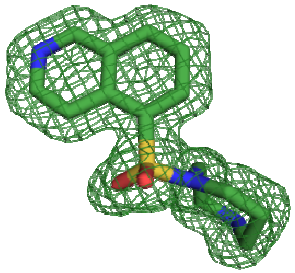
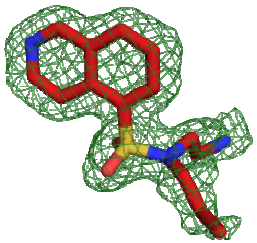
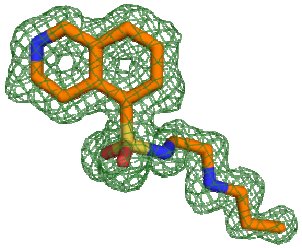
34 Neeb, M.; Hohn, C.; Ehrmann, F. R.; Härtsch, A.; Heine, A.; Diederich, F.; Klebe, G. *Bioorg Med Chem* **2016**, *24*, 4900–4910.

35 Rühmann, E.; Rupp, M.; Heine, A.; Klebe, G. *ChemMedChem* **2016**, *11*, 309–319.

36 Martin, S. F.; Clements, J. H. *Annu Rev Biochem* **2013**, *82*, 267–293.

2.10 Supplementary information

Table S1. All atoms of all ligands were 100% occupied. For visualiazation, σ-weighted mF_o-DF_c-densities for all five ligands are shown.

Ligand 1, at 3.0 σ 5LCU	Ligand 2, at 2.8 σ[a)] 5LCT	Fasudil (Ligand 3), at 3.0 σ 5LCP
		
Ligand 4, at 3.0 σ 5LCR	**Ligand 5, at 3.0 σ 5LCQ**	
		

a) Density map shown at 2.8 σ in order to provide full visibility oft the ligand.

Table S2. Crystallographic table for all crystal structures. Table spreads over two pages

Ligand, PDB entry →	**1, 5LCU**	**2, 5LCT**	**3, 5LCP**	**4, 5LCR**	**5, 5LCQ**
Data collection & processing					
No. Crystals used	1	1	1	1	1
Wavelength [Å]	0.91841	0.91841	0.91841	0.91841	0.91841
Space group	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)
Unit cell parameters: a, b, c [Å];	58.5, 73.0, 109.4	59.0, 73.7, 109.8	58.4 72.8, 109.3	58.3, 72.8, 108.7	57.9, 71.6, 109.7
Diffraction data[a)]					
Resolution range [Å]	50.00-1.58	50.00-1.62	50.00-1.43	50.00-1.56	50.00-1.42
Highest shell resolution range [Å]	1.68-1.58	1.71-1.62	1.52-1.43	1.66-1.56	1.51-1.42
Unique reflections	61633 (9566)	61770 (9549)	84420 (13201)	64829 (10151)	84798 (13274)
R(I)sym [%][b)]	5.8 (47.9)	4.8 (48.3)	4.1 (46.6)	6.8 (49.1)	5.7 (47.4)
Completeness [%]	94.6 (92.2)	99.4 (96.4)	97.9 (95.8)	98.3 (96.4)	98.4 (96.5)
Redundancy	3.4 (3.2)	6.5 (6.2)	4.0 (4.0)	4.0 (4.0)	4.0 (4.1)
I/σ (I)	12.4 (2.1)	24.0 (3.5)	17.1 (2.5)	11.2 (2.1)	15.3 (2.7)
Refinement					
Resolution range [Å]	43.77-1.58	46.1-1.62	39.90-1.43	39.75-1.56	34.04-1.42
Reflections used in refinement (work/free)	58551/3082	58679/3091	80197/4223	61585/3244	80556/4242
Final R values for all reflections (work[c)]/free[d)]) [%]	15.2/19.2	14.3/17.4	13.3/16.6	15.5/19.7	14.1/17.1

Amino acids (PKA/PKI)	353/20	353/20	353/20	353/19	353/18
Inhibitor atoms	20	20	20	20	20
Water molecules	417	412	448	392	551
RMSD from ideality					
Bond length [Å]	0.009	0.009	0.008	0.008	0.008
Bond angles [°]	0.95	0.96	0.94	0.98	0.96
Ramachandran plot[e)]					
Residues in favoured regions	91.5	92.7	92.4	93.0	92.1
Residues in additionally allowed regions [%]	8.5	7.3	7.6	7.0	7.6
Residues in generously allowed regions [%]	0.0	0.0	0.0	0.0	0.3
Mean B-factors [Å²]					
PKA (protein)/PKI (peptide)	23.8/24.8	25.4/26.9	24.3/23.8	25.3/23.6	16.6/14.9
Inhibitor	18.4	22.7	19.2	21.7	12.1
Water molecules	31.8	33.1	37.0	32.4	28.8

a)Numbers in parentheses are for the highest-resolution shell.

b)R_{sym}=$[\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}|I_{\mathrm{i}}(h)-\ \langle I(h)\rangle\ |/\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}I_{\mathrm{i}}(h)] \times 100$, $\langle I(h)\rangle$ is the mean of the $I(h)$ observation of reflection h.

c)R_{work}=$\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}-F_{\mathrm{c}}|/\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}$.

d)Calculation of R_{free} was performed as for R_{work} but on 5 % of the data which was excluded from the refinement.

e)Derived from Procheck[25]

Table S3. Thermodynamic data for all ligands including the respective standard deviations (titration without PKI).

Ligand	ΔG [kJ/mol]	ΔH [kJ/mol]	-TΔS [kJ/mol]
1	-31.3±0.3	-35.7±1.0	4.3±1.3
2	-31.9±0.1	-37.0±0.2	5.1±0.3
Fasudil	-36.1±0.3	-33.1±0.4	-3.1±0.7
4	-35.4±0.5	-35.1±0.9	-0.3±1.4
5	-32.0±0.3	-22.8±1.0	-9.2±1.3

Table S4. Thermodynamic data for all ligands including the respective standard deviations (titration with PKI).

Ligand	ΔG [kJ/mol]	ΔH [kJ/mol]	-TΔS [kJ/mol]
1	-28.5±0.0	-16.2±0.3	-12.3±0.3
2	-29.9±0.5	-17.9±0.6	-12.0±1.1
Fasudil	-32.8±0.1	-15.8±0.3	-17.0±0.4
4	-32.6±0.5	-15.7±1.0	-16.9±1.5
5	-31.5±0.4	-12.5±0.3	-19.1±0.7

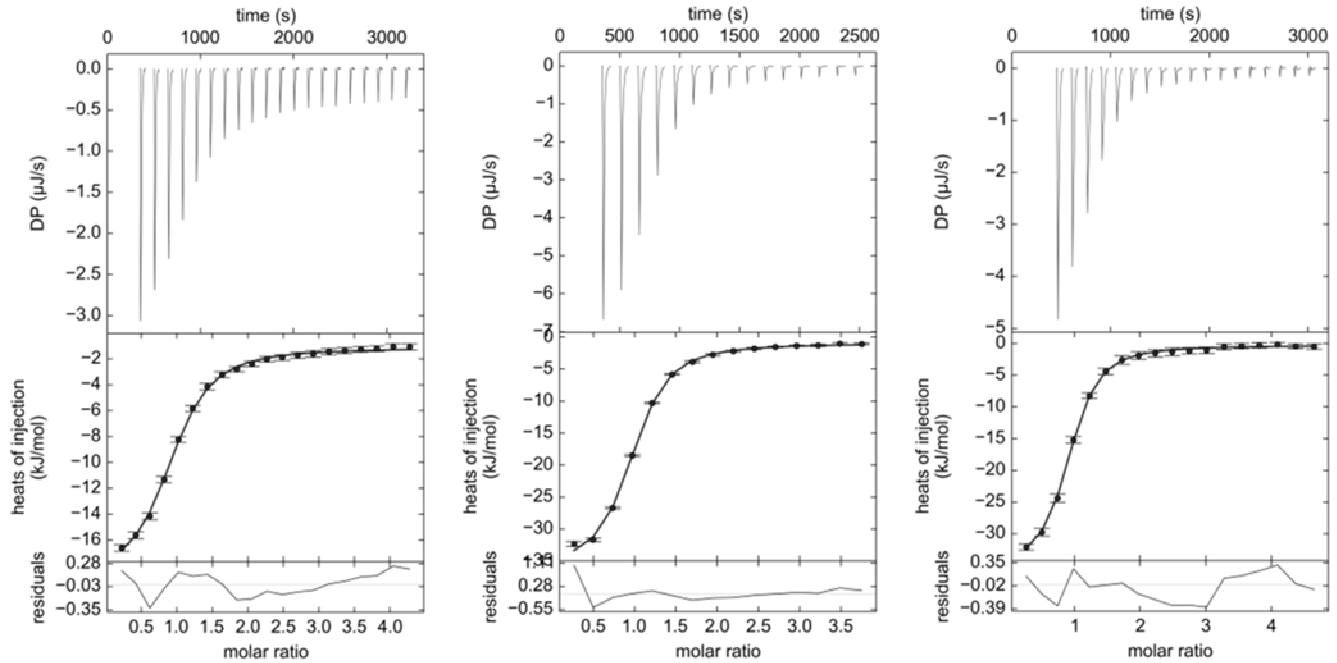


Figure S1. ITC-titration curves for ligand 01 without PKI in phosphate buffer as exemplary titration curves.

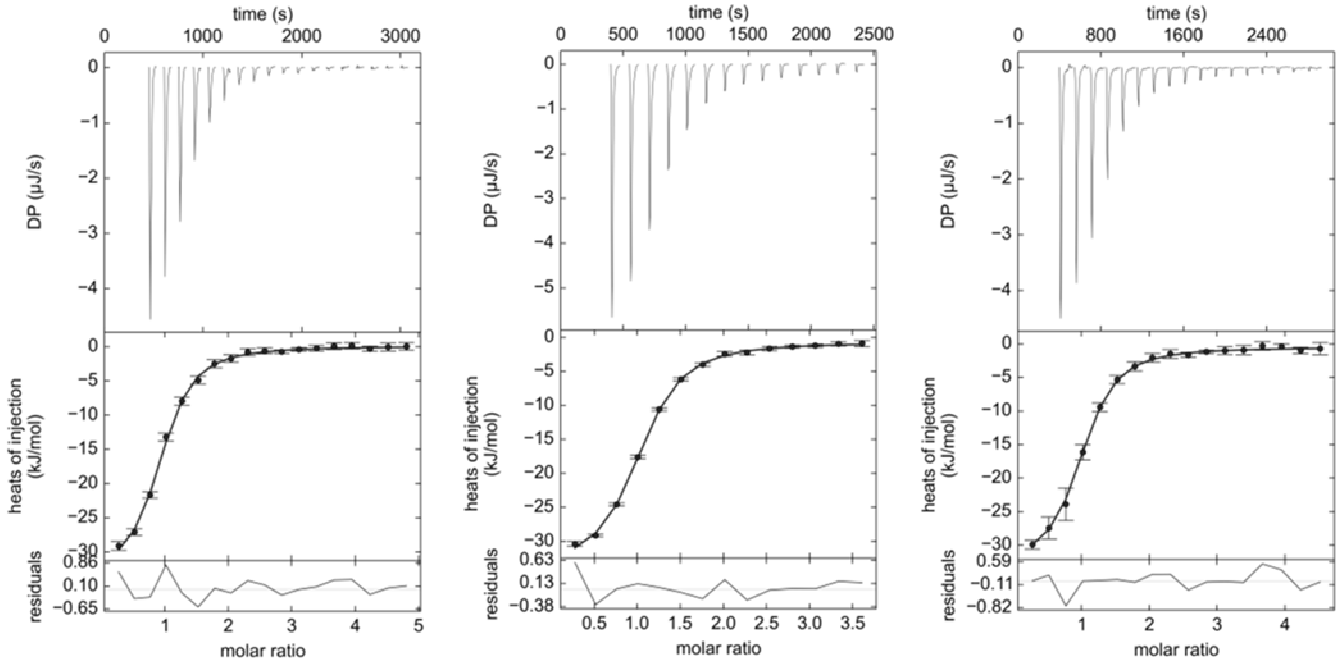


Figure S2. ITC-titration curves for ligand 02 without PKI in phosphate buffer as exemplary titration curves.

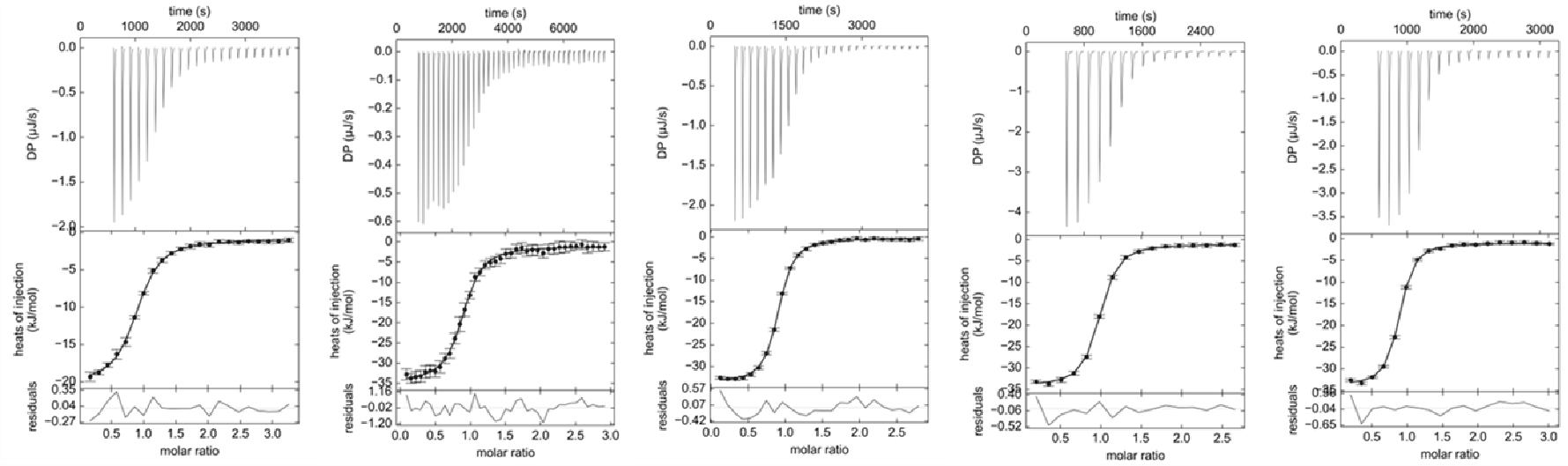


Figure S3. ITC-titration curves for Fasudil without PKI in phosphate buffer as exemplary titration curves.

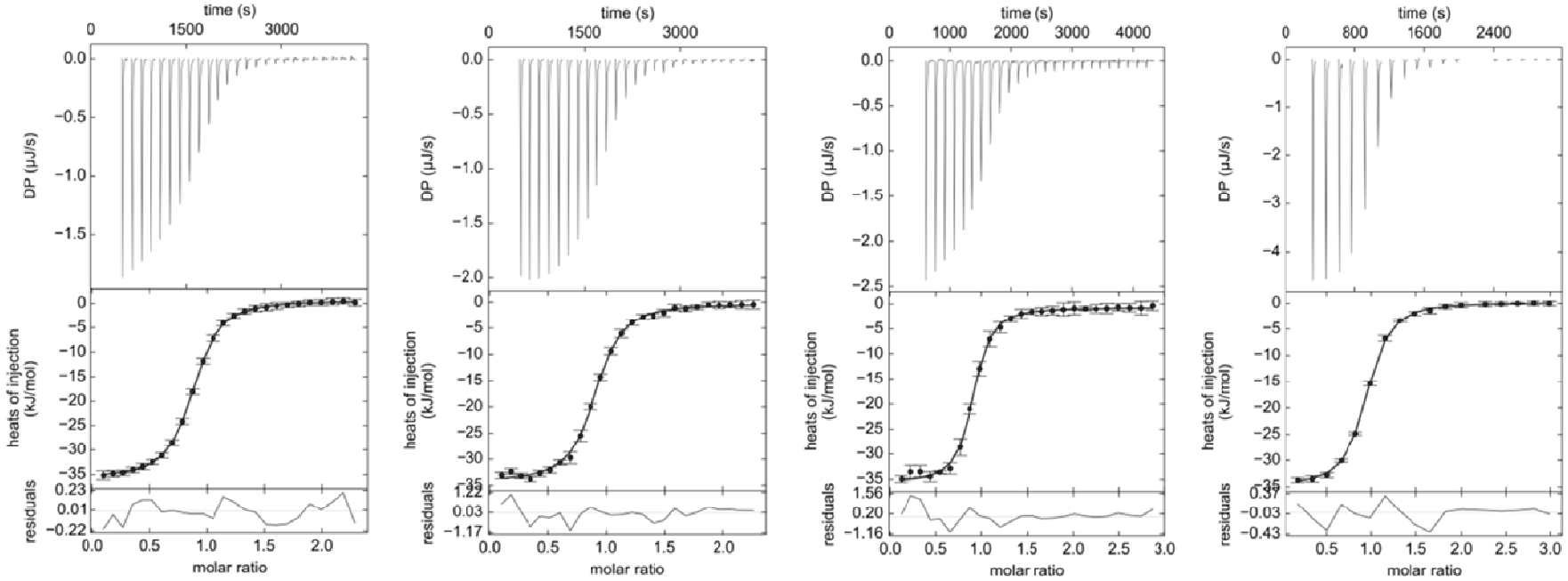


Figure S4. ITC-titration curves for ligand 04 without PKI in phosphate buffer as exemplary titration curves.

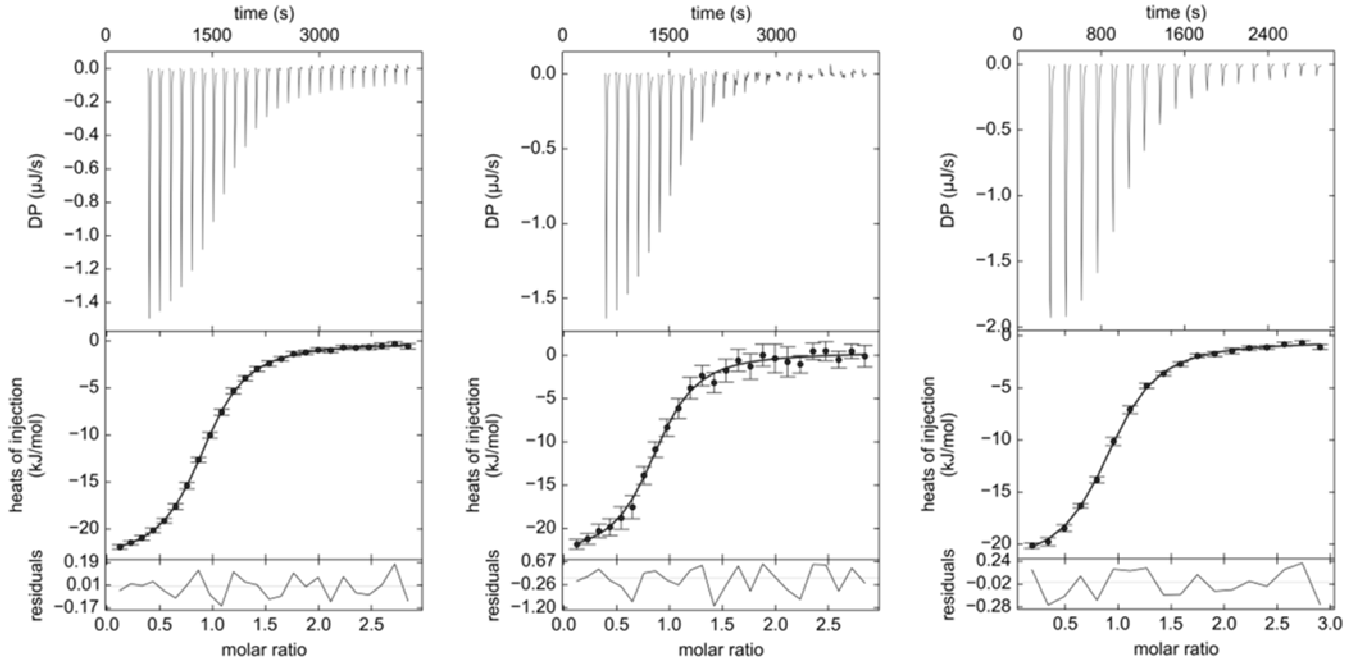


Figure S5. ITC-titration curves for ligand 05 without PKI in phosphate buffer as exemplary titration curves.

3 Enthalpy is Losing it: Stepwise Ligand Growth and its Influence on Kinase-Ligand Interaction

3.1 Annotations

The following chapter will be published. It is a cooperation of the groups of Prof. Gerhard Klebe and Prof. Harald Schwalbe. The following people contributed to the paper and will therefore be listed as authors on the publication:

Barbara Wienen-Schmidt[‡], Hendrik R. A. Jonker[†], Tobias Wulsdorf[‡], Hans-Dieter Gerber[‡], Krishna Saxena[†], Andreas Heine[‡], Harald Schwalbe[†]*, Gerhard Klebe[‡]*

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

[†]Institut für Organische Chemie und Chemische Biologie, Johann Wolfgang Goethe-Universität Frankfurt, Max-von-Laue-Straße 7, N160-3.14, 60438 Frankfurt am Main, Germany

* Corresponding author

Protein expression for ITC and crystallization was performed by Barbara Wienen-Schmidt as well as ITC-measurements, crystallization and crystallography. Computational analyses were performed by Tobias

Wulsdorf. Ligand qNMR analyses were performed by Hans-Dieter Gerber. Henry Jonker performed amide chemical shift perturbations analysis and amide ^{15}N-T2-relaxation measurements and related protein expression.

3.2 Abstract

Here we present a unique fragment study focusing on the thermodynamics of fragment growth. Four ligands with increasing molecular weight and their interaction with cAMP-dependent protein kinase (PKA) were thoroughly analyzed. Strikingly, these ligands display affinities between nanomolar and low micromolar potency on PKA despite their low molecular weight. This allows a direct measurement of the binding thermodynamics via isothermal titration calorimetry (ITC). A direct measurement can rarely be reliably performed for fragments. Therefore, the presented data provide a valuable insight into fragment thermodynamics. Furthermore, protein-ligand complexes were structurally analyzed using crystallography and NMR in order to explain the observed changes in thermodynamics. The clear trend towards more entropic and less enthalpic binding upon increasing molecular weight results from structural changes as well as differences of the uncomplexed ligands in solution. The latter being a factor that has been utterly underestimated.

3.3 Introduction

Fragment based drug design (FBDD) has been applied with increasing popularity for the development of new drugs in academia and the pharmaceutical industry. It is often considered as an improved approach compared to the extremely expansive high-throughput screening (HTS) in order to generate first hits for a subsequent individually tailored ligand

optimization.[1-4] Two frequently used strategies in FBDD are either fragment linking or fragment growing. In this contribution we present a systematic analysis of ligand properties following the process of ligand growing. Accordingly, four different ligands with stepwise increasing molecular weight are studied in detail in terms of their binding behavior (**Figure 1**). The cAMP-dependent protein kinase (PKA) is used as a model protein for our studies being a representative of the large clinically relevant class of protein kinases.

Binding mode analyses of the ligands as well as putative changes of the protein conformation were assessed using either X-ray crystallography or nuclear magnetic resonance spectroscopy (NMR). Moreover, thermodynamic analysis was performed providing detailed information about the stepwise changes of the thermodynamic signatures during the gradual process of fragment growing. Thermodynamic data were recorded using isothermal titration calorimetry (ITC) applying a protocol of direct titrations under the condition of c-values between 5 and 70. Such values indicate that a crude sigmoid curve can still be recognized.

The combined structural and thermodynamic data for fragment based approaches are still rare, but collection of such information is essential to improve our understanding of the ligand binding process.[5]

The data presented in this study demonstrate a classical example of partial enthalpy-entropy compensation. Structurally observable changes in ligand binding modes as well as surprising differences in the properties of the ligands in solution prior to protein binding appear to be responsible for the observed differences of the thermodynamic signatures upon ligand binding. These

factors are crucial in fragment growing. They have to be considered as determining factors in the design process upon growing of a ligand into a protein binding site.

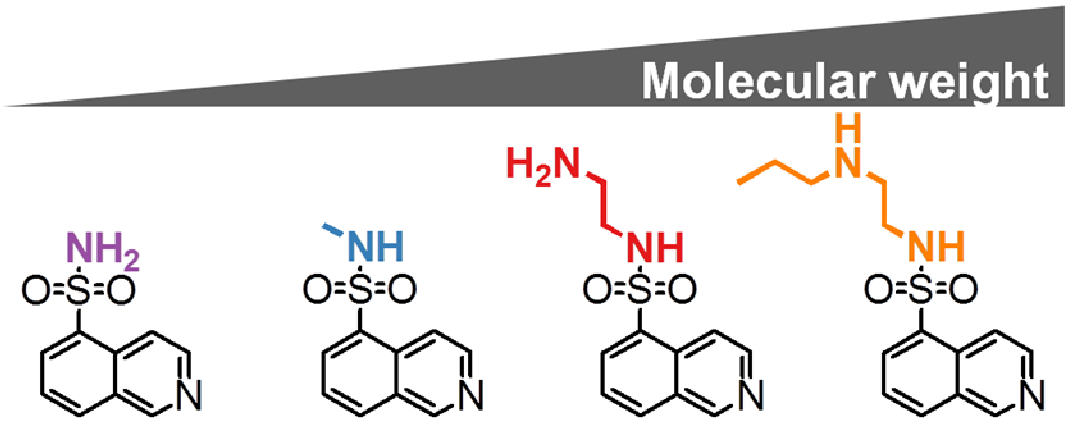


Figure 1. Chemical structure of the four ligands used in this study. Molecular weight of the ligands increases from left to right. In purple ligand **1**: isoquinoline-5-sulfonamide; in blue ligand **2**: the methylated N-methylisoquinoline-5-sulfonamide; in red ligand **3**: short chained N-(2-aminoethyl)isoquinoline-5-sulfonamide; in orange ligand **4**: long-chained N-[2-(propylamino)ethyl]isoquinoline-5-sulfonamide

It has been hypothesized that increasingly larger ligands in contrast to fragments favor a more entropic binding. Thus, the enthalpic component of fragment binding would not increase with ligand size.[6] There are several reasons to explain this observation that enthalpy dominated binding of fragments turns into a more entropy favored binding with growing ligand size. Discussed is for instance the impact of apolar desolvation that increases with ligand size due to a reduced probability for an optimal geometrical H-bond fit and an increased entropic penalty resulting from a mobility that decreases with the number of H-bonds between ligand and protein.[6] Furthermore, larger ligands do not only bind to protein hot-spots but also

beyond. These interactions outside protein hot-spots do not display as strong H-bond conservation as the hot-spots themselves. In fact, these interactions might play a role for the enthalpic component and binding selectivity.[6] Strikingly, even under consideration of the above discussed factors through size-independent ligand analysis methods such as the size-independent enthalpic efficiency-value (SIHE), the gain in enthalpy upon ligand growth remains noteworthy for the ligand set presented here.

3.4 Experimental section

Protein expression and purification

The catalytic subunit of cAMP-dependent protein kinase from Chinese hamster ovary cells was expressed with a His-tag in a modified pET16b-Vector with an introduced TEV-cleavage site between the protein N-terminus and His-tag. This plasmid was transformed into *E. coli* strain *Bl21 (DE3)/pLysS* (Novagen).[7]

Cell disruption was performed using a high-pressure homogenizer for multiple cycles. After centrifugation (1h at 30.000g) cell lysate supernatant was purified in a first step using a Ni-NTA column that binds the His-tag of the protein and was eluted by an imidazole gradient. The His-tag was then cleaved off by TEV-protease. Afterwards, an inverse Ni-NTA column was employed collecting PKA in the flow-through. Finally, ion exchange chromatography was performed using a MonoS column separating three-fold phosphorylated PKA from the four-fold phosphorylated form using a HEPES buffer with a sodium chloride gradient.[7]

Crystallization

Crystallization of **4** is discussed in the previous chapter. Co-crystallization was performed using the hanging drop method at 4 °C. The crystallization drops contained the following ingredients: 10 mg/mL PKA (240 µM), 30 mM MBT (MES/Bis-Tris Puffer pH 6.9), 1 mM DTT, 0.1 mM EDTA, 75 mM LiCl, 0.03 mM Mega 8, 0.07 mM PKI (Sigma: P7739), 1.2 mM ligand (120 µM for ligand 1) dissolved in DMSO from a 50-100 mM stock. The well contained a mixture of methanol in water with varying methanol concentrations (v/v) for the different ligands (**01**: 18% methanol; **02**: 18% methanol; **03**: 14% methanol). In the crystallization setup streak-seeding was performed with apo-crystals as seeds using a horse hair in order to initialize crystal growth. For crystal mounting, crystals were cryo protected in 5 mM MBT (MES/Bis-Tris Puffer pH 6.9), 1 mM DTT, 0.1 mM LiCl, 1.2 mM ligand (ligands 2 and 3) or 120 µM (ligand 1) dissolved in DMSO from a 50-100 mM stock, 16 % (v/v) methanol, 30% (v/v) MPD and flash frozen in liquid nitrogen.

Crystallography

Structures of **1** and **2** were collected at the storage ring Bessy II Helmholtz-Zentrum Berlin, Germany at Beamline 14.1 on a Pilatus 6M pixel detector. Structure of **3** was collected at the European Synchrotron Radiation Facility (ESRF) Grenoble, France at Beamline 14. Structure of **4** is discussed in the previous chapter.

The datasets were processed using XDS[8] and molecular replacement was performed using CCP4 Phaser[9] and PDB-structure of PKA from *bos taurus* 1Q8W as a model. This was followed by simulated annealing, multiple

refinement cycles of maximum likelihood energy minimization and B-factor refinement using Phenix[10]. Coot[11] was used to fit amino-acid side chains into σ-weighted 2Fo – Fc and Fo – Fc electron density maps. If appropriate electron density was observed, multiple side chain conformations were built into the model and maintained during the refinement if the minor populated side chain displayed at least 20 % occupancy. Ramachandran plots for structure validation were calculated using PROCHECK[12]. Data collection, unit cell parameters and refinement statistics are given in the supplementary information. Analysis of temperature factors was performed with Moleman[13]. Protein and PKI B-factors were anisotropically refined, water B-factors were isotropically refined for structure of **3**. Structures of **1** and **2** were TLS-refined. For the definition of the TLS groups the TLSMD server was used.[14, 15] Decision for anisotropic or TLS refinement was based on comparison of R_{free}. Anisotropic refinement was chosen over TLS if the achieved R_{free} values were at least 0.5% lower for anisotropic than for TLS refinement. R_{free} was calculated using 5% of all reflections which were randomly chosen and not used for the refinement. The required ligand restraint files were created using the Grade webserver[16, 17]. For figure preparation Pymol was used. Crystallographic tables as well as mFo-DFc-densities of the different ligands can be found in the supporting information.For ligand 2 a second, allosteric binding site was found which was also 100% occupied.

Calculation of polar and hydrophobic surfaces was done using the programme Molecular Surface.[18-20]

Isothermal titration calorimetry

The buffer used for the ITC experiments contained: 30 mM sodium phosphate buffer pH 7.2, 10 mM $MgCl_2$, 100 mM NaCl, 3% (v/v) DMSO. All measurements were repeated 4-7 times. Further buffers were used in order to check for protonation linkage. In these buffers 30 mM sodium phosphate buffer was replaced by 30 mM HEPES and 30 mM triethanolamine (TEA), respectively (both at pH 7.2). Buffer dependency was tested for all ligands and no significant buffer dependence could be detected. For the measurements expressed, purified and dialyzed PKA was used in the ITC-measuring cell. A 15-20 fold higher concentrated ligand solution, diluted in dialysis buffer, was then stepwise injected to the protein solution during the measurement. All measurements were performed at 25 °C. ITC data were analyzed using NITPIC and Sedphat[21, 22]. Raw data can be found in the supplementary information.

Compound purity was analyzed using quantitative nuclear magnetic resonance spectroscopy (qNMR) and in case of deviation, ligand concentration was corrected accordingly.[23]

Ligands

Ligands **1**, **2** and **4** were purchased from Uorsy (Ukraine). Ligand **3** was purchased from Toronto Research Chemicals (Canada).

NMR

All protein NMR measurements were performed by the group of Prof. Harald Schwalbe in Frankfurt/Germany.

Computational analysis

All atom MD (molecular dynamics) simulations were carried out for each compound. Each conformational ensemble was clustered into three clusters. Subsequently, MD simulations and GIST (grid inhomogeneous solvation theory)[24] analysis were carried out in order to study the thermodynamics of water molecules bound on the surface of the ligand molecule in solution.

MD simulation

Compounds 1-3 were extracted from their pdb files with fconv[25] and protonated with MOE (Molecular Operating Environment, Chemical Computing Group Inc., Montreal, Canada). Partial atomic charges were derived using multiconformational RESP fitting[26] and Gaussian09[27] on the HF/6-31+G level of theory. The programs antechamber and parmchk2[28] were used for assignment of gaff atom types and missing force field parameters. The compounds were solvated in a box of TIP4P-Ew water using tleap.

All following MD steps were carried out using Amber14 and AmberTools14 with periodic boundary conditions. Initially, each system was minimized with 1000 steps of steepest descend followed by 5000 steps of conjugate gradient keeping the solute heavy atoms fixed with a 25 kcal/mol/Å^2 force constant followed by second minimization similar to the first one, but with force constant decreased to 5 kcal/mol/Å^2. After heating the system to 300 K the system was equilibrated while lowering the force constant. A final 5 ns NVT equilibration was carried out before starting production MD. All production MDs were run for 100 ns at 300 K in NVT ensemble using Langevin thermostat with a collision frequency of 1.0 ps-1.

Clustering

The trajectories of the unbiased MD simulations were clustered into three distinct families using hierarchical agglomerative average-linkage clustering as implemented in cpptraj(V15.00)[24].

Restraint MD simulation

Each cluster centroid was used as starting structure for further MD simulations. All preparation steps were similar to the former MD simulations, except that counter ions were placed randomly in 10Å distance to the compound.

The minimization is similar to the one described before. During all following steps solute heavy atoms were constraint using a harmonic force constant of 25 kcal/mol/Å^2. After heating to 300 K, equilibration in NPT (constant particles, pressure and temperature) was carried out for 5 ns followed by 5 ns of NVT (constant particles, volume and temperature) equilibration. Productive MD runs were carried for 50 ns.

GIST analysis

The trajectories obtained from the restraint MD simulation were analyzed with GIST as implemented in cpptraj.

3.5 Results & discussion

Higher molecular weight correlates with increasing entropy and decreasing enthalpy. As pointed out, the thermodynamic data were recorded through direct ITC titrations with reasonable c-values. A study under such conditions is indeed rarely possible for fragments owing to their

low binding affinity combinded with a minor heat signal indicating binding. In the current ligand series measurements provided reliable information on the binding thermodynamics. The thermodynamic profiles of the four different ligands binding to PKA are presented in **Figure 2** and **Table 1**. A clear trend can be observed where entropy of binding increases with the molecular weight whereas enthalpy shows the reverse trend. The entropic contribution to the binding event is clearly increasingly unfavourable for **1** and **2** and becomes more favourable for **3** and **4**.

In terms of affinity, **3** is the most potent inhibitor of PKA. The enthalpic loss upon the addition of the propyl-chain in ligand **4** cannot entirely be counterbalanced by the entropic gain.

Table 1. Thermodynamic data and standard deviation for all four ligands.

Ligand	ΔG [kJ/mol]	ΔH [kJ/mol]	-TΔS [kJ/mol]
1	-29.7±0.4	-38.7±2.5	9.1±2.9
2	-27.9±0.4	-35.1±2.8	7.2±3.2
3	-35.4±0.7	-31.8±0.6	-1.3±1.4
4	-32.0±0.3	-22.8±1.0	-9.2±1.3

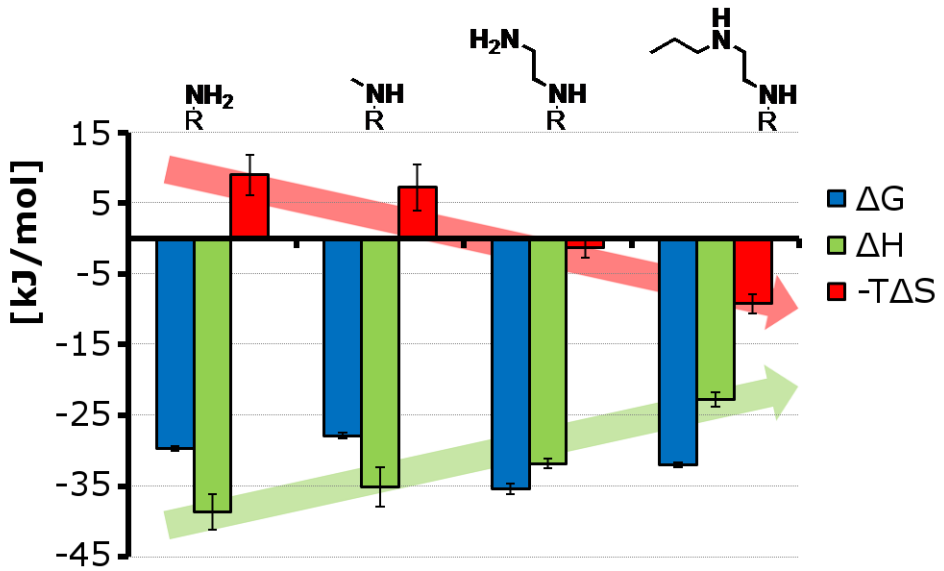


Figure 2. Thermodynamic signatures and observed trend (arrows) upon ligand growth. Upon ligand growth from left to right the entropy (red arrow) becomes more favourable whereas enthalpy (green arrow) turns less beneficial. Error bars indicate standard deviation.

Increase of buried polar surface of the protein upon ligand growth. **Figure 3** shows the development of various surfaces for the different ligands. Surprisingly, we do not observe a classical hydrophobic effect which is described as the amount of water being displaced from the hydrophobic protein surface. Instead the proportion of buried polar protein surface (green) augments in parallel with the increasing surface of the growing ligands. This might explain why we see an increase in entropy throughout the series but a decrease in enthalpy. The displacement of ordered water molecules which are mainly found in areas of the polar protein surface lead to a gain in entropy. However, the newly formed interactions between the growing part of the ligand and the protein might not be as geometrically and hence enthalpically

favorable as the original water-protein interactions. Hence a decrease in enthalpy throughout the series can be observed.

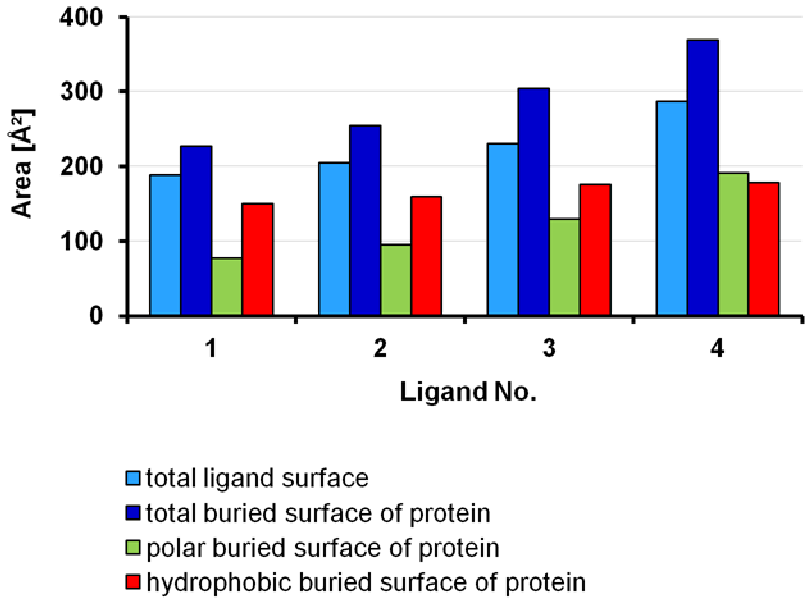


Figure 3. Development of ligand surface and buried protein surface during fragment growing. The increasing molecular weight leads to an increasing ligand surface (bright blue) which in turn leads to a larger amount of total buried surface of the protein (dark blue). The portion of polar buried protein (green) surface also augments upon fragment growth, however the increase in hydrophobic buried surface of the protein (red) is noticeably smaller.

Table 2. List of PDB-codes for the different ligand co-crystal structures

Ligand	PDB code
1	5M0C
2	5M0L
3	5M0B
4	5LCQ

Protein-ligand interaction pattern changes upon fragment growing. In order to fully understand the development of the thermodynamic signatures, a complete structural analysis of the entire system is necessary. Therefore, NMR spectroscopy as well as X-ray crystallography were used to receive a comprehensive structural picture. All crystal structures are deposited in the Protein Data Bank (PDB). The corresponding codes can be found in **Table 2**. Crystallographic results are shown in **Figure 4** and **Figure 5**. **Figure 4** visualizes the two different binding modes that can be observed for the four ligands. Interestingly ligands **1** and **2** share a common binding position (**Figure 4A**) where the nitrogen of the sulfonamide moiety forms a hydrogen bond to Asp127 (**Figure 5B**). Ligands **3** and **4** on the other hand not only display a slightly shifted binding position of the isoquinoline residue (**Figure 4B**) but show an entirely different interaction pattern of the sulfonamide moieties with the protein (**Figure 5B**). This sulfonamide portion is rotated around 80° compared to the orientation found for ligands **1** and **2** (**Figure 4C**). The longer alkyl chains, attached to the sulfonamides in ligands **3** and **4**, sterically hinder the interaction to Asp127. Instead the sulfonamide-nitrogen in both cases forms a hydrogen bond with Thr51 of the glycine rich loop (Gly-loop) thereby pulling the loop into the active site. This interaction and position of the Gly-loop for ligand **4** has been discussed in the previous chapter. The interaction pattern formed in these complexes might explain the less enthalpic binding signal of **3** and **4** compared to **1** and **2**.

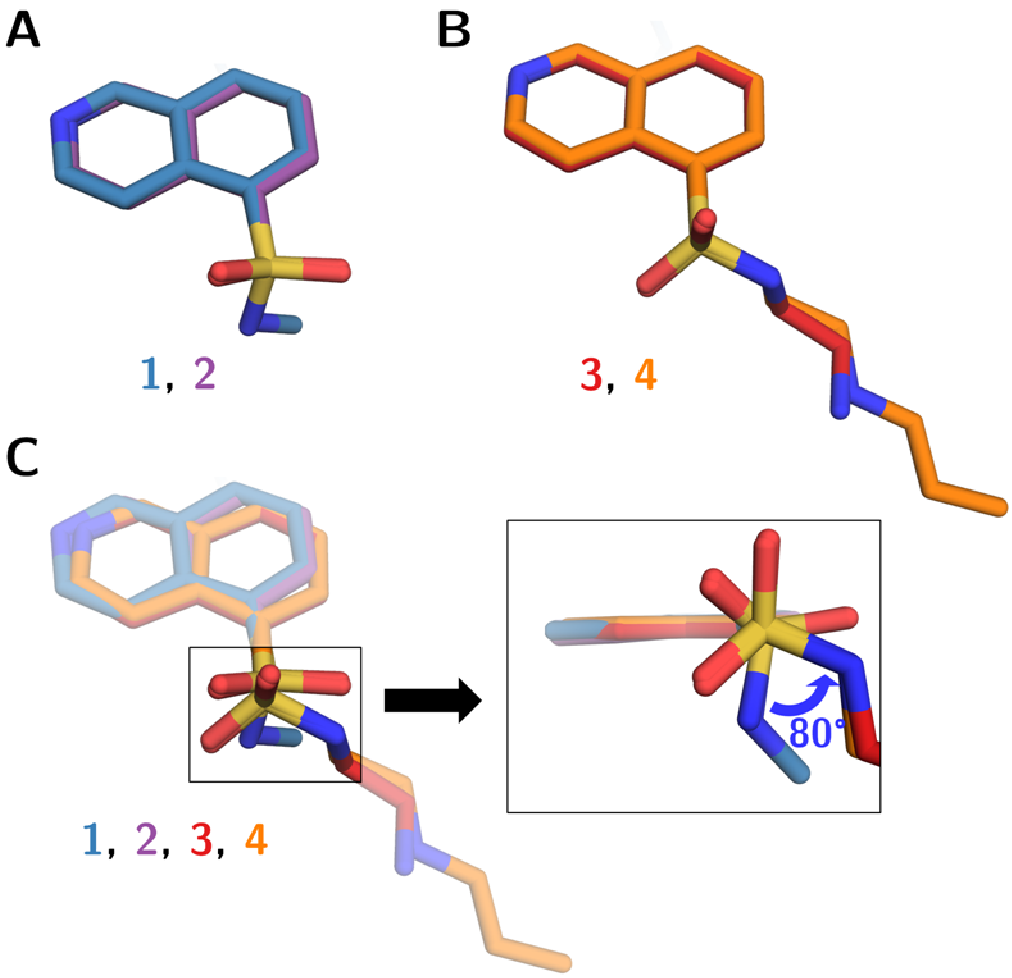


Figure 4: Comparison of ligand binding modes. A: Ligand **1** (purple) and **2** (blue) adopt a matching orientation of their corresponding atoms in the protein binding site. **B**: Ligand **3** (red) and **4** (orange) both have identical binding positions. **C**: Comparison and blow up of ligand **1** (purple) and **2** (blue) from A together with ligand **3** (red) and **4** (orange) from B show that the hinge binding isoquinoline portion is slightly shifted. More importantly the sulfonamide portion is rotated around 80° comparing binding modes from A and B (blue arrow).

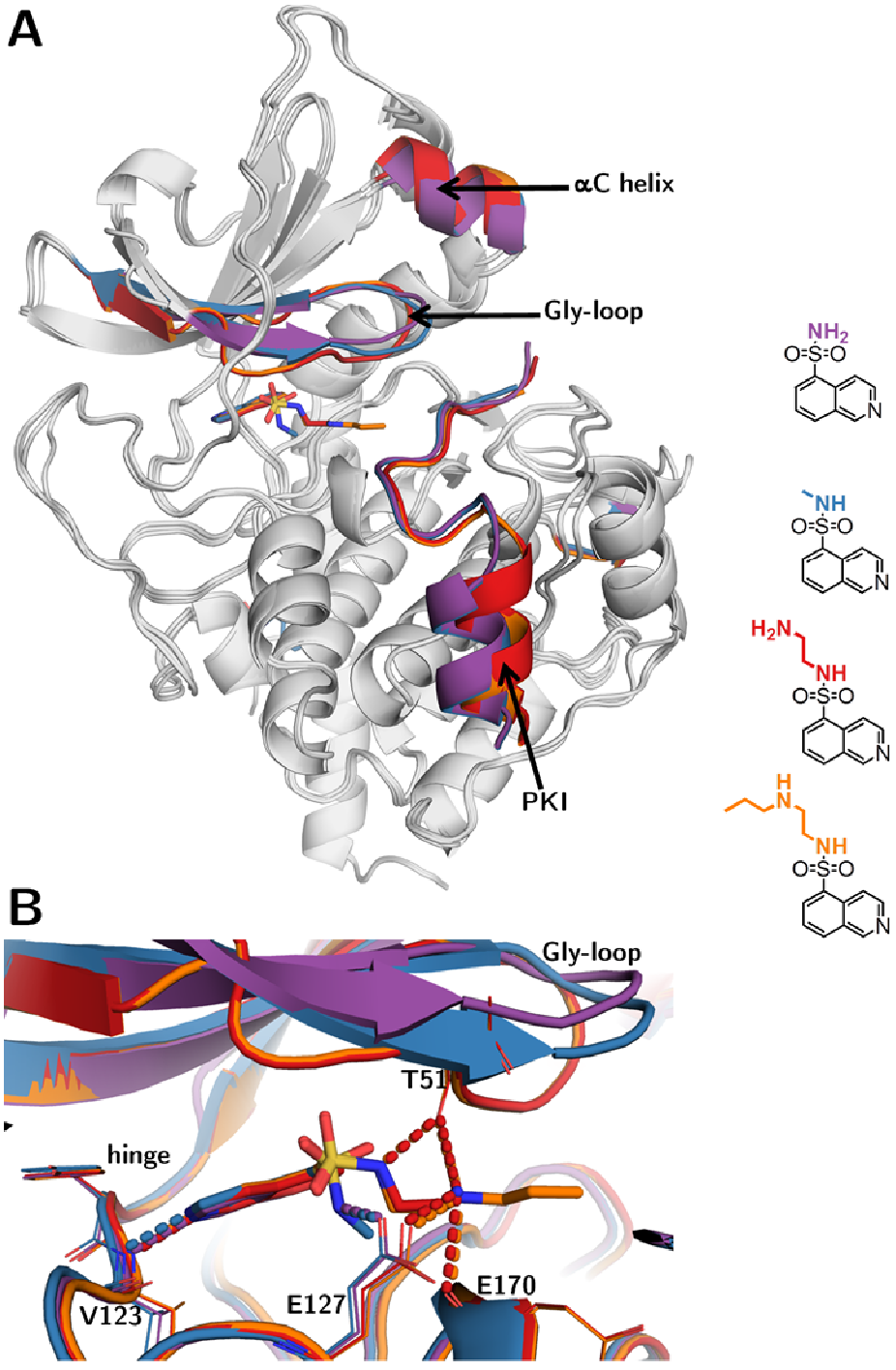


Figure 5. Superimposition of the co-crystal structures of all four ligands. **A**: overall view of the protein. **B**: Blow up of the active site and visualization of the polar protein-ligand interaction (dotted lines). The two binding modes described in **Figure 4** can be observed. Ligands with the same binding mode also form an analogous interaction pattern with the protein. In the complex structures of **3** and **4** the Gly-loop is pulled down into the active site through interactions with Thr51.

The structural overview of the protein-ligand complexes in **Figure 5A** depicts that not only the ligand binding mode but also the overall structure of the protein can be divided into the two groups described above. Outside the active site the largest structural difference can be observed for the position of the the αC helix, the APE motif (not visible in **Figure 5A**) and the position of the protein-kinase-inhibitor-peptide (PKI). Hence, it appears that both, ligand **3** and **4** trigger a stong induced fit that locks the Gly-loop in a closed conformation.

Unfortunately an extensive analysis of water molecules in the active site is not achievable in the presented cystal structures. Due to residual flexibility of the Gly-loop the electron densities of the water molecules in the active site are partially less well-defined. However it is clearly visible, that the additional propyl-sidechain of **4** leads to the displacement of a water molecule, which is well visible in the crystal structures of ligands **1**, **2** and **3** (**Figure 6B**). This is one explanation for the beneficial increase in entropy from **3** to **4**, as the release of a fixed and well-ordered water molecule into the bulk water phase results in an entropically favored binding singnature.

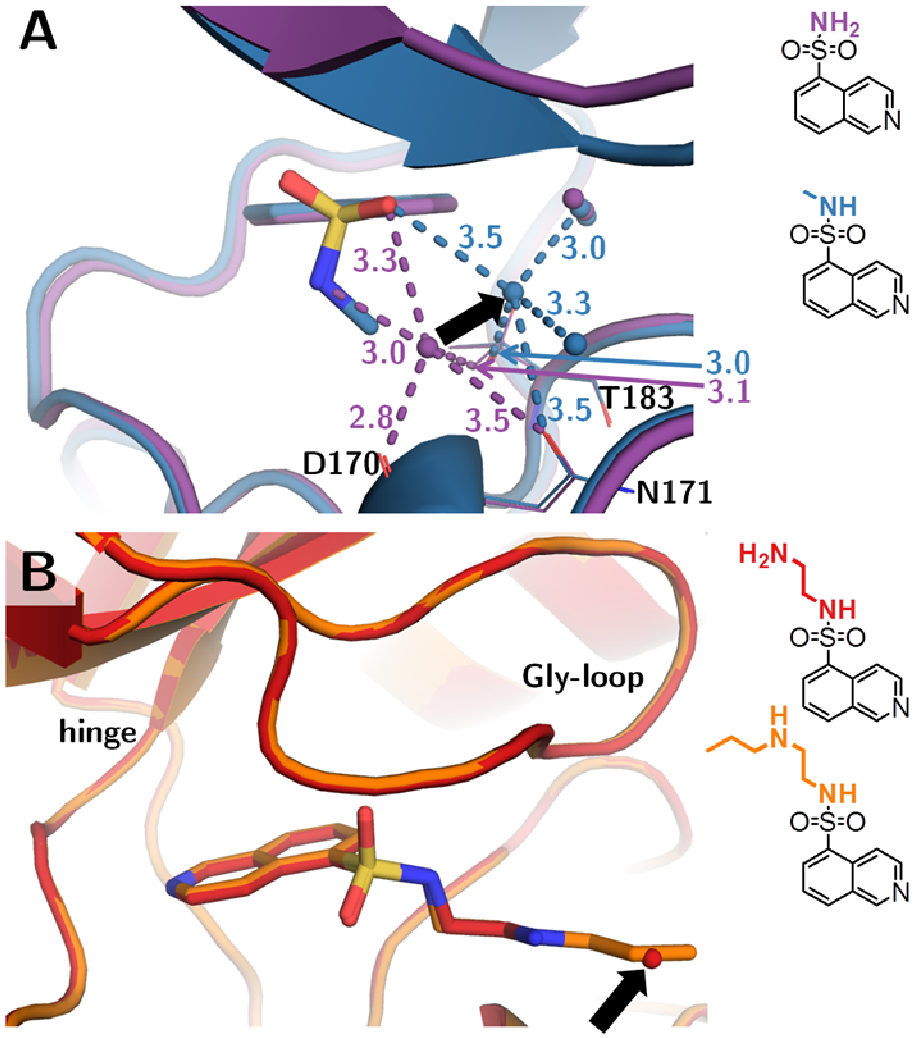


Figure 6. Crucial changes in the water structure. A: Comparison of **1** and **2**; dotted lines indicate polar interactions; numbers indicate distances in Å. The presence of the additional methyl group in ligand **2** leads to the shift of a water molecule. If remaining in its original position as observed in the complex of **1**, the methyl-carbon-to-water distance for ligand **2** would measure only 2.15 Å. The new position of the water molecule in the complex of **2** leads to a coordination with longer hydrogen bonds that is dominated by interactions with water molecules approaching the bulk phase whereas it is entirely coordinated through ligand and protein interactions in the position observed for **1**. The black arrow indicates the shift of the water molecule of 2.4 Å. **B:** Comparison of **3** and **4**; a clearly defined water molecule observed in the complex structure of **3** spacially overlaps with the additional propyl-sidechain introduced in **4**. Hence, the water molecule is displaced from its position (black arrow) and is replaced by the propyl-sidechain of **4**.

The additional methyl group attached to **2** compared to **1** does not displace a water molecule from the active site but nonetheless seems to slightly shift a water molecule from its original position (**Figure 6A**). As a result, this water molecule is kept in position by longer hydrogen bonds and is dominated by interactions formed to other water molecules. This is in contrast to the coordination of the water molecule observed in the complex with **1,** where all contacts are formed to either the ligand or the protein. Accordingly the binding of this water molecule in the complex of **2** approaches more the situation of its release into the bulk phase. Thus, the entropically more favored signature of this complex appears reasonable, even though other cases have been described where the sole attachement of a methyl group was responded by a more enthalpically and less entropically favored signal.[29, 30]

Amide T2 time does not indicate a difference in the dynamics of the different protein-ligand complexes. In order to analyse a potential difference in the residual flexibility of the different protein-ligand complexes ^{15}N-T_2-relaxation NMR measurements were performed. The amide T2-time can be used as an indicator for protein backbone dynamics. The lower the measured T2-time the more rigid is the protein. The T2-time for 75 amino acids of PKA could be assigned and analyzed. The mean of these values gives an impression of the overall dynamics of the different protein-ligand complexes as shown in **Figure 7**. Evidently, no significant difference can be detected for the four different ligands. Data for ligand **4** was discussed in the previous chapter. Obviously, in the present case residual flexibility does not differ across the four protein-ligand complexes and hence, cannot provide an

explanation for the deviations and trends between the thermodynamic profiles of the different protein-ligand complexes.

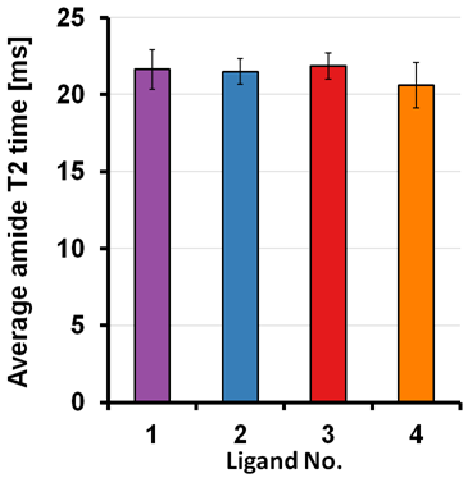


Figure 7: ^{15}N-T2-relaxation time in ms averaged over 75 amino acids for the different ligands, respectively. Overall differences are small. Error bars present the averaged standard deviation of the different amino acids.

MD simulations reveal ligand solvation patterns. As differences in the conformational and solvation properties of ligands in aqueous solution can be responsible for deviating thermodynamic profiles, a conformational ensemble of the unbound state was generated by 100 ns molecular dynamics simulations, separately for each ligand. It was found that the conformational diversity can be reasonably approximated by three distinct clusters for each ensemble. A subsequent analysis of solvation thermodynamics using the GIST approach[31, 32] revealed deviating solvation entropy contributions between the ligand molecules represented by their individual clusters. Specifically, the main clusters found for **3** and **4** seem to have a negative change in solvation entropy upon binding, which is triggered by a trapped water molecule in-between the sulfonamide and the terminal amine group.

However, there is a difference of this ligand-bound water molecule between **3** and **4.** The calculations revealed that ligand **3** entraps the water molecule in one of the three cluster representatives, corresponding to 40% of the simulation time. In the two other conformer clusters with 50% and 10% population, respectively, no water molecule is bound. Ligand **4** on the other hand, entraps a water molecule in all three major conformer clusters.

Once the ligands are accommodated at the binding site these ligand-bound water molecules are released to the bulk phase. This may contribute favorably to the entropy binding component of **3** and **4** in direct comparison to **1** and **2**. Furthermore, **3** entraps a water molecule less efficiently. Consistently, the entropic gain, as determined by ITC, for this ligand, is significantly smaller than for **4**.

Ligand performance put under the microscope. To analyze the development of ligand performance and characteristics upon ligand growth some descriptors, frequently consulted in FBDD were evaluated (**Figure 8**).[33, 34] In **Figure 8A** ligand efficiency (LE)[35] as a measure for affinity per heavy atom and the binding efficiency index (BEI)[36] as a measure of affinity per molecular weight are displayed. Both parameters are applied to define starting points in FBDD and record efficiency in fragment growth. No steady gain in affinity and LE or BEI is observed. In **Figure 8B** size-unbiased analyses of the ligands is displayed by fit quality (FQ)[37], percentage ligand efficiency (%LE)[38] and the size independent ligand efficiency (SILE)[39]. These evaluation methods are capable of comparing compounds at any stage of FBDD. Thereby the SILE value seems to be the most discriminating one, clearly pointing toward ligand **3** as the most favourable one. However, the

compound ranking appears to be similar for all three descriptors as they all score ligand **3** best. In **Figure 8C** lipophilic ligand efficiency (LLE)[40] and logP/LE (LELP)[41] values take the calculated logP values into account, thereby giving information on the lipophilicity in the process of lead optimization. For the LLE a higher value represents preferred characteristics whereas for the LELP a smaller value is advantageous. In both cases, ligand **3** ranks best and ligand **1** second best. For the LLE, compounds **2** and **4** rank equally unfavorable, whereas for the LEPL value ligand **4** ranks significantly worse than ligand **2**. Hence ligand **3** represents the ligand with the most beneficial affinity-lipophilicity compromise. If we compare ligand **3** to ligand **4**, the graph shows that the increased lipophilicity of ligand **4** through the addition of the propyl group is not overcompensated by a gain in affinity. Thus ligand **4** performs much worse than ligand **3** in this category. **Figure 8D** represents evaluation methods for entropic optimization. For this purpose the size-independent enthalpic efficiency (SIHE)[42], the enthalpic efficiency (EE)[43] and the specific enthalpic efficiency (specific EE)[43] were calculated. For SIHE and EE the same trend is observed, even though the relative grading appears smaller for the size-independent value. The SIHE value takes into account that the maximally achievable pK_H decreases with increasing molecular size. pK_H is taken as a measure of the enthalpic component of binding and is defined as $pK_H = -\Delta H/(2.303RT)$.

In conclusion ligand **3** performs best in most plots. Hence, ligand 1 is optimized and grown to ligand **3** but the additional introduction of the propyl group of ligand **4** appears inadvisable.

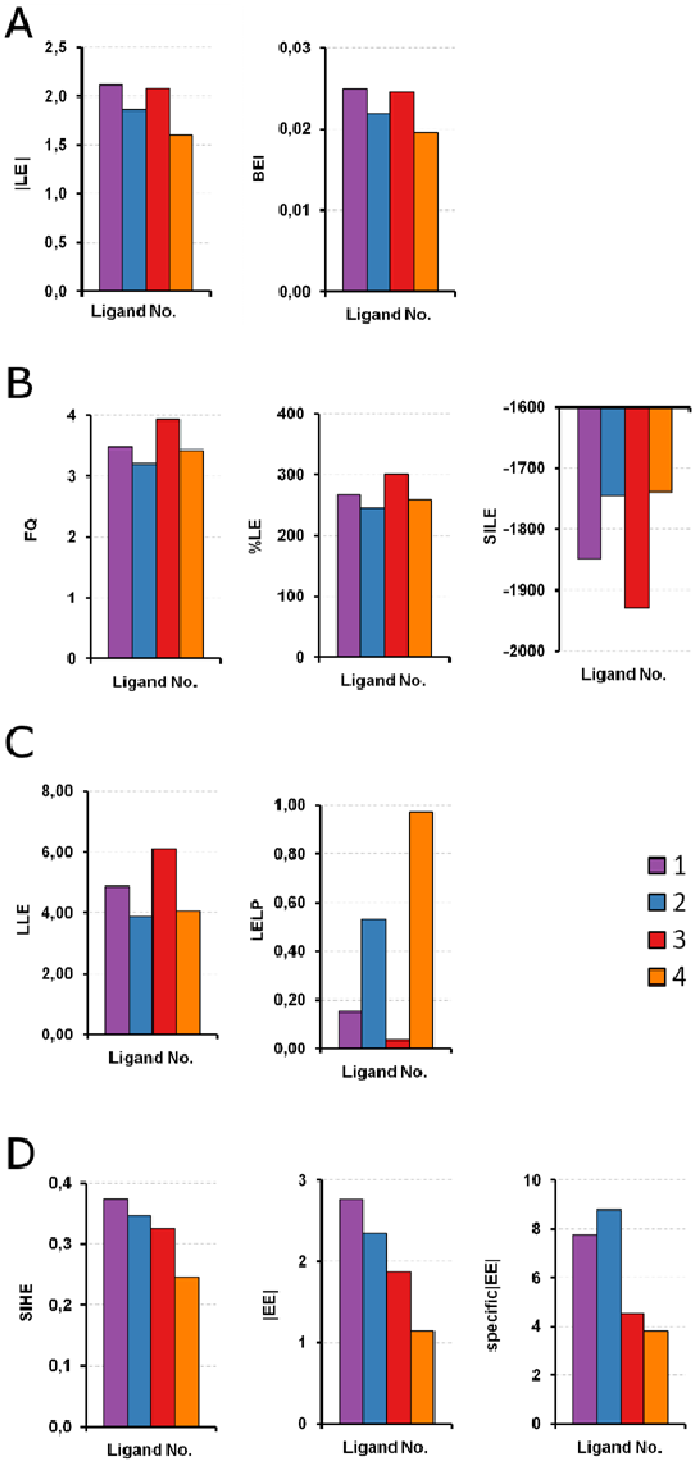


Figure 8. Different evaluation methods to analyze fragment development applied for all four different ligands. Calculations were made according to the references using ΔG [kJ/mol]. A: Classical Methods to prioritize the starting point in FBDD: Ligand

efficiency[35] and binding efficiency index[36]. **B:** Fit quality[37], percentage ligand efficiency[38], and size independent ligand efficiency[39] as size-unbiased comparison of early stage hits in FBDD. **C:** Lipophilicity development in the optimization process: Lipophilic ligand efficiency[40] and logP/LE[41]. **D:** Graphs focusing on enthalpic ligand optimization via Size independent enthalpic efficiency (SIHE)[42] and enthalpic efficiency (EE)[43] as well as specific enthalpic efficiency[43] where enthalpic potency optimization is assigned to the number of potential hydrogen bond donors and acceptors in the ligand.

3.6 Conclusion

The change in the thermodynamic signature from ligand **1** to ligand **2** might result from the increase in polar buried protein surface, a tendency observable throughout the entire series.

The entropically more favorable binding of **3** compared to **2** is most likely due to water displacement from the active site and water release from a conformer in solution prior to binding. MD simulations suggest that the ligand in solution partially entraps a water molecule. The release of this water molecule upon protein binding is entropically favorable. The explanation for the entropic gain of ligand **4** if compared to ligand **3** goes along the same direction. Calculations suggest that ligand **4** entraps this water molecule even more frequentely. Hence, the entropic benefit upon water release and ligand binding to the protein is even larger than for **3**.

The influence of the pronounced change in protein conformation observed for **3** and **4** compared to **1** and **2** on enthalpy and entropy is difficult to evaluate. However, no impact results from the overall protein-ligand complex dynamics.

In conclusion we observe a trend toward more entropic and less enthalpic binding upon ligand growing for the series of ligands presented in this study.

Thereby, affinity increases less due to partial enthalpy-entropy compensation. It has been suggested, that an enthalpically binding fragment is a more promising starting point in FBDD. Optimization for entropic reasons is usually more easily achieved than enthalpic improvement. It has been noted that fragments bind more enthalpy dominated, which is alluded to the formation of more optimal hydrogen bonds in an else hydrophobic environment (hot spot).[6]

However, the effect of minor affinity gain upon fragment growth has also been observed in this context, being the complication that results from enthalpy-entropy compensation.[5]

All in all, prediction of thermodynamic profiles remains challenging. This is also due to factors such as induced fits and changes in ligand binding modes upon ligand growth as presented in this study. Nonetheless it is highly interesting that we observe a gain in entropy of binding upon ligand growth even if the fragments are grown to relatively flexible ligands.

3.7 Abbreviations

Ala: alanine

Asp: aspartate

Arg: arginine

ATP: adenosine triphosphate

CHO: Chinese hamster ovary

CSP: chemical shift perturbation

DFG: aspartat-phenylalanin-glycin motif

DMSO: dimethyl sulfoxide

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

GIST: grid inhomogeneous solvation theory

Gly: glycine

Gly-loop: glycine-rich loop

Glu: glutamate

His-tag: histidine-tag

ITC: isothermal titration calorimetry

MBT: MES/Bis-Tris

MD: molecular dynamics

MPD: 2-methyl-2,4-pentanediol

Ni-NTA: nickel-nitrilotriacetic acid

NMR: nuclear magnetic resonance spectroscopy

NPT: constant particles, pressure and temperature

NVT: constant particles, volume and temperature

PDB: protein data bank

PKA: cAMP-dependent protein kinase

PKI: protein kinase inhibitor

qNMR: quantitative nuclear magnetic resonance spectroscopy

Ser: serine

TEA: triethanolamine

TEV: tobacco etch virus

Thr: threonine

TLS: translation/libration/screw

TROSY: transverse relaxation optimized spectroscopy

Val: valine

3.8 References

1 van Dongen, M.; Weigelt, J.; Uppenberg, J.; Schultz, J.; Wikström, M. *Drug Discov Today* **2002**, *7*, 471–478.

2 Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. *Nat Rev Drug Discov* **2004**, *3*, 660–672.

3 Hajduk, P. J.; Greer, J. *Nat Rev Drug Discov* **2007**, *6*, 211–219.

4 Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. *Nat Rev Drug Discov* **2016**, *15*, 605–619.

5 Edink, E.; Jansen, C.; Leurs, R.; de Esch, I. J. *Drug Discov Today Technol* **2010**, *7*, e147–e202.

6 Ferenczy, G. G.; Keserü, G. M. *J Chem Inf Model* **2012**, *52*, 1039–1045.

7 Kudlinzki, D.; Linhard, V. L.; Saxena, K.; Sreeramulu, S.; Gande, S.; Schieborr, U.; Dreyer, M.; Schwalbe, H. *Acta Crystallogr F Struct Biol Commun* **2015**, *71*, 1088–1093.

8 Kabsch, W. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132.

9 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658–674.

10 Adams, P. D. et al. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–221.

11 Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486–501.

12 Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J Appl Crystallogr* **1993**, *26*, 283–291.

13 Kleywegt, G. J.; Zou, J.-Y.; Kjeldgaard, M.; Jones, T. A. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; Springer Netherlands: Dordrecht, 2001; 353–356.

14 Painter, J.; Merritt, E. A. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 439–450.

15 Painter, J.; Merritt, E. A. *J Appl Crystallogr* **2006**, *39*, 109–111.

16 Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. http://www.globalphasing.com.

17 Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. *J Chem Inf Comput Sci* **2004**, *44*, 2133–2144.

18 Connolly, M. L. *QCPE Bull* **1981**, *1*, 75–83.

19 Connolly, M. L. *J Appl Crystallogr* **1983**, *16*, 548–558.

20 Connolly, M. L. *Science* **1983**, *221*, 709–713.

21 Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. *Anal Chem* **2012**, *84*, 5066–5073.

22 Houtman, J. C. D.; Brown, P. H.; Bowden, B.; Yamaguchi, H.; Appella, E.; Samelson, L. E.; Schuck, P. *Protein Sci* **2007**, *16*, 30–42.

23 Holzgrabe, U.; Deubner, R.; Schollmayer, C.; Waibel, B. *J Pharm Biomed Anal* **2005**, *38*, 806–812.

24 Roe, D. R.; Cheatham III, T. E. *J Chem Theory Comput* **2013**, *9*, 3084–3095.

25 Neudert, G.; Klebe, G. *Bioinformatics* **2011**, *27*, 1021–1022.

26 Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J Phys Chem* **1993**, *97*, 10269–10280.

27 Frisch, M. J. et al. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

28 Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* **2004**, *25*, 1157–1174.

29 Biela, A.; Nasief, N. N.; Betz, M.; Heine, A.; Hangauer, D.; Klebe, G. *Angew Chem Int Ed* **2013**, *52*, 1822–1828.

30 Krimmer, S. G.; Betz, M.; Heine, A.; Klebe, G. *ChemMedChem* **2014**, *9*, 833–846.

31 Lazaridis, T. *J Phys Chem B* **1998**, *102*, 3531–3541.

32 Nguyen, C. N.; Young, T. K.; Gilson, M. K. *J Chem Phys* **2012**, *137*, 044101.

33 Hann, M. M.; Keserü, G. M. *Nat Rev Drug Discov* **2012**, *11*, 355–365.

34 Ferenczy, G. G.; Keserü, G. M. *Drug Discov Today* **2010**, *15*, 919–932.

35 Hopkins, A. L.; Groom, C. R.; Alex, A. *Drug Discov Today* **2004**, *9*, 430–431.

36 Abad-Zapatero, C.; Metz, J. T. *Drug Discov Today* **2005**, *10*, 464–469.

37 Reynolds, C. H.; Tounge, B. A.; Bembenek, S. D. *J Med Chem* **2008**, *51*, 2432–2438.

38 Orita, M.; Ohno, K.; Niimi, T. *Drug Discov Today* **2009**, *14*, 321–328.

39 Nissink, J. W. M. *J Chem Inf Model* **2009**, *49*, 1617–1622.

40 Leeson, P. D.; Springthorpe, B. *Nat Rev Drug Discov* **2007**, *6*, 881–890.

41 Keserü, G. M.; Makara, G. M. *Nat Rev Drug Discov* **2009**, *8*, 203–212.

42 Ferenczy, G. G.; Keseru, G. M. *J Chem Inf Model* **2010**, *50*, 1536–1541.

43 Ladbury, J. E.; Klebe, G.; Freire, E. *Nat Rev Drug Discov* **2010**, *9*, 23–27.

3.9 Supplementary information

Table S1.Crystallographic table for all crystal structures. Table spreads over two pages.

Ligand, PDB entry →	**3, 5M0B**	**2, 5M0L**	**1, 5M0C**	**Apo, 5M0U**
Data collection & processing				
No. Crystals used	1	1	1	1
Wavelength [Å]	0.97625	0.91841	0.91841	0.91841
Space group	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)
Unit cell parameters: a, b, c [Å] β [°]	57.9, 71.7, 109.9	65.8, 79.7, 84.5	72.4, 75.4, 80.2	58.2, 72.0, 109.8
Diffraction data[a)]				
Resolution range [Å]	45.08-1.51	34.09-1.47	43.75-1.73	45.27-1.67
Highest shell resolution range [Å]	1.60-1.51	1.55-1.47	1.84-1.73	1.77-1.67
Unique reflections	73047(11656)	75918(11950)	45445(7171)	53774(8345)
R(I)sym [%][b)]	5.4(49.8)	4.1(49.5)	3.8(48.0)	4.9(50.0)
Completeness [%]	99.9(99.7)	98.7(97.4)	98.4(97.3)	98.4(96.0)
Redundancy	7.3(7.3)	4.1(4.2)	4.5(4.6)	7.3(7.4)
I/σ (I)	26.3(3.9)	19.1(2.7)	22.9(3.2)	22.4(3.7)
Refinement				
Resolution range [Å]	39.88-1.51	32.9-1.47	37.68-1.73	45.27-1.67
Reflections used in refinement (work/free)	69394/3653	72119/3799	43172/2273	52019/1755
Final R values for all reflections (work[c)]/free[d)]) [%]	13.2/15.4	15.3/17.3	17.3/20.0	15.2/18.9

Amino acids (PKA/PKI)	353/19	340/18	338/19	349/20
Inhibitor atoms	17	2 x 15	14	-
Water molecules	515	400	301	515
RMSD from ideality				
Bond length [Å]	0.008	0.008	0.006	0.009
Bond angles [°]	0.96	1.05	0.79	0.96
Ramachandran plot[e)]				
Residues in favoured regions [%]	91.5	90.6	91.5	91.2
Residues in additionally allowed regions [%]	8.5	9.1	8.5	8.8
Residues in generously allowed regions [%]	0.0	0.3	0.0	0.0
Mean B-factors [Å²]				
PKA (protein)/PKI (peptide)	15.6/15.9	23.3/21.4	26.6	15.9
Inhibitor	11.4	15.8/23.1	28.5	-
Water molecules	28.3	32.4	36.9	28.3

a)Numbers in parentheses are for the highest-resolution shell.

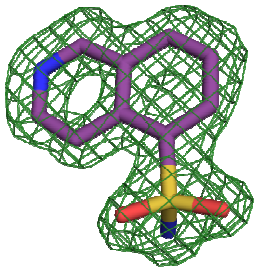
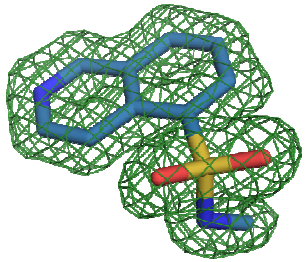
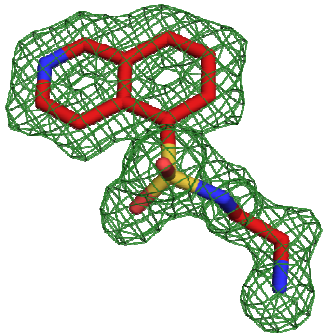
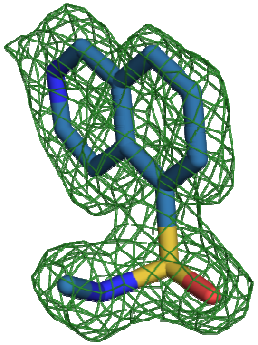
b)R_{sym}=$[\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}|I_{\mathrm{i}}(h)-\ \langle I(h)\rangle\ |/\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}I_{\mathrm{i}}(h)] \times 100$, $\langle I(h)\rangle$ is the mean of the $I(h)$ observation of reflection h.

c)R_{work}=$\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}-F_{\mathrm{c}}|/\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}$.

d)Calculation of R_{free} was performed as for R_{work} but on 5 % of the data which was excluded from the refinement.

e)Derived from Procheck.[12]

Table S2. All atoms of all ligands were 100% occupied. For visualiazation, σ-weighted mF_o-DF_c-densities for all three ligands are shown.

Ligand 1, at 3.0 σ	**Ligand 2, at 3.0 σ**	**Ligand 3, at 3.0 σ**
5M0C	**5M0L**	**5M0B**
	Active site	
		
	Allosteric:	
		

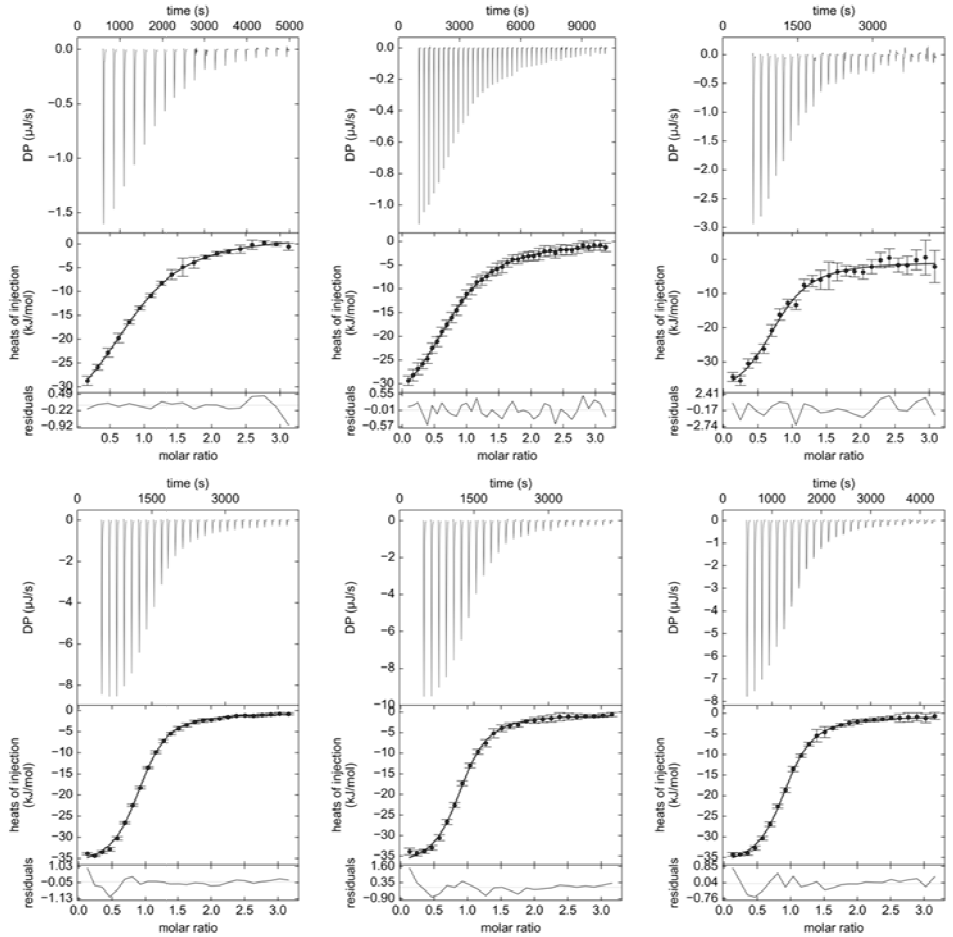


Figure S1. ITC-titration curves for ligand 01 in phosphate buffer as exemplary titration curves. Upper three curves result from measurements with c-values between 3 and 6, the lower three curves result from measurements with c-values between 19 and 20.

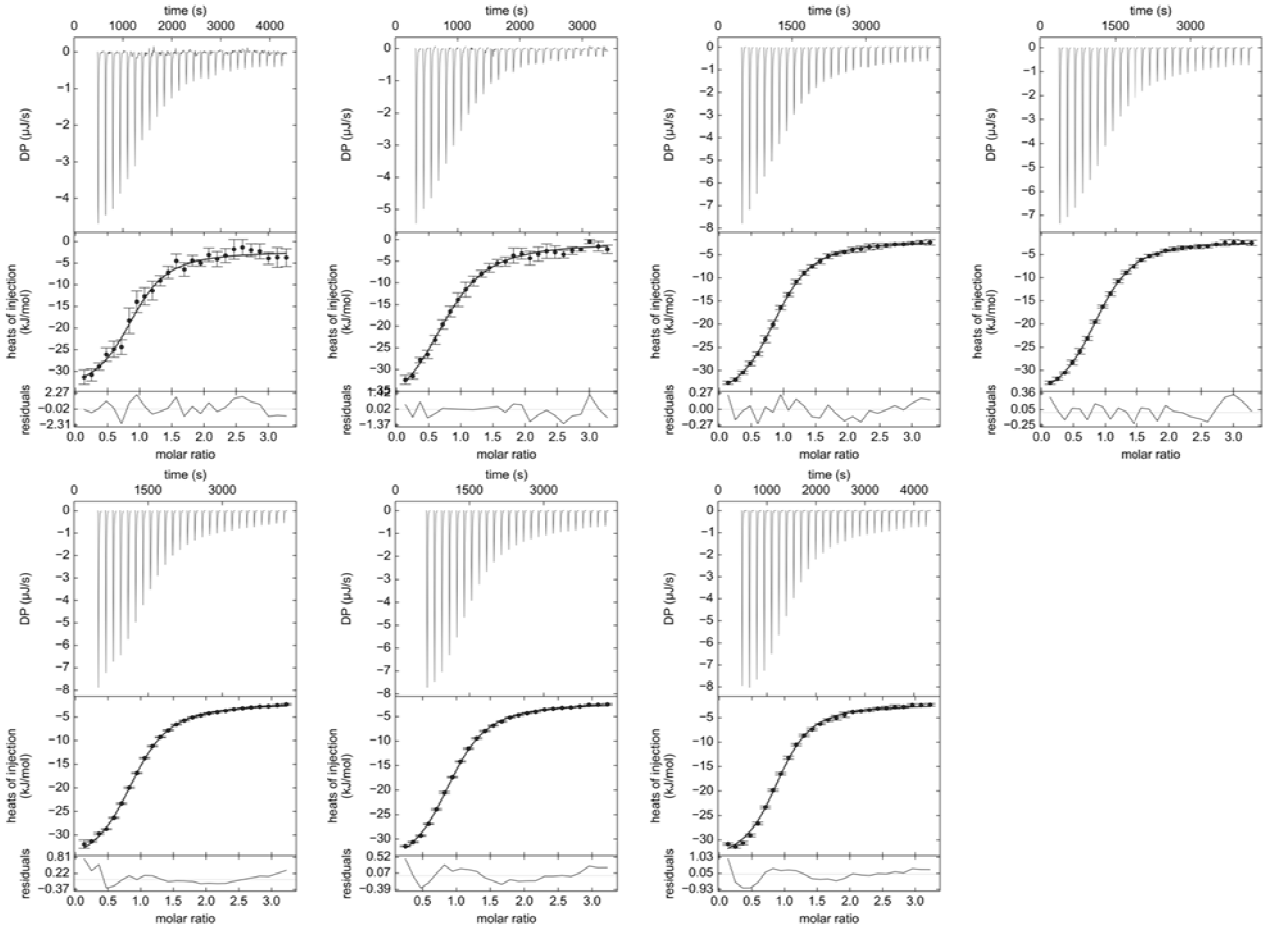


Figure S2. ITC-titration curves for ligand 02 in phosphate buffer as exemplary titration curves. The second titration curve (top, middle) results from a measurement with a c-value of 5. All other titrations result from measurements with c-values of 10.

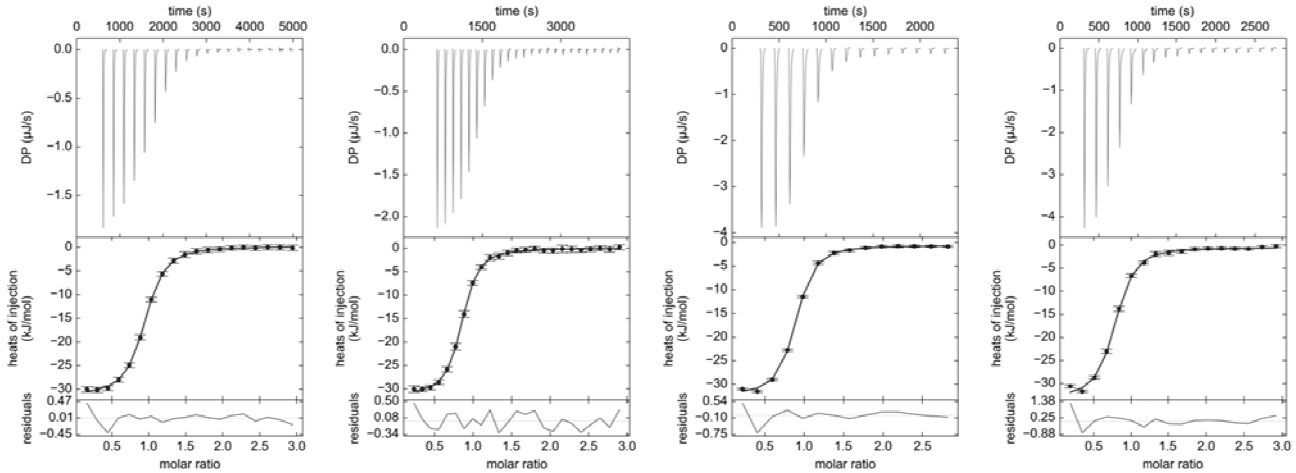


Figure S3. ITC-titration curves for ligand 03 in phosphate buffer as exemplary titration curves. All titrations result from measurements with c-values between 45 and 73.

4 Surprising Non-Additivity of Methyl-Groups in Drug-Kinase Interaction

4.1 Annotations

The following chapter will be published. The following people contributed to the paper and will therefore be listed as authors on the publication:

Barbara Wienen-Schmidt[‡], Denis Schmidt[‡], Hans-Dieter Gerber[‡], Andreas Heine[‡], Gerhard Klebe[‡]*

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

* Corresponding author

Protein expression for ITC and crystallization was performed by Barbara Wienen-Schmidt as well as ITC-measurements, crystallization and crystallography.

Ligand synthesis and qNMR measurements were performed by Hans-Dieter Gerber

4.2 Abstract

Drug optimization is guided by biophysical methods with increasing popularity. In the context of lead structure modifications, the introduction of methyl groups is a simple but potentially powerful approach. Hence, it is crucial to systematically investigate the influence of ligand methylation on biophysical characteristics such as thermodynamics. Here we investigate the influence of ligand methylation in different positions and combinations on drug-kinase interaction. Binding modes and complex structures were analyzed using protein crystallography. Thermodynamic signatures were measured via isothermal titration calorimetry (ITC). We found that not only position but also stereochemistry of the methyl group has an influence on binding potency as well as the thermodynamic signature of ligand binding to the protein. Strikingly, the combination of single methyl groups does not lead to additive effects. In our case the merger of two methyl groups in one ligand leads to an entirely new alternative ligand binding mode in the protein ligand complex. Moreover, the combination of the two methyl groups also resulted in a non-additive thermodynamic profile of ligand binding. This unexpected drastic change in protein ligand interaction highlights the importance of crystallographic control even for minor modifications such as the introduction of a methyl group.

4.3 Introduction

In the process of drug design and drug optimization, lead structures can be decorated with diverse chemical functions in order to improve binding potency and biophysical characteristics. As a matter of fact, the design of drugs with tailored biophysical properties has become increasingly popular

over the last years. [1-6] This development is based on the assumption that additional binding characteristics such as binding thermodynamics and kinetics might help to accelerate and improve the development and prediction of clinically successful drugs. Thereby, the introduction of a methyl group to a ligand is a very simple but potentially powerful approach. The importance but also the enormous popularity of methyl groups in drug design has been reported and reviewed previously.[7, 8] The replacement of a hydrogen by a methyl group can significantly alter structure-activity relationships. These changes are based on improved electrostatics, modulated polarity and steric complementarity as well as conformational energetics and restrictions of the ligand. Furthermore, ordered water molecules in the protein active site that are replaced or shifted in consequence of the presence of a methyl group can take a major impact on structure-activity relationships.[7] In rare cases, the introduction of an additional methyl group can even boost activity of a ligand 180 fold.[7] On the other hand it can also reduce activity drastically if added at the wrong position.[7]

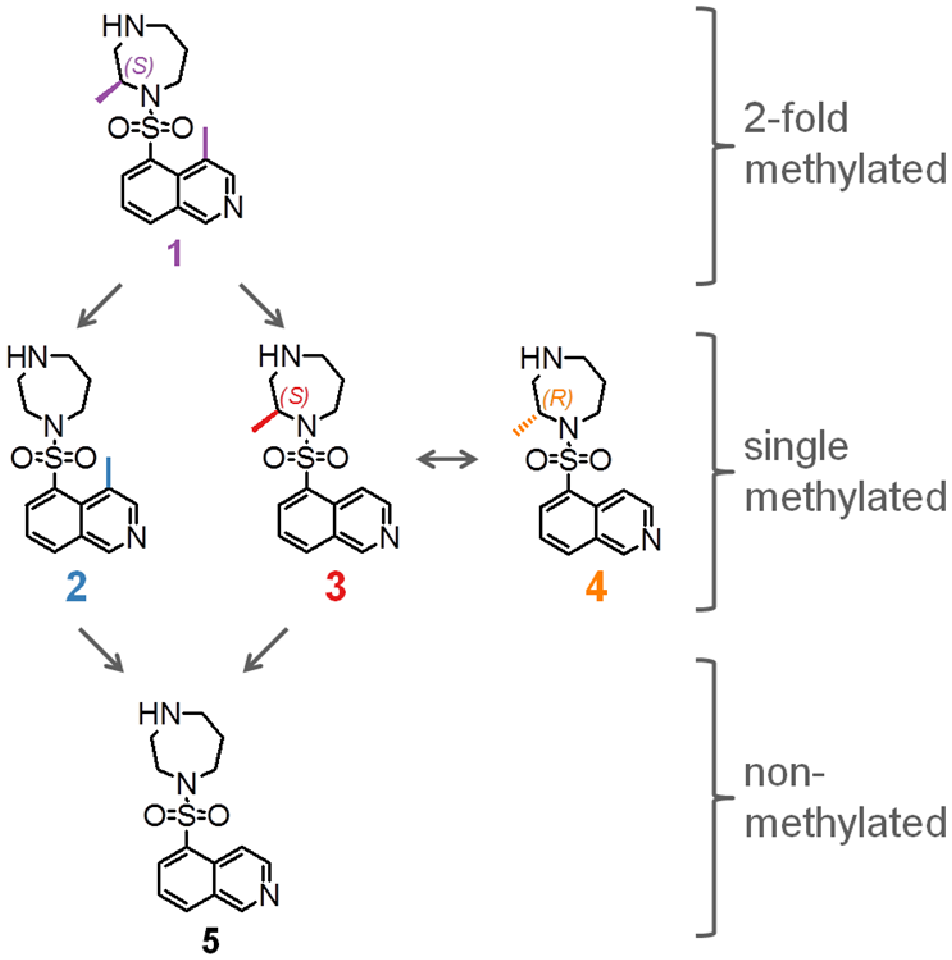


Figure 1. Overview of all ligands discussed. Different methylation positions were investigated. Ligand **H-1152** (**1**) displays two methylated sites, one at the isoquinoline moiety and a second at the homopiperazine portion. These two sites were separately investigated via analysis of ligand **2** and **3**. In order to examine the influence of the spatial position of the methyl group of **3**, its stereoisomer **4** was additionally studied. Fasudil (**5**) was used as the non-methylated reference.

Here we present a study where the influence of methyl groups on drug-kinase interactions is thoroughly investigated. Therefore, cAMP-dependent protein kinase (PKA) is used as a model protein, allowing the use of a wide range of experimental techniques. The ligands presented in this study comprise two

prominent kinase inhibitors: **Fasudil** and **H-1152**[9-17]. In addition, we studied "intermediate" ligands showing methyl groups at different positions of the parent scaffold. An overview of the chemical structures of all five ligands is given in **Figure 1**. **H-1152** (**1**) represents a dimethylated fasudil, **fasudil** (**5**) itself being the non-methylated reference ligand presented previously (**Chapter 2**). The two different single methylated ligands that derive from **H-1152** are ligand **2** and **3**. Ligand **2** is methylated at the 4 position of the isoquinoline portion and **3** is methylated at the 2 position of the homopiperazine ring. The latter methyl group in **3** introduces a stereogenic center which exhibits for both, **1** and **3**, *S*-configuration. In order to evaluate the influence of this methyl group in inverted configuration, ligand **4** with R-configuration was added to the panel.

Crystal structures of all ligands depict differences in binding modes and suggest non-additivity with respect to the geometric influence of the two methyl groups. Non-additivity is also observed for the binding thermodynamics. Thermodynamic data was measured using isothermal titration calorimetry (ITC). Not only the binding potency ΔG is influenced by the different methylation patterns but also the partitioning in enthalpic and entropic contribution is altered. The crystal structure of the dimethylated **H-1152** reveals an alternative binding mode compared to **fasudil**. Multiple crystal structures of bovine PKA, bovine PKA-mutants and rho-kinase 1 in complex with **H-1152** have been published, however none of them reports this alternative binding mode as described in this report.[18, 19, 13] Most interestingly, this binding mode is similar to that found with the kinase Roco4 in complex with **H-1152**. Roco4 is a model protein for

the human leucine-rich-repeat kinase 2 (LRRK2) which has been described to be involved in late-onset Parkinson.[20]

Strikingly, this alternative binding mode is not observed for any other ligand of this series in complex with PKA. This drastic change in protein-ligand interaction is likely to cause a strong difference in ligand binding properties towards PKA.

4.4 Experimental section

Protein expression and purification

The catalytic subunit of cAMP-dependent protein kinase from Chinese hamster ovary cells was expressed with a His-tag in a modified pET16b-Vector with an introduced TEV-cleavage site between the protein N-terminus and His-tag. This plasmid was transformed into *E. coli* strain *Bl21 (DE3)/pLysS* (Novagen).[21]

Cell disruption was performed using a high-pressure homogenizer for multiple cycles. After centrifugation (1h at 30.000g) cell lysate supernatant was purified in a first step using a Ni-NTA column that binds the His-tag of the protein and was eluted by an imidazole gradient. The His-tag was then cleaved off by TEV-protease. Afterwards, an inverse Ni-NTA column was employed collecting PKA in the flow-through. Finally, ion exchange chromatography was performed using a MonoS column separating three-fold phosphorylated PKA from the four-fold phosphorylated form using a HEPES buffer with a sodium chloride gradient.[21]

Crystallization

Co-crystallization of **5** is discussed in **Chapter 2**. Co-crystallisation was performed using the hanging drop method at 4 °C. The crystallization drops contained the following ingredients: 10 mg/mL PKA (240 µM), 30 mM MBT (MES/Bis-Tris Puffer pH 6.9), 1 mM DTT, 0.1 mM EDTA, 75 mM LiCl, 0.03 mM Mega 8, 0.07 mM PKI (Sigma: P7739), 1.2 mM ligand dissolved in DMSO from a 50-100 mM stock. The well contained a mixture of methanol in water with varying methanol concentrations (v/v) for the different ligands (**1**: 18% methanol; **2**: 20% methanol; **3**: 16% methanol; **4**: 19% methanol). In the crystallization setup, streak-seeding was performed with apo crystals by the help of a horse hair in order to initialize crystal growth. For crystal mounting, crystals were cryo protected in 5 mM MBT (MES/Bis-Tris buffer pH 6.9), 1 mM DTT, 0.1 mM LiCl, 1.2 mM ligand dissolved in DMSO from a 50-100 mM stock, 16 % (v/v) methanol, 30% (v/v) MPD and flash frozen in liquid nitrogen.

Crystallography

Structures from **2** and **3** were collected at the storage ring Bessy II Helmholtz-Zentrum Berlin, Germany at Beamline 14.1 on a Pilatus 6M pixel detector. Structure from **1** was collected at the European Synchrotron Radiation Facility (ESRF) Grenoble, France at Beamline 14. Structure from **5** is discussed in **Chapter 2**.

The datasets were processed using XDS[22] and molecular replacement was performed using CCP4 Phaser[23] and PDB-structure of PKA from *bos taurus* 1Q8W as a model. This was followed by simulated annealing, multiple refinement cycles of maximum likelihood energy minimization and B-factor

refinement using Phenix[24]. Coot[25] was used to fit amino-acid side chains into σ-weighted 2Fo – Fc and Fo – Fc electron density maps. If appropriate electron density was observed, multiple side chain conformations were built into the model and maintained during the refinement if the minor populated side chain displayed at least 20 % occupancy. Ramachandran plots for structure validation were calculated using PROCHECK[26]. Data collection, unit cell parameters and refinement statistics are given in the supplementary information. Analysis of temperature factors was performed with Moleman[27]. Protein and PKI B-factors were anisotropically refined, water B-factors were isotropically refined for all structures. Decision for anisotropic or TLS refinement was based on comparison of R_{free}. Anisotropic refinement was chosen over TLS if the achieved R_{free} values were at least 0.5% lower for anisotropic than for TLS refinement. R_{free} was calculated using 5% of all reflections which were randomly chosen and not used for the refinement. The required ligand restraint files were created using the Grade webserver[28, 29]. For figure preparation Pymol was used. Crystallographic tables as well as mFo-DFc-densities of the different ligands can be found in the supporting information.

Isothermal titration calorimetry

The buffer used for the ITC experiments contained: 30 mM sodium phosphate buffer pH 7.2, 10 mM $MgCl_2$, 100 mM NaCl, 3% (v/v) DMSO. All measurements were repeated 3-5 times. Further buffers were used in order to check for protonation linkage. In these buffers 30 mM sodium phosphate buffer was replaced by 30 mM HEPES and 30 mM triethanolamine (TEA), respectively (both at pH 7.2). Buffer dependency was tested for all ligands and no significant buffer dependence could be detected. For the

measurements expressed, purified and dialyzed PKA was used in the ITC-measuring cell. A 15-20 fold higher concentrated ligand solution, diluted in dialysis buffer, was then stepwise injected to the protein solution during the measurement. All measurements were performed at 25 °C. ITC data were analyzed using NITPIC and Sedphat[30, 31]. Raw data and exact values and standard deviations for ΔG, ΔH and -TΔS can be found in the supplementary information.

Compound purity was analyzed using quantitative nuclear magnetic resonance spectroscopy (qNMR) and in case of deviation, ligand concentration was corrected accordingly.[32]

Ligands

Ligands **5** was purchased from Uorsy (Ukraine) and ligand **1** from Toronto Research Chemicals (TRC) (Canada). Ligands **2**, **3** and **4** were synthesized.

4.5 Results and discussion

Crystal structures reveal surprising binding mode. Co-crystal structures with resolutions between 1.4 and 1.5 Å could be obtained. All structures were deposited in the protein data bank (PDB). The respective PDB-codes are listed in **Table 1**.

Table 1. PDB codes of all five co-crystal structures.

Ligand	PDB-Code
1	5M6V
2	5M6Y
3	5M75
4	5M71
5	5LCP

An overview and superposition of the different ligands is depicted in **Figure 2**. The hinge binding mode of ligands **2**, **3**, **4** and **5** are all, as expected, similar (**Figure 2A**). Surprisingly, ligand **1** deviates from this pattern and populates two distinct binding poses **(Figure 2B)**. Apart from the first orientation which agrees with the hinge binding mode observed for the other ligands, **1** exhibits a second, alternative interaction pattern with the hinge where the ligand is flipped over and rotated about 60°. **Figure 3** displays a schematic overview of the hinge binding portion of the different ligands for ease of visualization. Interestingly the alternative binding position of **1** is the only position where a hydrogen bond between the sulfonyl-oxygens and the protein is formed.

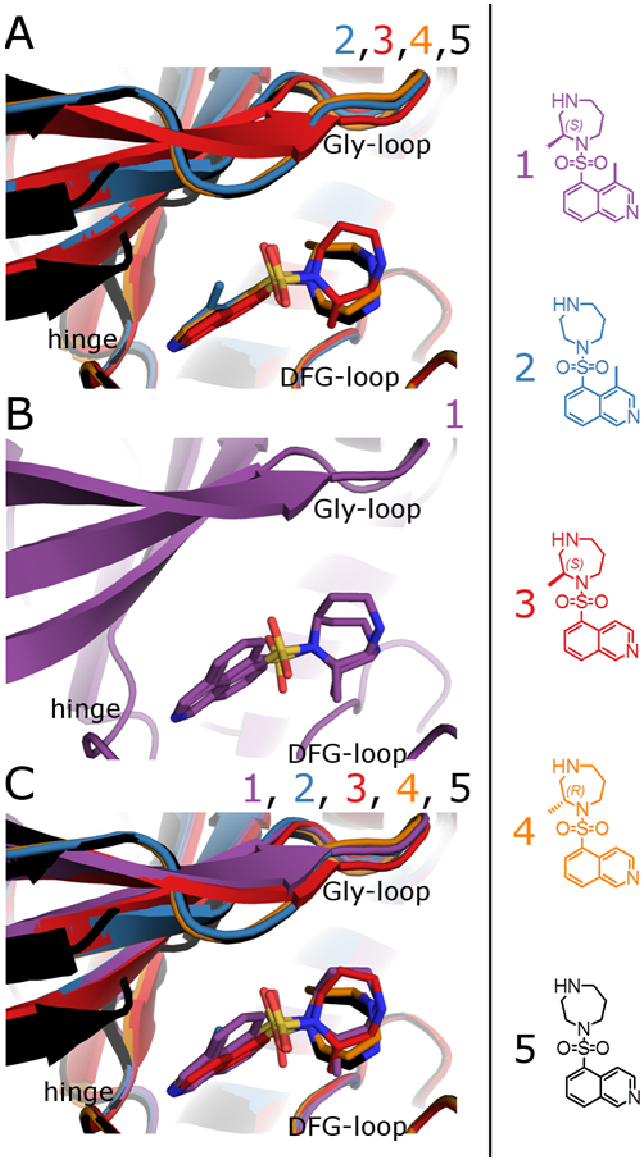


Figure 2. Cocrystal structures of the different ligands. A: Superposition of complex structures of ligands **2**, **3**, **4** and **5**. All ligands share a common hinge binding mode. **B:** Complex structure of **1**. Ligand **1** binds in two different conformations **C:** Superposition of all ligands **1-5**.

A superposition of all 5 ligands (**Figure 2C**) shows that also the homopiperazine moiety of the ligand adopts different ring conformations depending on the methylation. This, in turn, has severe consequences on the hydrogen bonding pattern formed by the respective ligands (**Figure 4**).

While the homopiperazine portions of **1** (**Figure 4A**), **2** (**Figure 4B**) and **5** (**Figure 4E**) form several direct hydrogen bonds to the protein, none are formed by ligands **3** (**Figure 4C**) and **4** (**Figure 4D**). However, all ligands form in all adopted poses a hydrogen bond between their respective isoquinoline nitrogen and the backbone nitrogen of Val123 which is part of the hinge region of the protein (**Figure 4F**). Interestingly, the conformation of the homopiperazine moiety of **3** corresponds exactly to that of **1**, considering that both exhibit the same stereochemistry of the methyl group. However, the hydrogen bond of **1** to Glu127 does not occur in the complex with **3**. This is due to the absence of the second alternative conformation of this amino acid in the complex of **3** which is found in the complex of **1** (**Figure 4A,C**).

For the other enantiomer of **3**, ligand **4**, the homopiperazine ring adopts a different conformation. In fact the ring conformation of **4** corresponds to that of **fasudil** (**5**) only having the nitrogen in another position. This altered nitrogen position of **4** leads to the disruption of the hydrogen bonds observed in the complex structure of **1**. This leaves both the complexes of ligand **3** and **4** with a total of a single hydrogen bond to the protein (**Figure 4F**).

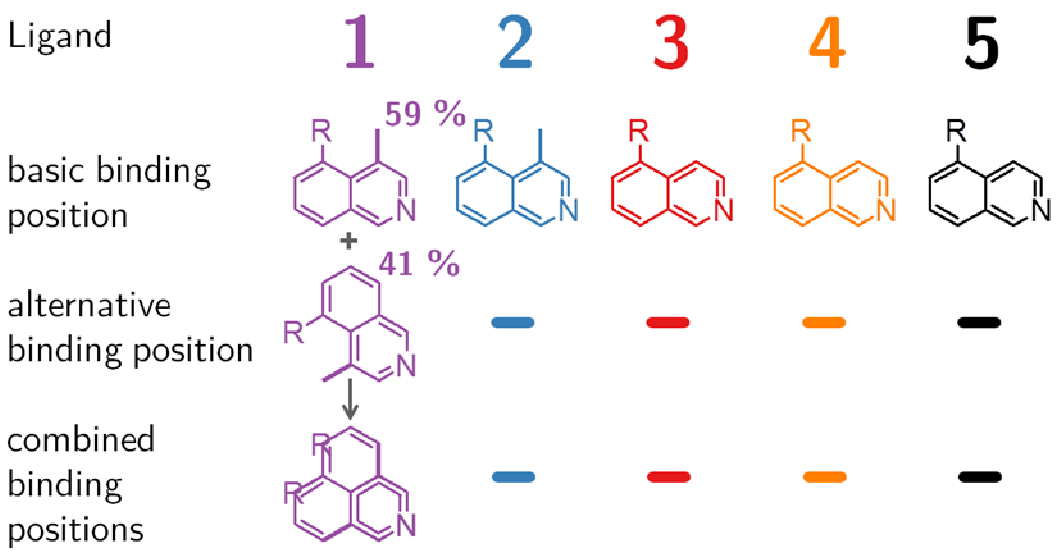


Figure 3. Schematic hinge binding pattern. All five ligands show a common basic hinge binding pose. However, ligand **1** shows an additional alternative pose where the ligand is flipped over and rotated by about 60° which is 41% populated. Color coding corresponds to the ligand numbering scheme.

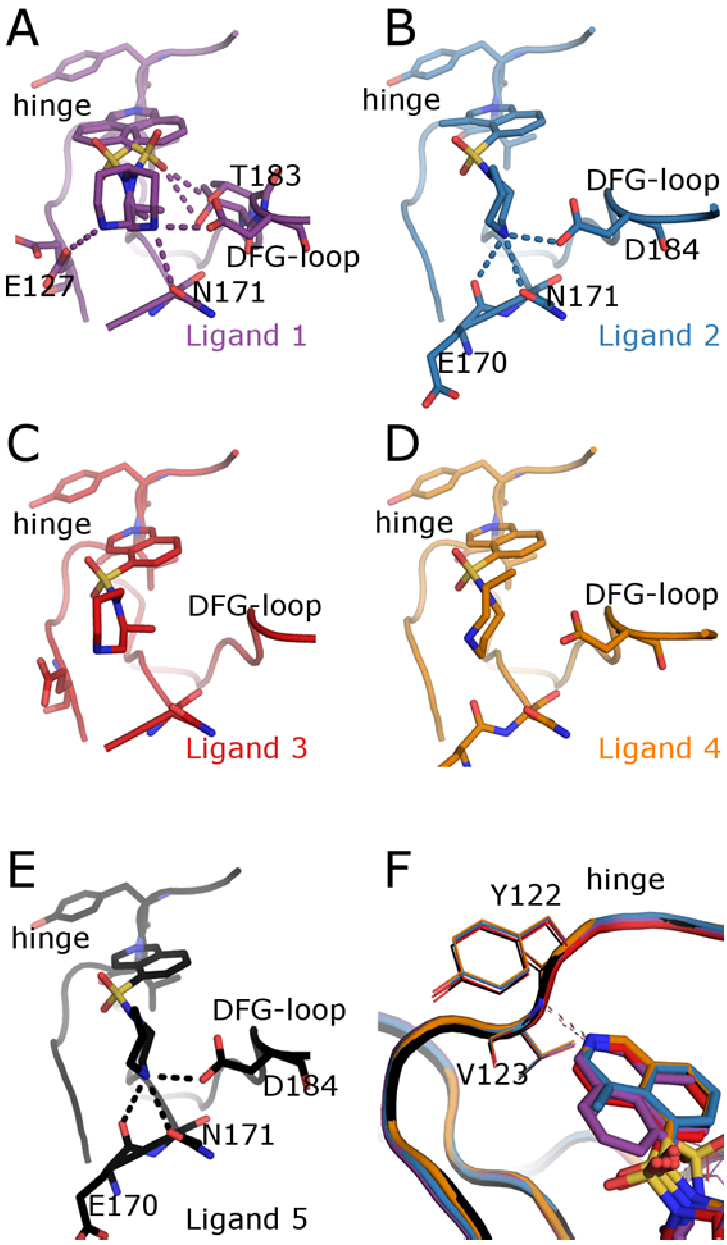


Figure 4. Protein-ligand hydrogen bonding patterns. Dotted lines indicate hydrogen bonds. **A:** 2-fold methylated ligand **1** in complex with PKA, both conformers form three different hydrogen bonds with the protein. Respective occupancies are 51% of conformation A and 49% of conformation B. **B:** Ligand **2** in complex with PKA. A total of four hydrogen bonds are formed. **C**: Complexe structure of *S*-configurated ligand **3**, only one hydrogen bond is formed to the hinge (see **F**). **D**: *R*-configurated ligand **4** complexed with PKA, also here one hydrogen bond is formed to the hinge (see **F**). **E**: Unmethylated ligand **5** forms four hydrogen bonds to PKA. **F**: Superposition of the hinge binding motif of all ligands.

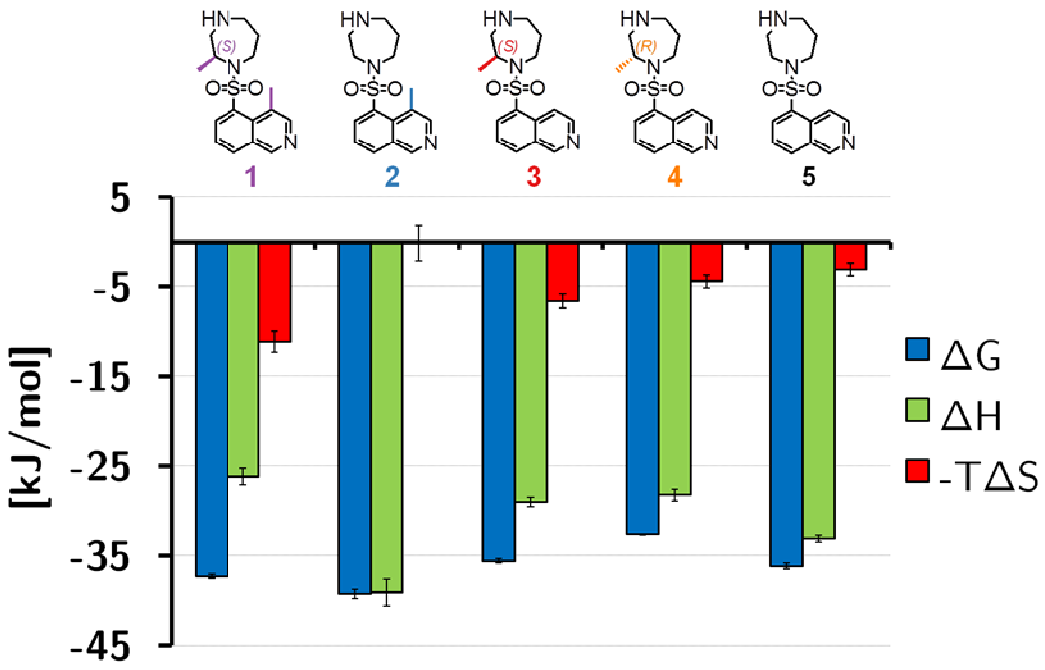


Figure 5. Thermodynamic signatures for ligands 1-5. Binding potency represented by Gibbs free energy (ΔG) and the partitioning in enthalpy (ΔH) and entropy (-TΔS) changes in consequence of the methyl groups. Ligand **2** is the most potent and enthalpic binder.

Non-additivity of the thermodynamic signatures. ITC data were used to analyze the binding thermodynamics of different ligands to the protein. **Figure 5** gives an overview of the measured thermodynamic signatures. The profile of the double methylated ligand **1** (**H-1152**) is not an additive combination of the profiles of the two single methylated ligands **2** and **3**. This is in accordance with the crystallographic data that demonstrate a distribution of **1** over two alternative binding modes whereas **2** and **3** each adopt only one orientation. This disorder provides ligand **1** with an entropic advantage over the other two ligands.

Methylation at isoquinoline portion is most favorable. If thermodynamic binding profiles of the three differently singly methylated ligands are compared to that of the non-methylated fasudil (**5**), modulations in affinity as well as enthalpy and entropy are observed. Methylation solely at the isoquinoline moiety leads to an improved affinity for enthalpy reasons. One explanation might be an inventory of preferred ligand conformations in solution governed by the presence or absence of the methyl group. The rotatable bond between the isoquinoline portion and the sulfonamide group of the ligand allows multiple orientations of the homopiperazine moiety relative to the isoquinoline scaffold. In the crystal structures however, this multiplicity is reduced to only one of these orientations. Upon introduction of the methyl group at the isoquinoline moiety of ligand **2**, the favored ligand conformation in solution might change. Due to steric reasons, the methyl group at the isoquinoline might promote a conformation where the homopiperazine portion is rotated away from the methyl group at the isoquinoline moiety. This conformation is close to the one in the crystal structure and hence makes the ligand conformation in the crystal structure enthalpically more favorable. Moreover, the methyl group is placed in a highly hydrophobic environment of the protein binding site (Leu49, Phe327). This enables the formation of additional van-der-Waals contacts between ligand and PKA without displacing a water molecule. Hence, this factor will promote enthalpy driven binding.

Ligands **3** and **4** bind enthalpically less favorable than ligand **2**. An explanation for this observation is the lack of polar H-bond interactions of the amino group in the methylated homopiperazine portion to the protein in both binding poses. Overall, **2** is the most potent inhibitor of this series. This

ligand profits from a beneficially positioned methyl group in the small hydrophobic niche, whereas the methyl groups introduced at the homopiperazine portion of **3** and **4** reduce binding potency, as the ring adopts in both cases an orientation that ruptures favorable H-bonds to be formed to Asp184 and Asn171 which is detrimental to the enthalpic contribution to binding. Remarkably, **fasudil** (**5**) which also establishes these H-bonds again experiences an enthalpic advantage compared to **3** and **4**. Ligand **1** exhibits also a profile with an enthalpically less favorable binding and an entropic benefit. Only in the partially populated binding pose the enthalpically favorable H-bonds are formed, whereas the disorder over two binding modes provides this ligand with an entropic advantage. Therefore, the affinity gain of **1** is primarily entropy driven.

4.6 Conclusion

Our study shows that not only the region but also the stereochemical attachement of methyl groups matter for binding potency and the thermodynamic inventory. Combination of different ligand methylation sites does not necessarily lead to an additive effect on binding properties observed for the single methylation sites. As observed in our case, a combination can alter the ligand binding mode in a hardly predictable way. However, this might be the result of a combination of a favorable with an unfavorable ligand methylation. All in all, the data reveal the importance of region and stereochemistry as well as the combination of both by methylating a ligand. These factors are cruicial to take into account when optimizing ligands. Hence, continuous control of binding modes by crystallography is highly

advised in drug design, even for such small ligand alterations as methylations.

4.7 Abbreviations

Ala: alanine

Asp: aspartate

Arg: arginine

ATP: adenosine triphosphate

CHO: Chinese hamster ovary

DFG: aspartate-phenylalanine-glycine motif

DMSO: dimethyl sulfoxide

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

Gly: glycine

Gly-loop: glycine-rich loop

Glu: glutamate

His-tag: histidine-tag

ITC: isothermal titration calorimetry

MBT: MES/Bis-Tris

MPD: 2-methyl-2,4-pentanediol

Ni-NTA: nickel-nitrilotriacetic acid

PDB: protein data bank

PKA: cAMP-dependent protein kinase

PKI: protein kinase inhibitor

qNMR: quantitative nuclear magnetic resonance spectroscopy

Ser: serine

TEA: triethanolamine

TEV: tobacco etch virus

Thr: threonine

Val: valine

4.8 References

1 Freire, E. *Drug Discov Today* **2008**, *13*, 869–874.

2 Ladbury, J. E.; Klebe, G.; Freire, E. *Nat Rev Drug Discov* **2010**, *9*, 23–27.

3 Kawasaki, Y.; Freire, E. *Drug Discov Today* **2011**, *16*, 985–990.

4 Núñez, S.; Venhorst, J.; Kruse, C. G. *Drug Discov Today* **2012**, *17*, 10–22.

5 Klebe, G. *Nat Rev Drug Discov* **2015**, *14*, 95–110.

6 Ákos Tarcsay,; Keserü, G. M. *Drug Discov Today* **2015**, *20*, 86–94.

7 Leung, C. S.; Leung, S. S. F.; Tirado-Rives, J.; Jorgensen, W. L. *J Med Chem* **2012**, *55*, 4489–4500.

8 Barreiro, E. J.; Kümmerle, A. E.; Fraga, C. A. M. *Chem Rev* **2011**, *111*, 5215–5246.

9 Bain, J.; Plater, L.; Elliott, M.; Shpiro, N.; Hastie, C. J.; Mclauchlan, H.; Klevernic, I.; Arthur, J. S. C.; Alessi, D. R.; Cohen, P. *Biochem J* **2007**, *408*, 297–315.

10 Luo, T.; Masson, K.; Jaffe, J. D.; Silkworth, W.; Ross, N. T.; Scherer, C. A.; Scholl, C.; Fröhling, S.; Carr, S. A.; Stern, A. M.; Schreiber, S. L.; Golub, T. R. *Proc Natl Acad Sci U S A* **2012**, *109*, 2860–2865.

11 Ikenoya, M.; Hidaka, H.; Hosoya, T.; Suzuki, M.; Yamamoto, N.; Sasaki, Y. *J Neurochem* **2002**, *81*, 9–16.

12 Nishio, M.; Watanabe, Y.; Hidaka, H. *J Pharmacol Exp Ther* **1998**, *287*, 1063–1067.

13 Jacobs, M.; Hayakawa, K.; Swenson, L.; Bellon, S.; Fleming, M.; Taslimi, P.; Doran, J. *J Biol Chem* **2006**, *281*, 260–268.

14 Sasaki, Y.; Suzuki, M.; Hidaka, H. *Pharmacol Ther* **2002**, *93*, 225–232.

15 Tamura, M.; Nakao, H.; Yoshizaki, H.; Shiratsuchi, M.; Shigyo, H.; Yamada, H.; Ozawa, T.; Totsuka, J.; Hidaka, H. *Biochim Biophys Acta* **2005**, *1754*, 245–252.

16 Yamaguchi, H.; Kasa, M.; Amano, M.; Kaibuchi, K.; Hakoshima, T. *Structure* **2006**, *14*, 589–600.

17 Shibuya, M.; Asano, T.; Sasaki, Y. In *Cerebral Vasospasm*; Seiler, R. W., Steiger, H.-J., Eds.; Springer Vienna: Vienna, 2001; 201–204.

18 Breitenlechner, C.; Gassel, M.; Hidaka, H.; Kinzel, V.; Huber, R.; Engh, R. A.; Bossemeyer, D. *Structure* **2003**, *11*, 1595–1607.

19 Bonn, S.; Herrero, S.; Breitenlechner, C. B.; Erlbruch, A.; Lehmann, W.; Engh, R. A.; Gassel, M.; Bossemeyer, D. *J Biol Chem* **2006**, *281*, 24818–24830.

20 Gilsbach, B. K.; Ho, F. Y.; Vetter, I. R.; van Haastert, P. J. M.; Wittinghofer, A.; Kortholt, A. *Proc Natl Acad Sci U S A* **2012**, *109*, 10322–10327.

21 Kudlinzki, D.; Linhard, V. L.; Saxena, K.; Sreeramulu, S.; Gande, S.; Schieborr, U.; Dreyer, M.; Schwalbe, H. *Acta Crystallogr F Struct Biol Commun* **2015**, *71*, 1088–1093.

22 Kabsch, W. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132.

23 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658–674.

24 Adams, P. D. et al. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–221.

25 Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486–501.

26 Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J Appl Crystallogr* **1993**, *26*, 283–291.

27 Kleywegt, G. J.; Zou, J.-Y.; Kjeldgaard, M.; Jones, T. A. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; Springer Netherlands: Dordrecht, 2001; 353–356.

28 Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. http://www.globalphasing.com.

29 Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. *J Chem Inf Comput Sci* **2004**, *44*, 2133–2144.

30 Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. *Anal Chem* **2012**, *84*, 5066–5073.

31 Houtman, J. C. D.; Brown, P. H.; Bowden, B.; Yamaguchi, H.; Appella, E.; Samelson, L. E.; Schuck, P. *Protein Sci* **2007**, *16*, 30–42.

32 Holzgrabe, U.; Deubner, R.; Schollmayer, C.; Waibel, B. *J Pharm Biomed Anal* **2005**, *38*, 806–812.

4.9 Supplementary information

Table S1.All atoms of all ligands were 100% occupied. For visualiazation, σ-weighted mF_o-DF_c-densities for all four ligands are shown.

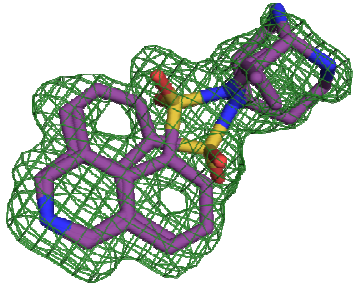
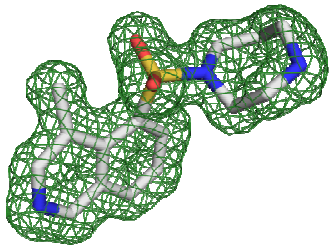
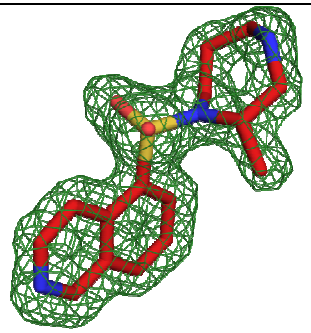
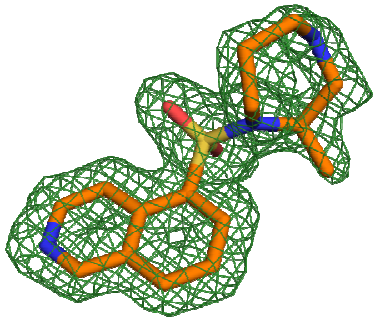
Ligand 1, at 3.0 σ 5M6V	Ligand 2, at 3.0 σ 5M6Y	Ligand 3, at 3.0 σ 5M75
		
Ligand 4, at 3.0 σ 5M71		
		

Table S2.Crystallographic table for all crystal structures. Table spreads over two pages.

PDB entry →	**1, 5M6V**	**2, 5M6Y**	**3,5M75**	**4, 5M71**
Data collection & processing				
No. Crystals used	1	1	1	1
Wavelength [Å]	0.976250	0.918409	0.918409	0.918409
Space group	19	19	19	19
Unit cell parameters: a, b, c [Å]	58.5, 73.2, 108.8	58.7, 73.1, 109.3	58.7, 73.1, 109.3	58.7, 72.9, 109.9
Diffraction data[a)]				
Resolution range [Å]	45.71-1.42	45.77-1.37()	45.46-1.54()	42.23-1.49()
Highest shell resolution range [Å]	1.50-1.42	1.45-1.37()	1.63-1.54()	1.58-1.49()
Unique reflections	88220(13855)	98699(15403)	68403(10782)	77336(12077)
R(I)sym [%][b)]	4.5(47.6)	3.3(48.4)	5.7(48.6)	5.0(48.2)
Completeness [%]	98.8(97.1)	98.7(96.5)	98.5(97.1)	99.0(97.0)
Redundancy	7.3(7.3)	4.0(3.9)	3.9(3.9)	4.3(4.1)
I/σ (I)	23.6(3.9)	20.1(2.3)	13.4(2.3)	15.20(2.4)
Refinement				
Resolution range [Å]	45.71-1.42	45.77-1.37	41.94-1.54	42.23-1.49
Reflections used in refinement (work/free)	83809/4410	93764/4935	64983/3420	73469/3867
Final R values for all reflections (work[c)]/free[d)]) [%]	14.4/17.4	14.4/16.6	14.5/17.4	14.7/17.6
Amino acids (PKA/PKI)	353/20	353/20	353/19	353/20
Inhibitor atoms	22	21	21	21
Water molecules	437	420	389	429
RMSD from ideality				
Bond length [Å]	0.008	0.008	0.008	0.008

Bond angles [°]	1.034	0.989	0.973	0.960
Ramachandran plot[e)]				
Residues in favoured regions [%]	91.5	97.7	91.5	91.8
Residues in additionally allowed regions [%]	8.5	7.3	7.9	8.2
Residues in generously allowed regions [%]	0.0	0.0	0.6	0.0
Mean B-factors [Å²]				
PKA (protein)/PKI (peptide)	21.6/23.6	24.1/24.2	24.0/23.2	24.8/24.8
Inhibitor	18.5	20.6	18.5	22.9
Water molecules	30.8	32.7	32.3	33.1

a)Numbers in parentheses are for the highest-resolution shell.

b)$R_{\mathrm{sym}}=[\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}|I_{\mathrm{i}}(h)-\langle I(h)\rangle|/\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}I_{\mathrm{i}}(h)]\times 100$, $\langle I(h)\rangle$ is the mean of the $I(h)$ observation of reflection h.

c)$R_{\mathrm{work}}=\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}-F_{\mathrm{c}}|/\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}$.

d)Calculation of R_{free} was performed as for R_{work} but on 5 % of the data which was excluded from the refinement.

e)Derived from Procheck.[26]

Table S3. Thermodynamic data and standard deviations for all five ligands.

Ligand	ΔG [kJ/mol]	ΔH [kJ/mol]	-TΔs [kJ/mol]
1	-37.3±0.2	-26.2±0.9	-11.1±1.2
2	-39.2±0.5	-39.1±1.5	-0.1±2.0
3	-35.6±0.2	-29.0±0.5	-6.5±0.8
4	-32.6±0.1	-28.2±0.6	-4.4±0.7
5	-36.1±0.3	-33.1±0.4	-3.1±0.7

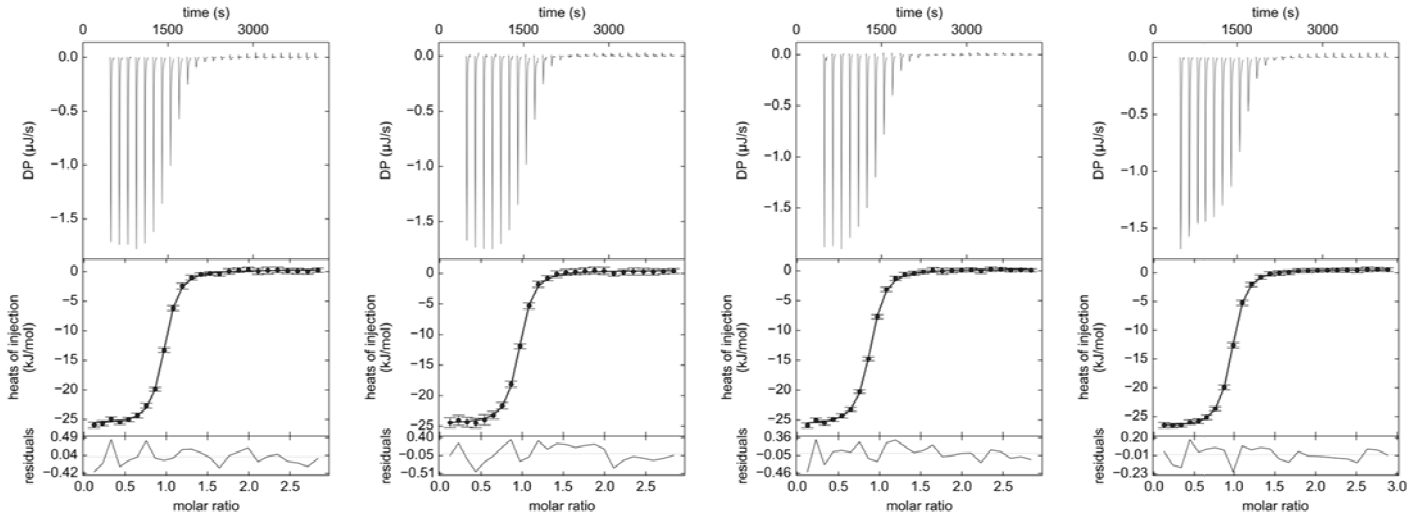


Figure S1. ITC-titration curves for ligand 01 in phosphate buffer as exemplary titration curves.

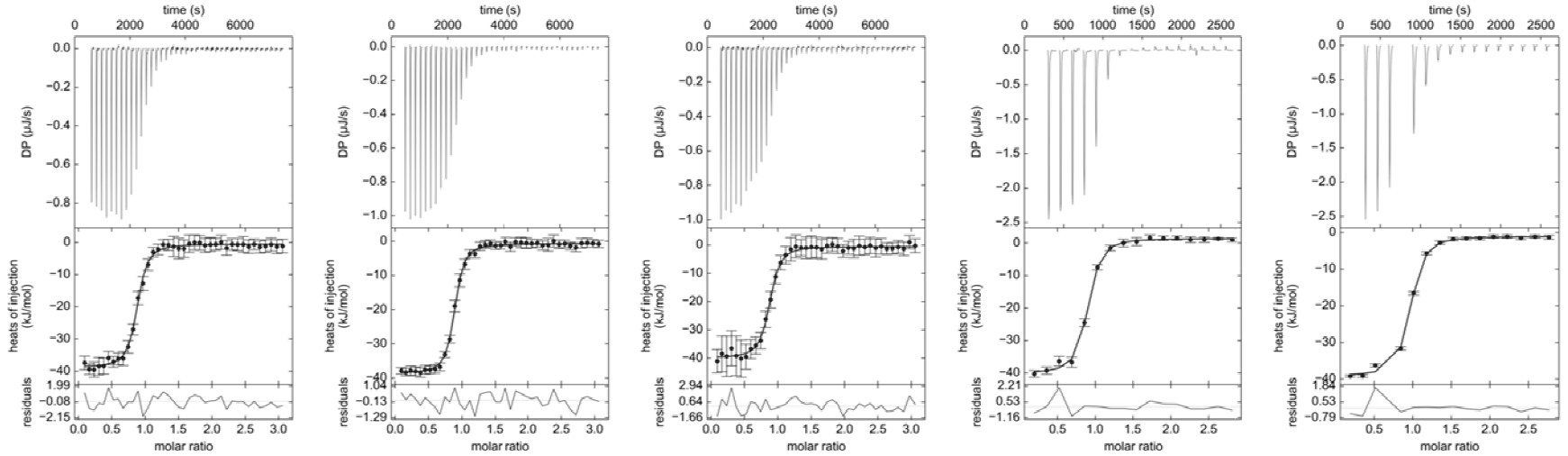


Figure S2. ITC-titration curves for ligand 02 in phosphate buffer as exemplary titration curves.

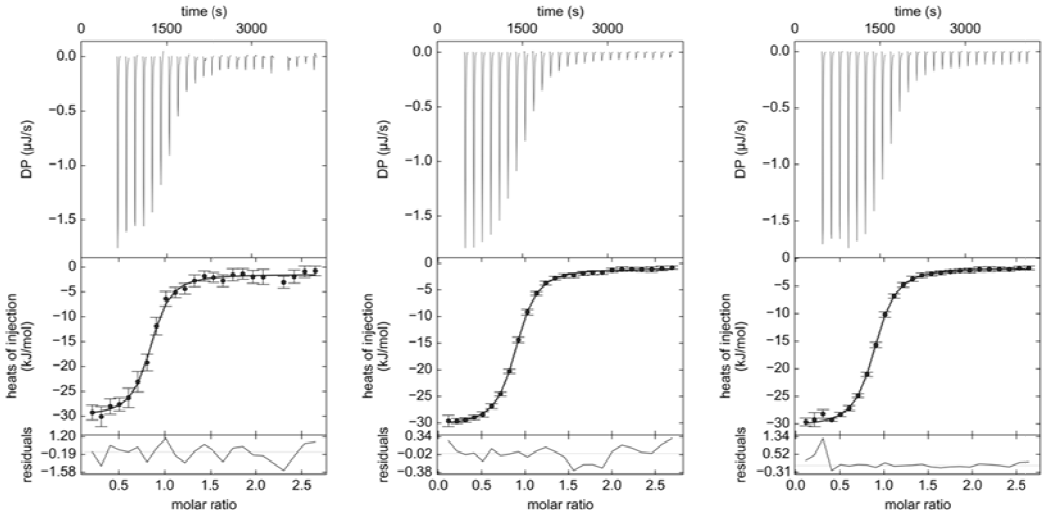


Figure S3. ITC-titration curves for ligand 03 in phosphate buffer as exemplary titration curves.

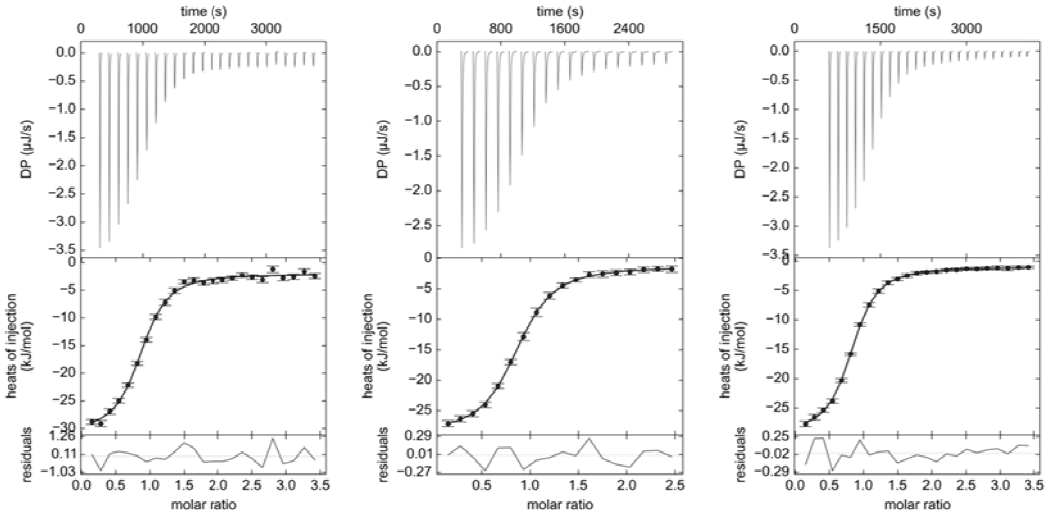


Figure S4. ITC-titration curves for ligand 04 in phosphate buffer as exemplary titration curves.

5 Small Changes, Big Effect: Variations in the Hinge Binding Modes

5.1 Annotations

Crystallization of all compounds and fragments was performed by Barbara Wienen-Schmidt. Structure refinement for all structures except for isoquinoline and adenine were entirely done by Barbara Wienen-Schmidt. For isoquinoline and adenine data processing and few cycles of automatic refinement were performed by Barbara Wienen-Schmidt. Final refinement was performed by Matthias Oebbeke in the framework of his Master thesis and crystallographic tables for these two compounds will be published in his thesis.

5.2 Abstract

Addressing the hinge binding motif is of utmost importance for the development of ATP-competitive kinase inhibitors. In this chapter, the influence of minor chemical changes and ligand fragmentation on the ligand-kinase complex formation is investigated. Therefore, a series of ten ligands was crystallographically analyzed. The resulting structures and ligand binding positions were compared and could be classified into four groups showing different hinge interaction pattern. Strikingly, for specific ligands minor changes in chemical structure resulted in major changes in the ligand's hinge binding mode. The same observation was made for fragments derived thereof. A subset of the fragments did not bind the hinge in the same way as the larger compound having the same fragment included. These findings are

highly relevant for drug optimization in a fragment-based drug discovery approach. Assumptions about conserved hinge binding modes upon ligand growing and evolvement have to be verified by crystallography. Moreover, the results of this study suggest that biophysical data such as thermodynamics or kinetics of protein-ligand complex formation should only be interpreted and correlated if a conserved position of the basic ligand binding scaffold has been crystallographically confirmed.

5.3 Introduction

As set out in the previous chapters, protein kinases are a major target in drug discovery. Different kinase inhibitor types have been developed, targeting active or inactive kinase conformation, in- or outside of the binding pocket.[1] However, the majority of kinase inhibitors target the hinge region of the protein.[2] The hinge region comprises the deeply buried far end of the active site of kinases. It connects the small N-terminal lobe with the bigger C-terminal lobe (**Figure 1**). Since the hinge binding portion of the ligand is crucial for ligand potency, most ligands that target the active site form at least one interaction with the hinge backbone. The hinge binding fragments themselves usually exhibit above average ligand efficiency compared to ligands evolved from this fragment as starting point. Therefore, the hinge binding motif is of utmost importance for affinity, but likely less for selectivity. Selectivity is predominately influenced by interactions to other protein areas such as the backpocket.[3] Nevertheless, novel hinge binding scaffolds are important to enable novel intellectual property, which is why new hinge binding motifs are of particular interest for pharmaceutical companies.[2]

Physiologically, the backbone amide and carbonyl functions of the hinge region interact with the adenine portion of adenosine triphosphate (ATP) to recognize the substrate during the catalytic phosphorylation process. The unique feature of the hinge is the virtually linear arrangement of three functionalities: One donor function that is framed by two acceptor functions (**Figure 2**).[4] The area around the hinge region is a hydrophobic cleft, which is why aromatic ring portions are preferred comprised in hinge binding motifs of potential drugs. Hydrogen-bond donors and acceptors within the heterocyclic ring system can form complementary polar interactions to the donor and acceptor functions of the hinge backbone.[4] The goal is the formation of optimal hydrogen bonds in an else hydrophobic environment.

This chapter focuses on the different hinge binding modes observed for different ligands that are derived from the drug fasudil as well as fragments thereof. Furthermore, hinge-binding modes of fragments deriving from ATP are discussed. Some of the ligands have already been presented in the previous chapters.

Here, as well as in the other chapters, cAMP-dependent protein kinase (PKA) is used as a model protein. For PKA, the crucial part of the hinge is formed by the backbone carbonyl of Glu121 (acceptor 1), the backbone amide nitrogen of Val123 (donor) and the backbone carbonyl of Val123 (acceptor 2) (**Figure 2**). The ligands presented here all interact with the donor function of Val123 and partially in addition with the acceptor function of Glu121.

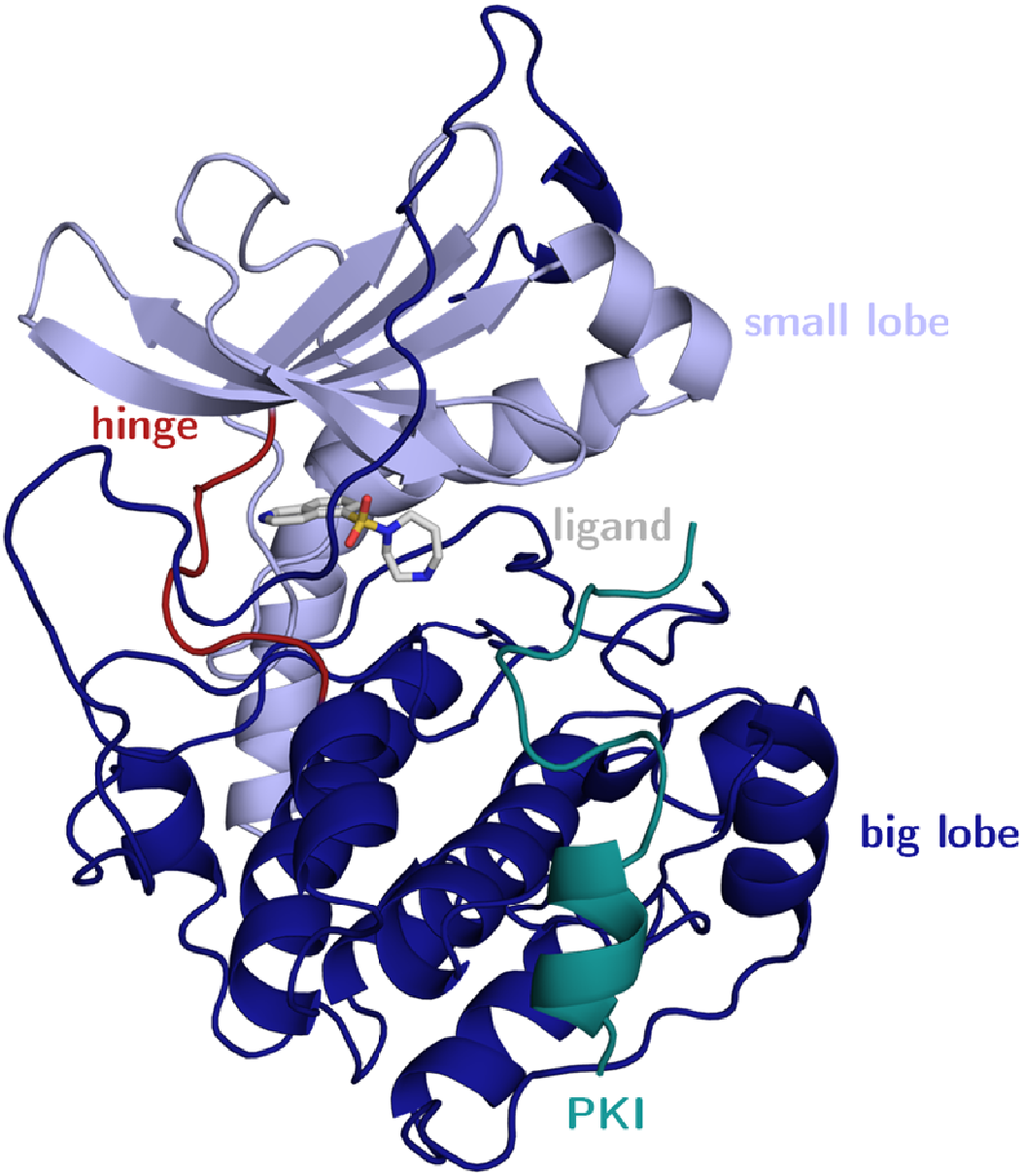


Figure 1. Overview of a crystal structure of PKA. The hinge region (red) connects the small lobe (light blue) with the big lobe (dark blue). It is an important part of the catalytic cleft between the two lobes. Physiologically, the adenine portion of ATP binds to the hinge. The γ-phosphate of ATP is then transferred to a substrate protein. The substrate protein binds in the region of PKI (teal), which is a peptidic inhibitor mimicking a protein substrate and facilitates protein crystallization.

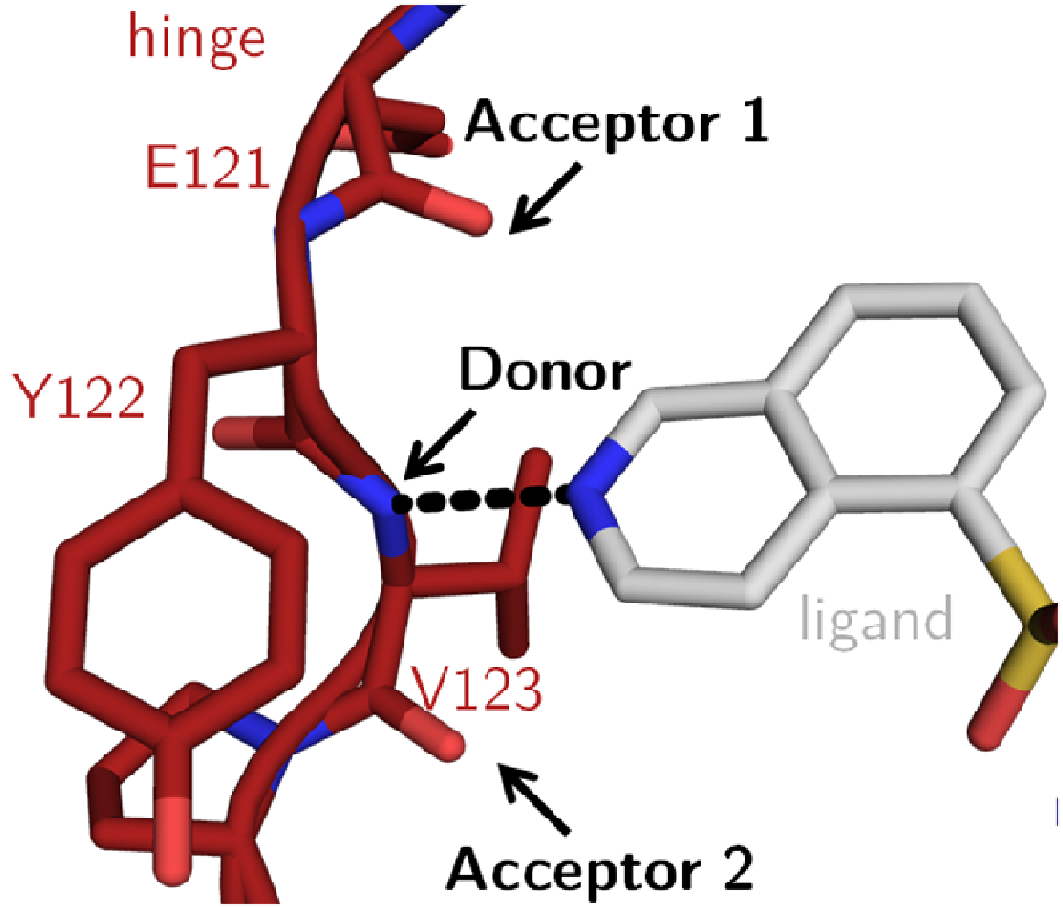


Figure 2. The hinge displays a nearly linear arrangement of three functionalities: The backbone carbonyl of Glu121 as an acceptor (1), the backbone amide nitrogen of Val123 as a donor and the backbone carbonyl of Val123 as a second acceptor.

Only minor chemical and structural modifications or fragmentation of the fasudil-and ATP-type ligands were made. Surprisingly, this led to strong differences in the hinge binding modes of several ligands. In the process of drug optimization and in fragment-based drug design (FBDD), changes in the hinge binding modes are an often disregarded aspect. For drug optimization, minor changes such as the addition of a hydroxyl or methyl group to the ligand are not expected to entirely change ligand binding

behavior. Also in FBDD, fragments that are merged or stepwise grown to form drug-sized molecules are assumed to roughly remain in their original binding position when embedded into a larger scaffold.[5-8] This study demonstrates that this assumption is too superficial. The crystallographic data presented here, demonstrates that fragments can have an unpredictable hinge binding behavior if compared to a larger scaffold having this fragment embedded as a key functionality.

5.4 Experimental section

Protein expression and purification

The catalytic subunit of cAMP-dependent protein kinase from Chinese hamster ovary cells was expressed with a His-tag in a modified pET16b-Vector with an introduced TEV-cleavage site between the protein N-terminus and His-tag. This plasmid was transformed into *E. coli* strain *Bl21 (DE3)/pLysS* (Novagen).[9]

Cell disruption was performed using a high-pressure homogenizer for multiple cycles. After centrifugation (1h at 30.000g) cell lysate supernatant was purified in a first step using a Ni-NTA column that binds the His-tag of the protein and was eluted by an imidazole gradient. The His-tag was then cleaved off by TEV-protease. Afterwards, an inverse Ni-NTA column was employed collecting PKA in the flow-through. Finally, ion exchange chromatography was performed using a MonoS column separating three-fold phosphorylated PKA from the four-fold phosphorylated form using a HEPES buffer with a sodium chloride gradient.[9]

Crystallization

Crystallization of **hydroxyfasudil**, **isoquinoline**, **1-aminoisoquinoline**, **4-quinazolamine** and **5-isoquinoline carboxylic acid** will be described in the following. Crystallization of all other ligands (**fasudil**, **H-1152**, **isoquinoline-5-sulfonamide**) has already been described in the previous chapters.

Co-crystallisation was performed using the hanging drop method at 4 °C. The crystallization drops contained the following ingredients: 5-10 mg/ml PKA (120-240 µM), 30 or 110 mM MBT (MES/Bis-Tris Puffer pH 6.9) (110 mM MBT buffer was only used for **5-isoquinolinesulfonic acid** and **5-isoquinolinecarboxylic acid** to buffer their acidity), 1 mM DTT, 0.1 mM EDTA, 75 mM LiCl, 0.03 mM Mega 8, 0.07 mM PKI (Sigma: P7739), 10 mM ligand dissolved in DMSO from a 50-100 mM stock. The well contained a mixture of methanol in water with varying methanol concentrations (v/v) for the different ligands (**hydroxyfasudil**: 14 % methanol; **isoquinoline**: 24 % methanol; **adenine**: 20 % methanol; **4-quinazolamine**: 20 % methanol; **1-aminoisoquinoline**: 20 % methanol; **5-isoquinolinesulfonic acid**: 11 % methanol; **5-isoquinolinecarboxylic acid**: 24 % methanol). In the crystallization setup streak-seeding was performed with apo-crystals as seeds using a horse hair in order to initialize crystal growth. For crystal mounting, crystals were cryo protected in 5 mM MBT (MES/Bis-Tris Puffer pH 6.9), 1 mM DTT, 0.1 mM LiCl, 10 mM ligand dissolved in DMSO from a 50-100 mM stock, 16 % (v/v) methanol, 30% (v/v) MPD and flash frozen in liquid nitrogen.

Crystallography

All cocrystal structures of **fasudil**, **H-1152** and **isoquinoline-5-sulfonamid** are discussed in the second, third and fourth chapter. All other structures were collected at the storage ring Bessy II Helmholtz-Zentrum Berlin, Germany at Beamline 14.1 on a Pilatus 6M pixel detector. The datasets were processed using XDS[10] and molecular replacement was performed using CCP4 Phaser[11] and PDB-structure of PKA from *bos taurus* 1Q8W as a model. This was followed by simulated annealing, multiple refinement cycles of maximum likelihood energy minimization and B-factor refinement with Phenix[12]. Coot[13] was used to fit amino-acid side chains into σ-weighted 2Fo – Fc and Fo – Fc electron density maps. If appropriate electron density was observed, multiple side chain conformations were built into the model and maintained during the refinement if the minor populated side chain displayed at least 20 % occupancy. Ramachandran plots for structure validation were calculated using PROCHECK[14]. Data collection, unit cell parameters and refinement statistics are given in the supplementary information. Analysis of temperature factors was performed with Moleman[15]. Protein and PKI B-factors were anisotropically refined, water B-factors were isotropically refined for the structure of **hydroxyfasudil**. Structures of **isoquinoline**, **adenine**, **1-aminoisoquinoline**, **4-quinazolamine**, **5isoquinoline sulfonic acid** and **5-isoquinoline carboxylic acid** were TLS-refined. For the definition of the TLS groups the TLSMD server was used.[16, 17] Decision for anisotropic or TLS refinement was based on comparison of R_{free}. Anisotropic refinement was chosen over TLS if the achieved R_{free} values were at least 0.5% lower for anisotropic than for TLS refinement. R_{free} was calculated using 5% of all reflections which were randomly chosen and not used for the refinement. The

required ligand restraint files were created using the Grade webserver[18, 19]. For figure preparation Pymol was used. Crystallographic tables can be found in the supporting information.

Ligands

Fragments **isoquinoline**, **1-aminoisoquinoline**, **4-quinazolamine** and **5-isoquinoline carboxylic** were purchased at sigma aldrich. **Hydroxyfasudil** was purchased at Toronto Research Chemicals (Canada). Vendors of all other ligands are listed in the previous chapters.

5.5 Results and discussion

A schematic overview of all ligands and their respective hinge binding pose relative to each other is presented in **Figure 3** and **Figure 4**. **Figure 3** displays the hinge-binding fragments from **fasudil**, namely **isoquinoline**, and that of **ATP**, which is **adenine**. In addition, two fragments with an "intermediate" chemical structure of isoquinoline and adenine were investigated (**1-aminoisoquinoline** and **4-quinazolamine**). For comparison, the crystal structure of ATP in complex with PKA from *mus musculus* deposited in the protein data bank (PDB) with the code 1ATP was analyzed.[20]

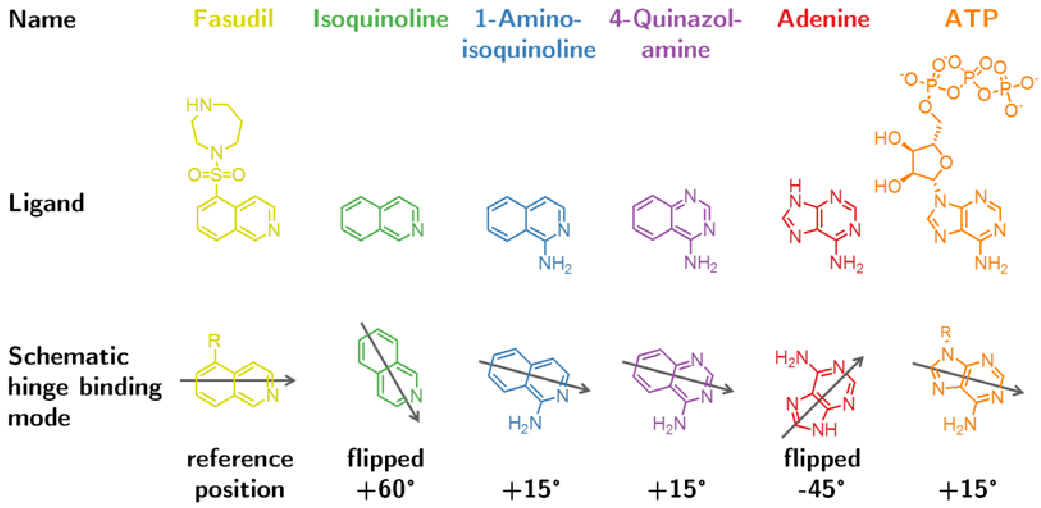


Figure 3. Schematic overview of the ligands and their respective hinge binding poses relative to each other. Fasudil and its hinge binding fragment **isoquinoline**, as well as **ATP** and its hinge binding fragment **adenine** are shown. In addition the two fragments with an "intermediate" chemical structure of isoquinoline and adenine, **1-aminoisoquinoline** and **4-quinazolamine**, were investigated. The binding mode of **fasudil** is defined as the reference. Angles of rotation refer to the orientation of its aromatic ring system as indicated by gray arrows.

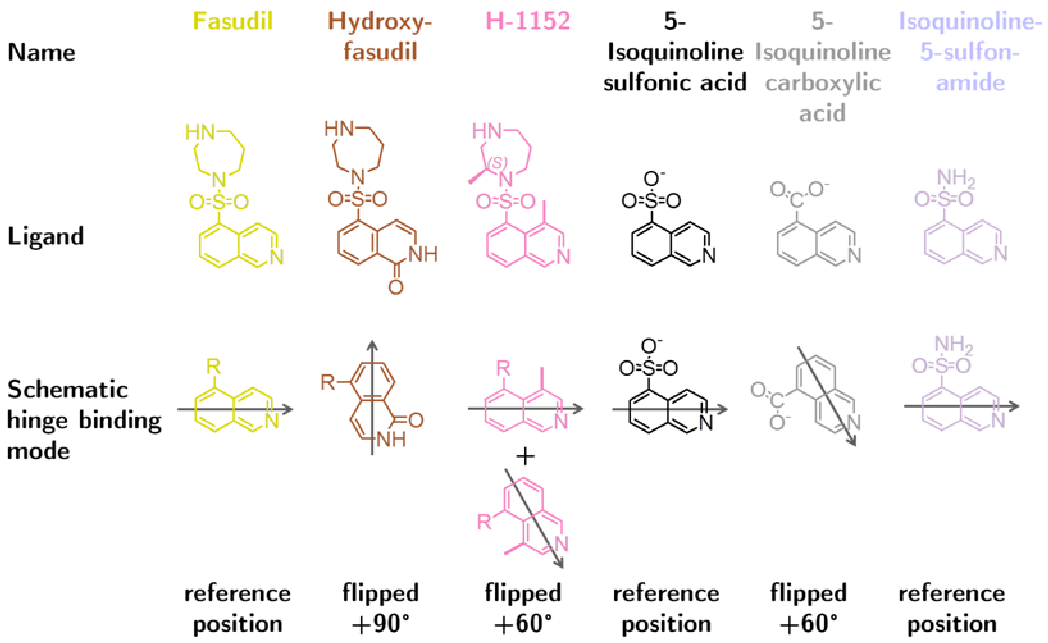


Figure 4. Schematic overview of the ligands and their respective hinge binding poses relative to each other. Hydroxyfasudil and **H-1152** as slightly modified fasudil-derivatives display an entirely different hinge binding mode. **Isoquinoline-5-sulfonamide** and **5-isoquinoline sulfonic acid** bind with a similar hinge binding pose as **fasudil**. If the sulfonic acid is replaced by a carboxylic acid, the binding mode changes severely. The binding mode of **fasudil** is defined as the reference. Angles of rotation refer to the orientation of its aromatic ring system as indicated by gray arrows. **Hydroxyfasudil** is presented as its isoquinolinone tautomer with a protonated nitrogen, this has been reported previously.[21]

ATP and adenine do not share a common hinge binding mode. The binding mode of the two hinge-binding elements of fasudil and **ATP** show a similar binding mode, the long axis through the adenine portion is only slightly rotated by about 15° compared to the orientation of the isoquinoline moiety of **fasudil** (**Figure 5**). Due to their similar position, these two hinge binding modes can be summarized as similar.

Both ligands form a hydrogen bond to the backbone nitrogen of Val123. In addition, **ATP** interacts with acceptor 1 via its exocyclic amino function. However, if we compare the binding mode of **ATP** to that of its fragment **adenine** (**Figure 6**) it becomes apparent, that they do not share a common binding orientation. **Adenine** is flipped and rotated about 30°along its long axis compared to the adenine portion of **ATP**. The more, it is striking that the hinge residues which interact with **ATP** and **adenine** are identical. The deviating orientation results from the fact that **adenine** forms hydrogen bonds via its endocyclic N3 and N9 amino functionality. While in **ATP** the corresponding amino groups of N1 and the exocyclic NH_2 function is used. Most likely due to the attached ribose-triphosphate moiety the binding mode adopted by adenine is not accessible for **ATP**.

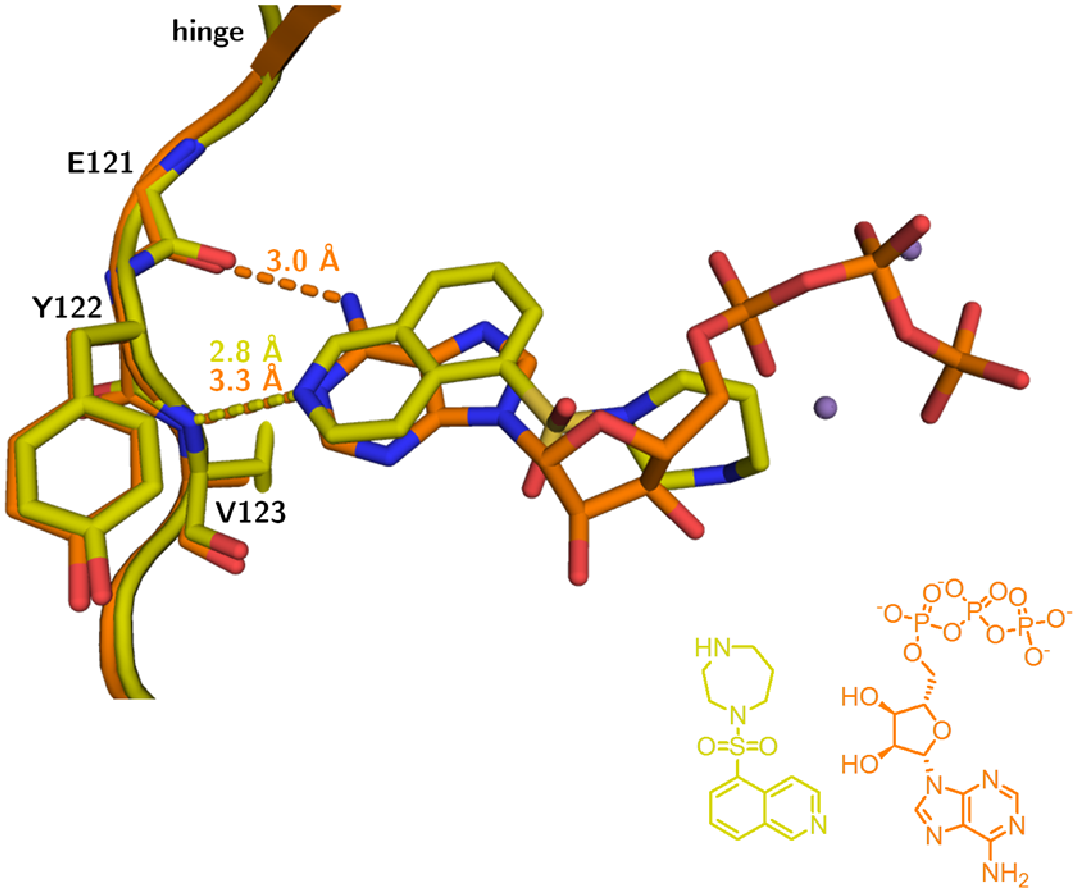


Figure 5. Superposition of the co-crystal structures of PKA with fasudil and ATP. Crystal structure of **ATP** in complex with PKA from *mus musculus* is taken from the protein data bank (PDB), 1ATP.[20] The hinge binding motifs of both ligands are similar. The adenine moiety is rotated about the long axis by 15° compared to the isoquinoline portion of **fasudil**. Hydrogen-bond formation to Val123 is identical for both ligands. In addition, ATP-adenine interacts with the backbone carbonyl of Glu121.

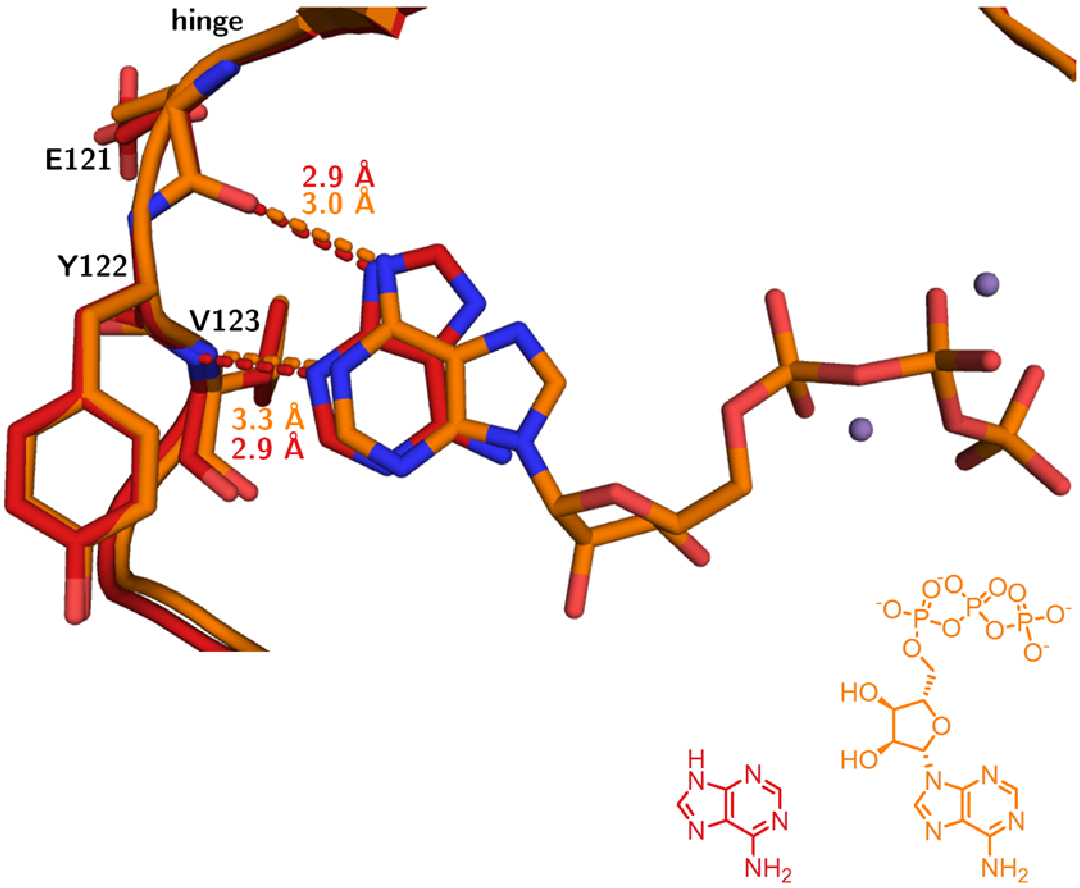


Figure 6. Superposition of the co-crystal structures of PKA with ATP and adenine. Crystal structure of **ATP** in complex with PKA from *mus musculus* is taken from the protein data bank (PDB), 1ATP.[20] Compared to the adenine portion of **ATP**, **adenine** is flipped and rotated by 30° about the long axis. However, hydrogen bonding partners provided by the hinge motif are equal in both cases.

Fragments with the "intermediate" chemical structure of adenine and isoquinoline display the same hinge binding mode as ATP. In contrast to **adenine**, the crystal structures of **1-aminoisoquinoline** and **4-quinazolamine** share a common hinge binding mode with **ATP** (**Figure 7**) and can be assigned to the same class of hinge binders. This results from the absence of an appropriate amino group at the position of N3 and N9 in

adenine. For **1-aminoisoquinoline** none of the hydrogen bonds could be formed if it adopted an orientation similar to that of **adenine**.

In comparison, **4-quinazolamine** does have the corresponding N3 nitrogen in position 1 of the quinazoline, yet the second nitrogen corresponding to the N9 position of adenine is absent. Hence, an analogous binding mode as observed for **adenine** would reduce the number of hydrogen bonds of **4-quinazolamine** to the protein. Thus, the binding mode adopted by adenine would also be unfavourable for **4-quinazolamine**.

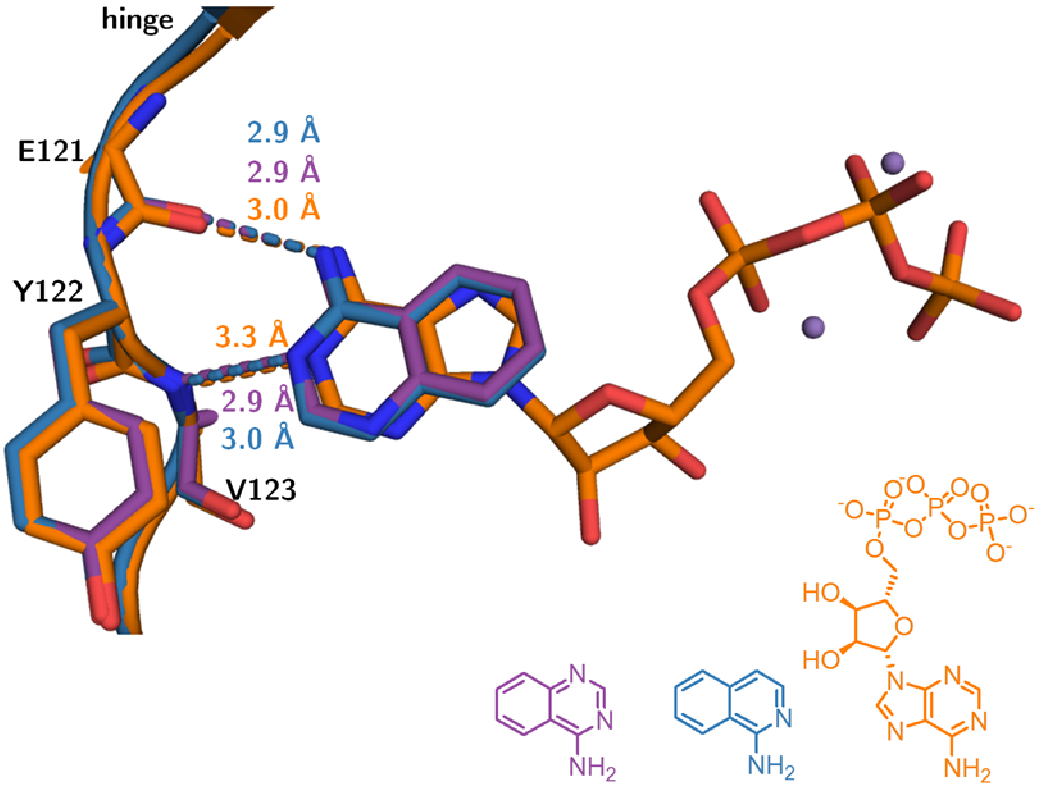


Figure 7. Superposition of the co-crystal structures of PKA with ATP, 1-aminoisoquinoline and 4-quinazolamine. Crystal structure of **ATP** in complex with PKA from *mus musculus* is taken from the protein data bank (PDB), 1ATP.[20] Both fragments share a common hinge binding mode with respect to the adenine portion of **ATP**.

Fasudil and part of its fragments share a common hinge binding mode. Upon comparison of the binding mode of **fasudil** and its two fragments **isoquinoline-5-sulfonamide** and **5-isoquinoline sulfonic acid**, an identical position of their respective hinge binding moieties can be observed (**Figure 8**). Only the position of the sulfonamide and sulfonic acid functions differ slightly from that in **fasudil**, but they match very closely for both fragments. The reason for this rotation of the sulfonamide group in the **isoquinoline-5-sulfonamide**-PKA complex has already been discussed in **chapter 3**: For steric reasons only the small fragments can form an interaction to Asp127 with their respective sulfonamide portions.

Nevertheless, it is interesting to see that the ligand with the sulfonic acid group as H-bond acceptor function, shares a common binding mode with the ligand possessing a sulfonamide as donor function at the same position.

Fasudil and isoquinoline do not share a common binding mode. The compared binding modes of the smaller fragment **isoquinoline** and the drug molecule **fasudil** possessing an embedded isoquinoline moiety is shown in **Figure 9**. Interestingly, the **isoquinoline** is flipped and rotated by 60° along its long axis compared to the position of the isoquinoline moiety in **fasudil**. Nevertheless, the isoquinoline nitrogen remains in position, thereby keeping the hydrogen bond to the hinge region intact. This binding mode establishes a further class of hinge binding modes.

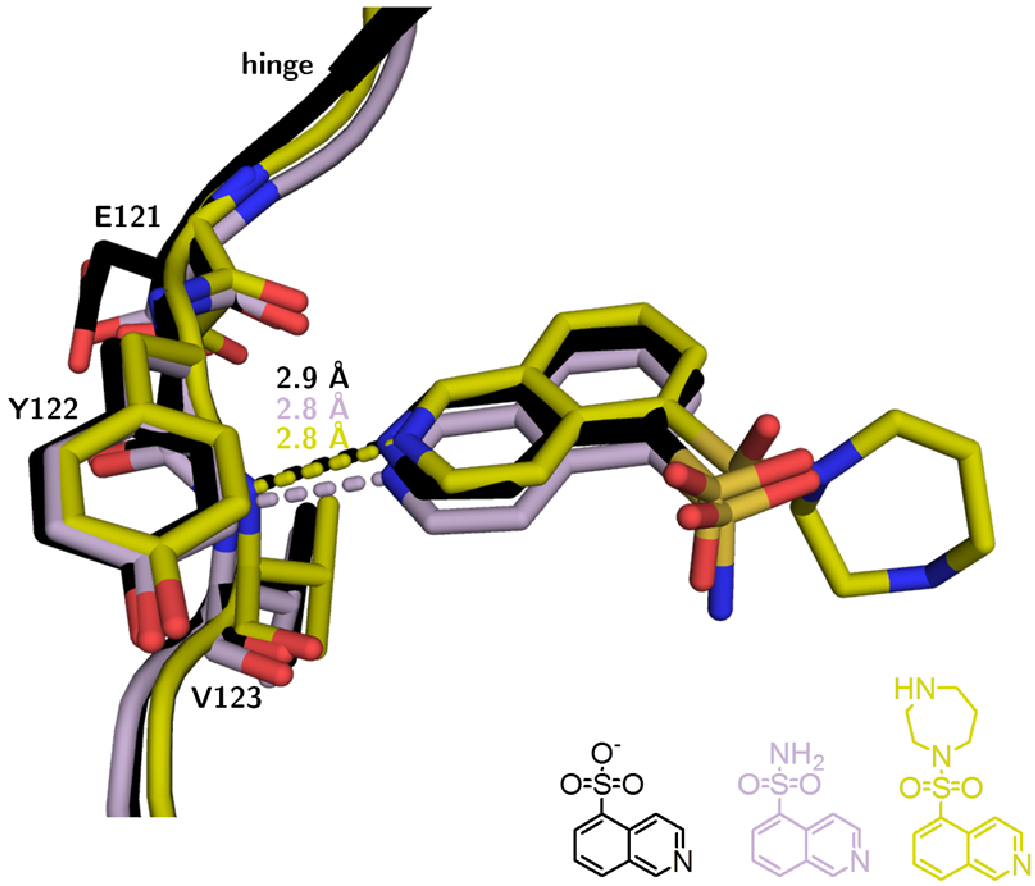


Figure 8. Superposition of the co-crystal structures of PKA with fasudil, isoquinoline-5-sulfonamide and 5-isoquinoline sulfonic acid. Both fragments occupy the same position using their hinge binding motif similarly to **fasudil**.

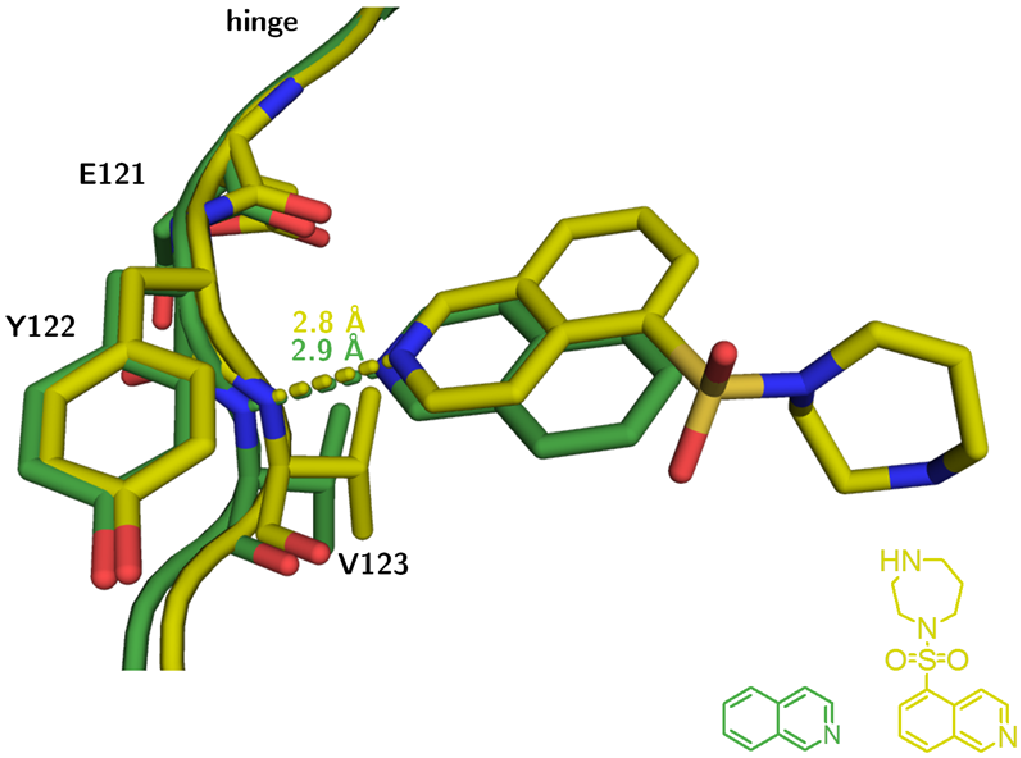


Figure 9. Superposition of the co-crystal structures of PKA with fasudil and isoquinoline. Isoquinoline adopts a hinge binding mode that is flipped and rotated 60° about the ligand's long axis compared to that of **fasudil**.

The hinge binding mode of isoquinoline is observed for other ligands. The comparison of the binding modes of **isoquinoline** and **5-isoquinoline carboxylic acid** reveals an analogous orientation of the isoquinoline moiety in both structures (**Figure 10**). This is particularly interesting, because **5-isoquinoline sulfonic acid**, an alternative fragment with an acceptor function in a similar position, does not share this hinge binding mode (**Figure 11**). This might be due to steric reasons: While the carboxylic acid is planar, the sufonic acid features a tetrahedral geometry. Strikingly, the binding modes of **isoquinoline** and **5-isoquinoline**

carboxylic acid are identical to the second, alternative binding mode observed for the methylated fasudil-derivative **H-1152** (**Figure 12**). This indicates that both, the basic hinge binding position and the flipped, 60° rotated position (**Figure 3** and **Figure 4**), appear to be relevant and energetically quite similar.

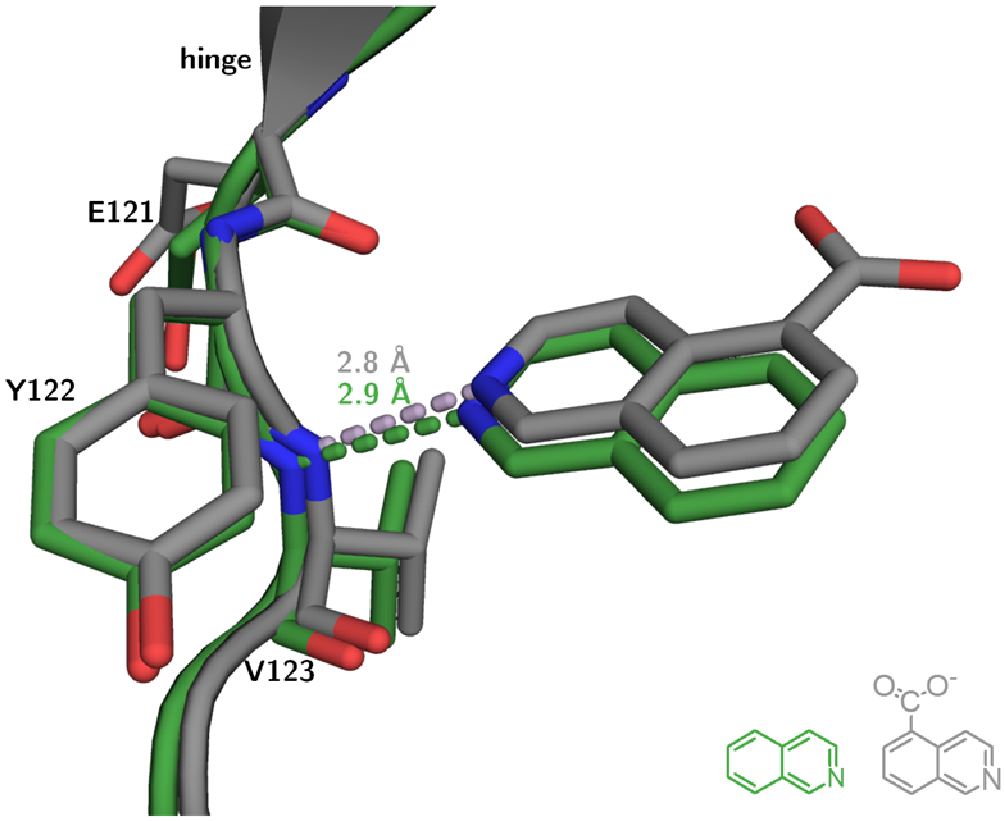


Figure 10. Superposition of the co-crystal structures of PKA with 5-isoquinoline carboxylic acid and isoquinoline. Both fragments share a common hinge binding mode.

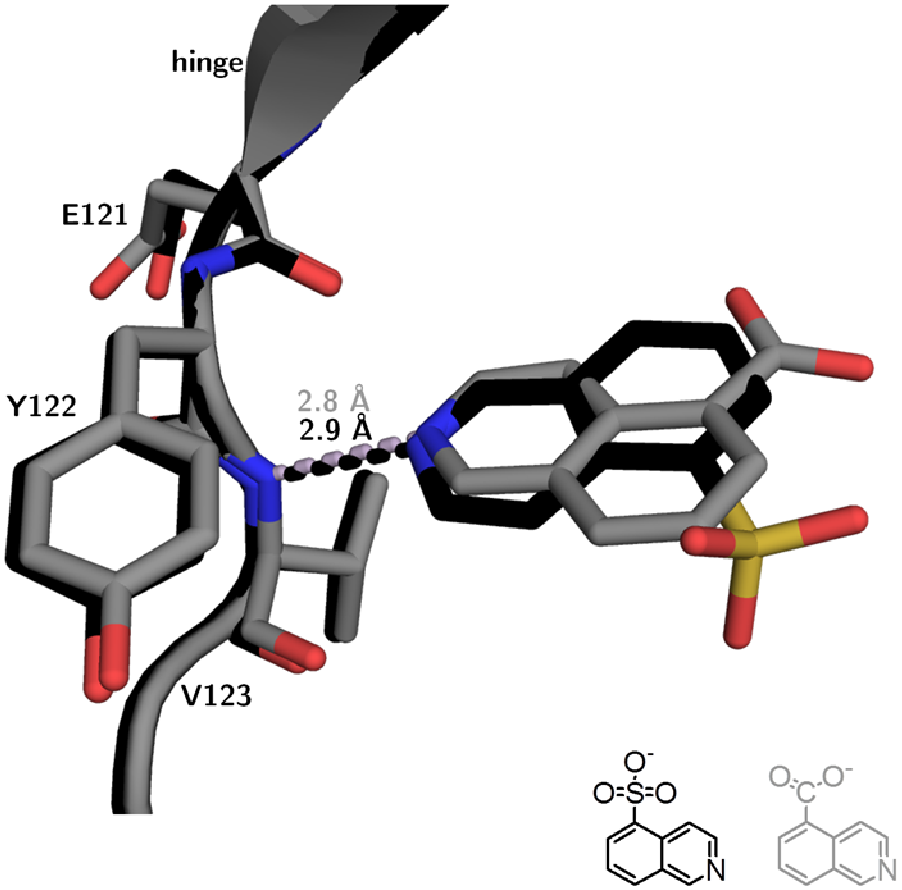


Figure 11. Superposition of the co-crystal structures of PKA with 5-isoquinoline sulfonic acid and 5-isoquinoline carboxylic acid. Both fragments comprise an isoquinoline moiety with an additional carboxylic or sulfonic acid as group H-bond donor function. Nevertheless, the observed binding modes are different from each other.

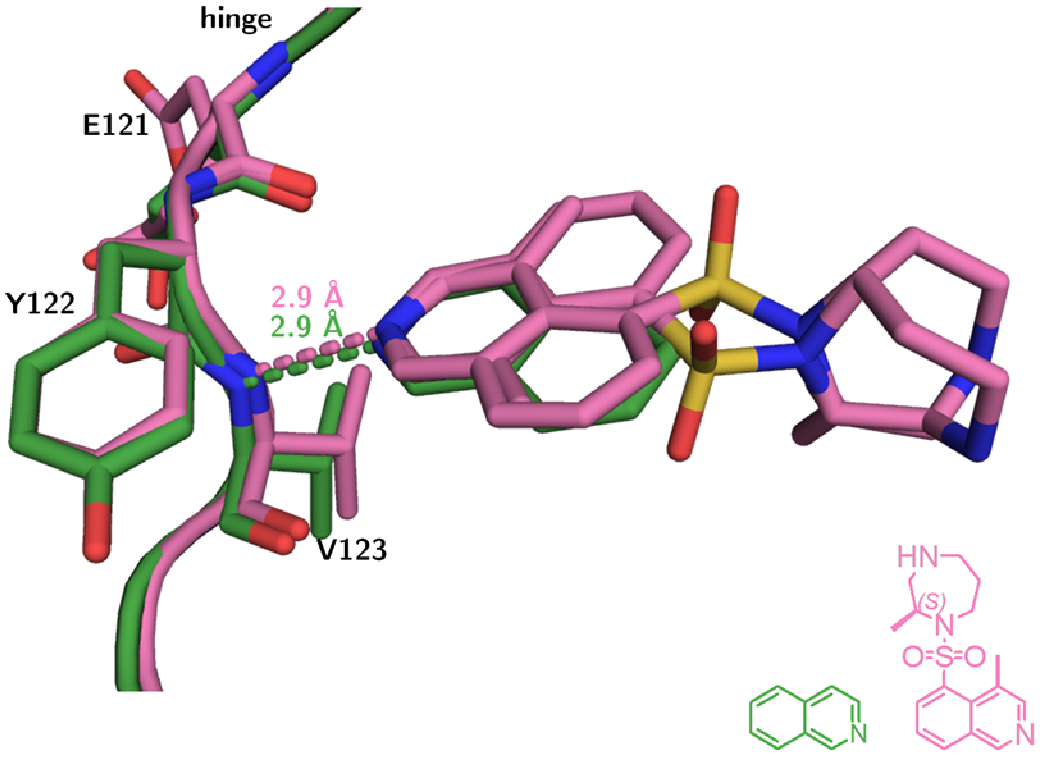


Figure 12. Superposition of the co-crystal structures of PKA with isoquinoline and H-1152. The ligand binding mode populated to 49% of **H-1152** corresponds to that of **isoquinoline**, while the 51% occupied one relates to the orientation of the isoquinoline portion of **fasudil**.

Hydroxyfasudil exposes a fourth hinge binding mode. **Hydroxyfasudil** binds in a new orientation, not observed for any of the other ligands. It has previously been reported to bind differently from **fasudil**.[21] The binding mode to PKA from *mus musculus* has already been published.[21] It corresponds to the binding mode we observed in the current study for PKA from Chinese hamster ovary (CHO). As **Figure 13** depicts, the hinge binding motif is flipped and rotated by about 90°. The hydrogen bond to the donor functionality of hinge region (Val123) is formed via the oxygen of the isoquinolinone moiety. The adjacent nitrogen of the

isoquinolinone-moiety of **hydroxyfasudil** must be protonated, since it donates the hydrogen to the H-bond with the backbone carbonyl acceptor of Glu121. The presence of this tautomer of **hydroxyfasudil**, in contrast to the fully aromatic isoquinolinol with the deprotonated nitrogen, has previously been reported for this protein-ligand complex.[21]

Exhibiting a distance of 2.8 Å and 2.7 Å, both of the hydrogen bonds formed by **hydroxyfasudil** are comparable to the one of **fasudil** (2.8 Å).

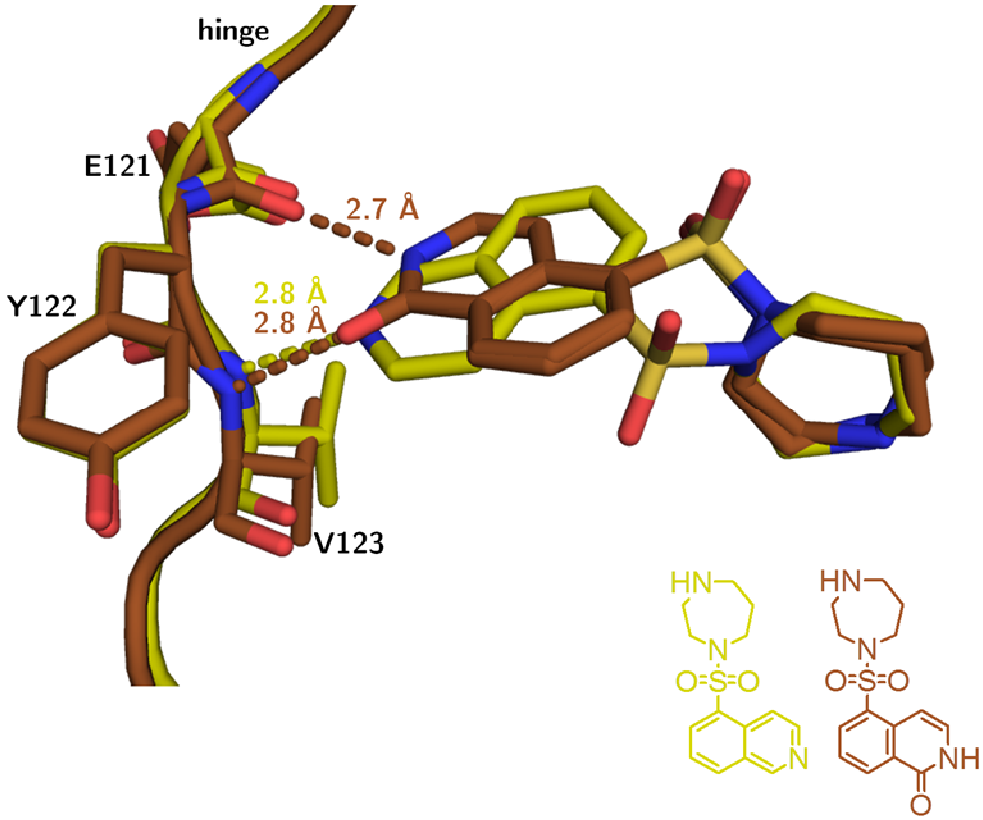


Figure 13. Superposition of the co-crystal structures of PKA with fasudil and hydroxyfasudil. The isoquinoline moiety of **hydroxyfasudil** is flipped and rotated 90° around the long axis through the isoquinoline portion. It is the only structure where the isoquinoline-nitrogen does not interact with the acceptor functionality of Val123. Instead a contact to Val123 is formed by the oxygen of the isoquinolinone moiety.

5.6 Conclusion

In total four different hinge binding modes could be defined for the structurally related scaffolds of quinoline or purine type. All of the scaffolds exhibit an isoquinoline, a quinazoline or purine-based moiety as a hinge binding motif. Binding mode one includes: **Fasudil**, **isoquinoline-5-sulfonamide**, **5-isoquinoline sulfonic acid**, **ATP**, **1-aminoisoquinoline**, **4-quinazolamine** and one of the two alternative binding modes of **H-1152**. Binding mode two is adopted by **isoquinoline**, **5-isoquinoline carboxylic acid**, and by the second binding mode of **H-1152**. The third binding mode is only represented by **adenine** and the fourth merely by **hydroxyfasudil**.

Despite their structural similarity in terms of chemical formular, even closely related ligands or fragments thereof can exhibit striking and novel hinge binding positions. It is crucial to consider such aspects in the process of FBDD and drug optimization. Linking or growing fragments into potential drug candidates can change the behavior of the ligands' key functions in a way that is still hardly predictable.

Vice versa discrepancies in the binding modes of fragments and already characterized compounds (detected by approaches other than FBDD) having these fragments embedded could be exploited to create novel drug candidates with improved affinity. If a known compound (e.g. a high-throughput screening hit) is virtually fragmented and these fragments are then crystallographically analyzed, the potentially different binding modes of the fragments can provide ideas for an alternative connection of the different functionalities in a way that does not violate the crystallographically discovered fragment binding modes. Taking the **isoquinoline** fragment and

fasudil as an example, the **isoquinoline** binding pose would suggest an attachement of the sulfonamide-homopierazine moiety at position C6 instead of C5 of isoquinoline. Using a novel topology possibly leads to new drug candidates with possibly improved or at least altered properties. Hence, new drug candidates could be developed from former high-throughput screening hits by fragmenting the latter and subsequently analyzing the corresponding fragment binding modes. Certainly, an optimized fragment linking strategy will not always be sterically possible and synthetically accessible. Nonetheless, in some cases this could even open a perspective for an easy road to success.

Furthermore, this study underscores the importance of continuously controlling binding modes by crystallographic analysis. Interpretation of biophysical data such as thermodynamics or kinetics is pointless without the verification of a conserved overall ligand binding mode. For instance, thermodynamic profiles of adenine and ATP cannot easily be compared, simply because the binding pose changes. The same is true for **isoquinoline** and **fasudil**. Only if the key functionalities of the grown ligands and the initial fragments remain in unchanged position, development of thermodynamic and kinetic properties can be thoroughly understood. If key elements of the ligand binding modes change, too many factors influencing biophysical parameters will possibly overlap and complicate the picture. So far, it is impossible to clearly differentiate to what extend all the overlapping factors distort the readout of such experiments.

5.7 Abbreviations

Ala: alanine

Asp: aspartate

Arg: arginine

ATP: adenosine triphosphate

CHO: Chinese hamster ovary

DMSO: dimethyl sulfoxide

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

FBDD: fragment-based drug design

Gly: glycine

Glu: glutamate

His-tag: histidine-tag

MBT: MES/Bis-Tris

MPD: 2-methyl-2,4-pentanediol

Ni-NTA: nickel-nitrilotriacetic acid

PDB: protein data bank

PKA: cAMP-dependent protein kinase

Ser: serine

TEV: tobacco etch virus

Thr: threonine

TLS: translation/libration/screw

Val: valine

5.8 References

1 Roskoski, R. *Pharmacol Res* **2016**, *103*, 26–48.

2 Akritopoulou-Zanze, I.; Hajduk, P. J. *Drug Discov Today* **2009**, *14*, 291–297.

3 Klebe, G. *Drug Design: Methodology, concepts, and mode-of-action*; Springer Publishing Company, Incorporated, 2013; 599–640.

4 Stroud, R., Finer-Moore, J., Eds. *Computational and Structural Approaches to Drug Discovery*; RSC Biomolecular Sciences; The Royal Society of Chemistry, 2007; 001–382.

5 Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. *Nat Rev Drug Discov* **2004**, *3*, 660–672.

6 Hajduk, P. J.; Greer, J. *Nat Rev Drug Discov* **2007**, *6*, 211–219.

7 de Kloe, G. E.; Bailey, D.; Leurs, R.; de Esch, I. J. P. *Drug Discov Today* **2009**, *14*, 630–646.

8 Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. *Nat Rev Drug Discov* **2016**, *15*, 605–619.

9 Kudlinzki, D.; Linhard, V. L.; Saxena, K.; Sreeramulu, S.; Gande, S.; Schieborr, U.; Dreyer, M.; Schwalbe, H. *Acta Crystallogr F Struct Biol Commun* **2015**, *71*, 1088–1093.

10 Kabsch, W. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132.

11 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658–674.

12 Adams, P. D. et al. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–221.

13 Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486–501.

14 Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J Appl Crystallogr* **1993**, *26*, 283–291.

15 Kleywegt, G. J.; Zou, J.-Y.; Kjeldgaard, M.; Jones, T. A. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; Springer Netherlands: Dordrecht, 2001; 353–356.

16 Painter, J.; Merritt, E. A. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 439–450.

17 Painter, J.; Merritt, E. A. *J Appl Crystallogr* **2006**, *39*, 109–111.

18 Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. http://www.globalphasing.com.

19 Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. *J Chem Inf Comput Sci* **2004**, *44*, 2133–2144.

20 Zheng, J.; Trafny, E. A.; Knighton, D. R.; Xuong, N.-H.; Taylor, S. S.; Ten Eyck, L. F.; Sowadski, J. M. *Acta Crystallogr D Biol Crystallogr* **1993**, *49*, 362–365.

21 Jacobs, M.; Hayakawa, K.; Swenson, L.; Bellon, S.; Fleming, M.; Taslimi, P.; Doran, J. *J Biol Chem* **2006**, *281*, 260–268.

5.9 Supplementary information

Table S1.Crystallographic table for all crystal structures. Table spreads over two pages.

Ligand →	**Hydroxyfasudil**	**4-Quinazol-amine**	**1-Amino-isoquinoline**	**5-Isoquinoline sulfonic acid**	**5-Isoquinoline carboxylicacid**
Data collection & processing					
No. Crystals used	1	1	1	1	1
Wavelength [Å]	0.918409	0.918409	0.918409	0.918409	0.918409
Space group	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)
Unit cell parameters: a, b, c [Å]	58.5; 72.9;109.5	58.1; 73.1;107.6	58.1; 73.1;107.6	58.3; 73.1;109.0	58.5; 73.3;109.5
Diffraction data[a)]					
Resolution range [Å]	45.61-1.53	45.47-1.71	45.48-1.69	45.601.79	45.70-1.88
Highest shell resolution range [Å]	1.53-1.63	1.82-1.71	1.79-1.69	1.90-1.79	1.88-2.0
Unique reflections	68873(10692)	45909(6206)	48217(6544)	43652(6891)	37996(5794)
R(I)sym [%][b)]	3.7(47.8)	4.5(39.6)	5.1(44.8)	5.1(47.8)	9.3(49.6)
Completeness [%]	97.3(94.7)	91.8(77.9)	92.6(78.9)	98.0(97.2)	98.0(94.0)
Redundancy	4.0(3.9)	4.1(3.1)	5.1(4.2)	4.2(3.7)	6.4(4.6)
I/σ (I)	23.0(2.8)	19.4(2.4)	19.3(2.6)	19.4(2.6)	12.8(2.6)
Refinement					
Resolution range [Å]	45.61-1.53	45.47-1.71	45.48-1.69	45.60-1.79	45.70-1.88
Reflections used in refinement (work/free)	65429/3444	43613/2296	45806/2411	41469/2183	36096/1900

Final R values for all reflections (work[c]/free[d]) [%]	14.7/17.7	17.7/21.3	18.2/21.9	18.4/22.5	19.4/24.2
Amino acids (PKA/PKI)	351/19	351/19	351/19	351/20	351/18
Inhibitor atoms	21	11	11	2x14	13
Water molecules	406	351	309	251	155
RMSD from ideality					
Bond length [Å]	0.009	0.006	0.009	0.011	0.011
Bond angles [°]	0.97	0.788	1.045	1.040	1.001
Ramachandran plot[e]					
Residues in favoured regions [%]	91.5	91.4	90.4	90.5	90.0
Residues in additionally allowed regions [%]	8.2	8.6	9.0	8.6	9.1
Residues in generously allowed regions [%]	0.3	0.0	0.6	0.9	0.9
Mean B-factors [Å²]					
PKA (protein)/PKI (peptide)	23.4/22.0	27.0/23.7	29.4/29.0	34.8/29.0	37.8/34.2
Inhibitor	26.5	21.8	22.9	40.2/44.9	41.2
Water molecules	32.7	32.5	32.4	33.0	30.9

a)Numbers in parentheses are for the highest-resolution shell.

b)$R_{\mathrm{sym}}=[\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}|I_{\mathrm{i}}(h)-\langle I(h)\rangle|/\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}I_{\mathrm{i}}(h)]\times 100$, $\langle I(h)\rangle$ is the mean of the $I(h)$ observation of reflection h.

c)$R_{\mathrm{work}}=\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}-F_{\mathrm{c}}|/\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}$.

d)Calculation of R_{free} was performed as for R_{work} but on 5 % of the data which was excluded from the refinement.

e)Derived from Procheck.[14]

6 Kinase Screen: Targeting Compound Selectivity

6.1 Abstract

Modern drug design, is frequently guided by the assumption that strong enthalpic binders are more selective than less enthalpic binding ones. Using the example of the cAMP-dependent protein kinase we characterized the thermodynamic binding profiles of 15 structurally related ligands using isothermal titration calorimetry. These 15 ligands were then tested in a kinase screen consisting of 39 human protein kinases. In addition staurosporine was used as an unspecific reference in the screen. Selectivity data from the kinase screen were then faced with the thermodynamic profiles against PKA. Strikingly no clear-cut correlation between ΔH and ligand selectivity emerged from the analysis. The more it is surprising that the most enthalpic binder displayed the worst selectivity apart from stauropsorine. However, a correlation between ΔG and ligands selectivity could be shown.

6.2 Introduction

Kinases are a protein family of major clinical relevance. The human genome migh encode for up to 518 protein kinases, several of which might be of pathological relevance.[1] In eukaryotes 30% of all proteins are temporarily reversibly phosphorylated, hence kinases are powerful switches to regulate activity regulation of signaling pathways. Anomalous kinase activity is particularly important in diseases that originate from inflammatory and proliferative responses. Examples are cancer, rheumatoid arthritis,

cardiovascular and neurological disorders, asthma and psoriasis. These implications make protein kinases a major target in drug design.[2-4]

Classical kinase inhibitors bind competitively to the ATP-binding site. However, not only kinases, but approximately 3000 proteins encoded in the human genome, use ATP as substrate in various ways. Unfortunately, this bears the risk that developed inhibitors bind unselectively in competitive manner to the classical ATP-binding sites.[5] This lack of selectivity is especially challenging across various members of the same kinase family as the target proteins. This is due to the pronounced structural similarity within the respective kinase families.

Over the last years, different types of inhibitors have been developed to overcome this lack of selectivity. Kinase inhibitors are subdivided into type I, type II, type I I/2, type III, type IV, type V and type VI. Type I inhibitors represent the group of classical ATP-competitive inhibitors that bind to the active protein, adopting the so-called DFG-in conformation. All ligands considered in this study are such type I inhibitors. Type II inhibitors have come up over the last couple of years. They bind to the protein's ATP-binding site in its inactive DFG-out conformation. As the inactive conformation is less geometrically conserved than the active conformation, these ligands are expected to be more selective than type I inhibitors.

For a DFG-out conformation, the DFG-loop and the corresponding phenylalanine are rotated away from the ATP binding site. Interestingly, not all kinases can adopt this conformation. So far no DFG-out conformation for PKA has been reported, thus it remains unclear if PKA can adopt this conformation. Type I I/2 inhibitors bind to an inactive conformation in the ATP-binding site that is, however, in DFG-in state. Type III inhibitors bind

adjacent to the ATP-binding site, whereas another allosteric inhibitor, type IV, binds far away from the active site. Type V inhibitors on the other hand, bind in a bivalent manner. The class of irreversible inhibitors is compiled through type VI inhibitors.[6]

Over the last years, biophysical properties of ligands have become increasingly popular as descriptors to explain failure and success in subsequent clinical studies. In consequence, drug candidates are frequently selected with respect to these properties and often kinetic rate constants or thermodynamic properties are consulted in this context. It has been hypothesized that one of such properties, namely the enthalpic contribution to the thermodynamic binding profile of a ligand, is favorably correlated with the selectivity of this ligand.[7]

Despite of this challenging hypothesis, the influence of the thermodynamic binding profile of putative drug candidates on their selectivity properties is still hard to estimate and needs a more in-depth investigation. This case study aims at providing data to shed some light on this promising hypothesis in terms of systematically collected experimental data. The hypothesis is tested for a series of compounds investigated in the previous chapters of this thesis. To do so, a kinase screen was performed, where 16 different ligands were tested (**Figure 1**) against 39 different kinases (**Figure 2**).

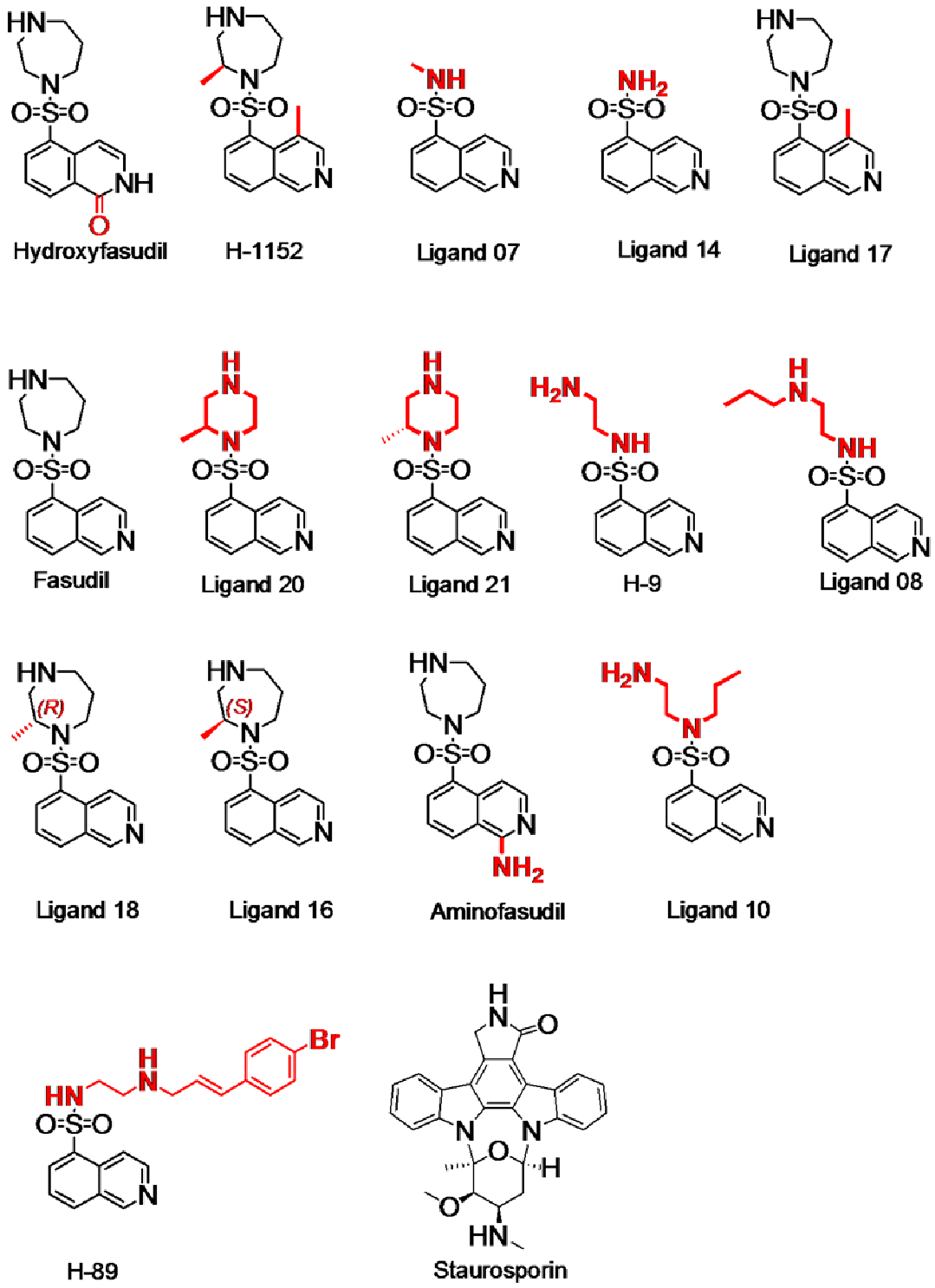


Figure 1: Overview of ligands tested in the kinase screen. The most discriminating structural moieties are highlighted in red.

The kinases selected for this screen are uniformly distributed over the entire human kinome. About 8-9% of each kinase family was selected as representatives (**Figure 2**). As PKA belongs to the family of AGC-kinases, we allowed a higher percentage of 19% of AGC-kinases in the screen. The results from the selectivity screen were then faced with the recorded profiles from the isothermal titration calorimetry (ITC) experiments. The measured and calculated values for ΔG, ΔH, $T\Delta S$ were each plotted against the results of the selectivity screen. **Staurosporine** was used as a known unselective reference in the kinase screen but has not been analyzed thermodynamically.

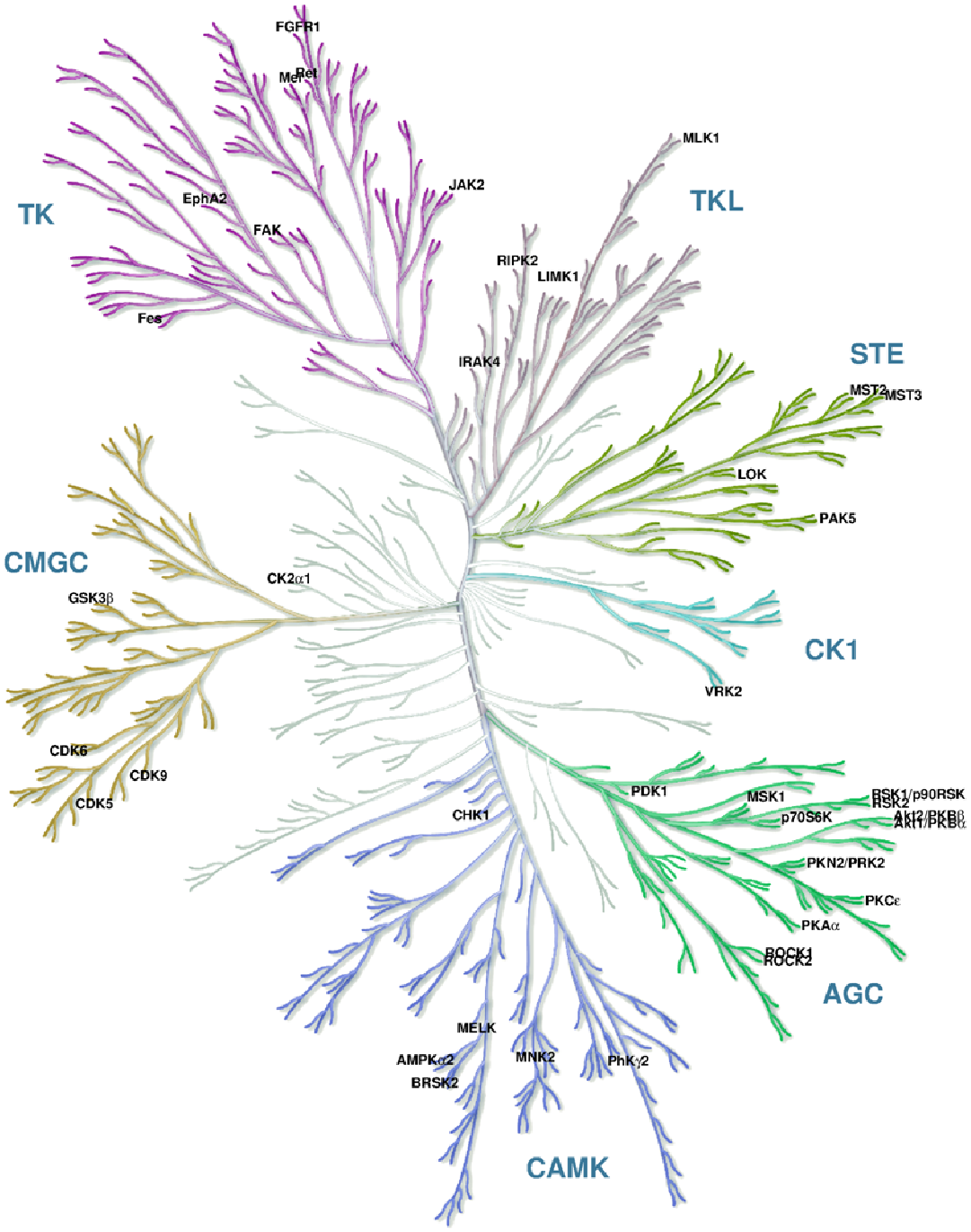


Figure 2: Overview of the human kinome. Only the kinases labeled in this image of the kinome were considered in the kinase screen. Kinase families are assigned and differ in color. Adapted from Manning 2002. Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com).[1, 8]

6.3 Experimental section

Protein expression and purification

The catalytic subunit of cAMP-dependent protein kinase from Chinese hamster ovary cells was expressed with a His-tag in a modified pET16b-Vector with an introduced TEV-cleavage site between the protein N-terminus and His-tag. This plasmid was transformed into *E. coli* strain *Bl21 (DE3)/pLysS* (Novagen).[9]

Cell disruption was performed using a high-pressure homogenizer for multiple cycles. After centrifugation (1h at 30.000g) cell lysate supernatant was purified in a first step using a Ni-NTA column that binds the His-tag of the protein and was eluted by an imidazole gradient. The His-tag was then cleaved off by TEV-protease. Afterwards, an inverse Ni-NTA column was employed collecting PKA in the flow-through. Finally, ion exchange chromatography was performed using a MonoS column separating three-fold phosphorylated PKA from the four-fold phosphorylated form using a HEPES buffer with a sodium chloride gradient.[9]

Thermodynamics

ITC measurements were performed as described in **Chapters 2-4**. Experimental values for all ligands can be found in the supplementary information. Detailed data including standard deviations and raw data for hydroxyfasudil is reported in the supplementary information as well.

Kinase Assay

Kinases selected for the screen were retrieved based on literature data[10-13] and a binding site similarity comparison using Cavbase[14]. Thereby, kinases with published crystal structures were selected. One criterion of the literature search was based on data reporting kinases to be inhibited by either fasudil, H-1152, H-89 or hydroxyfasudil. These four drug candidates were the only compounds from our subset that have been previously subjected in kinase inhibition screens. This protocol guaranteed that solely kinases were selected which had been shown to be inhibited by at least one member of our subset of compounds. Hence, receiving results showing a modulation of kinase inhibition levels appeared more likely and the risk of recording no inhibition at all and hence no difference in selectivity profiles was reduced. A comparable percentage of kinases from each family were chosen, however, a twice as large percentage of entries from the AGC-kinase family were considered.

Kinase inhibition assay was performed by *Eurofins Pharma Discovery Services UK Limited*. Details on the assay protocol are documented in the "KinaseProfiler Data Report Guide (v3)" as provided by the company.

The performed activity assay is a radiometric kinase assay. All ligands were measured at a concentration of 50 μM. The respective kinase is incubated with a substrate peptide and radioactive ATP (Adenosine triphosphate, radioactive gamma phosphate) at a concentration corresponding to that of the K_m-value of the respective kinase. In the catalytic process the radioactive gamma-phosphate of ATP is transferred to the peptidic substrate. The amount of transferred radioactive phosphate is then recorded. Upon inhibition of the kinase this catalytic process is hampered, resulting in a

signal decrease. All assays used for this inhibition screen involved direct measurement of substrate phosphorylation by the kinase under investigation and data was captured as duplicates. The raw data are measured by scintillation counting (in cpm). Positive control wells contain all components of the reaction except the inhibitor of interest; however, DMSO is included in these wells to control for solvent effects. Blank wells contain all components of the reaction with a reference inhibitor replacing the compound of interest. This calibrates kinase activity and establishes the base line (0 % kinase activity remaining). Results are expressed in terms of percentage of the mean kinase activity, with respect to the positive control samples. The positive control value is adjusted to be 100 %, and all test samples are reported relative to this value. For example, a result of 42 % means that, in comparison to the positive control, 42 % kinase activity remains in the presence of the test compound at a predefined concentration; or expressed differently, the test compound inhibits the kinase activity by 58 %. Residual activity percentages between 30 % and 70 % are considered partial hits. Beyond 30 % residual activity the tested compounds were classified as strong binders.

Experimental values from the screen can be found in the supplementary information.

6.4 Results

The original data from the duplicate measurements of the kinase screen refer to the residual catalytic activity (in %) of the different 39 kinases tested. The amount of strongly inhibited kinases (residual activity: 0 % to 30 %) for the individual test compounds varies between 3 (8 % of kinases tested) for the

most discriminating compound **7** and 37 (95 % of kinases tested) for the known unselective ligand **staurosporine**. Across the entire study an average of 14 (36 %) strongly inhibited kinases is found which underlines the purposeful selection of the study sample (**Figure 3**). When considering both, strongly and partially inhibited kinases (from 0 % to 70 % residual activity) this distribution scatters from 14 (36 %, ligand **14**) as minimum, to 39 (100 % of kinases tested, **staurosporine**) as maximum, with an average of 24 (62 %) inhibited kinases (**Figure 3**). Ligands that reduce activity of the majority of the regarded kinases (non-selective) and those that selectively block only few kinases are similar in both graphs. However, differences in selectivity between the ligands are more pronounced once the threshold is set to 30 %. According to this analysis the most selective inhibitors have to be assigned as **7** and **14** as both strongly inhibit no more than 15 of the studied kinases. In contrast ligands **H-1152**, **17**, **H-89** and unsurprisingly **staurosporine** display the most unspecific inhibition profiles as they strongly inhibit at least 50% of all 39 studied kinases.

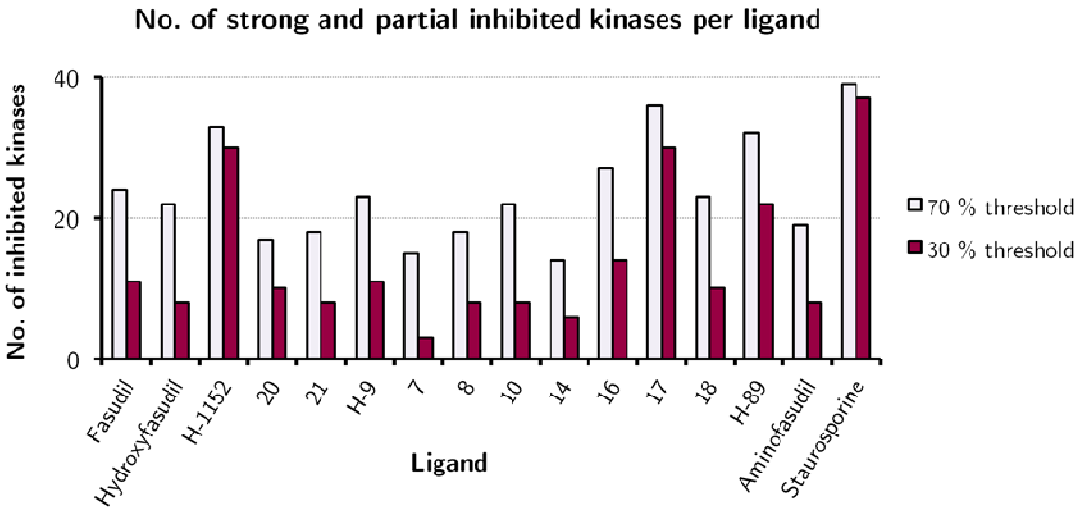
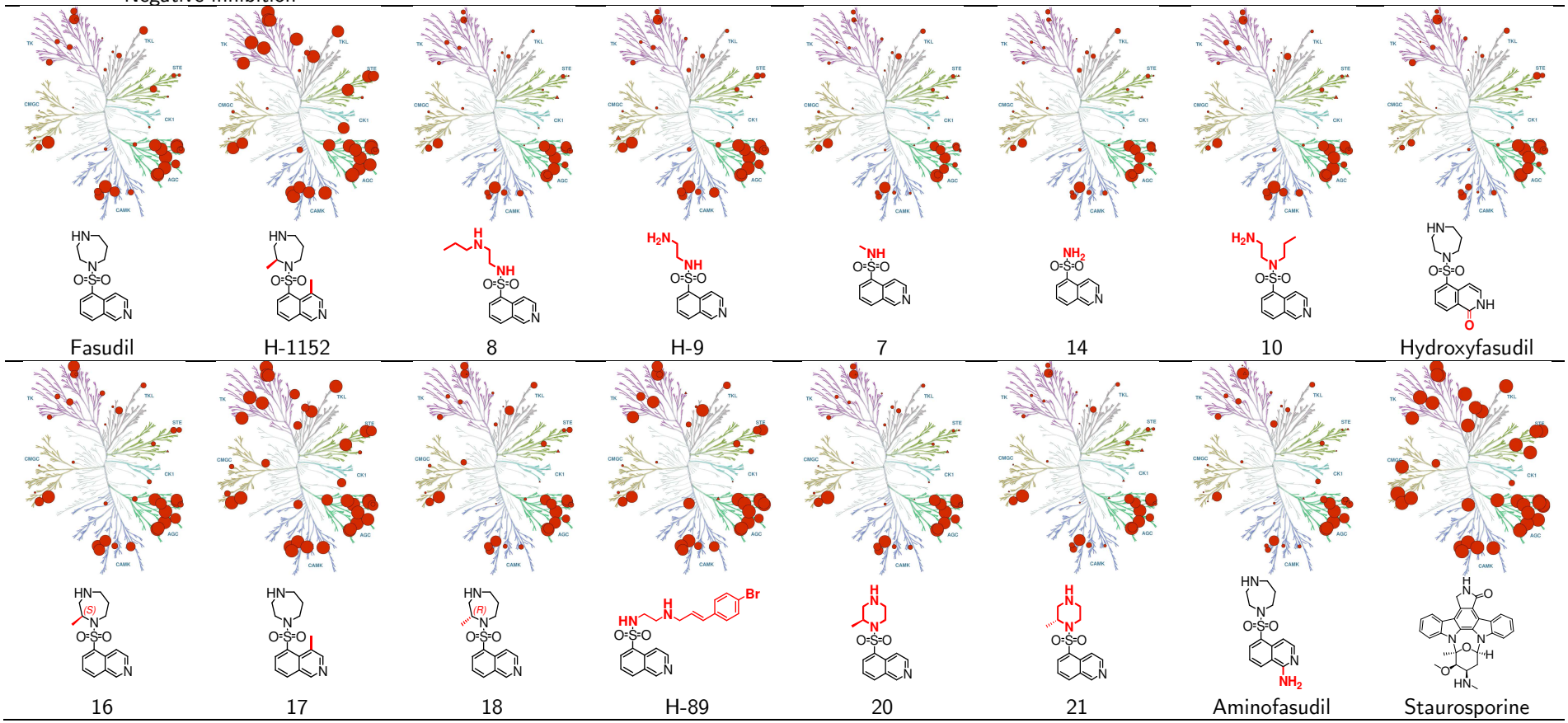


Figure 3. Overview of the number of kinases strongly inhibited (purple) and strongly and partilly inhibited (gray) by the respective compounds. The kinase panel consists of a total of 39 kinases.

From the values of residual kinase activity the percentage of inhibition can be calculated by subtracting the residual activity from 100%. These data can then be mapped on a kinome tree, which allows visualization of the data from the kinase screen distributed across a phylogenetic tree for each of the studied compounds, respectively. These kinome trees are visualized in **Table 1**. They provide a more detailed picture on the inhibitory properties of each ligand. This suggests that especially entries from the AGC- and CAMK-family are more frequently inhibited than kinases from other families. **H-1152**, **17** and **H-89** display similar, however, rather unselective profiles. A second group of similar profiles is formed by ligand **7**, **8** and **14** where stronger inhibition is largely restricted to AGC-, CAMK- and CMGC-kinases.

Table 1. Kinome trees based on inhibition data. All ligand structures and corresponding kinome trees are shown. Adapted from Manning 2002. Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com).[1, 8]

- 20% inhibition
- 60% inhibition
- 100% inhibition
- Positive inhibition
- Negative inhibition



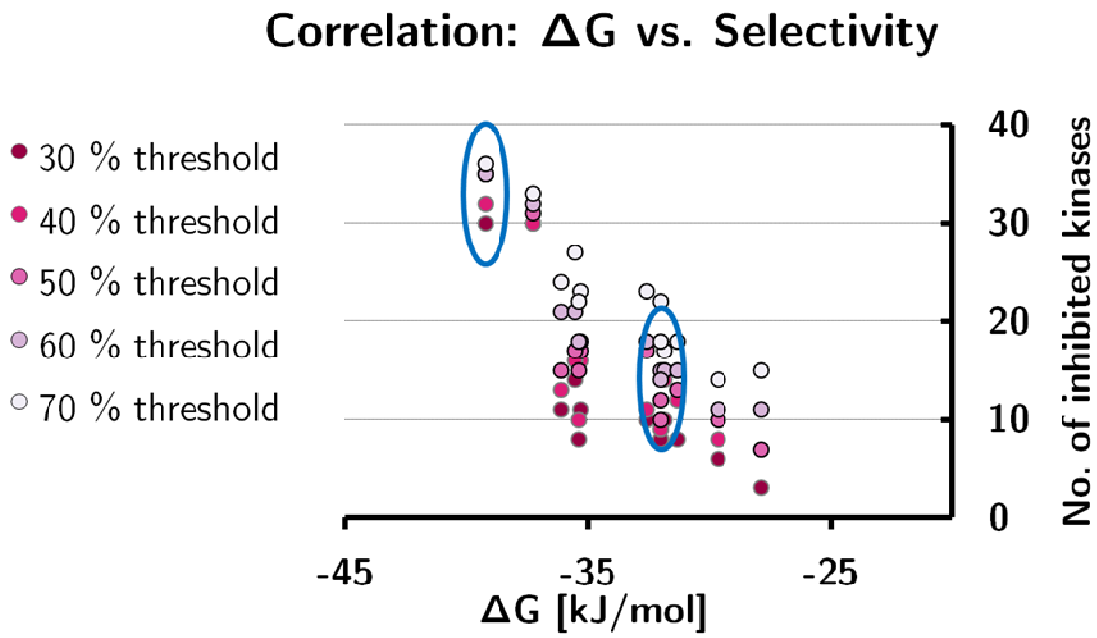


Figure 4. Graph depicts ΔG vs. selectivity for different thresholds of kinase inhibition. For example, a 70% threshold means that all kinases with a residual activity of 70 % or lower are considered inhibited. Blue circles indicate data for **8** and **17**

The interesting question can be put forward whether any correlation with the thermodynamic parameters is given. Therfore, ΔG, ΔH and TΔS as determined by previous ITC investigations, were plotted against the number of inhibited kinases. The thermodynamic profile of **H-89** could unfortunately not be measured via ITC which is due to a combination of low solubility and weak affinity toward PKA. Thermodynamic data for **Aminofasudil** will be measured by Matthias Oebbeke and published in his master thesis. Hence, data for **H-89** and **Aminofasudil** are not displayed in the correlation. Kinase inhibition is shown for different threshold levels of residual kinase activity (less than 30 %, 40 %, 50 %, 60 % or 70 %). Data from ΔG as a representative measure for ligand affinity to inhibit PKA is displayed in

Figure 4. Similar correlation diagrams are reported for ΔH and TΔS in **Figure 5** and **Figure 6**, respectively.

For ΔG and kinase selectivity a correlation is visible (**Figure 4**). The numerical correlation quality increases once data corresponding to 70 % residual activity are considered as threshold. This is indicated by an increasing R^2 value reported in **Table 2**. The R^2 value is the square of the Pearson product moment correlation coefficient. This value can be interpreted as the proportion of the variance in y (no. of inhibited kinases) attributable to the variance in x (ΔG, ΔH or TΔS respectively). Therefore, it is a measure for the numerical correlation. The closer R^2 approaches 1, the better the correlation can be estimated. A value close to zero indicates uncorrelated data. Usually R^2=0.5 is considered the critical threshold for a given intrinsic correlation.[15]

Table 2. R^2 values (square of the Pearson product moment correlation coefficient) for different potential correlations. R^2 is related to the quality of a correlation. Correlation quality of the different correlations of ΔG vs. selectivity increases with increasing threshold. The highest R^2 and hence the best correlation results from the data with the 70 % remaining kinase activity threshold. This means all kinases with a remaining activity of 70 % or lower are considered inhibited.

Threshold	R^2 of ΔG vs. selectivity correlation	R^2 of ΔH vs. selectivity correlation	R^2 of TΔS vs. selectivity correlation	R^2 of MW vs. selectivity correlation
30%	0.657	0.006	0.240	0.286
40%	0.644	0.000	0.158	0.221
50%	0.711	0.001	0.176	0.287
50%	0.802	0.003	0.267	0.345
70%	0.832	0.026	0.366	0.414

For the data shown in **Figure 4**, the R^2 value increases from 0.657 for the 30 % threshold data to 0.832 for the 70 % threshold data (**Table 2**). This indicates a good correlation between ΔG and selectivity. However, with reverse sign, meaning with increasing potency; the tested compounds lose selectivity.

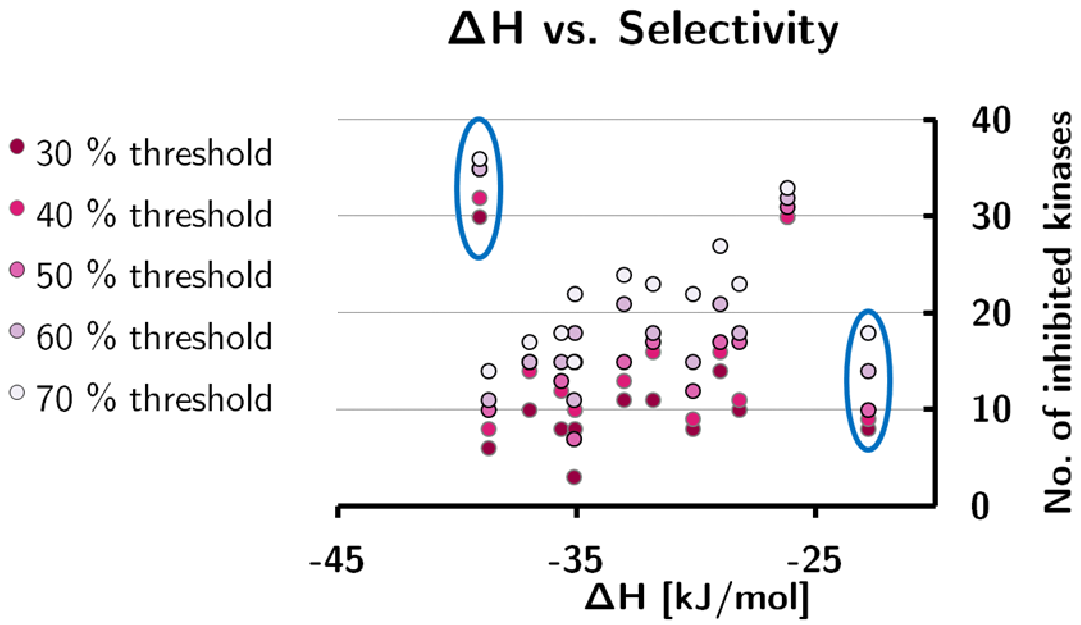


Figure 5. Graph depicts ΔH vs. selectivity for different thresholds of kinase inhibition. For example, a 70% threshold means that all kinases with a residual activity of 70 % or lower are considered inhibited. No significant correlation can be observed. Blue circles indicate data for **8** and **17** that perturb a potential correlation and suggest changes in the binding properties. Remarkably, in **Figure 4** they do not depart from the overall correlation, an indication that enthalpy-entropy compensation is in operation.

If we plot however, ΔH (**Figure 5**) or TΔS (**Figure 6**) against the number of inhibited kinases no real correlation can be observed. Correlation coefficents are reported in **Table 2**. R^2 values between 0.000 and 0.026 for the ΔH data indicate no correlation. For the TΔS dependent data R^2 values

between 0.1580 and 0.3662 are larger but nonetheless they do not indicate significant correlation. If, in both diagrams, **8** and **17** are arbitrarily omitted, significantely better R^2 values are obtained (blue cycles in **Figure 5** and **Figure 6** indicate removed data points, ΔH for 70% level: R^2=0.772, TΔS correlation for 70% level=0.966). Nonetheless, if a crude trend line is imagined for the two different graphs, interestingly the algebraic signs of the slopes are opposed. This observation points to the frequently discussed enthalpy-entropy compensation which, as a reasonable correlation of ΔG with selectivity is indicated, must suggest the observed vague anticorrelation in the diagrams with enthalpy and entropy.

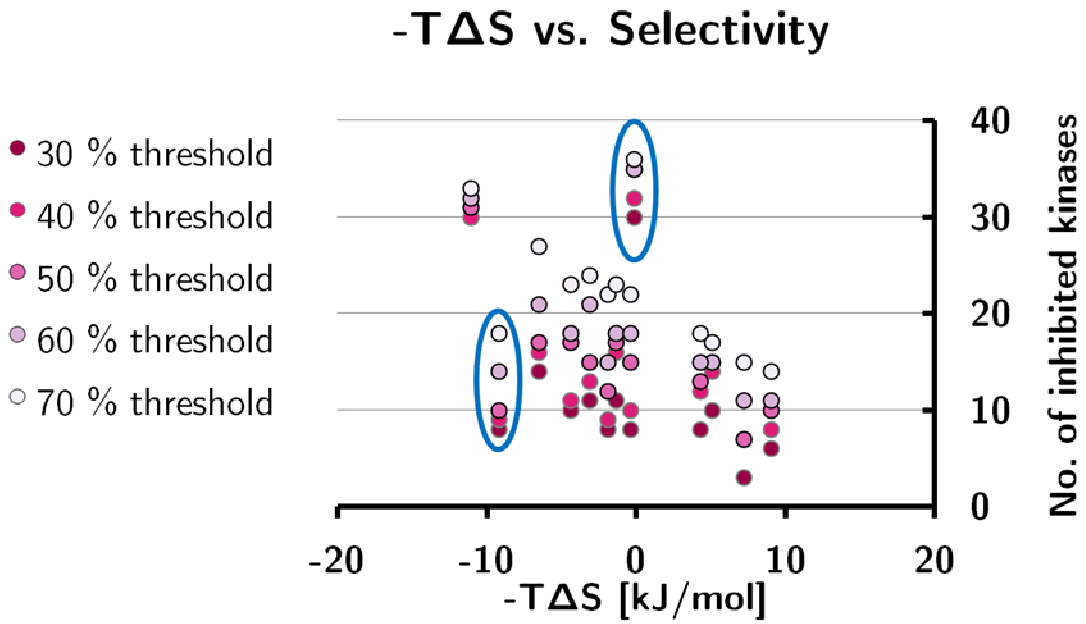


Figure 6. Graph depicts TΔS vs. selectivity for different thresholds of kinase inhibition. For example, a 70% threshold means that all kinases with a residual activity of 70 % or lower are considered inhibited. No significant correlation can be observed. Blue circles indicate data for **8** and **17** that perturb a potential correlation.

In addition, the number of inhibited kinases was plotted in dependence on the molecular weight (MW) of the different ligands (data not shown). The R^2 values are listed in Table 2 and indicate no significant correlation beyond $R^2>0.5$ .R.

6.5 Discussion

The results from the selectivity screen show pronounced differences in the selectivity profiles within the presented ligand panel (**Figure 3**). This is the basis for our analysis to correlate selected biophysical descriptors that might influence these profiles.

The data distribution across the kinase trees suggests that in particular members of the AGC-kinase familiy are inhibited by the ligands considered in this analysis. This is likely due to their similarity with PKA being a member of this family. Furthermore, in particular kinases from the CAMK- and CMGC-family were frequently inhibited.

The observed correlation between ΔG and selectivity is in fact a sign for increasing unselectivity of all ligands with growing potency. It indicates that the weaker a ligand binds to PKA, the weaker it also binds to all other kinases. Vice versa ligands with high affinity to PKA also inhibit a fair number of all other kinases in the screen.

The lack of correlation between MW and selectivity suggests that selectivity cannot be directly linked to the size or the degree of decoration of the studied ligands.

The lack of a clear correlation between ΔH and $T\Delta S$ versus selectivity is another important finding. To our opinion, as we observe a correlation with

ΔG, this lack of correlation for ΔH and ΔH is a direct consequence of enthalpy-entropy compensation indicated in our ITC data. Accordingly the very weak trends suggested by the reverse correlation of ΔH and TΔS supports this hypothesis. If we remove compounds **8** and **17** from the correlation the enthalpic binders show a trend toward higher selectivity. In contrast a more favorable entropic contribution to binding tends to yield less selective compounds. Remarkably ligand **17,** arbitrarily removed from the analysis, binds enthalpically most favorable and is at the same time the most unselective inhibitor of the series. Ligand **8**, another ligand deviating from the correlation, has been discovered as an entropically favored ligand. It induces a different conformation of the Gly-rich loop (**Chapter 2**), however, also ligand **5** induces this structural change (**Chapter 3**). Nevertheless, whereas **8** falls out of the series, ligand **5** matches well with trends suggested by the series. Thus, we believe a correlation between thermodynamic signature and selectivity suggested by an arbitrarily curated subset of the series is flawed. This observation finds its explanation in the fact that the thermodynamic profile is determined by multiple contributions arising from different steps of the complex formation process and not only from the interactions established by the ligands at the binding site.

6.6 Conclusion

Our aim was to collect some evidence for a putative correlation of thermodynamic properties of ligand-protein complex formation of one target kinase (PKA) with selectivity profiles of the respective ligands against a large panel of related kinases. We conclude that a significant correlation can only be observed between ΔG and selectivity. Remarkably, the overall selectivity

is reduced with increasing potency of the studied ligands against PKA. With some care, this might be interpreted as an indication for limited overall selectivity of the studied compound series. Nevertheless, it has to be reminded that particularly the regarded fasudil derivatives with attached methyl groups were developed and reported to show increasing selectivity.[16-18] On the other hand, at first glance, the most selective binders of the series are **7** and **14**, two ligands classified to be of fragment size. Obviously, they show only potent binding against a limited number of kinases. In the current compound series additional decorations at the parent scaffold of the initial fragments results in a reduced selectivity discrimination against the studied kinase panel.

ΔH and $T\Delta S$ do not show any clear-cut correlation, possibly only certain trends emerge, if two compounds are arbitrarily removed from the correlation. Admittedly, the dataset evaluated in this case study is limited and not largely diverse. Thus, more comprehensive investigations need to be performed to validate the question about putative correlations with descriptors on a broader scope. Based on the current results it appears hardly justified to hypothesize that higher selectivity would generally correlate with enthalpically more favored binding.

6.7 Abbreviations

AGC: PK**A**, PK**G**, PK**C** containing group

ATP: adenosine triphosphate

CAMK: calcium/**cal**modulin-dependent protein **k**inases

CK1: **C**asein **k**inase **1** containing group

CMGC: **C**DK, **M**APK, **G**SK3, **C**LK containing group

His-tag: histidine-tag

ITC: isothermal titration calorimetry

MW: molecular weight

Ni-NTA: nickel-nitrilotriacetic acid

R^2: square of the Pearson product moment correlation coefficient

TEV: tobacco etch virus

TK: **T**yrosin **k**inases

TKL: **T**yrosin **k**inase **l**ike group

STE: **Ste**ril kinase homologue containing group

6.8 References

1 Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912–1934.

2 Melnikova, I.; Golden, J. *Nat Rev Drug Discov* **2004**, *3*, 993–994.

3 Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. *Science* **2004**, *303*, 1800–1805.

4 Zhang, J.; Yang, P. L.; Gray, N. S. *Nat Rev Cancer* **2009**, *9*, 28–39.

5 Klebe, G. *Drug Design: Methodology, concepts, and mode-of-action*; Springer Publishing Company, Incorporated, 2013; 599–640.

6 Roskoski, R. *Pharmacol Res* **2016**, *103*, 26–48.

7 Freire, E. *Drug Discov Today* **2008**, *13*, 869–874.

8 Chartier, M.; Chénard, T.; Barker, J.; Najmanovich, R. *PeerJ* **2013**, *1*, e126.

9 Kudlinzki, D.; Linhard, V. L.; Saxena, K.; Sreeramulu, S.; Gande, S.; Schieborr, U.; Dreyer, M.; Schwalbe, H. *Acta Crystallogr F Struct Biol Commun* **2015**, *71*, 1088–1093.

10 Davies, S. P.; Reddy, H.; Caivano, M.; Cohen, P. *BiochemJ* **2000**, *351*, 95–105.

11 Anastassiadis, T.; Deacon, S. W.; Devarajan, K.; Ma, H.; Peterson, J. R. *Nat Biotechnol* **2011**, *29*, 1039–1045.

12 Jester, B. W.; Gaj, A.; Shomin, C. D.; Cox, K. J.; Ghosh, I. *J Med Chem* **2012**, *55*, 1526–1537.

13 Gao, Y.; Davies, S. P.; Augustin, M.; Woodward, A.; Patel, U. A.; Kovelman, R.; Harvey, K. J. *Biochem J* **2013**, *451*, 313–328.

14 Kuhn, D.; Weskamp, N.; Hüllermeier, E.; Klebe, G. *ChemMedChem* **2007**, *2*, 1432–1447.

15 Mukaka, M. M. *Malawi Med J* **2012**, *24*, 69–71.

16 Tanaka, H.; Ohshima, N.; Takagi, M.; Komeima, K.; Hidaka, H. Novel vascular relaxant, HMN-1152: its molecular mechanism of action. Naunyn Schmiedebergs Arch Pharmacol. 1998; R219–R219.

17 Sasaki, Y.; Suzuki, M.; Hidaka, H. *Pharmacol Ther* **2002**, *93*, 225–232.

18 Breitenlechner, C.; Gassel, M.; Hidaka, H.; Kinzel, V.; Huber, R.; Engh, R. A.; Bossemeyer, D. *Structure* **2003**, *11*, 1595–1607.

6.9 Supplementary information

Table S1. Overview of thermodynamic and inhibition data.

	MW [g/mol]	Phosphatpuffer ΔG [kJ/mol]	Phosphatpuffer ΔH [kJ/mol]	Phosphatpuffer -TΔS [kJ/mol]	No. of inh. kinases 70 % threshold	No. of inh. kinases 30 % threshold
Fasudil	291.4	-36.1	-33.1	-3.1	24	11
Hydroxyfasudil	307.4	-32.0	-30.1	-1.9	22	8
H-1152	319.4	-37.3	-26.2	-11.1	33	30
20	291.4	-31.9	-37.0	5.1	17	10
21	291.4	-31.3	-35.7	4.3	18	8
H-9	251.3	-35.4	-31.8	-1.3	23	11
7	222.3	-27.9	-35.1	7.2	15	3
8	293.4	-32.0	-22.8	-9.2	18	8
10	293.4	-35.4	-35.1	-0.3	22	8
14	208.2	-29.7	-38.7	9.1	14	6
16	305.4	-35.6	-29.0	-6.5	27	14
17	305.4	-39.2	-39.1	-0.1	36	30
18	305.4	-32.6	-28.2	-4.4	23	10
H-89	446.4	n.a.	n.a.	n.a.	32	22
Aminofasudil	306.4	n.a.	n.a.	n.a.	19	8
Staurosporine	466.5	n.a.	n.a.	n.a.	39	37

n.a.: Data not available

Table S2. Overview of inhibition data (% residual activity) for all compounds and kinases. Kinase families are assigned. Table spreads over two pages.

Kinase Family	Kinase	Fasudil	Hydroxy-fasudil	H-1152	20	21	H-9	7	8	10	14	16	17	18	H-89	Amino-fasudil	Stauro-sporine
CAMK	AMPK[alpha]2	51	63	4	48	65	40	72	64	47	75	14	4	42	5	76	-2
CAMK	BRSK2	32	43	4	38	53	39	60	62	41	58	17	5	40	18	63	0
CMGC	CDK5	57	67	11	36	39	29	54	54	69	61	36	3	42	31	103	0
CMGC	CDK6	88	99	100	91	104	126	95	100	101	110	94	60	93	87	99	-8
CMGC	CDK9	8	52	17	2	5	14	30	28	30	37	6	4	12	14	32	1
CAMK	CHK1	84	86	59	88	97	75	91	87	77	84	78	36	98	18	91	1
CMGC	CK2[alpha]1	97	97	90	90	91	88	107	88	95	91	86	42	61	90	86	42
TK	EphA2	70	96	1	79	85	82	91	72	65	85	67	2	75	39	73	-4
TK	FAK	65	69	3	82	86	71	86	78	67	85	54	5	72	11	62	-1
TK	Fes	57	54	7	73	78	68	83	78	80	83	70	5	63	12	58	1
TK	FGFR1	57	75	2	60	69	68	72	72	65	69	14	3	45	20	99	-1
CMGC	GSK3[beta]	101	97	84	97	102	92	88	90	92	92	95	70	82	91	93	0
TKL	IRAK4	85	102	87	87	99	75	95	92	91	72	93	41	79	65	52	-1
TK	JAK2	97	78	5	98	93	83	85	98	95	98	68	27	89	108	88	-4
TKL	LIMK1	67	79	6	71	59	72	61	77	53	71	50	10	35	8	71	-1
STE	LOK	53	85	12	86	71	72	80	93	73	87	73	8	63	53	71	0
CAMK	MELK	7	56	12	9	15	6	19	13	6	19	15	4	20	4	27	-1
TK	Mer	30	30	4	38	38	31	71	55	40	77	46	4	40	5	13	5
TKL	MLK1	43	63	12	79	76	78	69	82	83	64	65	4	57	50	66	-2
CAMK	MNK2	28	90	15	74	65	60	80	60	42	71	59	5	74	47	80	-1
AGC	MSK1	13	9	4	18	32	10	33	21	24	17	8	3	23	1	53	-1

Kinase Family	Kinase	Fasudil	Hydroxy-fasudil	H-1152	20	21	H-9	7	8	10	14	16	17	18	H-89	Amino-fasudil	Stauro-sporine
STE	MST2	75	49	13	86	76	58	78	77	84	75	59	9	79	44	94	2
STE	MST3	86	69	26	96	85	66	106	98	112	97	67	26	64	19	108	1
AGC	p70S6K	8	3	8	16	25	9	34	15	10	24	19	1	27	1	15	0
STE	PAK5	94	102	88	111	116	112	115	117	118	111	107	79	114	65	100	1
AGC	PDK1	76	84	12	102	107	66	104	99	88	94	67	17	86	115	107	-1
CAMK	PhK[gamma]2	63	67	3	72	85	28	90	59	50	70	35	4	70	26	80	0
AGC	PKA[alpha]	5	18	3	10	11	4	33	12	3	25	2	2	14	1	2	1
AGC	Akt1/PKB[alpha]	50	43	46	68	77	31	90	65	49	90	64	25	65	2	51	-1
AGC	Akt2/PKB[beta]	78	87	77	85	93	78	109	90	90	108	100	75	88	16	99	-1
AGC	PKC[epsilon]	40	68	8	38	44	36	68	57	59	81	24	8	45	36	28	0
AGC	PKN2/PRK2	3	4	3	5	7	3	20	11	10	24	2	3	6	1	4	1
TK	Ret	85	91	9	102	118	114	113	115	106	113	86	9	101	61	114	-3
TKL	RIPK2	92	98	65	83	92	99	88	91	99	79	91	45	99	89	90	4
AGC	ROCK1	21	16	3	20	32	30	54	47	33	45	7	9	28	8	7	0
AGC	ROCK2	1	0	2	2	3	2	4	7	2	4	0	2	1	1	1	1
AGC	RSK1/p90RSK	8	6	2	9	18	12	50	30	24	38	15	-1	22	2	69	0
AGC	RSK2	9	8	4	8	15	7	57	19	17	40	11	2	16	3	40	0
CK1	VRK2	93	89	16	97	107	96	93	97	101	98	72	37	102	100	67	61

Table S3. Thermodynamic data with standard deviations for hydroxyfasudil.

Ligand	ΔG [kJ/mol]	ΔH [kJ/mol]	-TΔS [kJ/mol]
Hydroxyfasudil	-32.0±0.5	-30.1±1.1	-1.9±1.6

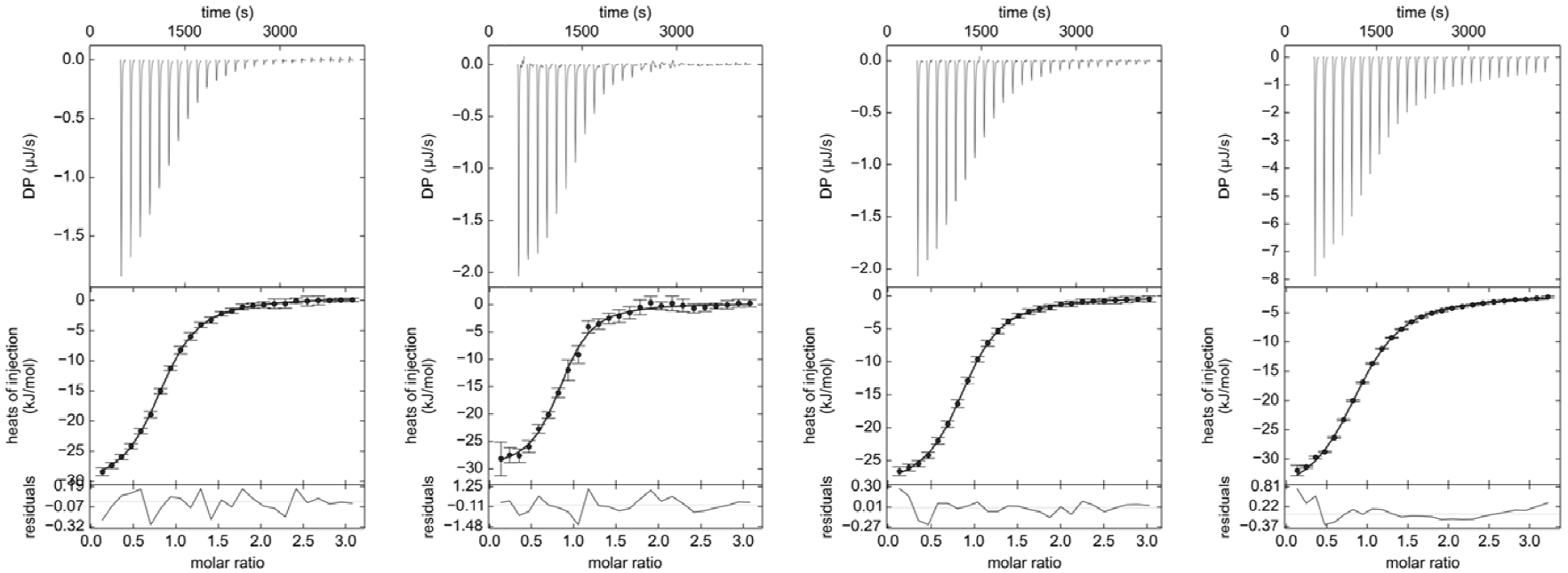


Figure S1. ITC-titration curves for hydroxyfasudil in phosphate buffer as exemplary titration curves.

7 Two Methods one Goal: Structural Differences between Results from Crystal Soaking and Co-crystallization

7.1 Abstract

This study focuses on the comparison of two popular protocols to produce crystals of protein-ligand complexes: soaking and co-crystallization. Both methods are applied to access information on protein-ligand interactions. This is of particular interest in structure-based drug design (SBDD). Thereby, soaking is the convenient and less time consuming method and hence widely applied. However, our study suggests, that crystal structures obtained by soaking can significantly differ from those received by co-crystallization experiments. Co-crystallized structures are considered to be the more accurate and relevant representation of the equilibrium state in solution. We demonstrate that results from soaking experiments bear the risk to not only falsely represent the experienced protein-ligand interaction patterns and adopted protein conformations but even the ligand orientation in the active site. The impact of applied crystallization protocols is likely to depend on the change in protein conformation upon ligand binding. For kinases, which are particularly flexible proteins, a large induced fit triggered by the bound ligand reduces the structural agreement between crystal structures from soaking and co-crystallization. We conclude that the importance of the applied crystallization protocol must not be

underestimated. It can take crucial impact and must therefore be considered in SBDD.

7.2 Introduction

The protein discussed in this thesis is cAMP-dependent protein kinase (PKA). It is considered a typical representative of the clinically highly relevant class of protein kinases and is frequently used as a model protein. [1]

Kinases in general and specifically PKA are highly flexible proteins.[2-4] In particular, the glycine-rich loop (Gly-loop) covering the active site can adopt multiple conformations.[5] This property will be a key aspect for the present study.

In the process of drug design and the search for new lead structures as well as lead optimization, protein crystallography is a commonly applied powerful tool to analyze drug binding to a target protein.[6] The binding mode as well as non-covalent interactions such as salt bridges, hydrogen bonds and van der Waals interactions can be characterized and evaluated in terms of distances and angular relationships. Such information is crucial to guide modeling of drug properties. There are two commonly applied protocols to produce crystals of protein-ligand complexes for subsequent crystallographic analysis, namely co-crystallization and soaking.[7]

For co-crystallization, protein and ligand are mixed in solution prior to the crystallization process. Upon crystallization the preformed protein-ligand complex from equilibrium conditions in solution assembles in the crystalline phase.[7, 8] Due to the presence of the ligand and usually the presence of varying amounts of e.g. DMSO, a popular solvent for small organic

molecules, crystallization conditions can differ significantly for a series of co-crystal trials using different ligands. Sometimes even screenings for a completely new condition are required. Therefore, the development of successful co-crystallization protocols can be a quite time and material intensive approach.

Thus, especially in industry with the typically imposed time restrictions, the much faster soaking method is frequently applied. Here, the protein is crystallized in its uncomplexed state lacking any bound ligand. The premanufactured apo crystal is then placed into a droplet containing a high concentration of the targeted ligand for which the binding pose should be elucidated. The ligand may then diffuse into the crystal and bind to the protein. Prerequisite for the success of this procedure is the presence of sufficiently large water-filled channels passing contiguously through the packed crystal and accessing the binding site of the protein. Soaking is thus a fast method. It requires significantly less material than co-crystallization and builds on a well-established crystallization protocol.[7, 8] Given that the apo protein crystallizes well, hundreds of differently soaked crystals can be generated from a single crystallization plate. Co-crystallization on the other hand can easily require multiple crystallization plates per ligand in order to find the optimized conditions, resulting in a high demand of protein and ligand material.

Differences between crystal structures obtained from soaked crystals and from co-crystallization have been reported.[9-13] However, the number of systematic and well-documented examples in literature is still surprisingly

small considering its importance for the relevance of the drug discovery pipeline.

Meaningful examples are crystal structures of glutathione S-transferase of the malarial parasite plasmodium falciparum (PfGST) and lymphocyte-specific kinase (Lyck).[12, 13]

Using a systematic approach, we will analyze the impact of the crystallization protocol on the protein-ligand complexation more thoroughly. Based on the obtained results we will put the important questions forward: How comparable are the binding modes received by the two different methods? Are they equally suited to represent the properties of the drug molecules in the bound state in solution? To tackle this question, we analyzed a series of ligands that are derived from the approved drug fasudil in complex with PKA. These ligands differ in their size as well as in their ability to trigger induced-fit adaptations of the protein.

In general, the following restrictions must be considered in soaking: Large conformational changes induced by ligand binding may be difficult to experience in the solid state, cracking crystals or even dissolution can occur or access to the ligand binding site in the apo-crystal may be blocked.[6] Accordingly, it is recommended to cross-validate results from soaking experiments with co-crystals so that the full range of conformational changes can be discovered, particularly as co-crystal structures are considered to be the more relevant representation of the protein-ligand complex in equilibrium.[7] Nevertheless, for most drug discovery studies results from soaking experiments are not cross-checked by co-crystallization attempts.

For our comparative study co-crystals as well as soaked crystals were prepared and crystallographically analyzed. Striking differences between the two approaches were discovered. Especially for ligands that trigger a strong induced fit, soaked structures represent the ligand with a misleading binding mode, deviating interaction patterns and conformational differences of the protein structures.

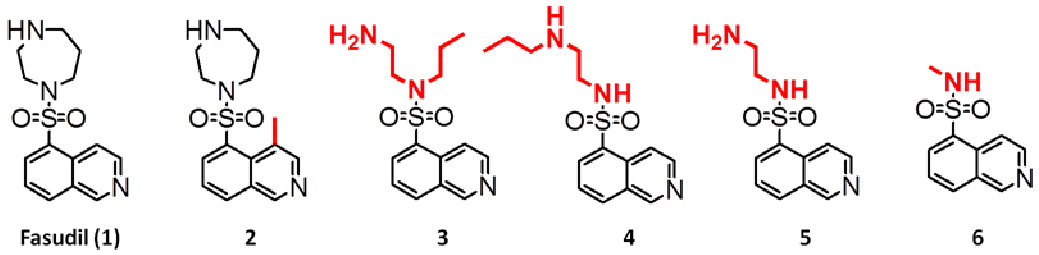


Figure 1. Overview of ligands used in this study. Ligand **1** is fasudil. **2**: a methylated form of fasudil. **3**: Open-chain N-(2-aminoethyl)-N-propylisoquinoline-5-sulfonamide. **4**: Long-chain N-[2-(propylamino)ethyl]isoquinoline-5-sulfonamide. **5**: short chained N-(2-aminoethyl)isoquinoline-5-sulfonamide. **6**: fragment sized methylated N-methylisoquinoline-5-sulfonamide.

7.3 Experimental section

Protein expression and purification

The catalytic subunit of cAMP-dependent protein kinase from Chinese hamster ovary cells was expressed with a His-tag in a modified pET16b-Vector with an introduced TEV-cleavage site between the protein N-terminus and His-tag. This plasmid was transformed into E. coli strain *Bl21 (DE3)/pLysS* (Novagen).[14]

Cell disruption was performed using a high-pressure homogenizer for multiple cycles. After centrifugation (1h at 30.000g) cell lysate supernatant was

purified in a first step using a Ni-NTA column that binds the His-tag of the protein and was eluted by an imidazole gradient. The His-tag was then cleaved off by TEV-protease. Afterwards, an inverse Ni-NTA column was employed collecting PKA in the flow-through. Finally, ion exchange chromatography was performed using a MonoS column separating three-fold phosphorylated PKA from the four-fold phosphorylated form using a HEPES buffer with sodium chloride gradient.[14]

Crystallization

All cocrystal protocols are also discussed in the previous chapters but are repeated in the following for improved comparability of the soaking and cocrystallization protocols. Crystallisation for soaking was performed using the hanging drop method at 4 °C. The crystallization drops contained the following ingredients: 10 mg/mL PKA (240 µM), 30 mM MBT (MES/Bis-Tris Puffer pH 6.2-6.9), 1 mM DTT, 0.1 mM EDTA, 75 mM LiCl, 0.03 mM Mega 8, 0.07 mM PKI (Sigma: P7739 for co-crystals; Sigma: SCP0064 for apo crystals for soaking), 120µM or 1.2 mM ligand dissolved in DMSO from a 50-100 mM stock for co-crystals but not for apo crystals. The well contained a mixture of methanol in water with varying methanol concentrations (v/v) for the different ligands (14-23% methanol). In the crystallization setup streak-seeding was performed with apo crystals as seeds using a horse hair in order to initialize crystal growth. Soaking was performed in a buffer containing 30 mM MBT (MES/Bis-Tris Puffer pH 6.9), 1 mM DTT, 0.1 mM EDTA, 75 mM LiCl, 16 % methanol (v/v), 120µM ligand dissolved in DMSO for 24 hours. For crystal mounting, crystals were cryo protected in 5 mM MBT (MES/Bis-Tris buffer pH 6.9), 1 mM DTT, 0.1 mM LiCl, 120µM or 1.2 mM

ligand dissolved in DMSO from a 50-100 mM stock, 16 % (v/v) methanol, 30% (v/v) MPD and flash frozen in liquid nitrogen.

Crystallography

All cocrystal structures are discussed in the previous chapters. All soaked structures were collected at the storage ring Bessy II Helmholtz-Zentrum Berlin, Germany at Beamline 14.1 on a Pilatus 6M pixel detector. The datasets were processed using XDS[15] and molecular replacement was performed using CCP4 Phaser[16] and PDB-structure of PKA from *bos taurus* 1Q8W as a model. This was followed by simulated annealing, multiple refinement cycles of maximum likelihood energy minimization and B-factor refinement with Phenix[17]. Coot[18] was used to fit amino-acid side chains into σ-weighted 2Fo – Fc and Fo – Fc electron density maps. If appropriate electron density was observed, multiple side chain conformations were built into the model and maintained during the refinement if the minor populated side chain displayed at least 20 % occupancy. Ramachandran plots for structure validation were calculated using PROCHECK[19]. Data collection, unit cell parameters and refinement statistics are given in the supplementary information. Analysis of temperature factors was performed with Moleman[20]. Protein and PKI B-factors were anisotropically refined, water B-factors were isotropically refined for all soaked structures. Decision for anisotropic or TLS refinement was based on comparison of R_{free}. Anisotropic refinement was chosen over TLS if the achieved R_{free} values were at least 0.5% lower for anisotropic than for TLS refinement. R_{free} was calculated using 5% of all reflections which were randomly chosen and not used for the refinement. The required ligand restraint files were created using the Grade webserver[21, 22]. For figure preparation Pymol was used. For RMSD calculation the iterative

alignment routine Matchmaker[23] implemented in Chimera[24] was used. Crystallographic tables of the different ligands can be found in the supporting information.

7.4 Results

Six different ligands were used in this comparative study. For each ligand complex a diffraction dataset could be collected using specimen obtained by the two crystallization protocols, resulting in a total of twelve structures. The superpositions of the respective structure pairs bound to the same ligand are shown in **Figure 2**-**Figure 7**. **Figure 2** depicts **fasudil** in complex with PKA. The position of the glycine-rich loop (Gly-loop) is more open in the co-crystal structure. In particular, this affects the residues Gly50 to Ser53, where backbone atoms are approximately shifted by 2 Å. Concerning the ligand, in particular the homopiperazine ring differs in position, mostly due to a deviating ring conformation. The root-mean-square deviation (RMSD) of the ligands between both structures amounts to 0.88 Å, the same value as for the CA-atoms of the protein. The two homopiperazine conformations lead to a difference in hydrogen bonding to the protein. In fact the co-crystal structure displays more polar interactions and involves Asp184 and Glu170 whereas the soaked structure suggests an H-bond to Asn171.

Figure 3 shows the binding modes for ligand **2**. Here, the positions of the ligand match closer (RMSD: 0.64 Å). The position of the Gly-loop is not fully defined in the electron density for the soaked structure. Nonetheless, the amino acid residues of the loop visible in both structures share common positions. The RMSD of the CA-atoms of the protein is 0.65. In contrast to

1, the positions and conformations of the homopiperazine moieties adopt quite similar orientations.

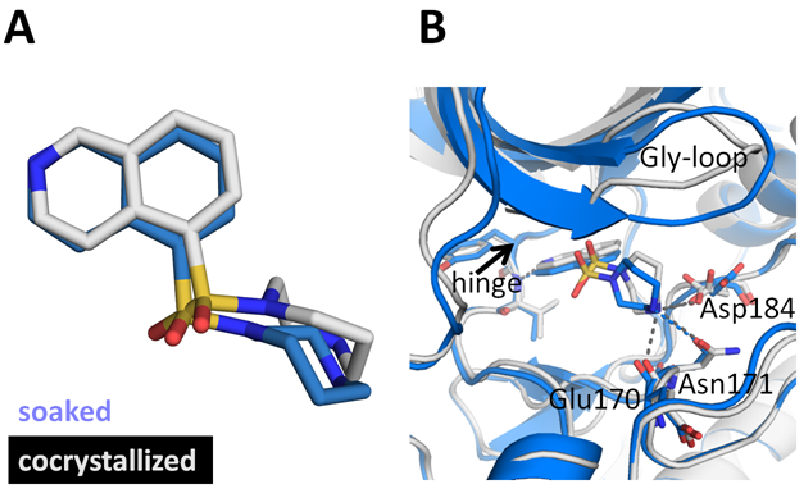


Figure 2. Crystal structures of fasudil (1) in complex with PKA from a soaked (blue) and co-crystallized structure (gray). A: Ligand superposition indicates a slight rotation of the ligand from one to the other structure (RMSD ligand: 0.88 Å, RMSD of the CA-atoms of the protein: 0.88). **B**: Active site superposition. Dotted lines indicate hydrogen bonds. The position of the Gly-loop is more open in the co-crystal structure. Furthermore, the co-crystal structure shows more hydrogen bonds between the ligand and PKA.

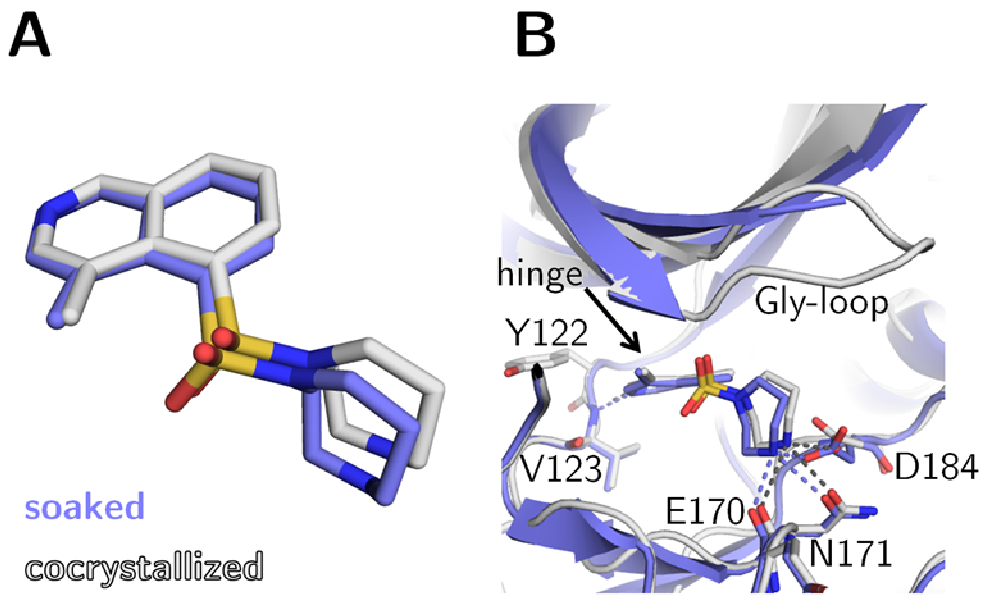


Figure 3. Crystal structures of methylated fasudil (2) in complex with PKA from a soaked (violet) and co-crystallized structure (gray). A: Ligand superposition indicates a slight rotation of the ligand from one to the other structure (RMSD ligand: 0.65 Å, RMSD of the CA-atoms of the protein: 0.64). **B:** Active site superposition. Dotted lines indicate hydrogen bonds. The position of the Gly-loop is not defined for the soaked crystal structure. Hydrogen bond formation is suggested as similar for both structures.

The observation of a partially disordered Gly-loop in the soaked crystal structure is also experienced for ligand **3** (**Figure 4**). In addition, an altered position of the aminoethyl portion of the ligand was striking. Therefore, an entirely different interaction pattern of the aminoethyl-nitrogen of the ligand with the protein is found for the two crystal structures resulting from the different crystallization protocols.

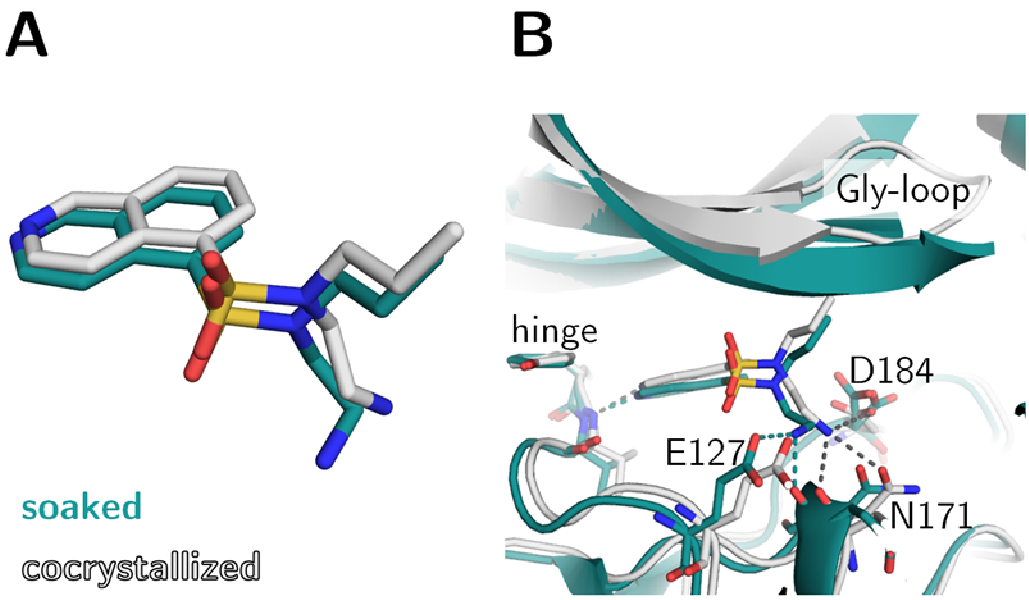


Figure 4. Crystal structures of open chain fasudil-derivative 3 in complex with PKA from a soaked (teal) and co-crystallized structure (gray). A: Ligand superposition indicates a slight shift of the whole ligand as well as an altered position of the aminoethyl moiety in the two structures (RMSD ligand: 1.20 Å, RMSD of the CA-atoms of the protein: 0.94). **B:** Active site superposition. Dotted lines indicate hydrogen bonds. The position of the Gly-loop is not defined for the soaked crystal structure. Hydrogen bonds are formed to different amino acids in the soaked and the co-crystal structure. In the co-crystal structure hydrogen bonds are formed to the hinge, the side chains of Asp 184 and Asn171 and the backbone of Glu170. In the soaked structure only the hinge and the backbone of Glu170 are addressed. An additional hydrogen bond is formed to the side chain of Glu127 in the soaked structure.

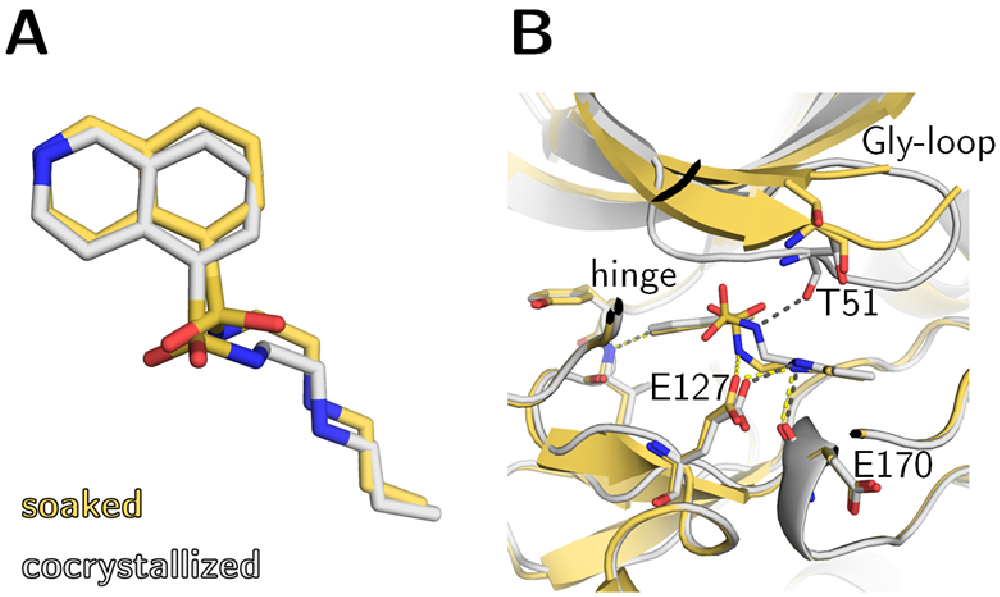


Figure 5. Crystal structures of long chain fasudil-derivative 4 in complex with PKA from a soaked (yellow) and co-crystallized structure (gray). A: Ligand superposition indicates a strong rotation of the whole sulfonamide portion and the attached substituent of the ligand between the two structures (RMSD ligand: 0.91 Å, RMSD of the CA-atoms of the protein: 0.65). **B:** Active site superposition. Dotted lines indicate hydrogen bonds. The position of the Gly-loop is not entirely defined for the soaked structure. Strikingly, the positions of those residues visible in the soaked structure differ greatly from their position in the co-crystal structure. Only in the co-crystal structure an induced fit can be observed. The Gly-loop is pulled down into the active site. The hydrogen bond to Thr51 is only formed in the co-crystal structure, whereas the hydrogen bond between the side chain of Glu 127 and the sulfonamide nitrogen of **4** is only present in the structure from the soaked crystal.

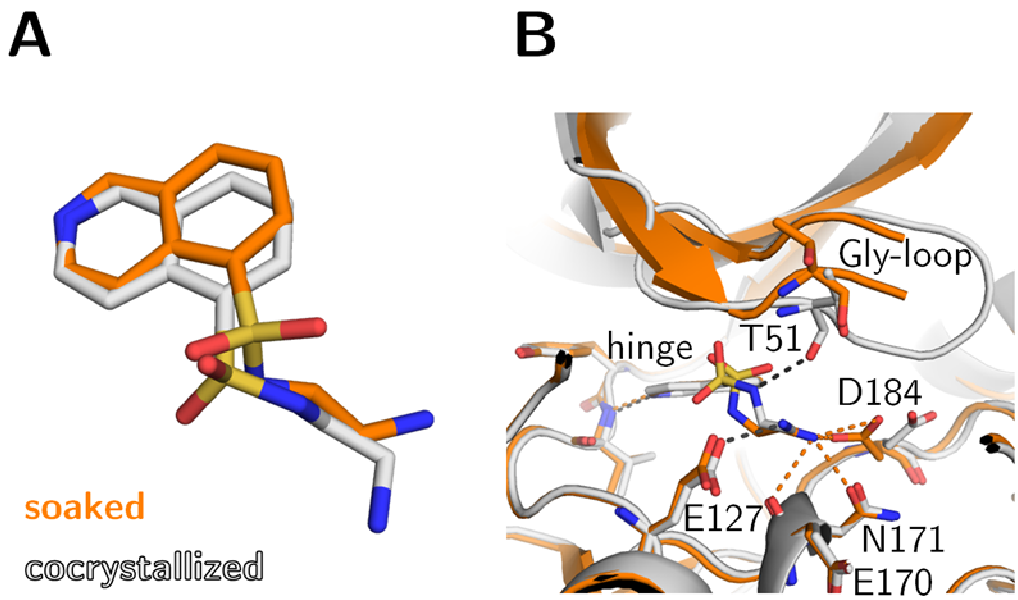


Figure 6. Crystal structures of short chain fasudil-derivative 5 in complex with PKA from a soaked (orange) and a co-crystal structure (gray). A: Ligand superposition indicates a strong rotation of the whole sulfonamide portion and the attached substituent of the ligand between the two structures (RMSD ligand: 1.26 Å, RMSD of CA-atoms of the protein: 0.61). **B:** Active site superposition. Dotted lines indicate hydrogen bonds. The position of the Gly-loop is not entirely defined for the crystal structure from the soaked crystal. Strikingly, the positions of those residues visible in the soaked structure (Thr51 and Gly52) differ 2.2-5.0 Å in their backbone position from the co-crystal structure. Merely in the co-crystal structure an induced fit can be observed. The Gly-loop is pulled down into the active site. Hydrogen bonds are formed to different amino acids for the soaked (Val 123, Asp184, Asn171, Glu170) and the co-crystal structure (Val 123, Thr51, Glu127).

The most prominent differences between crystal structures resulting from soaking and co-crystallization were observed for ligands **4** and **5**. Both ligands share are common binding mode. A detailed view of the binding mode of **4** is shown in **Figure 5**. In the structure resulting from the co-

crystallization protocoll, a strong induced fit takes place provoked by a pull-down of the Gly-loop. This movement is triggered by the formation of a hydrogen bond between the ligand and Thr51 of the protein. It is absent in the soaked crystal structure. This in turn leads to an altered position of the ligand, where in particular the sulfonamide is conformationally rotated. The angle between the plane through the atoms of the isoquinoline moiety and the sulfonamide nitrogen is a suitable descriptor for this rotation. For the structure obtained by soaking this angle is 94° while it is only 26° in the same direction for the ligand in the co-crystal structure. The very same observation is made for **5** (**Figure 6**). The differences between the soaked and co-crystal structures are even more pronounced for this ligand. The rotated sulfonamide (angle between isoquinoline plane and the sulfonamide nitrogen bond vectors is 25° for co-crystal and 103° for soaked crystal, thus a similar situation as observed for **4**) is accompanied by an aminoethyl moiety that points into the opposite direction. In the crystal structure resulting from soaking, only one hydrogen bond to the protein is formed whereas there are four present in the co-crystal structure.

The binding mode of fragment-sized ligand **6** in **Figure 7** is less influenced by the crystallization protocol. Both ligand binding mode and protein structure align well (RMSD ligand: 0.18 Å, RMSD of the CA-atoms of the protein: 0.44).

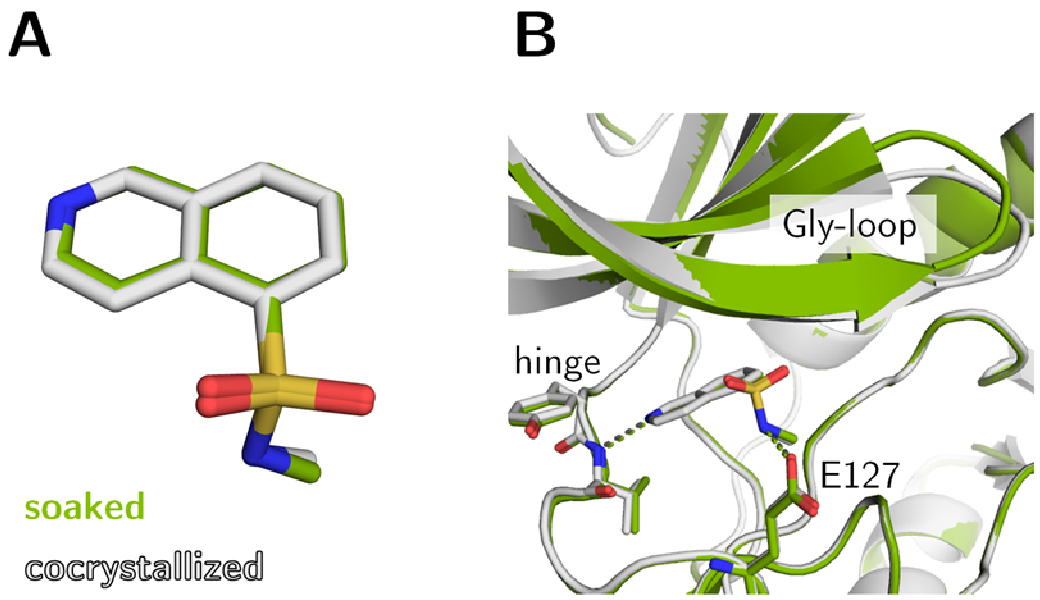


Figure 7. Crystal structures of fasudil-fragment 6 in complex with PKA from a soaked crystal (green) and a co-crystal (gray). A: Ligand superposition shows an equal ligand binding conformation and position in both structures (RMSD ligand: 0.18, RMSD of CA-atoms of the protein: 0.44). **B:** Active site superposition. Dotted lines indicate hydrogen bonds. The position of the Gly-loop differs only slightly. The hydrogen bonding pattern between ligand and PKA are identical in both structures.

7.5 Discussion

Obviously, the overall comparability of ligand binding modes and protein-ligand interaction patterns in crystal structures obtained by ligand soaking or co-crystallization is limited in case of PKA. This protein, like most kinases, is a flexible protein, especially in the active site. We assume that co-crystal structures are a better approximation of the protein-ligand complex in solution. In solution, the protein can perform an induced fit, thereby adopting a conformation that allows optimal protein-ligand interactions. Moreover, crystallization usually results from a state of equilibrium. These conditions are captured and crystallized during co-crystallization. For soaking

experiments this is different. Here the uncomplexed protein is arranged in premanufactured crystals and hence in a rather constrained environment when the ligand diffuses into the crystal with densely packed protein. The subsequent geometry established upon complexation and protein-ligand interactions are limited by this spatially restricted environment. Hence, an energetically favorable induced fit that requires an entire loop to alter its position cannot take place to the required extend.

Crystal structures based on soaking bear the risk to suggest misleading binding modes particularly for flexible proteins such as kinases. Ligands **2** and **3** demonstrate that flexible protein regions are more likely to be disordered in a soaked structure and accordingly not defined in the electron density. Ligands **4** and **5** indicate that a geometrically hampered protein adaption during the ligand binding process not only influences the number of observed interactions between ligand and protein but moreover influences the adopted bound conformation, the ligand binding mode and position. Therefore, crystal structures obtained by a soaking protocol can misdirect drug optimization and bias structure-based drug design (SBDD) by suggesting artificially distorted ligand binding modes. Co-crystallization on the other hand, depicts a ligand binding mode closer to the situation in solution: Flexible protein regions can adopt the optimal conformation yielding crystal structures with more reliable positions and improved visibility of these regions. Thereby they draw a more realistic picture of hydrogen bond formation in equilibrium between protein and ligand and more importantly ligand binding position. Our finding is strongly supported by one of our previous studies which analyzed the structural properties of the studied PKA complexes in solution by multidimensional NMR. Here we

recorded a convincing match between chemical shift perturbations of the kinase and geometrical differences observed among the complexes in cocrystallization.

Our findings agree with results from other authors. For the kinase Lyck in complex with staurosporine the crystal structure resulting from soaking also underestimated the movement of the Gly-loop. The comparative co-crystal structure revealed a more prominent movement in this highly flexible protein region.[12]

Ligand **6** on the other hand shows that the impact of the crystallization protocoll is minor for a fragment-sized ligand. We believe two reasons are important in that context. Firstly, due to their lower affinity fragments are usually not potent enough to induce larger changes in protein conformation. Second, the larger ligands **1-5** exhibit a fair number of torsional degrees of freedom that are rather soft, however allow the ligand to adopt to the packed protein environment in the premanufactured apo crystals during the soaking process. Fragments are usually selected to possess a much smaller number of torsional degrees of freedom. Hence, protein structures and ligand binding modes are more likely to be comparable in case of fragment complexes between crystal structures obtained by soaking and co-crystallization.

Therefore, effort has to be made to promote and ease high throughput cocrystallization experiments. Various aspects involved in the cocrystallization process can be optimized. The utilization of robotic systems to pipette crystallization conditions reduces material consumption.[25] Moreover, strategies to reduce or remove the solvents used to solubelize the respective compounds can help keeping the crystallization conditions similar

to those of the apo crystals. Solubilizing the compounds in methanol and subsequentely allowing the methanol to evaporate on a crystallization plate prior to the addition of the crystallization drop is one approach.[26] An analog dry-down procedure for the more popular solvent DMSO has also been decribed.[27] In case of too low compound solubility, additives can increase compound solubility for an additional soaking after a cocrystallization prodcedure in order to increase ligand population in the binding site without altering the crystallization condition through the presence of the additive.[28] Finally, high throughput cocrystallization for fragment screening using acoustic droplet ejection, even allowing crystal growth on data collection media such as micro meshes, as published by Yin et al is a promising approach to promote large scale cocrystallization making cocrystallization screenings rapid and economic.[29] An exemplified high throughput co-crystallization experiment has been made by Ember and coworkers[30]. Performing a robotic co-crystallization screening campaign using 518 compounds on the kinase BRD4-1 resulted in 377 successful crystallization approaches, yielding 194 structures leading to the identification of bound 14 kinase inhibitors.

7.6 Conclusion

Because of its time and cost effectiveness, soaking is the more popular method to produce crystal structures of protein-ligand complexes. Here, we demonstrate that the more laborious method of co-crystallization is however the superior approach. In particular for flexible proteins, such as kinases, only co-crystallization can capture the actual ligand binding position and induced protein conformation. As we can take reference to NMR data collected on

some of the complexes reported in this study, we are very confident about this conclusion. The geometric discrepancies between structures generated from soaked crystals and co-crystallized ones appear to be smaller for fragment-sized ligand molecules with limited degrees of torsional freedom. However, for ligands that trigger conformational changes in the protein structure, soaking is definitely a misleading method. Soaking is likely to underestimate the number of polar interactions between protein and ligand due to inadequate, highly impaired positions of protein amino acid side chains and main chains. All in all, our study suggests that co-crystallization should be the gold standard to study protein-ligand complex formation. It provides structural information of the equilibrated protein-ligand interactions and complex conformation. This will be the only relevant basis to plan drug optimization as well as SBDD. Application of co-crystallization experiments can be promoted through optimization of co-crystallization protocols aiming at a reduction of material consumption and compound solvent content.

7.7 Abbreviations

Ala: alanine

Asp: aspartate

Arg: arginine

BRD1-4: first bromodomain of bromodomain containing protein 4

DMSO: dimethyl sulfoxide

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

Glu: glutamate

Gly: glycine

Gly-loop: glycine-rich loop

His-tag: histidine-tag

Lyck: lymphocyte-specific kinase

MBT: MES/Bis-Tris

MPD: 2-methyl-2,4-pentanediol

Ni-NTA: nickel-nitrilotriacetic acid

PDB: protein data bank

PfGST: glutathione S-transferase of the malarial parasite plasmodium falciparum

PKA: cAMP-dependent protein kinase

RMSD: root-mean-square deviation

SBDD: structure-based drug design

Ser: serine

TEV: tobacco etch virus

Thr: threonine

TLS: translation/libration/screw

Val: valine

7.8 References

1 Taylor, S. S.; Zhang, P.; Steichen, J. M.; Keshwani, M. M.; Kornev, A. P. *Biochim Biophys Acta* **2013**, *1834*, 1271–1278.

2 Wong, C. F. *Biochim Biophys Acta* **2008**, *1784*, 244–251.

3 Huang, Z.; Wong, C. F. *J Comput Chem* **2009**, *30*, 631–644.

4 Kornev, A. P.; Taylor, S. S. *Trends Biochem Sci* **2015**, *40*, 628–647.

5 Taylor, S. S.; Kornev, A. P. *Trends Biochem Sci* **2011**, *36*, 65–77.

6 Danley, D. E. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 569–575.

7 Hassell, A. M. et al. *Acta Crystallogr D Biol Crystallogr* **2007**, *63*, 72–79.

8 Klebe, G. *Drug Discov Today* **2006**, *11*, 580–594.

9 Steuber, H.; Zentgraf, M.; Gerlach, C.; Sotriffer, C. A.; Heine, A.; Klebe, G. *J Mol Biol* **2006**, *363*, 174–187.

10 Rauh, D.; Klebe, G.; Stubbs, M. T. *J Mol Biol* **2004**, *335*, 1325–1341.

11 Rauh, D.; Klebe, G.; Stürzebecher, J.; Stubbs, M. T. *J Mol Biol* **2003**, *330*, 761–770.

12 Zhu, X.; Kim, J. L.; Newcomb, J. R.; Rose, P. E.; Stover, D. R.; Toledo, L. M.; Zhao, H.; Morgenstern, K. A. *Structure* **1999**, *7*, 651–661.

13 Hiller, N.; Fritz-Wolf, K.; Deponte, M.; Wende, W.; Zimmermann, H.; Becker, K. *Protein Sci* **2006**, *15*, 281–289.

14 Kudlinzki, D.; Linhard, V. L.; Saxena, K.; Sreeramulu, S.; Gande, S.; Schieborr, U.; Dreyer, M.; Schwalbe, H. *Acta Crystallogr F Struct Biol Commun* **2015**, *71*, 1088–1093.

15 Kabsch, W. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132.

16 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658–674.

17 Adams, P. D. et al. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–221.

18 Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486–501.

19 Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J Appl Crystallogr* **1993**, *26*, 283–291.

20 Kleywegt, G. J.; Zou, J.-Y.; Kjeldgaard, M.; Jones, T. A. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; Springer Netherlands: Dordrecht, 2001; 353–356.

21 Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. http://www.globalphasing.com.

22 Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. *J Chem Inf Comput Sci* **2004**, *44*, 2133–2144.

23 Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Huang, C. C.; Ferrin, T. E. *BMC Bioinformatics* **2006**, *7*, 339.

24 Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J Comput Chem* **2004**, *25*, 1605–1612.

25 Stewart, L.; Clark, R.; Behnke, C. *Drug Discov Today* **2002**, *7*, 187–196.

26 Davies, D. R.; Mamat, B.; Magnusson, O. T.; Christensen, J.; Haraldsson, M. H.; Mishra, R.; Pease, B.; Hansen, E.; Singh, J.; Zembower, D.; Kim, H.; S., K.; B., B. A.; E., G. M.; Stewart, L. J. *J Med Chem* **2009**, *52*, 4694–4715.

27 Benson, N.; Boyd, H. F.; Everett, J. R.; Fries, J.; Gribbon, P.; Haque, N.; Henco, K.; Jessen, T.; Martin, W. H.; Mathewson, T. J.; Sharp, R. E.; Spencer, R. W.; Stuhmeier, F.; Wallace, M. S.; Winkler, D. *J Biomol Screen* **2005**, *10*, 573–580.

28 Steuber, H.; Zentgraf, M.; Podjarny, A.; Heine, A.; Klebe, G. *J Mol Biol* **2006**, *356*, 45–56.

29 Yin, X.; Scalia, A.; Leroy, L.; Cuttitta, C. M.; Polizzo, G. M.; Ericson, D. L.; Roessler, C. G.; Campos, O.; Ma, M. Y.; Agarwal, R.; Jackimowicz, R.; Allaire, M.; Orville, A. M.; Sweet, R. M.; Soares, A. S. *Acta Crystallogr D: Biol Crystallogr* **2014**, *70*, 1177–1189.

30 Ember, S. W.; Zhu, J.-Y.; Olesen, S. H.; Martin, M. P.; Becker, A.; Berndt, N.; Georg, G. I.; Schönbrunn, E. *ACS Chem Biol* **2014**, *9*, 1160–1171.

7.9 Supplementary information

Table S1. Crystallographic table for soaked crystal structures of Fasudil (1), 5 and 6. Table spreads over two pages.

Ligand →	**Fasudil (1)**	**5**	**6**
Data collection & processing			
No. Crystals used	1	1	1
Wavelength [Å]	0.918409	0.918409	0.918409
Space group	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)
Unit cell parameters: a, b, c [Å]	58.3; 72.5;109.0	58.1; 72.3;108.8	58.3; 73.1;109.4
Diffraction data[a)]			
Resolution range [Å]	45.45-1.47	43.47-1.72	45.58-1.40
Highest shell resolution range [Å]	1.56-1.47	1.82-1.72	1.48-1.40
Unique reflections	79455(12522)	49381(7745)	91422(13988)
R(I)sym [%][b)]	3.5(47.8)	5.1(49.6)	4.9(46.7)
Completeness [%]	99.7(98.4)	99.3(97.5)	97.9(93.7)
Redundancy	6.6(6.5)	5.3(5.2)	6.5(6.2)
I/σ (I)	27.6(3.8)	20.0(2.9)	19.0(3.0)
Refinement			
Resolution range [Å]	30.85-1.47	41.81-1.72	36.56-1.40
Reflections used in refinement (work/free)	75482/3973	46912/2469	86850/4572
Final R values for all reflections (work[c)]/free[d)]) [%]	16.2/19.0	16.1/19.6	14.8/17.4
Amino acids (PKA/PKI)	353/13	353/13	353/13

Inhibitor atoms	20	17	15
Water molecules	306	257	380
RMSD from ideality			
Bond length [Å]	0.006	0.005	0.008
Bond angles [°]	1.078	0.908	1.186
Ramachandran plot[e)]			
Residues in favoured regions [%]	92.9	92.1	92.5
Residues in additionally allowed regions [%]	7.1	7.9	6.9
Residues in generously allowed regions [%]	0.0	0.0	0.6
Mean B-factors [Å²]			
PKA (protein)/PKI (peptide)	28.3/28.6	31.5/35.7	22.7/23.7
Inhibitor	33.7	29.5	19.0
Water molecules	36.3	35.8	32.4

a)Numbers in parentheses are for the highest-resolution shell.

b)R_{sym}=$[\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}|I_{\mathrm{i}}(h)-\ \langle I(h)\rangle\ |/\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}I_{\mathrm{i}}(h)] \times 100$, $\langle I(h)\rangle$ is the mean of the $I(h)$ observation of reflection h.

c)R_{work}=$\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}-F_{\mathrm{c}}|/\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}$.

d)Calculation of R_{free} was performed as for R_{work} but on 5 % of the data which was excluded from the refinement.

e)Derived from Procheck.[19]

Table S2. Crystallographic table for soaked crystal structures of 4, 3 and 2. Table spreads over two pages.

Ligand →	**4**	**3**	**2**
Data collection & processing			
No. Crystals used	1	1	1
Wavelength [Å]	0.918409	0.918409	0.918409
Space group	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)	19 ($P2_12_12_1$)
Unit cell parameters: a, b, c [Å]	58.1; 72.4;108.7	58.3; 72.7;109.4	58.2; 73.2;108.5
Diffraction data[a)]			
Resolution range [Å]	45.34-1.42	45.46-1.58	43.58-1.67
Highest shell resolution range [Å]	1.50-1.42	1.67-1.58	1.77-1.67
Unique reflections	86529(13515)	62707(8401)	53731(8549)
R(I)sym [%][b)]	3.8(46.2)	5.6(49.5)	6.8(49.0)
Completeness [%]	98.5(96.4)	96.6(81.1)	98.2(98.1)
Redundancy	6.6(6.2)	4.9(4.6)	5.4(5.4)
I/σ (I)	22.7(3.1)	15.7(2.4)	14.3(2.8)
Refinement			
Resolution range [Å]	36.24-1.42	36.33-1.58	42.001.67
Reflections used in refinement (work/free)	82202/4327	59571/3136	51044/2087
Final R values for all reflections (work[c)]/free[d)]) [%]	15.1/18.0	15.7/19.7	16.0/20.0
Amino acids (PKA/PKI)	353/13	353/13	353/13
Inhibitor atoms	20	20	21
Water molecules	303	347	247
RMSD from ideality			

Bond length [Å]	0.008	0.010	0.009
Bond angles [°]	1.144	1.183	1.015
Ramachandran plot[e)]			
Residues in favoured regions [%]	92.5	93.4	90.6
Residues in additionally allowed regions [%]	6.9	6.6	9.1
Residues in generously allowed regions [%]	0.6	0.0	0.3
Mean B-factors [Å²]			
PKA (protein)/PKI (peptide)	28.5/32.0	27.5/29.2	29.6/31.3
Inhibitor	29.0	31.9	23.2
Water molecules	36.1	35.1	33.2

a)Numbers in parentheses are for the highest-resolution shell.

b)$R_{\mathrm{sym}}=[\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}|I_{\mathrm{i}}(h)-\langle I(h)\rangle|/\Sigma_{\mathrm{h}}\Sigma_{\mathrm{i}}I_{\mathrm{i}}(h)]\times 100$, $\langle I(h)\rangle$ is the mean of the $I(h)$ observation of reflection h.

c)$R_{\mathrm{work}}=\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}-F_{\mathrm{c}}|/\Sigma_{\mathrm{hkl}}|F_{\mathrm{o}}$.

d)Calculation of R_{free} was performed as for R_{work} but on 5 % of the data which was excluded from the refinement.

e)Derived from Procheck.[19]

8 Summary (English)

Over the last 20 years, the consideration of biophysical parameters such as kinetics and thermodynamics has been used to guide modern drug design. In contrast to the classical approach that mainly relies on affinity optimization, biophysical parameters allow a further discrimination of potential drug candidates with comparably high affinity. However, there is still a lack of systematic studies analyzing the impact of chemical ligand structure on, for example, thermodynamics. The studies presented in this thesis aim to bridge this gap.

In this thesis a model protein, namely cAMP-dependent protein kinase (PKA) is used to gain particular insights into kinase behavior and thermodynamics upon ligand binding. Due to their implication in various diseases such as cancer, kinases are of utmost importance in drug design.

The ligands used in this study were derived from the approved drug fasudil. They differ in their ligand degrees of freedom, molecular weight and decorations.

Analyzing the impact of ligand degrees of freedom on the thermodynamic signatures, crystal structures were determined. The crystallographic analysis confirmed the flexible nature of the kinase. Particularly the position of the Gly-rich loop differs in the complex structures of ligands with varying ligand degrees of freedom. Thermodynamic signatures were determined using isothermal titration calorimetry. Remarkably, the ligand with the largest amount of internal degrees of freedom appeared to be the binder with the most beneficial entropic contribution. This counterintuitive observation is most likely the result of water displacement from the active site upon ligand

binding and due to a higher ordered local water structure of the ligand in solution prior to protein binding.

For the series of ligands with increasing molecular weight, differences in the ligand coordination with the protein could be observed. A clear trend toward a more entropic and less enthalpic binding upon increasing molecular weight could be observed. Again, this results from structural changes and probably from the state of the uncomplexed ligand in solution, an utterly underestimated factor.

For ligand decoration, introduction of methyl groups is a simple but potentially powerful approach. For differentely methylated ligands not only position but also stereochemistry of the methyl group has an influence on binding potency as well as the thermodynamic signature of ligand binding. Strikingly, the combination of single methyl groups does not lead to additive effects, neither in the binding mode visible in the crystal structure nor in the thermodynamic profile.

Further decorations and fragments of fasudil and adenosine triphosphate (ATP) were crystallographically analyzed focusing on their interaction with the hinge region of PKA. It is a key point of attack of ATP-competitive kinase inhibitors. Even minor changes in chemical ligand or fragment structure, result in severe changes of the hinge binding pose of the respective binders.

A kinase screen testing 16 ligands against 39 different kinases was performed in order to evaluate if thermodynamic properties can be correlated to the selectivity profile of a potential drug. Especially for kinases, selectivity is challenging but of utmost importance. Remarkably, only ΔG correlated well with the determined selectivity profiles.

Finally a methodology approach is presented comparing results from soaking and co-crystallization protocols. The results suggest that structural data from soaking experiments should ideally be verified by cocrystallization, since strong differences between the structures could be observed.

There is still a lack of systematical studies correlating structural data, biophysical parameters and selectivity profiles of closely related ligand series. This is the only way to understand the interplay of these different factors, and only then biophysical parameters exceeding affinity information can reveal their full potential and predictive power for the selection of drug candidates.

9 Zusammenfassung (Deutsch/German)

Seit rund 20 Jahren werden biophysikalische Parameter aus Thermodynamik und Kinetik immer häufiger verwendet, um die moderne Wirkstoffforschung zu verbessern. Im Gegensatz zum klassischen Wirkstoffdesign bei dem nach Affinitäten priorisiert wird, erlauben biophysikalische Methoden potentielle Wirkstoffkandidaten weitergehend zu differenzieren, auch wenn sie gleiche Affinitäten aufweisen. Leider werden mehr systematische Studien benötigt, die den Einfluss der chemischen Ligandstrukturen auf z.B. Thermodynamik untersuchen. Die vorliegende Arbeit zielt darauf ab diese bestehende Lücke zu verkleinern.

In dieser Arbeit wird die cAMP-abhängige Proteinkinase (PKA) als Modellprotein verwendet um Details über das Verhalten von Kinasen und der Thermodynamik bei der Ligandbindung zu untersuchen. Da Kinasen eine wichtige Rolle in verschiedensten Krankheiten spielen (z.B. Krebs), sind sie von höchster Wichtigkeit in der Wirkstoffentwicklung.

Die Liganden die in dieser Studie verwendet werden sind von dem zugelassenen Wirkstoff Fasudil abgeleitet. Sie unterscheiden sich von Fasudil in ihrem Molekulargewicht, der Anzahl ihrer Freiheitsgrade oder durch zusätzliche kleinere Substituenten.

Um die Relevanz von Liganden Freiheitsgraden auf die thermodynamischen Profile bei der Protein-Ligand Bindung zu untersuchen, wurden Kristallstrukturen dieser Komplexe bestimmt. Die Kristallstrukturanalyse bestätigte die bereits bekannte strukturelle Flexibilität der Proteinkinase. Insbesondere die Gly-reiche Schleife zeigt in allen Kristallstrukturen eine gesonderte räumliche Position.

Die thermodynamischen Signaturen wurden mit Hilfe der isothermen Titrationskalorimetrie gemessen. Bemerkenswerterweise zeigt der Ligand mit den meisten intrinsischen Freiheitsgraden, die entropisch vorteilhafteste Bindung an PKA. Dieses kontraintuitive Ergebnis ist vermutlich auf eine höher geordnete lokale Wasserstruktur des Liganden vor der Bindung, in Lösung, zurückzuführen.

Für die Serie, die auf Liganden mit zunehmendem Molekulargewicht fokussiert ist, konnten Unterschiede in der Koordination des Liganden mit dem Protein beobachtet werden. Dabei zeigte sich in den thermodynamischen Profilen ein klarer Trend, bei dem mit zunehmendem Molekulargewicht der entropische Bindungsbeitrag stieg, während der enthalpische Bindungsbeitrag sank. Der Grund hierfür ist vermutlich ebenfalls der Zustand des Liganden in Lösung, ein Faktor der leider häufig unbeachtet bleibt.

Methylgruppen sind einfache Substituenten, die dennoch eine große Wirkung haben können. Bei der Untersuchung von Liganden mit zusätzlichen Methylgruppen in verschiedenen Positionen stellte sich heraus, dass diese nicht nur einen Einfluss auf die Affinität sondern auch auf die thermodynamischen Signaturen haben. Interessanterweise hat die Kombination von verschiedenen Methylgruppen keinen additiven Effekt, weder bezüglich der Thermodynamik noch bezüglich des Bindemodus in der Kristallstruktur.

Andere Substituenten und Fragmente von Fasudil und Adenosintriphosphat (ATP) wurden kristallographisch auf ihre Interaktion mit der Scharnierregion der PKA untersucht. Die Scharnierregion ist der wichtigste Angriffspunkt für ATP-kompetitive Hemmstoffe. Kleinste Änderungen der chemischen

Ligandstruktur führten zu starken Änderungen der Bindeposition und veränderten Interaktionen mit der Scharnierregion.

Zur abschließenden Analyse wurde außerdem eine Kinase-Selektivitätsuntersuchung mit 39 Kinasen und 16 Liganden durchgeführt. Die so erhaltenen Selektivitätsprofile wurden dann auf eine mögliche Korrelation mit den thermodynamischen Daten hin überprüft. Für Kinaseinhibitoren ist eine hohe Selektivität einer der herausforderndsten und wichtigsten Parameter. Bei der Analyse zeigte sich, dass ausschließlich ΔG gut mit den Selektivitätsprofilen korreliert.

Des Weiteren wird in der Arbeit noch eine Methodenuntersuchung diskutiert, bei der Cokristallisation mit der sogenannten „*soaking*“ (engl. Durchtränken) Methode verglichen wird. Dabei konnten signifikante Unterschiede zwischen den resultierenden Kristallstrukturen beobachtet werden. Dies weist darauf hin, dass Ergebnisse aus „*soaking*“ Experimenten idealerweise durch Cokristallisation bestätigt werden sollten.

Der Bedarf an weiteren systematischen Studien, die das Zusammenspiel von Struktur, biophysikalischen Daten und Selektivitätsprofilen von Ligandenserien untersuchen, ist nach wie vor hoch. Solche Daten sind die notwendige Voraussetzung um die Abhängigkeiten dieser Parameter voneinander zu verstehen. Nur so können biophysikalische Parameter wirklich zielgerichtet und gewinnbringend eingesetzt werden um verlässliche Vorhersagen über die Auswahl von potentiellen Wirkstoffkandidaten zu treffen.

10 Danksagung (Deutsch/German)

Als ich meine Diplomarbeit in der Gruppe von Herrn **Prof. Gerhard Klebe** begann, war mir noch nicht klar, dass ich auch für die Promotion bleiben würde. Das kristallisierte sich erst zum Ende meiner Diplomarbeit heraus. Maßgeblich für diese Entscheidung war, dass ich Herrn **Prof. Gerhard Klebe** als Chef sehr zu schätzen wusste. Ich bin ihm sehr dankbar für eine Arbeitsatmosphäre die sowohl menschlich, als auch fachlich angenehm, konstruktiv, von Neugierde und wissenschaftlichem Freiraum geprägt und im Ganzen wunderbar war.

Herrn **Prof. Andreas Heine** möchte ich für seine hervorragende Supervision im Bereich der Kristallographie danken. Mit Geduld, einer wunderbar interaktiven Didaktik und Humor, hat er mir die Grundlagen des kristallographischen Arbeitens beigebracht.

Herrn **Prof. Harald Schwalbe**, **Dr. Henry Jonker**, **Dr. Krishna Saxena** und **Dr. Denis Kudlinzki** möchte ich für eine gute und ergiebige Zusammenarbeit mit stets produktiven Treffen danken.

Außerdem gilt mein Dank den Systemadministratoren **Felix Terwesten**, **Dr. Denis Schmidt**, **Tobias Wulsdorf**, **Felix Gut**, **Sven Siebler**, **Dr. Andreas Spitzmüller**, **Thomas Rickmeyer** und **Dr. Michael Betz,** die mir über die Zeit stets hilfbereit zur Seite standen und für eine funktionierende Rechnerinfrastruktur und reibungslose Arbeitsabläufe gesorgt haben. **Tobias Wulsdorf** danke ich im speziellen auch für die GIST-analysen, die er durchgeführt hat.

Meinem Ehemann **Dr. Denis Schmidt** möchte ich darüber hinaus für viele gute Diskussionen, das Korrekturlesen meiner Arbeit und seine stetige Hilfe bei allen Problemen und Herausforderungen, die die Arbeit mit Computern mit sich bringt, danken.

Besonders möchte ich mich auch bei **Dr. Johannes Schiebel** für eine schöne gemeinsame Zeit im Büro mit vielen fruchtbaren Diskussionen, und für seine große Hilfsbereitschaft bedanken.

Stephanie Dörr hat mit Ihrer Unterstützung bei der Proteinexpression ermöglicht, dass ich ausreichend Protein hatte um über 200 Kristallplatten anzusetzen und außerdem massenhaft ITC-Messungen machen konnte.

Mein besonderer Dank gilt außerdem **Hans Dieter Gerber**, der meine Ligandenserien mit seiner Synthesearbeit vervollständigt hat. Keine Synthese war zu schwer, Hans-Dieter hat immer einen Weg gefunden, wenn ich mit neuen Vorschlägen zu ihm kam.

Außerdem bedanke ich mich bei **Lydia Hartleben**, die immer versucht hat, mir und den anderen Doktoranden den bürokratischen Alltag so einfach wie möglich zu gestalten. Auch unsere vielen netten Gespräche werden mir immer in guter Erinnerung bleiben.

Für viele Gespräche und Tipps sowie für einige unterhaltsame Mittagspausen und eine schöne Büronachbarschaft bedanke ich mich auch bei Herrn **Prof. Peter Kolb**.

Auch den vielen Vertiefern die ich während meiner Zeit betreut habe, danke ich, dies sind: **Olga Pelikh**, **Marcel Köster**, **Levke Andersen**, **Oxana**

Litvinova und **Matthias Oebbeke** der auch als mein Nachfolger gerade seine Masterarbeit in der Arbeitsgruppe macht.

Prof. Klaus Reuter hat mit seiner herzlichen Art und Großzügigkeit den Alltag im Labor und in der Praktikumsbetreuung oft erleichtert.

Christian Sohn möchte ich für die Wartung und Hilfe beim Röngengerät danken.

Helene Köster als meine Vorgängerin auf dem Projekt der Protein Kinase A, gilt ebenfalls mein Dank. Sie hat mich während meiner Diplomarbeit eingearbeitet.

Für eine schöne gemeinsame Zeit und viele Kaffeepausen und Fachdiskussionen danke ich **Dres. Inna** und **Adam Biela**, **Dr. Stephan** und **Maren Jakobi**, **Dr. Manuel Neeb**, **Felix Terwesten**, **Dr. Johannes Schiebel** und **Dr. Alexander Metz**. Auch **Bettina Wilke**, **Dr. Michael Leitner**, **Dr. Eva-Lisa Bodmann**, **Dr. Alexandra Birk, Vroni Thallmair**, **Christian Göcke**, **Thomas Gerlach** und **Michael Klüver** haben für eine wunderbare Zeit in Marburg gesorgt und die ein oder andere Diskussion über die Arbeit hat mir neue Perspektiven eröffnet.

Mein herzlichster Dank gebührt vor allem auch meinen Eltern **Angelika** und **Ulrich Wienen**. Sie haben mich stets unterstützt und mir den Weg geebnet. Mit zwei naturwissenschaftlich ausgebildeten Eltern habe ich Neugierde („Wie funktioniert das eigentlich?“) und präzises analytische Denken und Reden schon in der Kinderstube erfahren und so gelernt. Die Sicherheit, die ich durch meine Eltern erfahren habe, war und ist für mich unverzichtbar. Auch meinen Schwestern **Annette Fronzeck** und **Christina Gierkens**

haben immer ein offenes Ohr für alle Probleme gehabt. Meinem Schwager **Thomas Gierkens** danke ich außerdem für den Druck der Arbeit.

11 Eidesstattliche Erklärung (Deutsch/German)

Erklärung

Gemäß §10 der Promotionsordnung vom 22.04.2009 versichere ich, dass ich meine Dissertation mit dem Titel:

Beyond Affinity: Drug-Kinase Interaction Put under the Microscope

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

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

Marburg, den

(Barbara Wienen-Schmidt)

12 Ausgewählte Forschungspräsentationen (Deutsch/German)

2010 "*RSC-BMCS Fragment-based Drug Discovery Meeting*", Stevenage/England, Mündliche Posterpräsentation, Stipendium zur Konferenzteilnahme

2011 "*Developments in Protein Interaction Analysis*" (DiPIA), Peking/China, Posterpräsentation

2015 „*NovAliX Conference Biophysics in Drug Discovery*", Straßburg/Frankreich, Posterpräsentation, Gewinn des Posterpreises

2015 „*European MicroCal Users' Meeting*", Martinsried/Deutschland, 30-minütiger Vortrag

13 Curriculum Vitae (Deutsch/German)

Aus Gründen des Persönlichkeitsschutzes wird von der elektronischen Veröffentlichung des Lebenslaufes abgesehen.