MASTERARBEIT | MASTER'S THESIS

Titel | Title

Investigating the Underappreciated Role of Weak Interactions in Biotin's Affinity for Streptavidin by NMR

verfasst von | submitted by

Dipl.-Ing. Sarah Elisabeth Kratzwald BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 834
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von | Supervisor:	Univ.-Prof. Dr. Robert Konrat

Der experimentelle Teil dieser Arbeit wurde in der Zeit von Juli 2023 bis Februar 2024 im Mag-Lab GmbH in 1030 Wien unter Supervision von Prof. Dr. Robert Konrat und Dr. Gerald Platzer durchgeführt.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass die vorliegende Arbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen von mir selbsständig erstellt wurde. Alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, sind in dieser Arbeit genannt und aufgelistet. Die aus den Quellen wörtlich entnommenen Stellen, sind als solche kenntlich gemacht.

Das Thema dieser Arbeit wurde von mir bisher weder im In- noch Ausland einer Beurteilerin/einem Beurteiler zur Begutachtung in irgendeiner Form als Prüfungsfach vorgelegt. Diese Arbeit stimmt mit der von den Begutachterinnen/Begutachtern beurteilte Arbeit überein.

Acknowledgements

First and foremost, I want to thank Professor Robert Konrat for allowing me to work on such an interesting and challenging topic. Your guidance and support have been invaluable, and I'm grateful for all the help and support you've given me.

A huge thanks to Gerald Platzer, who supervised me. Your expertise, enthusiasm about (weak) H-bonds, and unwavering support have been crucial throughout this journey. I also highly value your many uplifting words, which always motivated me and kept my spirits high during challenging times.

I also want to thank Katharina Siess for introducing me to wet lab work and for her patience and guidance during the early stages of my research.

Big thanks to Sven Brüschweiler for answering all my NMR-related questions and helping me understand this complex area.

Thanks to Matus Hlavac for the much-needed entertainment and for keeping things light during the tough times.

And to Roman Lichtenecker, thank you for your valuable scientific input and for opening doors into the protein NMR world in the first place.

A special shoutout to Aleksandra Ptaszek for the great collaboration and for being such a fantastic smart person.

I also want to acknowledge Alisa Wimmer, who has been a fellow master's thesis student during this time. Your camaraderie and shared experiences have been a great source of support.

I'm also grateful to the whole group of Prof. Dr. Robert Konrat and everyone at Mag-Lab. Your support and encouragement have made this journey much more enjoyable.

Lastly, I want to thank my boyfriend, friends, and family. Your love, support, and encouragement have been my rock through all of this.

Thank you!

Abstract

The streptavidin-biotin system is known to have one of the strongest non-covalent interactions known in nature. Among classical hydrogen bonds, the complex features several weak non-classical H-bonds like CH-π CH-O, and CH-S bonds. Especially, CH-π interactions are not well investigated yet. NMR spectroscopy is a great tool to study these interactions. By utilizing NMR spectroscopy, we investigated the role of CH-π interactions between biotin and tryptophan residues of streptavidin. In order to do so, we perdeuterated a monovalent streptavidin mutant (mSA2) and extracted chemical shift perturbations of biotin upon binding to streptavidin. Since the π -electrons of tryptophan's indole moiety shield the H-atoms of biotin, we could observe upfield shifts in the ^{1}H NMR spectrum of -0.49 ppm to -3.17 ppm. Additionally, we correlated our experimentally determined CSPs with calculated CSP values from Pople's formalism. Overall, we demonstrated that the extraction of ligand ^{1}H chemical shifts provides insights into biotin's binding mode at atomistic resolution. Moreover, the determination of CSPs allows us to roughly quantify the strength of the formed CH-π interactions. This work offers valuable insights into the contribution of non-classical H-bonds such as CH-π interactions to the overall binding affinity of biotin to streptavidin, which might be further exploited in the design of drugs aiming to target proteins as high-affinity binders.

Zusammenfassung

Die Streptavidin-Biotin Wechselwirkung ist bekannt als eine der stärksten nicht-kovalenten Wechselwirkungen in der Natur. Mittels NMR Spektroskopie untersuchten wir die Rolle von CH-π Interaktionen, die sich zwischen Biotin und den Tryptophan Seitenketten von Streptavidin ausbilden. Dazu haben wir eine monovalente Streptavidin Mutante deuteriert um die ^{1}H Signale des Proteins zu eliminieren. Anschließend wurden die Veränderungen der chemischen Verschiebungen von Biotin im gebundenen vs. ungebundenen Zustand bestimmt. Einige C-gebunden Protonen von Biotin verschieben sich im ^{1}H NMR um mehrere ppm ins Hochfeld, da sie durch das Aromatensystem der Tryptophane abgeschirmt werden. Die Bestimmung der Veränderungen der chemischen Verschiebungen erlaubt uns eine grobe Quantifizierung der gebildeten CH-π Interaktionen.

Im Anschluss korrelierten wir die experimentell bestimmten Werte der Veränderungen der chemischen Verschiebung mit berechneten Werten mittels der Pople-Gleichung. Somit konnten wir zeigen, dass die Extraktion der ^{1}H chemischen Verschiebungen am Ligaden wertvolle Information liefert um den Bindungsmodus von Ligand zu Protein auf atomarer Ebene in nahezu physiologischen Bedingungen genauer zu verstehen.

Diese Arbeit bietet wertvolle Einblicke auf die Wichtigkeit von nicht-klassischen, schwachen H-Brücken (CH-π Interaktionen) im Bezug darauf, wie Liganden eine hohe Affinität zu einem Protein erreichen können.

Contents

1 Introduction

1.1 Streptavidin-Biotin

The tetrameric protein streptavidin, isolated from *Streptomyces avidinii* and its homolog avidin - naturally found in egg whites - binds biotin (vitamin B_7) with an extraordinarily high affinity (dissociation constant $K_d = 10^{-15}$). Each monomer has an eight-stranded β-barrel motive, and each of them binds to one biotin molecule.[1–4] The understanding of the properties that drive this interaction has been the subject of numerous studies. Hence, Weber et al.[4] and Hendrickson et al.[3] first revealed the crystal structure of the complex in the late 1980s. The X-ray structure of the streptavidin-biotin complex demonstrated the involvement of multiple hydrogen-bonds (H-bonds), van der Waals interactions and dipole arrays, which stabilize oxyanion holes in the biotin-binding pocket.[3,4] As the crystal structure of the wt streptavidin in Figure 2A and B shows, biotin is surrounded by four tryptophan residues. Three of these residues (Trp-79, Trp-92, Trp-108) are within the same subunit that is binding biotin. The fourth tryptophan residue surrounding biotin (Trp-120) sits in an adjacent subunit. Later, it was shown, that the contact of biotin to the Trp-120 of the adjacent subunit also contributes to the high biotin-binding affinity, as well as the tight association of the streptavidin subunits.[5] More recently, Ozawa et al.[6] have investigated the role of these tryptophan residues as contributors to the strong biotin-streptavidin interaction and could first computationally demonstrate the contribution of CH-π interactions to the strong affinity of the streptavidin-biotin complex.[6]

Monovalent Streptavidin

In many bioanalytical applications, such as NMR spectroscopy, the oligomeric form of a protein - specifically, the tetrameric form in this case - presents certain limitations compared to its monomeric counterpart. In the context of NMR spectroscopy, the monomeric form offers more favorable relaxation properties than the larger tetrameric protein. Consequently, the larger protein complex can cause line broadening, a reduced signal-to-noise ratio, and increased signal overlap.[7] Therefore, utilizing monomeric streptavidin can be advantageous in various applications, including NMR spectroscopy.

Recently, streptavidin and its rhizavidin sequences have been combined and engineered to a monomeric streptavidin (mSA) mutant. Rhizavidin is a homolog of streptavidin and binds to biotin as a dimer with high affinity. Unlike streptavidin, rhizavidin binds biotin with a single subunit. The binding pocked was engineered by replacing amino acids in the biotin-binding pocket of streptavidin with those from rhizavidin, as highlighted in Figure 2E in blue (streptavidin amino acids) and

magenta (rhizavidin amino acids). For instance, the streptavidin residues between Ser-45 and Tyr-54 were replaced with the residues of rhizavidin Asn-45 through Tyr-55 (see Figure 1) constituting the loop 3,4 of the protein. Additionally, loop truncations were implemented, and surface residues of the protein were mutated to enhance solubility and prevent tetramer formation of the protein. Notably, many of these residues are common to both proteins as illustrated by the grey-colored amino acids in Figure 2E. Importantly, the tryptophan amino acids in the biotin binding pocket (Trp-79, Trp-92 and Trp-108) are the same in the wt and mutant. The Trp-120 residue, which belongs to the adjacent subunit in the wt streptavidin however is missing in the mSA mutant.[8] The affinity of the wt streptavidin to biotin however remains orders of magnitudes higher with a K_d = 400 pM versus K_d = 2.8 nM for mSA. In mSA, the binding pocket is more accessible to solvents, making it easier for biotin to dissociate as water molecules can compete with streptavidin-biotin hydrogen bonds. To address this, further development of mSA aimed at reducing the biotin off-rate. This was achieved by substituting Ser-25 with His-25, resulting in monomeric streptavidin 2 (mSA2), which displays slower biotin dissociation kinetics.[9]

Regarding experimental aspects, the expression and purification of mSA2 have been optimized in *E. coli* to circumvent purification from inclusion bodies.[10]

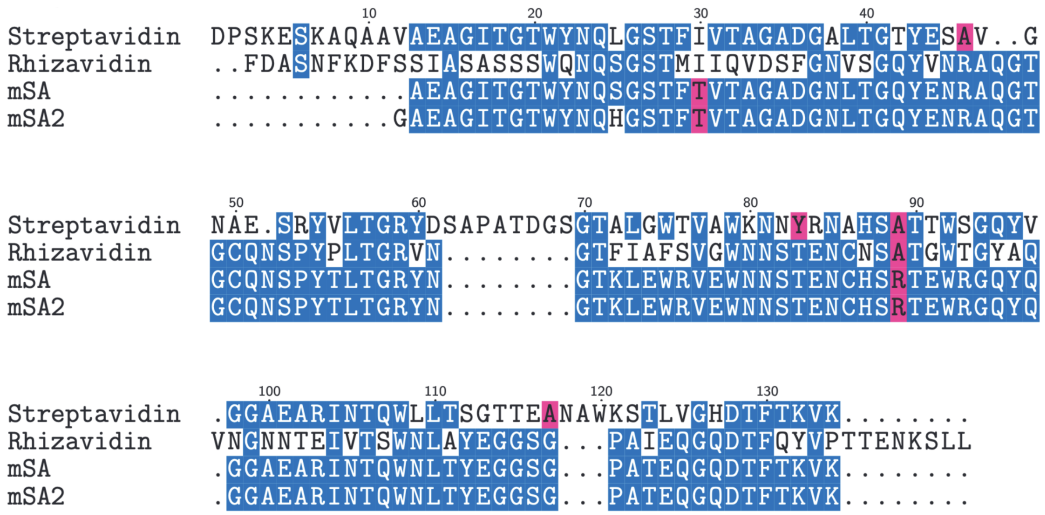

Figure 1: Amino acid sequences of wt streptavidin, wt rhizavidin, monovalent streptavidin (mSA) and monovalent streptavidin 2 (mSA2)

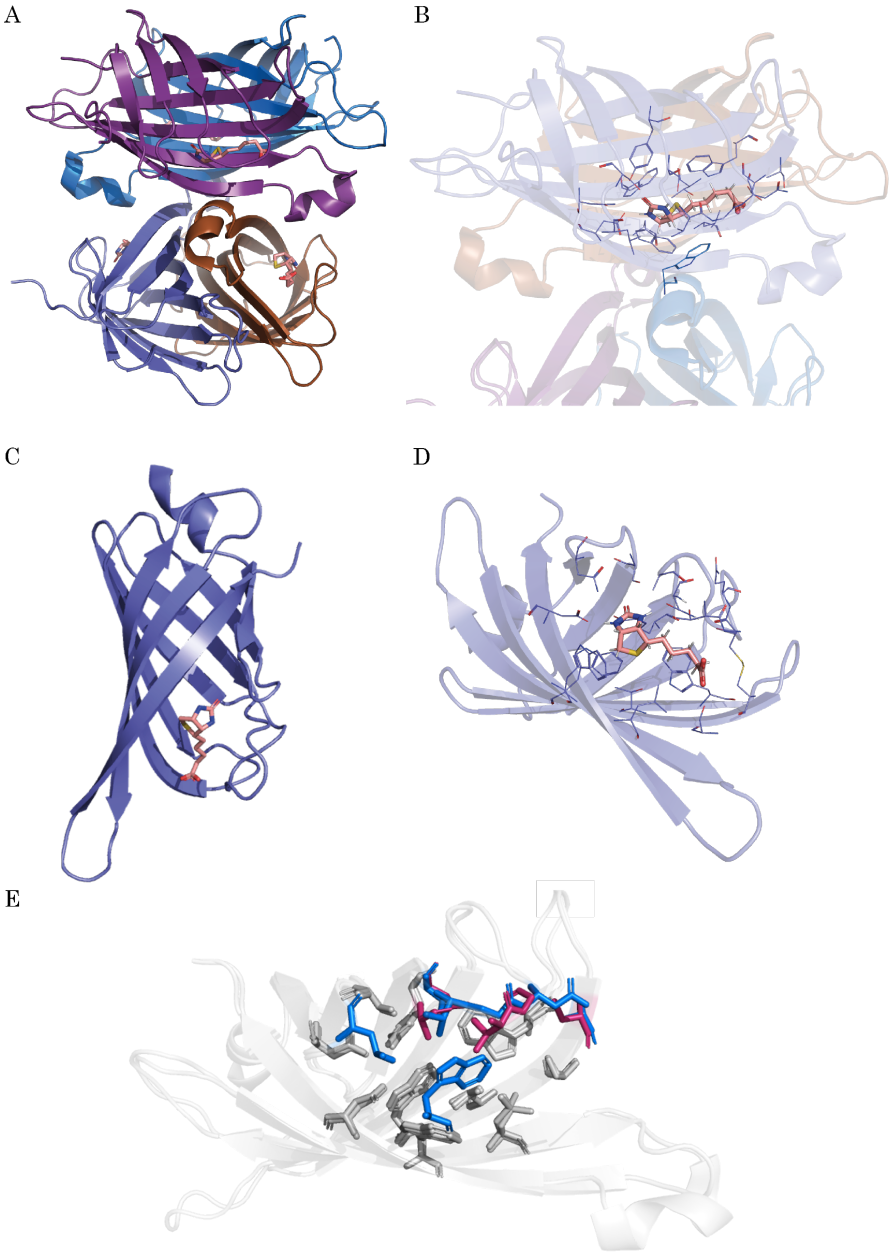

Figure 2: Crystal structures of biotin: A) wt streptavidin as a homotetrameric protein; B) amino acids in the biotin binding pocket of wt streptavidin (*PDB ID 3RY2*); C) The mSA protein as β-barrel structure, D) biotin in the binding pocket of mSA (*PDB ID 4JNJ*); E) aligned crystal structures of the WT and mSA highlighting differences in amino acids in the biotin binding pocket; the magenta colored amino acid belong to mSA and the blue colored ones belong to the WT, whereas the blue Trp belongs to the adjacent subunit.

1.2 Protein-Ligand Interactions

Protein-ligand interactions are driven by the affinity of one binding partner to the other. In order to understand the underlying mechanisms of such binding events, it is necessary to understand the physicochemical principles driving the interactions.[11]

1.2.1 Protein-Ligand Binding Kinetics

The protein-ligand binding kinetics describe the rate at which the association and dissociation between a protein and a ligand takes place. It can simply be described with the following equation:

$$P + L \underset{\mathrm{k}_{off}}{\overset{\mathrm{k}_{on}}{\rightleftharpoons}} PL \tag{1}$$

Thereby, PL is the protein-ligand complex, k_{on} and k_{off} are the kinetic rate constants for the association and dissociation of the complex, respectively. In equilibrium, the forward reaction is in balance with the reverse unbinding reaction an written as:

$$k_{on}\,[P]\,[L] = k_{off}\,[PL] \tag{2}$$

Hence, with the square brackets defining the concentration at equilibrium, the binding konstant K_b in M^{-1} as unit is defined as:

$$K_b = \frac{k_{on}}{k_{off}} = \frac{[PL]}{[P]\,[L]} = \frac{1}{K_d} \tag{3}$$

Whereas, K_d in M as unit is referred to as dissociation constant. With that, a fast binding rate together with a slow dissociation rate results in a low dissociation constant and therefore in a high binding affinity.[11]

1.2.2 Thermodynamic Aspects in Protein-Ligand Interactions

When looking at the specific and affine binding process between a protein and a ligand, the system - the solvent - in which the binding takes place has to be taken into consideration. Regarding the thermodynamic laws of such a binding process, it involves enthalpic (ΔH) and entropic (-TΔS) parts. Binding between the protein and a ligand takes place if ΔG (Gibbs' free binding energy) < 0.[11,12]

$$\Delta G = \Delta H - T\Delta S \tag{4}$$

The enthalpic term ΔH of the free binding energy includes ionic and H-bonds, electrostatic (Coulomb) and van der Waals interactions, as well as polarization of the interacting functional groups. Within a binding process, ΔH reflects the change in energy of the system when the ligand binds to the protein. This change in energy

not only includes the formation of favorable interactions, but it also includes the loss of H-bonds between protein and solvent molecules as well as between ligand and solvent molecules.[11]

The entropic term ΔS generally indicates the overall increase (positive value) and decrease (negative value) of degrees of freedom of the system. The overall change in entropy upon protein-ligand binding can be re-written as the sum of:

$$\Delta S = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{r/t} \tag{5}$$

ΔS_{solv} describes the release of H_2O molecules tightly bound to the apo-protein upon ligand, which usually contributes favorably to the overall ΔS.[12,13] ΔS_{conf} denotes the conformational entropy, reflecting the change in the conformational freedom of the protein and ligand upon binding. This contribution to the overall binding entropy can be either favorable or unfavorable. $\Delta S_{r/t}$ represents the reduction in translational and rotational degrees of freedom for the protein and ligand when they form a complex. When a ligand binds to a protein, the entropic penalty (negative $\Delta S_{r/t}$) needs to be overcome. That can happen due to a large solvent entropy gain (ΔS_{solv}) or favorable interactions in the protein-ligand binding event (negative ΔH).

Biotin binds to streptavidin with a ΔG of - 18.1 kcal/mol.[14] Thereby, it is an enthalpic binder, implying that the strong streptavidin-biotin affinity is predominantly driven predominantly by specific interactions such as H-bonds.

1.2.3 Hydrogen-Bonds

H-bonds are prevalent in nature and play a crucial role in facilitating protein-ligand interactions. The International Union of Pure and Applied Chemistry (IUPAC) defines an H-bond as "an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X-H, where X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, with evidence of bond formation".[15]

One intriguing aspect of biotin's high affinity for streptavidin is its relatively small molecular size, with a molecular weight (MW) of 244 g/mol. Biotin consists of 31 atoms, 15 of which are hydrogen atoms at physiological pH. This high hydrogen atom content significantly contributes to its strong binding affinity through multiple hydrogen bonds being formed with streptavidin. The following paragraphs will discuss general common hydrogen bonds and those specifically occurring in the biotin-streptavidin interaction.

Classical H-Bonds

Classical H-bonds involve an H-atom which is covalently bound to an electronegative atom as an H-bond donor and another electronegative atom as an H-bond acceptor. For reference, the electronegativity on the Pauling scale for the hydrogen is 2.20 and for the oxygen and nitrogen are 3.44 and 3.04, respectively.[16] Thus, classical H-bonds are referred to as strong H-bonds. Common classical H-bonds with prominent examples and typical binding energies are:

- O-H$\cdots$O Bonds

  The most prominent example of O-H$\cdots$O are the H-bonds formed between water molecules and have a binding energy ranging from 3.6 - 9.5 kcal/mol.[16]

- N-H$\cdots$O Bonds

  An example for N-H$\cdots$O H-bonds are the ones in proteins between amides and carbonyl oxygens with a binding energy ranging from 2.8 - 7.2 kcal/mol.[17]

- N-H$\cdots$N Bonds

  The N-H$\cdots$N H-bond is among others responsible for DNA base pairing and contributes with a binding energy of 2.8 - 6.0 kcal/mol.[16]

- O-H$\cdots$N Bonds

  The O-H$\cdots$N can be found between hydroxyl groups of e.g. serine and the N-atom of the protein backbone and the binding energy typically corresponds to 2.8 - 6.0 kcal/mol.[16]

Figure 3 depicts the H-bonds between biotin and streptavidin, which are evident in the crystal structure. In Figure 3A and B the classical H-bonds of biotin's ureido moiety are depicted, whereas in A the formed H-bonds between the -NH protons of the ureido moiety with the aspartate side-chain and the serin side-chain are depicted. Figure 3B shows the oxyanion H-bonds of biotin's carbonyl group with streptavidin's asparagine, serine, and tyrosine side-chain are shown. (Figure 3B also illustrates the charge-dipole interaction of biotin's carboxylate oxygens with streptavidin's serine and asparagine side chain.)

Non-Classical H-Bonds

Non-classical H-bonds are often referred to as weak H-bonds.[16] Thus, they may involve hydrogen donors and/or acceptors that are less electronegative, like carbon or sulfur with an electronegativity of 2.55 and 2.58 on the Pauling scale, respectively.[18] Some abundant non-classical H-bonds are listed with examples and corresponding binding energy:

- C-H$\cdots\pi$ Bonds

  C-H$\cdots\pi$ interactions are prevalent in a huge number of protein-ligand interactions. One CH-π interaction contributes about 0.5 - 2.0 kcal/mol to the overall stability of a protein-ligand complex.[19]

- O-H$\cdots$S Bonds

  O-H$\cdots$S interactions are common in proteins between a hydroxyl group from a serine, threonine or tyrosine with a sulfur from cysteine or methionine and contribute with a binding energy of 0.2 - 1.2 kcal/mol.

- C-H$\cdots$O Bonds

  C-H$\cdots$O interactions take place between a C-H group of an aliphatic side chain group with a carbonyl oxygen and contribute with a binding energy of 0.5 - 1.5 kcal/mol.[20]

- C-H$\cdots$N Bonds

  An example of C-H$\cdots$N interactions is the one formed between the C-H of a methylene group and the nitrogen in a histidine with a binding energy of 0.4 - 1.0 kcal/mol.[21]

- C-H$\cdots$S Bonds

  An example of C-H$\cdots$N interactions are the ones between the sulfur of a cysteine or methionine with aliphatic C-H groups with a binding energy of 0.5 - 1.2 kcal/mol.[22]

- C-H$\cdots$F Bonds

  Since fluorine does not naturally occur in biological systems, the interaction solely takes place between aliphatic amino acid side chains and a F-atom of an organic compound (the ligand). Their binding energy ranges from 0.5 - 1.0 kcal/mol.[23]

As illustrated in Figure 3C biotin forms four apparent CH-π interactions with two intramolecular tryptophan residues (Trp-79 and Trp-108) and in Figure 3D another two with the tryptophan residue (Trp-120) of the adjacent streptavidin subunit. Additionally, biotin's sulfur forms an O-H$\cdots$S H-bond with the hydroxyl group of Thr-90 as illustrated in Figure 3B.

CH-π interactions are weak H-bonds, where the C-H serves as the proton-donor, and the π -electrons of the aromatic ring as the proton-acceptor group. However, they are often misclassified as hydrophobic interactions.[24] Due to the magnetic ring current of the aromatic system, the proton is shielded, resulting in an upfield shift in the NMR spectrum. In contrast, the proton in a classical hydrogen bond is deshielded by the electronegative acceptor atom, such as oxygen or nitrogen, causing a downfield shift in the NMR spectrum. Although the deshielding effect of the CH-π interaction due to H-bond formation cannot be observed in an NMR spectrum, computational methods can correct for the shielding effect of the aromatic π-system. Once this correction is applied, the remaining deshielding effect confirms the formation of a hydrogen bond, which is consistent with the behavior of classical hydrogen bonds.[25]

The importance of these weak interactions in protein-ligand recognition becomes evident when looking at a systematic analysis of more than 11 000 protein-ligand interactions. This research has highlighted that the interactions between C-bound protons with aromatic ring systems (CH-π H-bonds) are the most common interactions found between high-efficiency ligands and their protein targets.[24]

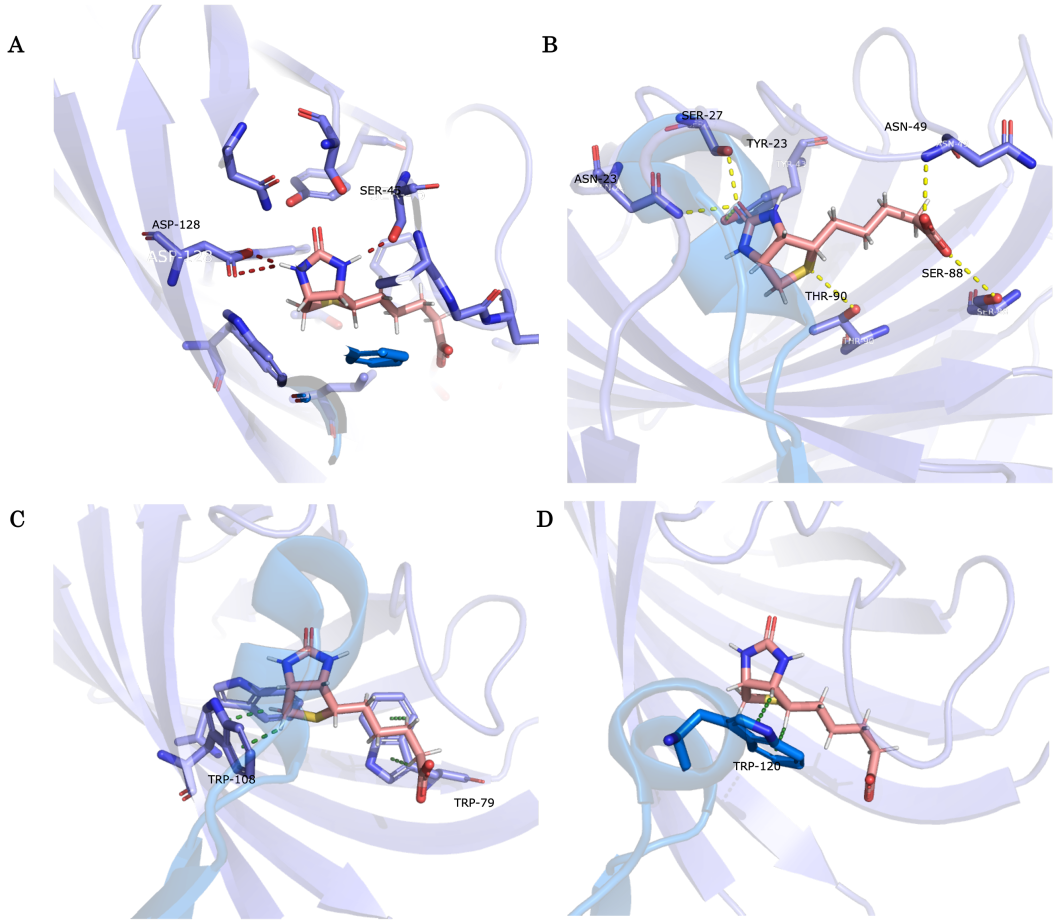

Figure 3: Biotin´s interactions with streptavidin: **A** Classical H-bonds formed between biotin's -NH protons and streptavidin **B** Classical H-bonds formed between the ureido's carbonyl with streptavidin, sulfur centered H-bond and charge-dipole interactions of biotin's carboxylate with streptavidin. **C** CH-π interactions with intramolecular Trp-residues, **D** CH-π interactions with intermolecular Trp-residues of the adjacent streptavidin subunit. *PDB ID 3RY2*

1.3 Protein NMR

Deciphering 3D structural information of proteins and protein-ligand interactions can be achieved via several methods and techniques. Among these techniques, X-ray crystallography is the main method to elucidate the structure of a biomolecule. However, X-ray crystallography shows one rigid snap-shop of a protein and is not capable of capturing the dynamical behavior of a protein in solution.[26]

An alternative or complementary approach is Nuclear Magnetic Resonance (NMR) spectroscopy. It relies on the chemical shift (CS) of NMR-active nuclei such as ^{1}H, ^{13}C or ^{15}N. In the simplest terms, the CS represents the resonance frequency of a nucleus dependent on its chemical environment and is measured relative to a standard. When a sample is placed in a magnetic field (B_0), nuclei with a spin of 1/2 align with or against the field. A 90° pulse (B_1) perturbs the alignment and causes the nuclei to process around B_0 at the Larmor frequency. The precession results in a free induction decay (FID) signal, which, after Fourier transformation, generates an NMR spectrum containing the CS information. Upfield shifts (lower ppm values) indicate a higher electron density around the nucleus, while downfield shifts (higher ppm values) indicate a lower electron density caused by the proximity of electron-withdrawing groups that deshield the nucleus.[27]

The atomistic resolution of NMR allows to show structural information of proteins. Hence, NMR is an invaluable tool for studying protein dynamics, interactions and changes in solution - at (almost) physiological conditions.[28]

Several NMR methods have been developed to allow for the investigation of protein-ligand interactions, such as Saturation Transfer Difference (STD) experiments[29] or the usage of Chemical Shift Perturbations (CSP)[30].

1.3.1 Chemical Shift Perturbations

CSP describe the change in chemical shift of a protein upon the addition of a ligand. Usually, this is achieved by expressing the protein of interest uniformly ^{15}N-labeled and monitoring the change of chemical shift in a 2D ^{1}H-^{15}N HSQC spectrum upon addition of the ligand in titration steps. Hence, information about involvement of each amino acid to the binding can be gained.[30] In previous literature, another example of measuring CSP as reporters for CH-π interactions has been shown. In that case, tryptophan residues of the protein have been selectively labeled by the addition of metabolic precursors containing an isolated ^{1}H-^{13}C spin system. That allows the recording of a ^{1}H-^{13}C HSQC spectrum only displaying peaks of the tryptophan residues of the protein. Upon addition of a ligand that binds to the protein in a nanomolar range, CSPs up to -2.74 ppm in the ^{13}C dimension were observed. Hence, that work demonstrated that CSPs can serve as a direct experimental output to identify CH-π interactions formed between protein and ligand.[31] Recently, this

approach was developed further. Thereby, the CSPs were measured on the ligand upon binding to a deuterated protein. As in the previously mentioned approach, these CSPs act as reporters for CH-π interactions between a protein-ligand complex.[32] When CSPs are measured on the ligand site, it involves protein deuteration and thereby depletion of ^{1}H signals from the protein. It is based solely on the peak assignment of the ligand´s H-atoms.[32,33] In contrast to the previous method of detecting CSP on the protein site, the detection of CSP on the ligand site involves - as mentioned - protein deuteration. Thus, the standard experiment would not be an HSQC experiment, which can be done in a time of 60 min for a 100 μM sample, but rather ^{1}H-^{1}H Nuclear Overhauser Enhancement Spectroscopy (NOESY) and Total Correlation Spectroscopy (TOCSY) NMR experiments are used. These experiments potentially allow the correlation of unbound ligand ^{1}H signals with the ^{1}Hs of the ligand when bound to the protein.[32]

The following sections briefly describe NMR experiments that have been used throughout this thesis.

1.3.2 Two-Dimensional NMR Spectroscopy

Heteronuclear Single Quantum Coherence Spectroscopy

Within protein NMR applications, the 2D Heteronuclear Single-Quantum Coherence (HSQC) experiments are the most utilized ones. Often, ^{1}H-^{15}N HSQC spectra serve as 'fingerprint' of the desired protein. The production of ^{15}N labeled proteins is straighforward and relatively cost-effective, since it only requires the addition of $^{15}NH_4Cl$ as nitrogen source to the expression medium, and is therefore usually acquired first when working with a new protein.[34]

Moreover, ^{1}H-^{15}N experiments provide valuable information about the folding state of the protein. If the ^{1}H CS are well dispersed, the protein most likely has a defined secondary structure. On the other hand, if the ^{1}H CS display a narrow dispersion, the protein is likely intrinsically disordered. This makes HSQC experiments a powerfull tool for quickly assessing the structural state of a protein.[28]

In a HSQC spectrum one-bond (^{1}J) scalar couplings of two different nuclei are correlated. This technique was first established for ^{1}H-^{15}N correlations[35], but can be performed as ^{1}H-^{13}C[31] or even ^{19}F-^{13}C.[36] The advantage for ^{1}H-^{15}N HSQC experiments lies in the scalar $^{1}J_{NH}$ coupling (90 Hz).

The general scheme for a HSQC experiment is as follows: magnetization is transferred from the more sensitive nucleus I (^{1}H) to the less sensitive nucleus S (^{15}N, ^{13}C, ...) like in an Insensitive Nuclear Enhancement by Polarization Transfer (INEPT).

In the following time delay (t_1), spin magnetization for the heteronucleus S evolves. Then, the magnetization is transferred back from the heteronucleus S to the ^{1}H I for detection.[37]

Correlation Spectroscopy and Total Correlation Spectroscopy

Amongst 2D-NMR experiments, the homonuclear Correlation Spectroscopy (COSY) experiment is one of the most popular ones. It transfers magnetization from one spin to another if there is scalar coupling. The diagonal peaks of a COSY spectrum have the same shifts in both dimensions. The symmetrical crosspeaks in a COSY spectrum indicate the couplings (^{2}J and ^{3}J) of nuclei within the molecule.[38]

Similar to a COSY experiment, in a TOCSY experiment cross-peaks display nuclei that have a scalar coupling. However, as the name suggests, coupling between nuclei can be observed if they are connected over more than 3 bonds. This extended range of coupling makes TOCSY particularly useful for studying larger molecular fragments and providing comprehensive information about the spin system within the molecule.[38]

Nuclear Overhauser Enhancement Spectroscopy

The nuclear Overhauser effect (NOE) is an effect that is based on the transfer of magnetization of spin active nuclei (usually ^{1}H) that does not only occur through bonds via scalar coupling, but through spatial proximity (< 5 Å). In a 1D NOE spectrum, the change in the intensity of one NMR resonance, when another resonance is held in saturation with a radio-frequency field, indicates spatial proximity between the nuclei. If the intensity of a nucleus changes under these conditions, it implies that this nucleus is close to the saturated one. The underlying relaxation mechanism is referred to as the cross-relaxation.[39]

Figure 4 illustrates the energy diagram for a two-spin system SI with four energy levels $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, $\beta\beta$ (from lowest to highest). The W_0, W_1, and W_2 are the probabilities for a zero-, single-, or double-quantum spin-state transition to take place. When the spin state populations are not in their thermodynamic equilibrium, relaxation processes bring the system back to thermal equilibrium. Thereby, cross-relaxation is the underlying relaxation mechanism, where a transition of the spin I transfers magnetization to the spin S.

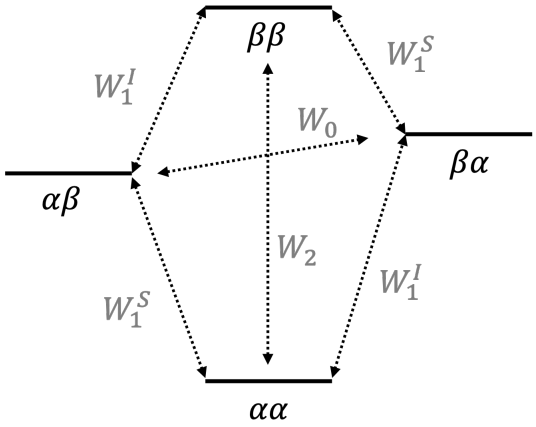

Figure 4: Solomon diagram for a two-spin system AX with four energy levels: $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, $\beta\beta$. W_0, W_1 and W_2 are the probabilities for a transition to take place.

The pulse sequence for a 2D NOESY experiment starts with a 90° pulse ($I_z \rightarrow I_y$) followed by the incremented evolution time t_1. Then a second 90° ($I_y \rightarrow I_{-z}$) marks the start of the mixing phase τ_m. In the time of τ_m the NOE takes place, hence cross-relaxation from spin I to spin S occurs ($I_{-z} \rightarrow S_z$). The third 90° pulse brings magnetization back to the transverse plane ($S_z \rightarrow S_y$) and then the FID can be detected. In the resulting 2D spectrum the diagonal peaks correspond to spins I and S and the crosspeaks in the spectrum correspond to the magnetization that has been transferred referred to as NOE peaks.

2 Aim

The high affinity of biotin to streptavidin is harnessed within a wide range of biotechnological applications and there has been a long-lasting interest in deciphering the driving forces of this interaction. Previous research has focused on describing the classical H-bonds which are contributing to the enthalpic binding energy. However, non-classical H-bonds, which are often mistermed as hydrophobic effects, have only been discussed rarely.[6] This thesis aims to investigate non-classical hydrogen bonds using NMR spectroscopy. Chemical shifts are sensitive to the chemical environment and can provide information about the formation of H-bonds. The subsequent detection of ligand-based CSPs reveals details about both the free and bound states of the ligand, allowing for straightforward analysis.

One aim was to optimize the expression and purification protocol for the monovalent streptavidin mutant (mSA2). Subsequently, we aimed to measure NMR experiments to allow for the determination of CSPs of biotin upon binding to streptavidin as reporters for the formation of favorable CH-π interactions. Moreover, we aimed formation to compare experimentally determined CSPs with mathematically calculated ones, whereas we extracted geometric parameters for the mathematical formalism from the X-ray structure of the streptavidin-biotin complex.

Overall, we aimed to gain a deeper understanding of the driving forces for the streptavidin-biotin interaction by utilizing NMR spectroscopy. Whereas NMR spectroscopy harbors the advantage of providing information in atomistic resolution and about a protein-ligand complex

3 Results

3.1 Expression and Purification of mSA2

Protein expression and purification was optimized from previously reported conditions.[10] To ensure for high protein yield, cells were induced at an Optical Density at 600 nm (OD_{600}) of 0.94 ± 0.2. Furthermore, biotin was added to the expression medium as well as to the lysis buffer.[10] An expression test, where different *E. coli* strains, temperatures and timeframes between induction and harvesting was performed. Hence, cells were harvested after 18 h of induction with 0.4 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C yielding the highest H6-MBP-mSA2 expression levels. In previous literature, the authors reported purification of mSA2 after protease cleavage simply by heating to elevated temperatures. Hereby, the solubility-tag MBP as well as the uncleaved protein construct H6-MBP-mSA2 was reported to precipitate.[10] However, we could not reproduce this purification, since impurities remained. Therefore, after 3C protease cleavage, a reverse Ni-NTA column was performed, followed by a size exclusion chromatography which successfully removed impurities, that mainly consisted of the MBP-tag as seen in the corresponding SDS page (Figure 9). For the perdeuterated mSA2 expression, one sample with a volume of 300 μl and 860 μM was obtained from 1 l of expression media.

Furthermore, we tried to generate the apo-mSA2 by removal of biotin from the mSA2-biotin complex for ^{15}N-uniformly labeled samples. Different conditions were tried to keep the apo-form of mSA2 in solution, like the addition of 3% glycerol (v/v) or the equimolar addition of aspartate and glutamate (50 mM)[40] to the final sample. However, mSA2 did not remain in solution upon biotin removal.

Hence, we expressed H6-MBP-mSA2 without the addition of biotin to the expression medium or lysis buffer. Since the stability of mSA2 was limited under these conditions, NMR measurements were taken right after concentrating the protein sample after the reverse Ni-NTA column to confirm CSPs of the apo-mSA2 vs the mSA2-biotin complex within a ^{1}H-^{15}N HSQC experiment. Although this approach did not give a perfectly pure sample, it allowed the measurement of a ^{1}H-^{15}N HSQC spectrum.

3.2 ^{1}H-^{15}N HSQC Measurement

The ^{1}H-^{15}N HSQC spectrum of the apo-mSA2 protein sample was recorded in tris-buffer. After spectrum acquisition of the apo-form, biotin was added to the apo-mSA2 protein sample to a final concentration of 1 mM. Subsequently, another ^{1}H-^{15}N HSQC spectrum was recorded. Hereby, spectra of mSA2-biotin (1:1 complex), apo-mSA2, and the regenerated mSA2-biotin (with biotin excess) were recorded. The spectrum of the apo-mSA2 and the subsequently regenerated mSA2-biotin spectrum

are depicted in Figure 5. Spectra of the mSA2-biotin and the regenerated mSA2-biotin show identical peaks in the ^{1}H-^{15}N HSQC spectrum.

Although no assignment for the mSA2 protein was performed, the CSPs from the ^{1}H-^{15}N HSQC confirm the binding of biotin to mSA2.

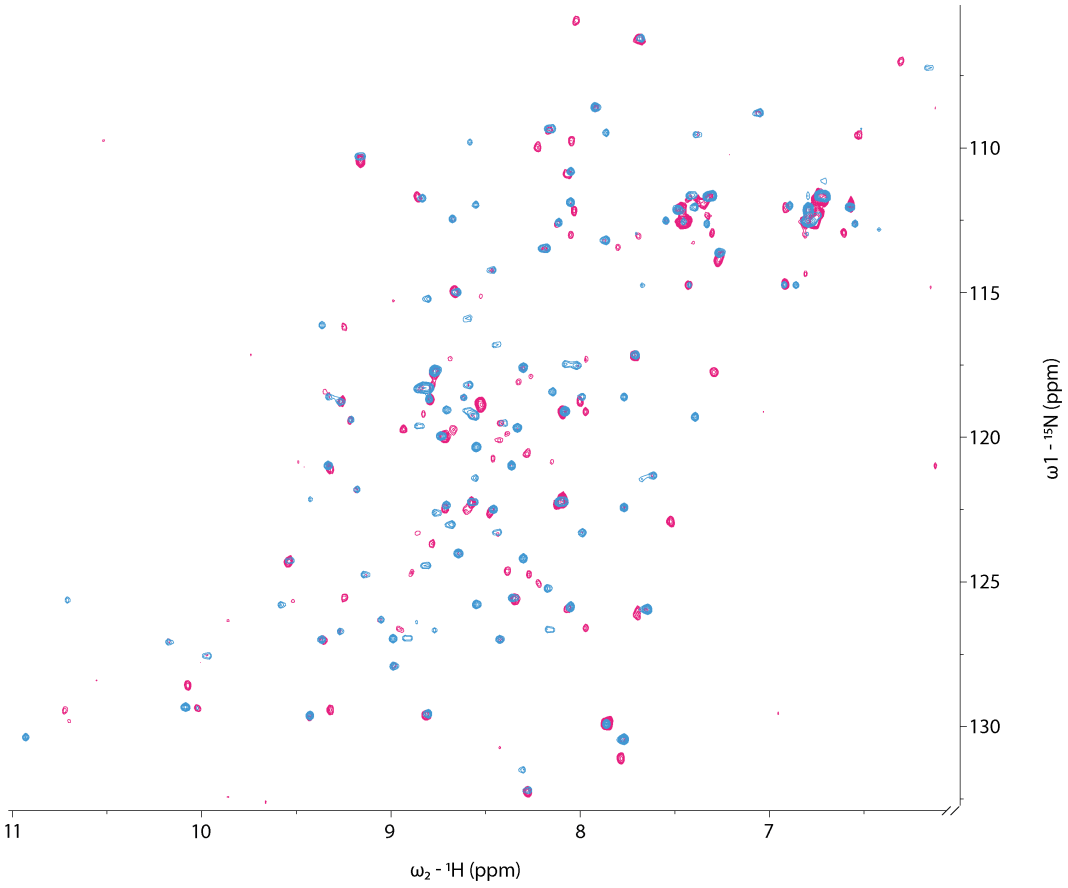

Figure 5: Overlay of ^{1}H-^{15}N HSQC spectra of apo-mSA2 (pink) and mSA2-biotin (blue). The observed CSPs confirm the binding of biotin to mSA2.

3.3 Biotin-Based Chemical Shift Perturbations upon Streptavidin Binding

Similar as reported in recent literature, protein deuteration was performed to allow for the detection of ^{1}H ligand-based CSPs.[32] All ^{1}H signals of the protein are depleted in ^{1}H NMR spectra, and biotin´s ^{1}H signals can easily be detected.

In previous literature, bound ligand signals were assigned by the usage of ^{1}H-^{1}H NOESY experiments and an excess ligand concentration. Thereby unbound ligand to bound ligand signals could be correlated and the experiment allowed for direct extraction of CSP values.[32] We recorded ^{1}H-^{1}H NOESY spectra with different mixing times (100 ms, 150 ms, 300 ms) which display crosspeaks of protons that are spatial proximity to each other. We recorded ^{1}H-^{1}H COSY as well as ^{1}H-^{1}H TOCSY spectra with different mixing times (7 ms, 25 ms, 50 ms, and 70 ms) to extraxt through-bond connectivity. ^{1}H peak assignment was planned to be done via trough-bond connectivity information from the COSY and TOCSY experiments and spatial proximity information of protons from the NOESY experiments. However, the information content of the COSY and TOCSY experiments was limited, and the TOCSY spectra with mixing times of 50 ms and 70 ms barely showed peaks.

We were able to observe signals in the ^{1}H-^{1}H NOESY spectrum with crosspeaks from the bound ^{1}H ligand signals to the unbound ^{1}H ligand signals. These crosspeaks were unexpected due to the high biotin affinity with a K_D in the low nanomolar range and the usage of the mSA2-biotin complex at a 1:1 ratio. However, since some protein precipitation was observed after the first 2D measurements (approximately after 72 h), the free ligand concentration has increased. Thereafter, ^{1}H ligand signals could be assigned from one single ^{1}H-^{1}H NOESY spectrum. Figure 6A shows the chemical structure of biotin with the used proton labels. Figure 6B displays the scheme for the assignment strategy of the NOESY crosspeaks. The dashed red lines illustrate the ^{1}H ligand signals h and h' of the free form, and the dashed blue lines illustrate the ^{1}H ligand signals h and h' of the bound form. Moreover, not only crosspeaks from the free to the bound form were observed, but also from the free to neighboring bound form (h_f - h_b - h'_b) - so-called relay crosspeaks. Figure 6C and D show the correlation of free ^{1}H signals a, a', b, b' (in C) and c, c', d, d' and e' (in D) to the bound ^{1}H signals a, a', b, b', f (in C) and c, c', d, d' and e' (in D). The ^{1}H CS of free and bound form, as well as the observed CSPs are summarized in Table 1.

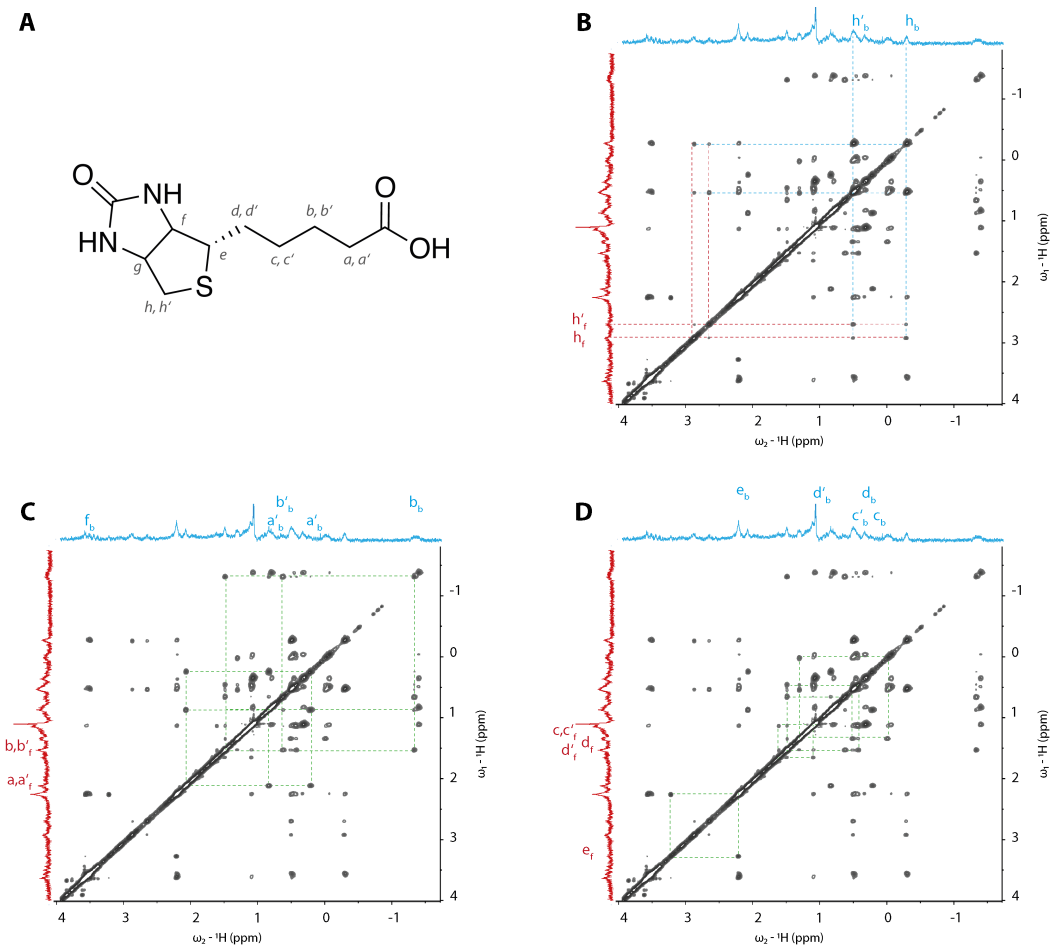

Figure 6: ^{1}H-^{1}H NOESY spectra of biotin in complex with mSA2 with c = 0.85 mM. Spectra were recorded as mSA2-biotin 2:3 complex, due to partial precipitation of mSA2, biotin is in 1.5-fold excess. 1D peak suffixes b and f refer to the free and bound ligand species, respectively. **A** Chemical structure of biotin with the corresponding proton labeling *a* - *h'*. **B** Scheme for the assignment strategy of NOE cross peaks showing exchange peaks between the free signals (red) and the bound signals (blue), NOE between two bound signals, and NOE between free to relay bound signals. **C** & **D** CSPs between free and bound signals are shown in dotted green lines.

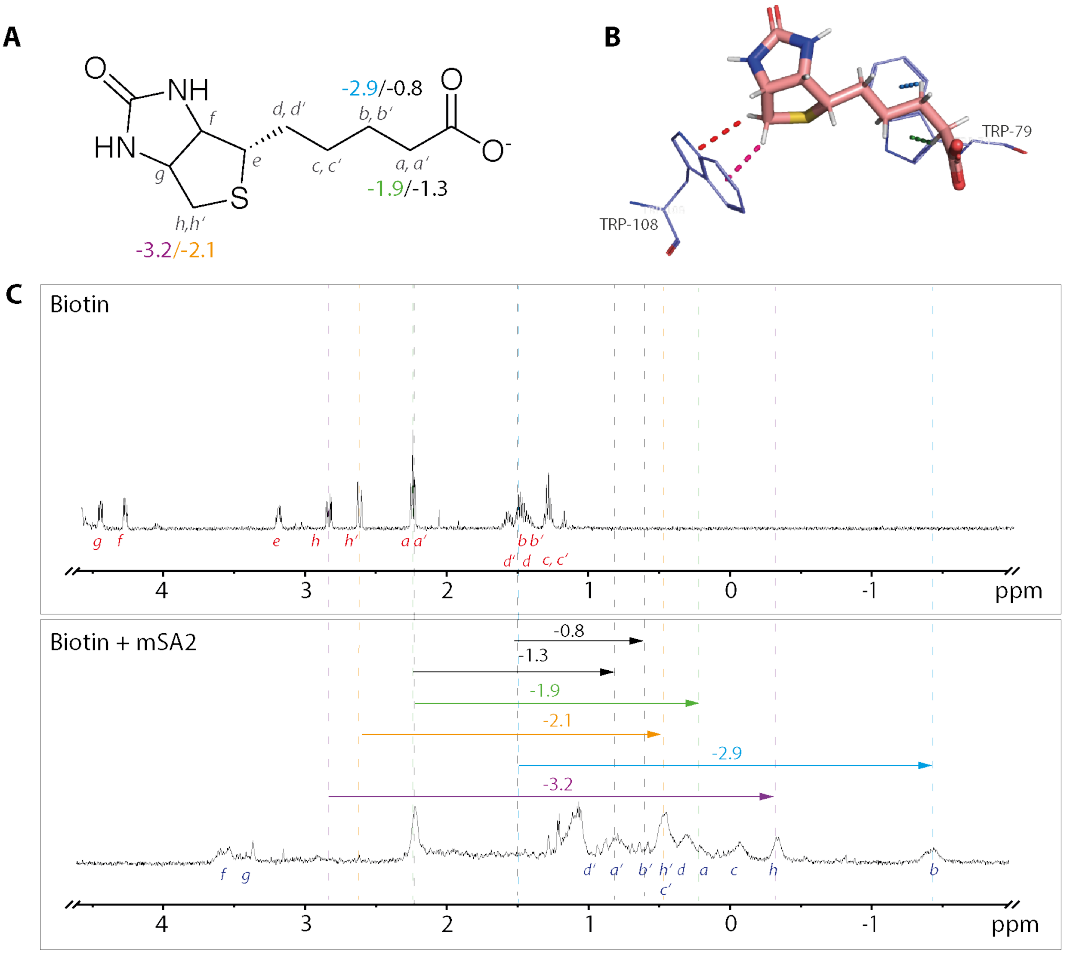

Figure 7: **A** Chemical structure of biotin. The colored numbers correspond to the CSP observed for the respective proton. **B** X-ray structure of the biotin:mSA2 complex, all amino acids other than Trp-79 and Trp-108 are hidden. **C** ^{1}H NMR spectra of free Biotin (top) and Biotin:mSA2 (bottom). CSPs of Biotin upon binding to mSA2 are indicated with arrows and correspond to the respective protons of biotin as labeled.

Figure 7 highlights 4 protons of biotin that undergo the strongest shielding effect. As the dashed lines in Figure 7C illustrate, CSPs of -3.2, -2.9, -2.1 and -1.9 ppm are amongst the four most prominent changes in chemical shifts that could be observed for biotin. These CSPs are assigned to the protons h, h', b and a, respectively. Furthermore, biotin´s protons a´, c, and d undergo CSPs of -1.28 ppm, -1.31 ppm, and -1.17 ppm, respectively. The CSPs of biotin´s remaining protons b´, c´, d´, e, f, and g have values between -0.49 ppm and -0.98 ppm.

Thereby, the large shielding effect of biotin's protons a, b, h, and h' can be explained by their proximity to aromatic π -systems. The x-ray structure of biotin in the mSA2 binding pocket in Figure 7B demonstrates the geometry and proximity of the listed protons to the aromatic side-chains of Trp-79 and Trp-108. There

the dashed lines visualize the protons´ proximities to the aromatic ring center and exemplify their position in regards to the aromatic ring, as they sit right on top of it.

To further analyze these large upfield CSPs, we correlated them with a formula that includes the underlying geometric dependencies, as it was shown in previous literature.[31] According to Pople's mathematical formalism, there is a dependence between the geometric parameters of the H-atom to the aromatic ring system. Hence, Pople's formalism allows to correlate the experimentally determined CSPs with the mathematically calculated value for the change in chemical shift $\Delta\sigma$.[41] In Pople's equation, the center of the aromatic ring induces a magnetic field according to Equation 6:

$$\Delta\sigma = 10^6 \times \frac{ne^2a^2}{4\pi mc^2} \times \frac{3cos^2\theta - 1}{r^3} \tag{6}$$

There, $\Delta\sigma$ represents the change in the chemical shift in ppm, n equals the number of πelectrons (6 for each of the aromatic rings of Trp), e equals the elementary charge in Franklin (1 e = 4.803× 10^{10} Franklin), a equals the radius of the aromatic rings in Trp (6-ring: 1.39×10^{-8} cm; 5-ring: 1.22×10^{-8} cm), m equals the electron mass in gram (9.109×10^{-28} g), c is the speed of light (2.998×10^{10} m×s^{-1}), θ represents the angle between the normal vector of the plane of one of the aromatic rings of Trp and the proton to ring center vector in rad, and r represents the distance from the proton to either the 5-ring or 6-ring center of the aromatic ring. Values (r, θ and a) were extracted from the crystal structure (PDB ID 4JNJ) and are listed in detail for each of biotin´s C-bound proton in Table 4 to Table 16.

Figure 8B illustrates a schematic depiction of the geometries involved in Equation 6 for the 5-ring in the indole moiety of tryptophan. The H illustrates the proton of the ligand, θ is the angle between the normal vector of the plane and the plane of the aromatic system and r are distances of the proton H to the ring center as mentioned above.

In the case of tryptophan's indole moiety, $\Delta\sigma$ was calculated with parameters for the indol's 5-ring and the parameters for the indol's 6-ring and the resulting values were simply added.

The resulting $\Delta\sigma$ values (CSP_{Pople}) are listed in Table 1 together with the experimentally determined values. The resulting correlation between these values is depicted in Figure 8C with an R^2 value of 0.86. Figure 8D shows the error bars of the experimentally determined CSPs relative to the calculated ones. Notably, CSP values of proton a and h are overestimated within the Pople formulism, all other protons show higher CSP values in the experimental determination relative to the calculated ones.

Table 1: Experimentally determined CS, CSP and calculated CSP for biotin´s carbon-bound H-atoms

Proton	δ Biotin $[ppm]$	CSP_{Pople} $[ppm]$	$\text{CSP}_{Experimental}$ $[ppm]$	δ Biotin:mSA2 $[ppm]$
a	2.13	-2.40	-1.88	0.25
a´	2.13	-0.84	-1.28	0.85
b	1.49	-2.12	-2.89	-1.41
b´	1.49	-0.70	-0.82	0.67
c	1.30	-0.86	-1.31	-0.01
c´	1.30	-0.48	-0.82	0.48
d	1.52	-0.51	-1.17	0.35
d´	1.61	-0.40	-0.49	1.12
e	3.25	-0.56	-0.98	2.27
f	4.35	-0.27	-0.70	3.65
g	4.50	-0.64	-0.94	3.56
h	2.89	-3.94	-3.17	-0.28
h´	2.66	-1.86	-2.15	0.51

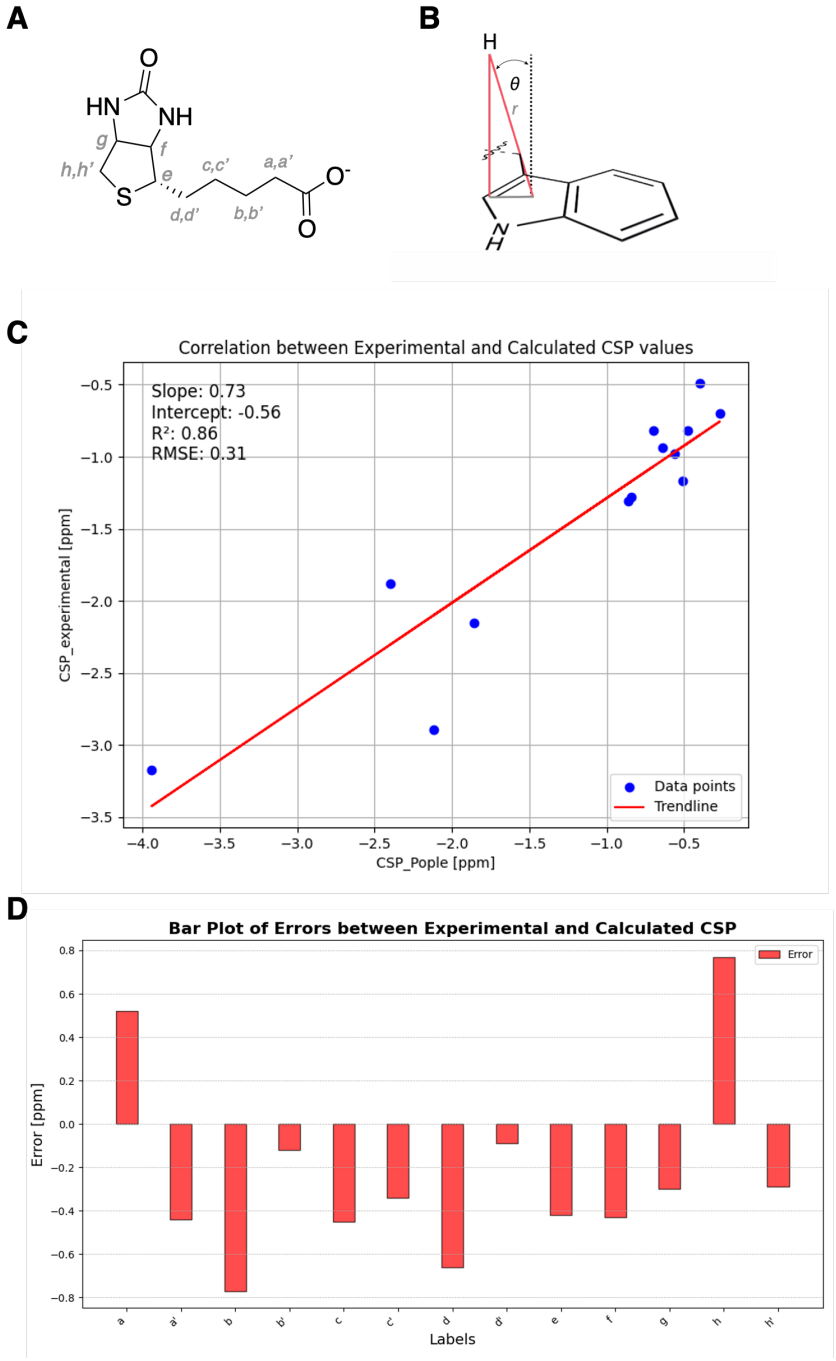

Figure 8: **A** Structure of biotin with the proton labels; **B** Geometric parameters extracted from X-ray crystal structure: proton-to-ring-center distance (r) and angle (θ) between the ring normal through the aromatic center; **C** Correlation between the experimentally determined CSPs and the calculated CSPs; **D** Error bars indicating the difference between the experimentally determined CSPs to the calculated CSPs.

4 Discussion

4.1 Large Chemical Shift Perturbations as Reporter for Favorable CH-π Interactions

The observed large upfield CSPs of most of biotin´s protons can be attributed to their formation of CH-π interactions with surrounding tryptophan residues of streptavidin. Thereby, the π -electrons of the indole moiety lead to shielding of the H-atom resulting in an upfield shift in the ^{1}H NMR spectrum.

Recently, the PI by NMR method has been shown to extract site-specific information on the formation of CH-π hydrogen bonds in protein-ligand interactions.[31] Within this thesis, we show that the site-specific information can be extracted on the ligand as ^{1}H chemical shifts, similar to as reported in previous literature.[33] Although ligand ^{1}H chemical shifts are rarely looked at, they have the advantage of straightforward peak assignment within a ^{1}H-^{1}H NOESY spectrum, as it allows the correlation of bound to unbound ligand proton signal. Peak assignment in the protein NMR spectrum, which can be quite crowded and thus time-consuming, can be avoided.[42]

The extraction of ligand ^{1}H CSPs allows the quantification of favorable CH-π interactions, when the ligand H-atoms serve as H-bond donors and the aromatic amino acid side-chains as H-bond acceptors. In the streptavidin-biotin system, the most prevalent protons of biotin forming CH-π interactions with streptavidin's tryptophan side-chains are a, b, h and h' as described in the Results section. In more detail, proton a has a distance to the Trp-79 6-ring of 3.14 Å and a θ of 12.34°. Proton b has a distance of 3.14 Å to the Trp-79 6-ring-center, and, it has a nearly optimal angle θ of 4.32° which explains the remarkable upfield shift from 1.49 ppm to -1.40 ppm ($\Delta\sigma$ = -2.89 ppm). Similar geometric parameters can be observed for proton h, which has a distance to the Trp-108 6-ring of only 2.55 Å and a θ of only 12.63° explaining the huge upfielding effect ($\Delta\sigma$ = -3.17 ppm) of this proton. Also proton h', which has a distance of 3.06 Å to the ring-center of the Trp-108 5-ring and a θ of only 4.00°.

Also other protons undergo a remarkable upfield shift of -1.28, -0.82, -1.31, -0.82, -1.17, -0.98, and -0.94 ppm (for a', b', c, c', d, e and g) respectively indicating the formation of CH-π interactions. The other protons d' and f which undergo a rather small upfield shift upon binding of -0.49 and -0.70 ppm still experience shielding from the aromatic ring. We hypothesize that they are too far away from the π -system to form actual H-bonds (r > 5 Å).

The geometric parameters of biotin's protons a-h' that were mentioned and used for the calculation of the $\Delta\sigma_{Pople}$ values are listed in Table 4 - Table 16.

Compared to classical H-bonds, CH-π interactions are relatively weak with a binding energy of 0.5 - 2.0 kcal/mol. It might be surprising that a weak interaction can contribute to such a high binding affinity, as seen in the biotin-streptavidin system. This becomes more understandable when considering previous systematic analyses of protein-ligand interactions. That analysis has shown that hydrophobic interactions, which include CH-π interactions as a simplified classification, are among the most significant contributors to binding affinity in protein-ligand interactions.[24] As we demonstrate here by utilizing CSPs, streptavidin forms multiple favorable CH-π interactions with biotin. In wild-type streptavidin, where the affinity for biotin is even higher, an additional tryptophan residue from an adjacent subunit acts as an additional hydrogen bond acceptor and resulting in the formation of additional CH-π interactions. It is important to note that many of these interactions involve the C-H protons of biotin's valeric acid tail.

The observation that an aliphatic tail can significantly contribute to favorable protein-ligand interactions could be particularly insightful for medicinal chemists designing drugs aimed at achieving high affinity and specificity in their target binding. Traditionally, medicinal chemists focus on classical H-bonds when designing drugs targeting a protein. And weak H-bonds like CH-π interactions are often not considered when designing high-affinity binders to a drug target. By understanding the relevance of these non-classical, weak H-bonds and their contribution to high binding affinity, medicinal chemists might utilize these strategies to enhance protein-ligand interactions.[43]

4.2 Correlation of Calculated vs Experimental CSPs

As the correlation of the calculated vs the experimentally determined CSP illustrates in Figure 8 the R^2 value of 85.80% seems rather low. In previously published work, experimentally determined CSPs were correlated to Pople-based CSPs and gave a R^2 value of 94.0%[44] or 92.0%[31]. Whereas in the described scenario, the CSPs were measured within a ^{1}H-^{13}C HSQC spectrum on a selectively labeled protein.[31,44] However, in the mSA2-biotin system, CSPs were measured on the ligand. When recording a ^{1}H NMR spectrum of a ligand, the ligand is placed in a solvent, which in our case was D_2O-based PBS buffer. In this environment, water molecules can form hydrogen bonds with biotin's protons, leading to a deshielding effect. If the compound could hypothetically be measured in a vacuum, these hydrogen bonds would not form. We hypothesize that as a result, the ^{1}H chemical shifts in the solvent are potentially more downfield in the NMR spectrum compared to the hypothetical ones in vacuum. When biotin is tightly bound within the streptavidin binding pocket, there is hardly any bulk solvent present. We hypothesize, that this might contribute to the slightly overestimated calculated CSPs of Pople's equation when

they are measured on the ligand.

However, as listed Table 1 the calculated CSPs of proton $b´$ and $d´$ only differ slightly from the experimentally determined ones. A look at the X-ray structure reveals the proximity of proton $b´$ and $d´$ to the -OH group of Thr-48, whereas the oxygen might form weak H-bonds to these protons, whereas the strength of these H-bonds might be similar to the ones formed in water.

Another reason for the rather low correlation of calculated and experimental CSPs might be, that in the Pople's equation, the aromatic πsystem is a benzene ring containing one perfect robust aromatic ring where no heteroatom is involved. Whereas, in our system, the aromatic πsystem is an indole ring. Unlike benzene, the indol ring is an aromatic heterocycle consisting of two aromatic rings: the six-membered benzene ring and the five-membered pyrrole ring. Whereas, the pyrrole ring contains nitrogen as a heteroatom in the aromatic ring.

5 Conclusion and Outlook

Within this thesis, we successfully expressed and purified perdeuterated mSA2 and extracted the CSPs of biotin upon binding to streptavidin. This allowed us to experimentally demonstrate the formation of CH-π interactions by NMR, which provides site-specific information at atomistic resolution and in solution at (nearly) physiological conditions. Furthermore, we could highlight the importance of these weak non-classical interactions for the high-affinity binding of biotin to streptavidin.

In the future, we aim to integrate the information of these ^{1}H chemical shifts with a QM/MM MD workflow to enable the construction of protein-ligand systems that accurately reflect the native state of the complex in solution.

6 Experimental Part

6.1 General Remarks

The 3C protease used for cleavage was produced in-house. For purification GST-beads (Pierce™ Glutathione Agarose) were used. Competent *E. coli* cells (*E. coli* BL21(DE3)) were purchased from a commercial supplier (New England Biolabs). Proteins were purified on an ÄKTA pure™ chromatography system.

NMR spectra were recorded on a Bruker 600 MHz, or Bruker 800 MHz, or a Bruker 600 MHz equipped with a cryoprobe. Samples with a volume >0.4 ml were measured in 5 mm NMR tubes, and samples with a volume <0.3 ml were measured in 5 mm Shigemi Advanced microtube set matched with D_2O.

6.2 Materials, Buffers, and Solutions

Buffers

Table 2: Buffers with corresponding concentrations used throughout the protein purification (TCEP = Tris-(2-carboxyethyl)-phosphin hydrochloride, CHAPS = 3-[(3-Cholamidopropyl)dimethylammonio]-1-propansulfonat, PMSF = Phenylmethansulfonylfluoride, PIC = Protease Inhibitor Cocktail)

Reagent	Concentration			
	LysisBuffer	**HisBufferA**	**HisBufferB**	**3C-CleavageBuffer**
TrisHCl pH 7.5	20 mM	50 mM	50 mM	20 mM
NaCl	300 mM	500 mM	500 mM	150 mM
TCEP	1 mM	-	-	0.5 mM
Imidazole	10 mM	-	500 mM	-
CHAPS	0.5% (v/v)	-	-	-
PMSF	1 μM	-	-	-
PIC	1:1000	-	-	-
Biotin	1 mM	-	-	-

Phosphate buffered saline (PBS) buffer was prepared by dissolving commercially available (Sigma Aldrich) tablets in H_2O yielding 10 mM phosphate buffer, 2.7 mM KCl, 137 mM NaCl pH 7.4.

M9 Minimal Media

Table 3: Minimal Media M9

Reagent	Concentration
KH_2PO_4	22 mM
Na_2HPO_4	42 mM
NaCl	12.5 mM
D-Glucose **or** D-Glucose-1,2,3,4,5,6,6-d7	22 mM **or** 14 mM
NH_4Cl **or** $^{15}NH_4Cl$	19 mM
$CaCl_2$	0.1 mM
$MgSO_4$	2 mM
Kanamycin	0.1 mM

Trace elements were prepared by dissolving 1 multivitamin tablet (Centrum) in 40 ml of ddH_2O and mixing it on the rotator for 30 min at 4°C. The solution was centrifuged for 5 min at 4500 rpm, RT, and filtered through a syringe filter. For 1 l of expression medium, 4 ml of the prepared trace elements solution were used.

Isotope-enriched media was prepared as described in Table 3 using the corresponding reagents.

LB Media

Highly-referenced nutrient-rich microbial growth powder medium was purchased from Sigma Aldrich, and 25 g of the powder were dissolved in 1 l of ddH_2O and autoclaved for 15 min at 121°C. Corresponding antibiotics were added to a final concentration of 25 - 50 μg/L.

6.3 Methods

6.3.1 Transformation of *E. coli*

Chemically competent cells (*E. coli* BL21 (DE3)) were thawed for 5 min on ice. An aliquot of 100 ng of the plasmid DNA carrying the gene of interest was added to 20 μl of competent cells, and gently mixed by pipetting. The mixture was placed on ice for 5 - 10 min. Heat shock transformation was done at 42°C for 30 - 45 s. Then, 250 μl of LB Media were added and the mixture was allowed to stand on ice for 5 min. Then the cells were put on 37°C for 45 - 60 min with shaking of 300 rpm. Afterwards, the cells were plated on agar plates containing the corresponding antibiotic and incubated at 37°C overnight.

6.3.2 Protein Expression

^{15}N labeled H6-MBP-3C-mSA2 was expressed by inoculating 100 ml of minimal media M9 (no isotope-enriched material) with one colony from the transformation plates. The culture was incubated overnight at 37°C with 200 rpm. The next day, the cells were pelleted (4500 rpm, 20°C, 10 min), the supernatant decanted and the pellet dissolved in 1 l of minimal media. The bacterial cells were grown at 37°C with 180 rpm. Biotin (1 mg in 1 ml PBS buffer pH 7.4) was added at an OD_{600} of 0.4 - 0.5. At an OD_{600} of 0.9, protein expression was induced with 0.4 mM IPTG and temperature was set to 18°C.

Perdeuterated H6-MBP-3C-mSA2 was expressed by transferring 20 μl of *E. coli* BL21 (DE3) cells harboring the plasmid from M9 medium (no labeled material, 100% H_2O to $M9^{50\%D2O}$ medium (no labeled material, 50% D_2O, 50% H_2O) to $M9^{100\%D2O}$ medium (no labeled material, 100% D_2O) and finally to 1 l of $M9^{100\%D2O}$ medium (1 g/l $^{15}NH_4Cl$, and D-glucose-1,2,3,4,5,6,6-d_7, 100% D_2O). At an OD_{600} of 0.4 - 0.5, biotin (1 mg) was added and protein expression was induced with 0.4 mM IPTG and temperature was set from 37°C to 18°C.

6.3.3 Protein Harvesting and Lysis

After 16 - 18 h, the cell suspension was centrifuged at 4500 rpm for 20 min at 20°C. The supernatant was decanted and the pellet was resuspended in 30 ml of lysis buffer and transferred to a sonication vial. Sonication was performed on ice for 4x3 min (1 s on, 1 s off; 50% amplitude). The lysed cells were centrifuged for 30 - 40 min at 18000 rpm at 4°C and the supernatant was filtered through a 0.45 μm filter.

6.3.4 Immobilized Ni-Affinity Chromatography I

The lysate supernatant was applied to 2x HisTrap 5 ml FF crude columns connected in series with a flow rate of 2.5 ml/min via the S1 syringe pump. Washing with HisBuffer A was performed until a stable UV baseline was observed. A subsequent washing step using HisBuffer A/ HisBuffer B (96:4 v/v, containing 20 mM Imidazole) was performed. The protein was eluted using His Buffer B (containing 500 mM Imidazole). Fractions containing the protein of interest were analyzed by SDS page and subsequently combined for the next step.

6.3.5 Buffer Exchange and Protease Cleavage

The combined fractions were applied to the HiPrep column on the Äkta via the S1 syringe pump and a flow rate of 5 ml/min. The buffer was exchanged for the 3C cleavage buffer. Then, in-house 3C protease was added to the protein solution, which was then rotated (10-20 rpm) at 4°C for 16 - 18 h to allow for cleave of the H6-MBP-tags.

6.3.6 Immobilized Ni-Affinity Chromatography II

The sample was applied to 2x HisTrap 5ml FF crude columns connected in series with a flow rate of 2.5 ml/min via the S1 syringe pump. Washing was conducted with the HisBufferA until a stable UV baseline was observed. The His-Tag-containing proteins were eluted with HisBufferB. Collected fractions were identified by SDS page. Fractions containing the cut protein of interest were combined and concentrated to a final volume of 1 ml by amicons (3 kDa = MW cutoff).

6.3.7 Size Exclusion Chromatography

The column was equilibrated with 120 ml PBS buffer pH 7.4 containing 1 mM TCEP. Then, the protein sample was applied to the HiLoad 16/600 Superdex 75 pg column via a 2 ml sample loop. Eluted fractions were analyzed by SDS page and fractions containing the purified protein were combined and concentrated to a final concentration of 100 - 200 μM. Figure 9 shows the chromatogramm of the deuterated mSA2 purification. The peak with an elution volume at 80 ml (marked in blue) corresponds to the purified mSA2 fractions. The purity of the fractions was confirmed via SDS page.

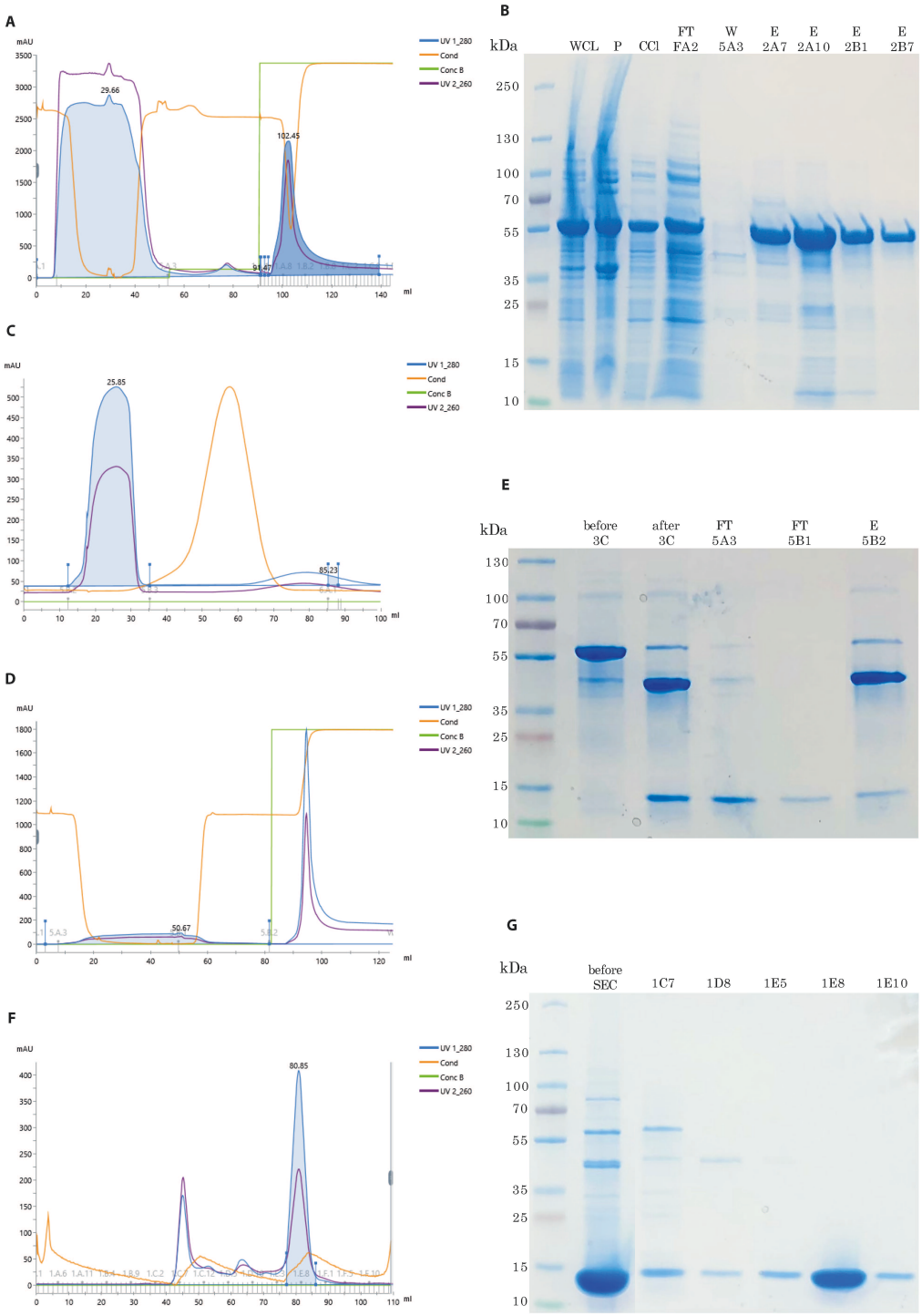

Figure 9: **A** Chromatogramm of first IMAC; **B** SDS-PAGE of **A** (WCL = whole cell lysate, P = pellet, CCL = clear cell lysate, FT = flowthrough, W = wash, E = eluate), band at 55 kDa corresponds to H6-MBP-mSA2; **C** Chromatogramm of buffer exchange chromatography (five runs were performed in total);

Figure 9: **D** Chromatogramm of second (reverse) IMAC after H6-MBP cleavage; **E** SDS-PAGE of **D** (before 3C = before addition of 3C protease, after 3C = after stirring 20 h, at 4°C upon addition of 3C protease, FT = flowthrough, E = eluate), bands at 12.7 kDa corresponds to mSA2, 40 kDa corresponds to H6-MBP, 55 kDa corresponds to non-cleaved H6-MBP-mSA2; **F** Chromatogramm of SEC (Superdex S75); **G** SDS-PAGE of **F**

Abbreviations

COSY Correlation Spectroscopy

CS chemical shift

CSP Chemical Shift Perturbations

E. coli Escherichia Coli

FID free induction decay

H-bonds hydrogen-bonds

HSQC Heteronuclear Single-Quantum Coherence

INEPT Insensitive Nuclear Enhancement by Polarization Transfer

IPTG Isopropyl β-D-1-thiogalactopyranoside

IUPAC International Union of Pure and Applied Chemistry

MBP Maltose Binding Protein

MW molecular weight

mSA monomeric streptavidin

mSA2 monomeric streptavidin 2

NMR Nuclear Magnetic Resonance

NOE nuclear Overhauser effect

NOESY Nuclear Overhauser Enhancement Spectroscopy

OD_{600} Optical Density at 600 nm

PBS Phosphate buffered saline

PDB Protein Database

SAR structure-activity relationship

SDS Sodiumdodecyl sulfate

STD Saturation Transfer Difference

TOCSY Total Correlation Spectroscopy

References

(1) Chaiet, L.; Wolf, F. J. *Archives of biochemistry and biophysics* **1964**, *106*, 1–5.

(2) Green, N. M. In *Advances in protein chemistry*; Elsevier: 1975; Vol. 29, pp 85–133.

(3) Hendrickson, W. A.; Pähler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P. *Proceedings of the National Academy of Sciences* **1989**, *86*, 2190–2194.

(4) Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J.; Salemme, F. *Science* **1989**, *243*, 85–88.

(5) Sano, T.; Cantor, C. R. *Proceedings of the National Academy of Sciences* **1995**, *92*, 3180–3184.

(6) Ozawa, M.; Ozawa, T.; Nishio, M.; Ueda, K. *Journal of Molecular Graphics and Modelling* **2017**, *75*, 117–124.

(7) Cavanagh, J., *Protein NMR spectroscopy: principles and practice*; Academic press: 1996.

(8) Lim, K. H.; Huang, H.; Pralle, A.; Park, S. *Biotechnology and bioengineering* **2013**, *110*, 57–67.

(9) DeMonte, D.; Drake, E. J.; Lim, K. H.; Gulick, A. M.; Park, S. *Proteins: Structure, Function, and Bioinformatics* **2013**, *81*, 1621–1633.

(10) Demonte, D.; Dundas, C. M.; Park, S. *Applied microbiology and biotechnology* **2014**, *98*, 6285–6295.

(11) Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. *International journal of molecular sciences* **2016**, *17*, 144.

(12) Matesanz, R.; Pera, B.; Díaz, J. F., *Thermodynamics as a Tool for the Optimization of Drug Binding, Thermodynamics-Interaction Studies-Solids, Liquids and Gases*; InTech: 2011.

(13) Dunitz, J. D. *Science* **1994**, *264*, 670–670.

(14) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. *Protein Science* **2006**, *15*, 459–467.

(15) Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P., et al. *Pure and applied chemistry* **2011**, *83*, 1637–1641.

(16) Jeffrey, G. A.; Jeffrey, G. A., *An introduction to hydrogen bonding*; Oxford university press New York: 1997; Vol. 12.

(17) Scheiner, S., *Hydrogen bonding: a theoretical perspective*; Oxford University Press, USA: 1997; Vol. 7.

(18) Pauling, L. *Ithaca, New York* **1960**.

(19) Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. *Journal of molecular biology* **2001**, *307*, 357–377.

(20) Derewenda, Z. S.; Lee, L.; Derewenda, U. *Journal of Molecular Biology* **2000**, *302*, 99–115.

(21) McKee, M. L.; Clemmer, D. E. *Journal of Physical Chemistry B* **2007**, *111*, 1189–1195.

(22) Cotton, F. A.; Falvello, L. R.; Schwotzer, W. *Inorganic Chemistry* **2006**, *25*, 4239–4243.

(23) Dunitz, J. D.; Taylor, R. *Chemistry–A European Journal* **1995**, *1*, 291–293.

(24) De Freitas, R. F.; Schapira, M. *Medchemcomm* **2017**, *8*, 1970–1981.

(25) Scheiner, S. *Molecules* **2016**, *21*, 1426.

(26) Burnley, B. T.; Afonine, P. V.; Adams, P. D.; Gros, P. *Elife* **2012**, *1*, e00311.

(27) Claridge, T. D., *High-resolution NMR techniques in organic chemistry*; Elsevier: 2016; Vol. 27.

(28) Wüthrich, K. *Europhysics News* **1986**, *17*, 11–13.

(29) Mayer, M.; Meyer, B. *Angewandte Chemie International Edition* **1999**, *38*, 1784–1788.

(30) Williamson, M. P. *Progress in nuclear magnetic resonance spectroscopy* **2013**, *73*, 1–16.

(31) Platzer, G.; Mayer, M.; Beier, A.; Brüschweiler, S.; Fuchs, J. E.; Engelhardt, H.; Geist, L.; Bader, G.; Schörghuber, J.; Lichtenecker, R., et al. *Angewandte Chemie* **2020**, *132*, 14971–14978.

(32) Platzer, G.; Mayer, M.; Böttcher, J.; Geist, L.; Fuchs, J. E.; McConnell, D. B.; Konrat, R. *bioRxiv* **2021**, 2021–11.

(33) Platzer, G.; Ptaszek, A. L.; Böttcher, J.; Fuchs, J. E.; Geist, L.; Braun, D.; McConnell, D. B.; Konrat, R.; Sánchez-Murcia, P. A.; Mayer, M. *ChemPhysChem* **2024**, *25*, e202300636.

(34) Keeler, J., *Understanding NMR spectroscopy*; John Wiley & Sons: 2010.

(35) Bodenhausen, G.; Ruben, D. J. *Chemical Physics Letters* **1980**, *69*, 185–189.

(36) Toscano, G.; Holzinger, J.; Nagl, B.; Kontaxis, G.; Kählig, H.; Konrat, R.; Lichtenecker, R. J. *Journal of Biomolecular NMR* **2024**, 1–9.

(37) Jacobsen, N. E., *NMR spectroscopy explained: simplified theory, applications and examples for organic chemistry and structural biology*; John Wiley & Sons: 2007.

(38) Ernst, R. R.; Bodenhausen, G.; Wokaun, A., *Principles of nuclear magnetic resonance in one and two dimensions*; Oxford university press: 1990.

(39) Anet, F. A.; Bourn, A. *Journal of the American Chemical Society* **1965**, *87*, 5250–5251.

(40) Shukla, D.; Trout, B. L. *The Journal of Physical Chemistry B* **2011**, *115*, 11831–11839.

(41) Pople, J. *The Journal of Chemical Physics* **1956**, *24*, 1111–1111.

(42) Cavanagh, J.; Fairbrother, W.; Palmer, A. G.; Rance, M.; Skelton, N. *Academic Press* **2007**.

(43) McConnell, D. B. *Journal of Medicinal Chemistry* **2021**, *64*, 16319–16327.

(44) Höfurthner, T.; Toscano, G.; Kontaxis, G.; Beier, A.; Mayer, M.; Geist, L.; McConnell, D. B.; Weinstabl, H.; Lichtenecker, R.; Konrat, R. *Journal of Biomolecular NMR* **2023**, 1–8.

Supplementary Material

^{1}H-^{1}H Homonuclear COSY NMR Spectrum

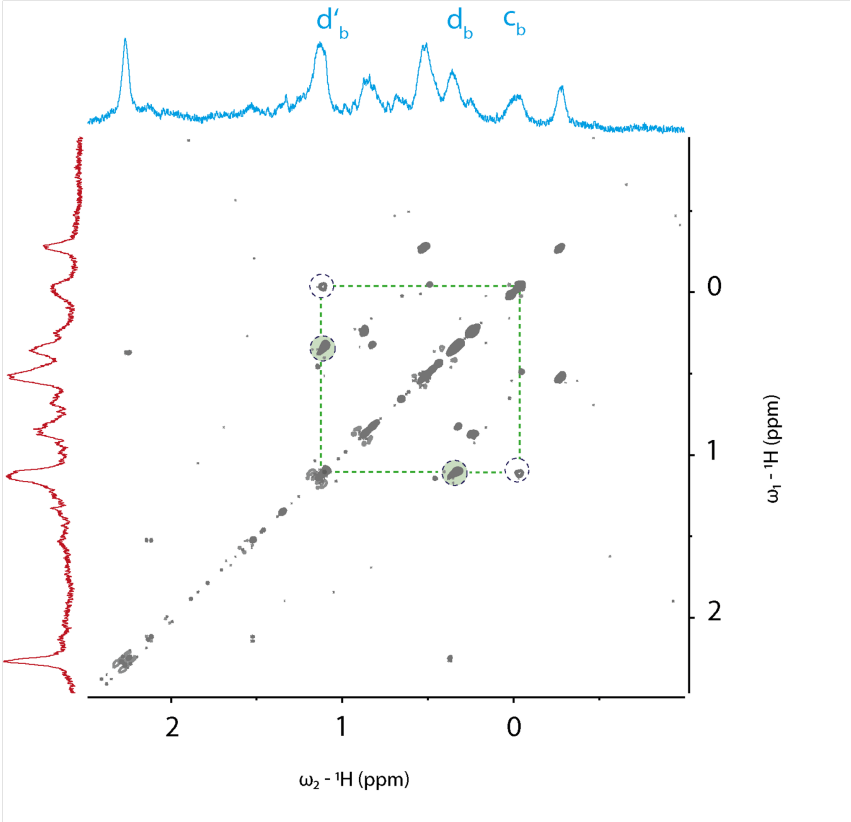

Figure 10: COSY spectrum of deuterated mSA2-biotin complex displaying the bound biotin protons c, d and d´. The $^{2}J_{HH}$ coupling between protons d´ and d can be seen. The corresponding crosspeaks are circled in black dashed lines and highlighted in green. The $^{3}J_{HH}$ between protons c and d´ can be seen as well. The corresponding crosspeaks are circled in black dashed lines.

Parameters for Pople´s Equation

Table 4 - Table 16 display the extracted values for θ (angle between the normal vector of the plane of the aromatic ring to the ring center) and r (distance from the proton to the 5- or 6-ring center of the aromatic ring) for biotin´s protons a, a´, b, b´, c, c´, d, d´, e, f, g, h and h´ that were used for the calculation of the CSP_{Pople} with Equation 6.

Table 4: Parameters Proton a

Trp-79, 5-ring	
θ	12.45°
θ	0.22 rad
$cos\theta^2$	0.95
r	$1.64{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	2.20 ppm
Trp-79, 6-ring	
θ	47.15°
θ	0.82 rad
$cos\theta^2$	0.46
r	$3.66{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.21 ppm
$\Delta\sigma$ total	2.41 ppm

Table 5: Parameters Proton a´

Trp-79, 5-ring	
θ	24.96°
θ	0.44 rad
$cos\theta^2$	0.82
r	$3.77{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.53 ppm
Trp-79, 6-ring	
θ	36.94°
θ	0.64 rad
$cos\theta^2$	0.64
r	$4.28{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.31 ppm
$\Delta\sigma$ total	0.84 ppm

Table 6: Parameters Proton b

Trp-79, 5-ring	
θ	32.94°
θ	0.58 rad
$cos\theta^2$	0.70
r	$3.70{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.42 ppm
Trp-79, 6-ring	
θ	4.32°
θ	0.08 rad
$cos\theta^2$	0.99
r	$3.14{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	1.69 ppm
$\Delta\sigma$ total	2.11 ppm

Table 7: Parameters Proton b´

Trp-79, 5-ring	
θ	16.20°
θ	0.28 rad
$cos\theta^2$	0.92
r	$5.00{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.27 ppm
Trp-79, 6-ring	
θ	9.30°
θ	0.16 rad
$cos\theta^2$	0.97
r	$4.90{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.43 ppm
$\Delta\sigma$ total	0.70 ppm

Table 8: Parameters Proton c

Trp-79, 5-ring	
θ	40.70°
θ	0.71 rad
$cos\theta^2$	0.57
r	$3.14{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.45 ppm
Trp-79, 6-ring	
θ	43.80°
θ	0.76 rad
$cos\theta^2$	0.52
r	$4.90{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.41 ppm
$\Delta\sigma$ total	0.86 ppm

Table 9: Parameters Proton c´

Trp-79, 5-ring	
θ	28.13°
θ	0.49 rad
$cos\theta^2$	0.78
r	$4.58{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.27 ppm
Trp-79, 6-ring	
θ	35.76°
θ	0.62 rad
$cos\theta^2$	0.66
r	$4.97{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.21 ppm
$\Delta\sigma$ total	0.48 ppm

Table 10: Parameters Proton d

Trp-79, 5-ring	
θ	35.80°
θ	0.63 rad
$cos\theta^2$	0.66
r	$5.90{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.26 ppm
Trp-79, 6-ring	
θ	25.60°
θ	0.45 rad
$cos\theta^2$	0.81
r	$5.30{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.25 ppm
$\Delta\sigma$ total	0.51 ppm

Table 11: Parameters Proton d´

Trp-79, 5-ring	
θ	51.20°
θ	0.89 rad
$cos\theta^2$	0.39
r	$4.9{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.03 ppm
Trp-79, 6-ring	
θ	38.30°
θ	0.67 rad
$cos\theta^2$	0.62
r	$3.9{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.37 ppm
$\Delta\sigma$ total	0.40 ppm

Table 12: Parameters Proton e

Trp-79, 5-ring			Trp-108, 5-ring
θ	44.00°		37.10°
θ	0.77 rad		0.65 rad
$cos\theta^2$	0.64		0.52
r	$6.10{\cdot}10^{-8}$ cm		$5.95{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.05 ppm		0.08 ppm
Trp-79, 6-ring			Trp-108, 6-ring
θ	44.10		18.00°
θ	0.77 rad		0.31 rad
$cos\theta^2$	0.52		0.90
r	$6.08{\cdot}10^{-8}$ cm		$4.99{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.06 ppm		0.36 ppm
$\Delta\sigma$ total	0.56 ppm		

Table 13: Parameters Proton f

Trp-108, 5-ring	
θ	39.36°
θ	0.69 rad
$cos\theta^2$	0.0.59
r	$5.94{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.07 ppm
Trp-108, 6-ring	
θ	31.00°
θ	0.54 rad
$cos\theta^2$	0.73
r	$5.40{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.20 ppm
$\Delta\sigma$ total	0.27 ppm

Table 14: Parameters Proton g

Trp-108, 5-ring	
θ	42.40°
θ	0.74 rad
$cos\theta^2$	0.54
r	$3.70{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.24 ppm
Trp-108, 6-ring	
θ	41.30°
θ	0.72 rad
$cos\theta^2$	0.56
r	$3.60{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.39 ppm
$\Delta\sigma$ total	0.64 ppm

Table 15: Parameters Proton h

Trp-92, 5-ring			Trp-108, 5-ring
θ	44.28°		32.70°
θ	0.77 rad		0.57 rad
$cos\theta^2$	0.51		0.71
r	$5.70{\cdot}10^{-8}$ cm		$2.96{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.05 ppm		0.84 ppm
Trp-92, 6-ring			Trp-108, 6-ring
θ	42.99		12.63°
θ	0.75 rad		0.22 rad
$cos\theta^2$	0.53		0.95
r	$5.55{\cdot}10^{-8}$ cm		$2.55{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.09 ppm		2.95 ppm
$\Delta\sigma$ total	3.94 ppm		

Table 16: Parameters Proton h´

Trp-92, 5-ring			Trp-108, 5-ring
θ	53.70°		4.00°
θ	0.94 rad		0.07 rad
$cos\theta^2$	0.35		0.99
r	$4.15{\cdot}10^{-8}$ cm		$3.06{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.01 ppm		1.34 ppm
Trp-92, 6-ring			Trp-108, 6-ring
θ	53.30		35.93°
θ	0.93 rad		0.63 rad
$cos\theta^2$	0.36		0.65
r	$4.08{\cdot}10^{-8}$ cm		$3.77{\cdot}10^{-8}$ cm
$\Delta\sigma$ contribution	0.03 ppm		0.47 ppm
$\Delta\sigma$ total	1.86 ppm		