

Technische Universität Dortmund
Dissertation

“Structure and dynamics studies of the
enzymes hCAII and GlpG via NMR
spectroscopy”

01.06.19 - 15.11.23

AG Linser, Department for Chemistry and Chemical Biology,
Physical Chemistry II, TU Dortmund University

by Julia Kotschy
born in Mettmann

First reviewer: Prof. Dr. Rasmus Linser
Second reviewer: Prof. Dr. Daniel Rauh

This dissertation was published by the Faculty of Chemistry and Chemical Biology at TU Dortmund University in Dortmund in February 2024 after the oral examination had been passed successfully in November 2023.

Acknowledgements

First, special thanks go to Prof. Dr. Rasmus Linser for giving me the fantastic opportunity to conduct my research in his group. Thanks for helping me navigate the various challenges of my projects. Stepping into a highly interdisciplinary area and learning new techniques with complex theoretical backgrounds takes patience and support, which you always offered me during my PhD. I very much enjoyed working independently and want to thank you for allowing me to do so.

Moreover, I want to thank Prof. Dr. Daniel Rauh for evaluating my dissertation and disputation as part of the committee. I am grateful to Prof. Dr. Wolf Hiller for taking the lead of the disputation committee. Thanks to Dr. Malte Gersch and Prof. Dr. Clemens Glaubitz, who gave valuable advice during three TAC meetings as part of the IMPRS-LM.

I am grateful for the financial support by the IMPRS-LM and Fonds der Chemischen Industrie. Furthermore, all my research projects benefited from collaborations with brilliant scientists. Thanks to Suresh Vasa for supporting me in learning how to acquire and process NMR spectra and collaborating with me on both the hCAII and GlpG projects. Prof. Dr. Kvido Strisovsky and Kathrin Bach from Prague were part of a fruitful and productive collaboration on GlpG and kindly hosted me shortly in Prague. Thanks to Dr. Alexander Klein for diving deeply into MethylFLYA assignments for hCAII and for sharing fun moments at the office. Benedikt Söldner kindly shared his expertise in MD simulations on our collaboration on the hCAII project. Dr. Himanshu Singh supported the hCAII project especially in the initial time of my PhD. Thanks to Dr. Raphael Gasper for aiding with the X-ray data acquisition and processing. I am grateful for the NMR spectrometer measurement time kindly offered by Prof. Dr. Wolf Hiller as well as Forschungszentrum Jülich. In addition, I would also like to thank Sara, Laura, Katja, Hanna, Suchandra, Christine, Vanessa, and Maria, who accompanied me during my PhD as members of the AG Linser. I genuinely enjoyed working and spending time together. Thanks to Lydia, Annette, and Mathias for proof-reading my thesis. Thanks to Julia and Damian for sharing competent cells and productive discussions.

Finally, I would like to thank my partner Mathias, my parents and sister, other family members and friends for encouraging and motivating me throughout this joyful despite challenging time.

Abstract

Solution and solid-state NMR are valuable tools to characterize the chemical environments and dynamics of molecules like proteins with atomic resolution. Here, the effects of protein-ligand interactions on structural features and flexibility, strategies for NMR sample optimization, as well as different tools for peak assignments were elucidated.

When looking at snapshots of protein-ligand structures, especially high-affinity complexes, it seems likely that such structures are rigid and have exceedingly low flexibility. Intriguingly, dynamics of an isotopically unlabelled ligand in protein-bound state were detected for a high-affinity complex via dedicated solution and solid-state NMR experiments. The enzyme hCAII is a well-studied drug target involved in pH regulation in human cells, among other tasks. The 30 μM affinity ligand SBR has a rather extended structure with less spatial restrictions for an increasing distance to the Zn^{2+} ion in the catalytic center. Together with slow-exchange behavior observed during ligand titrations and X-ray crystallography studies, the approach allowed to put specific protein-ligand interactions in context with ligand flexibility, which might be of interest for medicinal chemistry and lead optimization.

Furthermore, time-shared 3D NOESY/RFDR experiments were utilized both in solution and the solid state to assign methyl groups of hCAII based on dipolar couplings between amide and/or methyl groups. Due to line broadening and a low number of cross peaks in the solid-state spectra, the methyl assignment success was lower than the more straight-forward assignments in solution. In total, 52 out of 86 detected methyl peaks were assigned.

Intra-membrane proteases of the rhomboid-like family are involved in processes associated with devastating conditions like cancer and neurodegeneration^[1]. Although GlpG from *E. coli* is a well-studied representative of the family, it is not yet fully understood how GlpG recognizes and binds its substrates. Among others, trans-membrane helix 5 and loop 5 have been suggested to orchestrate substrate binding. After optimizing several expression and purification conditions, selected methyl groups were mutated and assigned to study chemical shift perturbations and microsecond to millisecond motion via solution NMR spectroscopy, comparing GlpG in apo and inhibitor-bound state. Our data indicate that upon binding of inhibitor STS1775, the conformational ensemble changes despite of up to 18 Å distance from the active site. Consequently, the occupancy of the active site modulates conformations close to but also far away from the substrate binding site.

Zusammenfassung

Lösungs- und Festkörper-NMR-Spektroskopie eignen sich gut, um Moleküle wie Proteine mit atomarer Auflösung zu charakterisieren. Hier wurden insbesondere strukturelle und dynamische Änderungen durch die Interaktion mit Inhibitoren, Strategien zur NMR-Probenoptimierung und verschiedene Methoden zur Zuordnung von NMR-Peaks eruiert.

Protein-Ligand-Komplexe mit hoher Bindungsaffinität werden oftmals als rigide Strukturen wahrgenommen. Hier wurden Lösungs- und Festkörper-NMR-Experimente entwickelt, um die Flexibilität von SBR in hCAII-gebundenem Zustand zu messen. Das Enzym hCAII reguliert osmotische Prozesse in humanen Zellen und ist deshalb ein Zielprotein für Arzneimittel. Der Ligand SBR interagiert durch seine Sulfonamid-Gruppe mit dem Zn^{2+} -Ion im katalytischen Zentrum von hCAII. Die entwickelte Methode zeigte, dass sich der Ligand trotz hoher Affinität in der Bindungstasche bewegt. Dies könnte von Interesse für die Wirkstoffentwicklung sein.

Des Weiteren wurden sogenannte „time-shared“ 3D Lösungs- und Festkörper-NMR-Experimente für die Zuordnung von Methyl-Gruppen durch die Korrelation von Amid- und Methyl-Protonen in räumlicher Nähe zueinander genutzt. Während eine starke Linienverbreiterung der Peaks und eine geringe Anzahl an Kreuzpeaks in den Festkörper-NMR-Spektren problematisch bei der Zuordnung der Methyl-Gruppen waren, gestaltete sich die Zuordnung anhand Lösungs-NMR-Spektren als deutlich erfolgreicher und unkomplizierter. Insgesamt konnten 52 von 86 Methyl-Peaks von hCAII zugeordnet werden.

Intramembranproteasen der Rhomboid-ähnlichen Familie sind an Prozessen beteiligt, die mit Krebs- und neurodegenerativen Erkrankungen^[1] assoziiert sind. Trotz intensiver Forschung ist für die Intramembranprotease GlpG aus *E. coli* noch nicht vollständig bekannt wie Substrate erkannt und gebunden werden, wobei u. A. Transmembranhelix 5 und Loop 5 daran beteiligt sein sollen. Nachdem die Expression und Aufreinigung von GlpG optimiert worden waren, wurden die NMR-Peaks der Methyl-Gruppen durch Punktmutationen zugeordnet. Diese wurden dazu genutzt lokale Änderungen der chemischen Umgebung und der Flexibilität der Seitenketten im Mikro- bis Millisekunden-Bereich für GlpG ohne und im Komplex mit Inhibitoren mittels Lösungs-NMR-Spektroskopie zu verfolgen. Die Ergebnisse deuten darauf hin, dass die Interaktion mit dem Inhibitor STS1775 Konformationsänderungen bzw. Änderungen in der Dynamik von Methyl-Gruppen induziert, die bis zu 18 Å weit vom aktiven Zentrum entfernt sind.

Abbreviations

Table 1: Abbreviations used subsequently.

Abbreviations	Full names
aa	Amino acids
BMRD	Bloch-McConnell relaxation dispersion
BSA	Bovine serum albumin
CEST	Chemical-exchange saturation transfer
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHAPSO	3-[(3-Cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate
CMC	Critical micelle concentration
CP	Cross polarization
CPMG	Carr-Purcell-Meiboom-Gill relaxation dispersion
CSA	Chemical shift anisotropy
CSPs	Chemical shift perturbations
DDM	<i>n</i> -Dodecyl- β -D-maltopyranoside
DM	<i>n</i> -Decyl- β -D-maltopyranoside
d ₆ -DMSO	Deuterated dimethyl sulfoxide
d ₃₈ -FC-12	Deuterated <i>n</i> -dodecylphosphocholine
d ₅ -TRIS	Deuterated tris(hydroxymethyl)aminomethane buffer
ddH ₂ O	Double distilled water
df	Degrees of freedom
DMPC	Dimyristoyl phosphatidylcholine
DNA	Deoxyribonucleic acid
DSS	3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt
dNTPs	Deoxynucleotide triphosphates
EDTA	Ethylendiaminetetraacetic acid
EM	Electron microscopy
EPR	Electron paramagnetic resonance spectroscopy
<i>E. coli</i>	<i>Escherichia coli</i>
EXSY	Exchange spectroscopy
FID	Free induction decay
FRET	Fluorescence-resonance energy transfer
FT	Fourier transform
GABA _A	γ -Aminobutyric acid receptors
GST-tag	Glutathione S-transferase-tag
hCAII	Human carbonic anhydrase II
HMQC	Heteronuclear multiple quantum coherence
HSQC	Heteronuclear single quantum coherence
ILV-methyl labelling	Isoleucine-, leucine-, and valine-methyl labelling
INEPT	Insensitive nuclei enhancement by polarization transfer
iRhom	Inactive rhomboid protein
kDa	Kilodaltons
LacYTM2	TM2 helix of the lactose permease LacY
LDAO	<i>n</i> -Dodecyl- <i>N,N</i> -dimethylamine- <i>N</i> -oxide
LMNG	2,2-Didecylpropane-1,3-bis- β -maltopyranoside
MAS	Magic angle spinning
mAU	Milli-absorption units
MD simulation	Molecular dynamics simulation
MW	Molecular weight
MWCO	Molecular weight cut-off
n.d.	No data
NERRD	Near-rotary resonance relaxation dispersion
NG	<i>n</i> -Nonyl- β -D-maltopyranoside
NMR spectroscopy	Nuclear magnetic resonance spectroscopy
NOE	Nuclear Overhauser effect

NOESY	Nuclear Overhauser effect spectroscopy
OD ₆₀₀	Optical density at 600 nm
OG	<i>n</i> -Octyl- β -D-maltopyranoside
P1-5	Position 1-5
PCR	Polymerase chain reaction
PDB	Protein database
pg	Prep grade
PMSF	Phenylmethylsulfonyl flouride
ppm	Parts per million
PARL	Presenilin-associated rhomboid-like protein
RAP	Reduced adjoining protonation
RFDR	Radio-frequency-driven recoupling
RFU	Relative fluorescence units
RHBDL	Rhomboid-related protein
rpm	Rotations per minute
SAIL	Stereo-array isotope labelling
Sc	Sidechain
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SLS	Swiss Light Source
Sofast-HMQC	Band-selective optimized flip angle short transient-HMQC
TM1-6	Transmembrane helices 1-6
TOCSY	Total correlation spectroscopy
TROSY	Transverse relaxation-optimized spectroscopy

Table of contents

Acknowledgements.....	2
Abstract.....	3
Zusammenfassung	4
Abbreviations.....	5
Table of contents	7
1. General introduction.....	10
1.1 Characterizing proteins and their interaction partners via NMR spectroscopy.....	10
1.1.1 The main concepts involved in NMR spectroscopy	10
1.1.2 The central principles of solid-state NMR	15
1.1.3 The contribution of NMR spectroscopy to the field of structural biology	17
1.1.4 The characterization of protein-ligand complexes.....	20
1.1.5 NMR spectroscopy as a reporter on dynamics.....	22
1.1.6 Isotopic labelling and selective methyl labelling	26
2. The Zn ²⁺ -complexing drug target hCAII.....	30
2.1 Introduction	30
2.1.1 Known characteristics of hCAII	30
2.1.2 The binding site of hCAII and binding inhibitors.....	33
2.2 Results and discussion	36
2.2.1 Screening high-occupancy conditions for hCAII-SBR crystals via X-ray crystallography	38
2.2.2 Chemical shift perturbations for hCAII induced by ligands in solution and solids	45
2.2.3 Ligand motion in hCAII-bound state detected in solution and solids.....	52
2.2.4 hCAII backbone dynamics in the hCAII-SBR complex in solution	59
2.2.5 Methyl chemical shift assignments for the hCAII sidechains	61
2.3 Conclusion and outlook	75
3. The intramembrane protease GlpG.....	77
3.1 Introduction	77
3.1.1 Membrane proteins and their isolation from the native environment	77
3.1.2 The family of rhomboid-like proteases.....	80
3.1.3 The structure and proteolytic activity of GlpG	81
3.1.4 Substrates and inhibitors interacting with GlpG	84
3.2 Results and discussion	87

3.2.1 Optimizing the preparation of full-length GlpG NMR samples	88
3.2.2 Amide and ILV-methyl sidechain peak assignments for full-length GlpG	98
3.2.3 Chemical shift perturbations for full-length GlpG induced by substrate-like inhibitors	110
3.2.4 Methyl sidechain dynamics of full-length GlpG detected via solution NMR.....	120
3.2.5 Crystallizing GlpG core for solid-state NMR studies.....	125
3.3 Conclusion and outlook	128
4. Materials and methods.....	131
4.1 Instrumentation	131
4.2 Biological and chemical substances.....	132
4.3 Consumables.....	135
4.4 DNA and proteins.....	136
4.5 Cell lines	139
4.6 Buffer and media composition and preparation	140
4.7 Software.....	141
4.8 Molecular biology and cell biology methods.....	141
4.8.1 Estimation of DNA concentrations	141
4.8.2 Plasmid DNA amplification and isolation.....	141
4.8.3 DNA sequencing.....	141
4.8.4 Single Site-directed mutagenesis.....	142
4.8.5 Multi site-directed mutagenesis.....	142
4.8.6 Preparation of chemically competent cells	143
4.8.7 Transformation of competent 5-alpha and XL-10 cells	144
4.8.8 Transformation of competent BL21-Condon Plus(DE3)-RIL cells	144
4.8.9 Transformation of competent Rosetta 2 cells	144
4.8.10 Transformation of competent Lemo21(DE3) cells	144
4.8.11 Transformation of competent LOBSTR(DE3) cells	145
4.9 Protein expression	145
4.9.1 Expression of hCAII or GlpG in LB and M9 or M9* medium.....	145
4.9.2 Adapting and growing cells in D ₂ O	146
4.9.3 ILV-methyl labelling	147
4.10 Biochemistry methods.....	148
4.10.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)	148
4.10.2 Cell lysis via high pressure homogenization	149

4.10.3 hCAII purification	149
4.10.4 Full-length GlpG and GlpG core purification	150
4.10.5 Estimation of protein concentrations.....	151
4.10.5.1 Estimation via BioDrop Duo.....	151
4.10.5.2 Estimation via Bradford assay.....	152
4.10.5.3 Estimation via Pierce 660 nm assay.....	152
4.10.5.4 Estimation via Äkta purification system	152
4.11 Activity assays	153
4.11.1 Promo Cell activity assay for hCAII	153
4.11.2 Activity assay for GlpG ^[169]	153
4.12 Crystallization and X-ray diffraction of protein crystals	154
4.12.1 Preparation of hCAII-SBR crystals	154
4.12.2 X-ray diffraction and data processing for hCAII-SBR crystals	155
4.12.3 Crystallization of GlpG core	155
4.13 NMR spectroscopy	156
4.13.1 Solution NMR sample preparation	156
4.13.2 Solid-state NMR sample preparation	156
5. Appendix	158
5.1 Precursors, ligands, additional titrations, and water contacts for hCAII.....	158
5.2 SEC column calibration	161
5.3 X-ray crystallography data	161
5.4 Expression, activity, temperature sensitivity, and assignments of GlpG	162
5.5 Raw data of relaxation dispersion plots for SBR ligand dynamics.....	167
5.6 CPMG plots for hCAII and GlpG in apo and inhibitor-bound state.....	172
5.7 Chemical shifts for hCAII and its ligand SBR as well as GlpG	188
5.8 FLYA assignments.....	199
5.9 Solution NMR pulse sequences	205
5.10 Solid-state NMR pulse sequences.....	211
References	221
List of publications	229
Conferences and workshops.....	230
Eidesstattliche Versicherung (Affidavit)	231

1. General introduction

1.1 Characterizing proteins and their interaction partners via NMR spectroscopy

Protein structures, dynamics, and function are closely related. Transitions between protein conformations are often connected to protein function, for example, by catalyzing a chemical reaction or regulating the access of binding partners to interaction surfaces. These transitions are associated with structural rearrangements and a time scale. The importance of protein research is highlighted by its impact on the development of new drugs to treat devastating diseases. Knowledge on the structures and the binding kinetics of protein-ligand complexes is essential for rational drug design. However, the dynamics of ligands in protein-bound state are usually not yet studied in detail, although such information could specify which ligand structure is required for a high-affinity interaction.

The next few chapters draw connections between protein complex structures and dynamics to explore the means to detect them with Nuclear Magnetic Resonance (NMR) spectroscopy, and to discuss the protein-inhibitor complexes of the enzyme human carbonic anhydrase II (hCAII) and the membrane protein GlpG.

1.1.1 The main concepts involved in NMR spectroscopy

This chapter aims to introduce and summarize the basics of NMR spectroscopy as described in several excellent text books^[2] to which one can refer to for more detailed explanations.

In an NMR experiment, molecules are subjected to an external magnetic field. As the atomic nuclei of the molecules possess a spin, they can be imagined as bar magnets themselves and therefore interact with other magnetic fields. The nuclear spin quantum number I has $2I + 1$ possible states and determines the nuclear spin angular momentum, which is an intrinsic feature of most nuclei. In a static, external magnetic field B_0 , spins precess around the axis of B_0 at their individual Larmor frequency ν_0 (Hz) or ω_0 (rad s⁻¹).

$$\nu_0 = -\frac{\gamma B_0}{2\pi} \quad (1)$$

Most nuclei in biomolecules have a spin quantum number of $I = 1/2$ with a spherical charge distribution and are suited well for NMR studies. For example, ¹H nuclei have a half-integer spin quantum number, one of the highest gyromagnetic ratios, and are prevalent in biomolecules compared to e.g. its isotope deuterium (Table 2). For other atoms like carbon

or nitrogen, molecules can be artificially enriched with NMR-active isotopes, preferably with the spin quantum number $I = 1/2$, via isotopic labelling (chapter 1.1.6).

Table 2: Characteristics of nuclei occurring in biomolecules which are relevant for NMR spectroscopy^[2a].

Nucleus	Nuclear spin quantum number I	Gyromagnetic ratio $\gamma / (\text{Ts})^{-1}$	Natural abundance / %
¹ H	1/2	$2.675 \cdot 10^8$	99.99
² H/D	1	$4.107 \cdot 10^7$	0.01
¹² C	0	-	98.93
¹³ C	1/2	$6.728 \cdot 10^7$	1.07
¹⁴ N	1	$1.934 \cdot 10^7$	99.63
¹⁵ N	1/2	$-2.713 \cdot 10^7$	0.37
¹⁷ O	5/2	$-3.628 \cdot 10^7$	0.04
¹⁹ F	1/2	$2.518 \cdot 10^8$	100.00
³¹ P	1/2	$1.084 \cdot 10^8$	100.00

Like other spectroscopic methods, NMR measurements rely on the distinct electronic energy levels of different spins. In a static B_0 field, nuclear spins with $I = 1/2$ can occupy two energy levels, $m = +1/2$ called α -state, which is aligned in parallel to B_0 and represents the energetically more favorable state, and $m = -1/2$ called the β -state (Figure 1, panel A), which is oriented anti-parallel to B_0 . The population excess of α - over β -spins is only 1 : 10,000 at room temperature in a 10 Tesla magnetic field, thus the sensitivity of NMR spectroscopy is inherently low^[3]. Several hardware-related specifications are necessary to exploit the already low sensitivity. For example, cryogenically cooled probes (cryoprobes) can augment the sensitivity of NMR spectra by up to four times, reducing the thermal noise during signal detection^[4]. Moreover, spectrometers with higher magnetic fields improve the experimental sensitivity with $B_0^{3/2}$.

The phenomenon of quantized energy states within a static, external magnetic field is named Zeeman splitting. The higher the gyromagnetic ratio γ of a nucleus, the larger is the energy difference between α - and β -spins. This energy difference ΔE corresponds to the Larmor frequency of the specific nuclear spin and can be calculated by equation (2) with the gyromagnetic ratio γ , the chemical shielding σ , the reduced Planck's constant \hbar , as well as the external magnetic field B_0 ^[3]. Transitions between these energy levels can be induced with electromagnetic pulses with the frequency ν_{rf} , which accounts for the energy difference between the two spin states.

$$\Delta E = \gamma \hbar (1 - \sigma) B_0 = \hbar 2\pi \nu_{\text{rf}} \quad (2)$$

A relative measure for this situation in thermal equilibrium is the Boltzmann distribution shown below with k_B being the Boltzmann constant and T representing temperature. The number of nuclei in state m is defined as N_m out of the entire population of nuclei N . With increasing B_0 and gyromagnetic ratio, the energy difference as well as the population difference between the states will be increased.

$$\frac{N_m}{N} = \frac{1}{2I + 1} \left(1 + \frac{m\hbar\gamma B_0}{k_B T} \right) \quad (3)$$

Moreover, orientation-dependent and -independent contributions relative to B_0 , called isotropic or anisotropic components, influence how spins behave during NMR experiments. For example, the chemical shift has an isotropic and an anisotropic component of up to 1 kHz, while homo- and heteronuclear dipolar couplings of up to 20 kHz and chemical shift anisotropy (CSA) solely represent anisotropic contributions^[3]. The strength of the local field B_1 generated by dipole-dipole interactions depends on the distance between the spins, their gyromagnetic ratios, and the relative orientations of the spins. CSA describes the dependence of the induced field B_1 on the orientation of the spin relative to the external field B_0 . The orientation, the molecular tumbling speed, and the gyromagnetic ratio of this spin influence how strongly CSA can contribute to e.g. relaxation. In proteins, ^1H - ^1H dipolar couplings are very strong with couplings of around 100 kHz, while ^1H - ^{13}C and ^1H - ^{15}N dipolar couplings are weaker by five- to ten-fold (Figure 3). In addition, protons have a naturally high abundance in proteins. Due to the nature of covalent bonds in proteins, ^{13}C - ^{13}C scalar couplings are very abundant with roughly 50-30 Hz and an amide group for each amino acid with ^1H - ^{15}N scalar couplings around 92 Hz. In solution, anisotropic interactions like dipolar couplings or CSA are averaged out due to molecular tumbling.

One of the central ideas of NMR experiments is to establish and transfer coherences between spins. The loss of coherence is described by T_2 relaxation, also called spin-spin or transverse relaxation with the rate $R_2 = 1/T_2$, manifesting itself in a decrease of detectable net magnetization (Figure 1, panel A). Coherences can be transferred, for example, via scalar couplings between two neighboring spins after developing anti-phase magnetization. On the other hand, T_1 relaxation is called spin-lattice or longitudinal relaxation with the rate $R_1 = 1/T_1$, which describes spins returning to the equilibrium state aligned along the z-axis. Pulses which equal the Larmor frequency of a spin are called on-resonance pulses. The phase of the pulse defines the destination of the spins (Figure 1, panel C) and whether they are

susceptible to the pulse. The flip angle β describes how much the spin is moved away from the z-axis and depends, for a given pulse strength, on the pulse length τ_{rf} . As soon as the spin is moved away from the z-axis, the x- and y-components of the overall magnetization vector will not equal zero anymore, enabling the detection of magnetization via a coil wrapped around the sample.

$$\beta = \gamma B_1 \tau_{\text{rf}} \quad (4)$$

As different nuclei with their individual chemical environment have different Larmor frequencies, pulses often need to cover a range of resonance frequencies. These pulses are called hard pulses, which affect e.g. all proton spins in the sample. For hard and on-resonance pulses, B_1 equals the effective field a spin experiences, B_{eff} . Assuming a 90° pulse, magnetization can be moved from the z-axis to the xy-plane. For soft pulses and off-resonance pulses instead, selective excitation of nuclei within a small range of chemical shifts can be achieved. Consequently, the low-power pulse needs to be longer to, for example, move magnetization from the xy-plane back to the z-axis. During the delay τ , the magnetization rotates and evolves in the xy-plane (Figure 2, panel B).

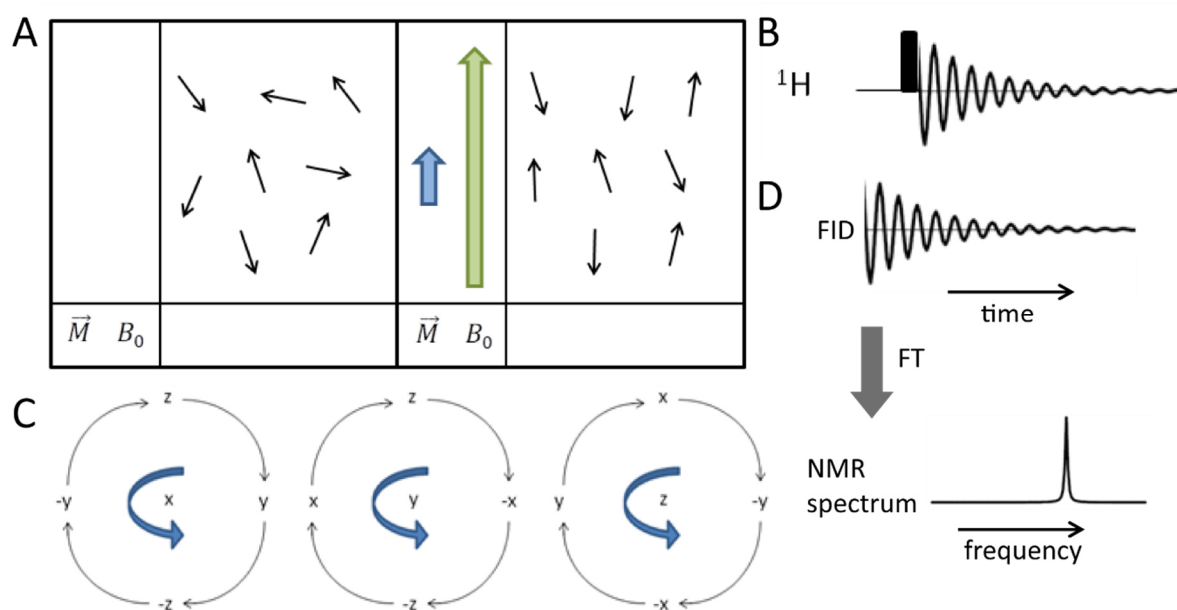


Figure 1: The net magnetization, the manipulation of spins with radiofrequency pulses, and the detection of spin Larmor frequencies. **A.** Spins in α - and β -state in an external magnetic field accounting for the net magnetization vector \vec{M} . **B.** One-pulse experiment for protons. **C.** Scheme displaying the rotation of magnetization upon the application of a pulse with the different phases x, y, or z. For positive pulse phases, e.g. x, the arrows are followed, for negative pulses, e.g. $-x$, the opposite direction should be chosen. **D.** Illustrated treatment of the free induction decay (FID) with the Fourier transform (FT), converting time-resolved into frequency-resolved data.

The net magnetization \vec{M} is detected as a series of discrete data points of the Free Induction Decay (FID) over time. Quadrature detection allows to determine the sign and the magnitude of the frequency offset by recording both the cosine and the sine components. To unravel the frequencies that account for the FID, the Fourier transform (FT) is applied, which identifies the intensity and wavelength of each oscillating frequency that contributes to the FID. The final spectrum is generated by plotting the peak intensity against the frequency. Here, slowly decaying FIDs with slower T_2 relaxation will result in sharper peaks in the final spectrum than fast decaying FIDs.

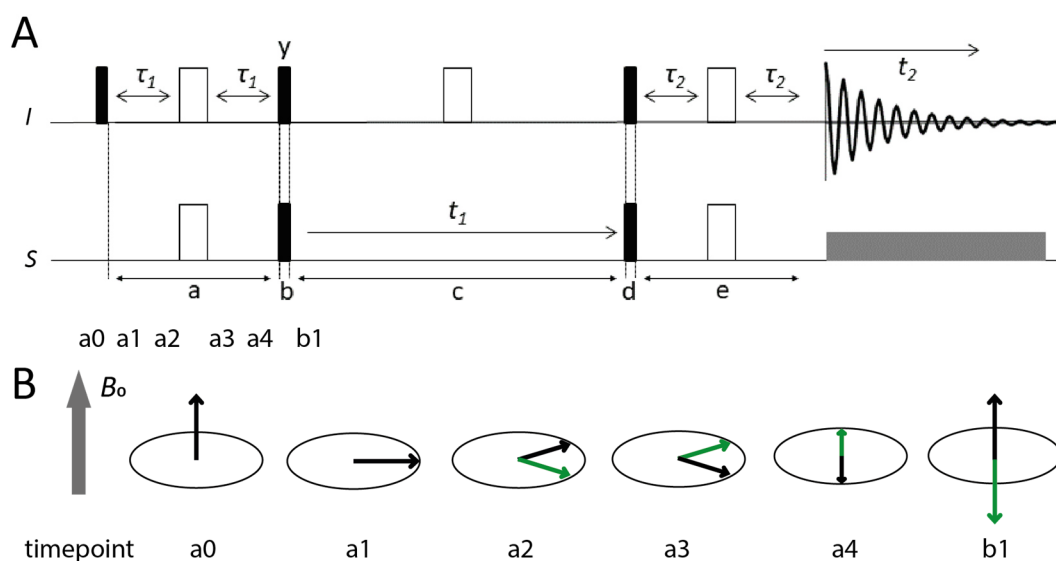


Figure 2: The HSQC experiment. **A.** Spins I and S are excited by pulses with the phase x , if not indicated differently. The black and white rectangles represent 90° and 180° pulses, respectively. With the spin echo (a), anti-phase magnetization is generated refocussing the evolution of the chemical shift but not the spin coupling. The two 90° pulses enable the coherence transfer from spin I to spin S (b). The S spins evolve the chemical shift for a time t_1 (c), during which the spin coupling is refocused, but the chemical shift is not. Finally, another spin echo transfers the coherence back to the I spins for signal detection (d, e). **B.** Vector models representing the evolution of the I spins, which are J -coupled to S spins in either α - and β -state (coloured in black and green), in the rotating frame at the timepoints indicated in panel A.

More complex molecules like proteins require sophisticated multi-dimensional experiments. For example, the heteronuclear single-quantum correlation (HSQC) experiment detects the chemical shifts of two neighboring nuclei (Figure 2). Historically, experiments based on ^{15}N -detection were commonly used but recent technological advances enable similar results with ^1H -detection, gaining higher experimental sensitivity due to the higher gyromagnetic ratio of protons. This way, the proton magnetization is detected directly and the chemical shifts of ^{15}N or ^{13}C are indirectly detected. During the preparation step, the spins in the sample are converted to the desired starting conditions. This is shown for the insensitive nucleus

enhanced by polarization (INEPT) module (period a in Figure 2), during which anti-phase magnetization is generated. The maximal peak intensity is reached for $\tau_{1,2} = 1/4J$. Next, period c allows the spins to evolve at their individual Larmor frequencies. For this period, t_1 is incremented over several scans of the experiment to indirectly detect the chemical shift of spin S . Then, magnetization is transferred back to spin I (period e) for detection during t_2 . In the final HSQC spectrum, only signals of I spins J -coupled to S spins are detected. In panel B, vector models are shown for the spin I over the first part of the experiment, during which anti-phase magnetization is prepared.

Assignment- and NOESY-based experiments for larger biomolecules are usually acquired in 3D or even higher dimensionality. The so-called “backbone walk” can be achieved with a set of J -coupling-based experiments, which allow to draw unique connections within one and between neighbouring amino acids. The classical set of experiments usually includes HNCACB, CBCAcoNH, HNCA, HNcoCA, HNcaCO, and HNCO experiments^[5]. Again, ^1H -detection provides higher sensitivity compared to ^{13}C -detected experiments due to the higher gyromagnetic ratio. Yet, proton-detected methods suffer if less amide protons H^{N} are present, for example, due to incomplete H/D back exchange or proline-rich sequences like for intrinsically disordered proteins. In the latter case, ^{13}C -detected experiments like CON, hnCOcaNCO, and hnCocaNCO experiments serve as alternative to achieve backbone assignments^[6]. NOESY experiments report on cross-relaxation between spins with dipolar couplings, which is called the nuclear Overhauser effect (NOE). This effect is modulated by the distance r between the spins with r^{-6} . Like J -coupled spins, the dipolar couplings manifest themselves as cross-peaks in the resulting NMR spectra. Both backbone-walk- and NOESY-based experiments can deliver peak assignments manually or automatically. Several automated assignment algorithms, e.g. FLYA^[7] and MethylFLYA^[8], have been developed to facilitate peak assignments.

1.1.2 The central principles of solid-state NMR

Solid-state NMR spectroscopy shares many characteristics with solution NMR spectroscopy in terms of theory, pulse sequences, detection, and data analysis. In the following chapter, the most crucial differences will be discussed.

Solid-state NMR samples are, as the name already indicates, in a solid phase, which can consist of protein crystals, pellets resulting from fast centrifugation, condensates, powders, or precipitates. Solution NMR sensitivity suffers if the protein of interest is large (>40 kDa)

and therefore tumbles more slowly. Here, T_2 relaxation increases and high-dimensionality experiments, which allow peak separation and assignment, are more difficult to perform. In solids, the orientation of molecules and their spins are fixed due to the absence of molecular tumbling. Anisotropic components like dipolar couplings and CSA are thus not averaged out like in solution due to fast molecular tumbling. The resulting static solid-state spectra show so-called powder patterns with highly overlapping, broadened peaks. To isolate single peaks in solids, magic angle spinning (MAS) was introduced achieving high-resolution spectra (Figure 3). For MAS, the samples are spun at 54.7° and very high frequencies of up to 100 kHz, averaging out most contributions from CSA and dipolar couplings^[2a]. The solid-state NMR samples thus need to be packed into rotors with diameters down to 0.7 mm. The smaller the rotor size, the faster the samples can be spun. Smaller rotor sizes are also accompanied by lower sample amount and thus lower net magnetization. In particular, the very strong ^1H - ^1H dipolar couplings require fast MAS to achieve narrow peak line widths. Deuteration can contribute to narrower line widths by decreasing the proton density in the sample. Still, the resolution of proton spectra in solution can be up to 10-100-fold better than in solids^[3].

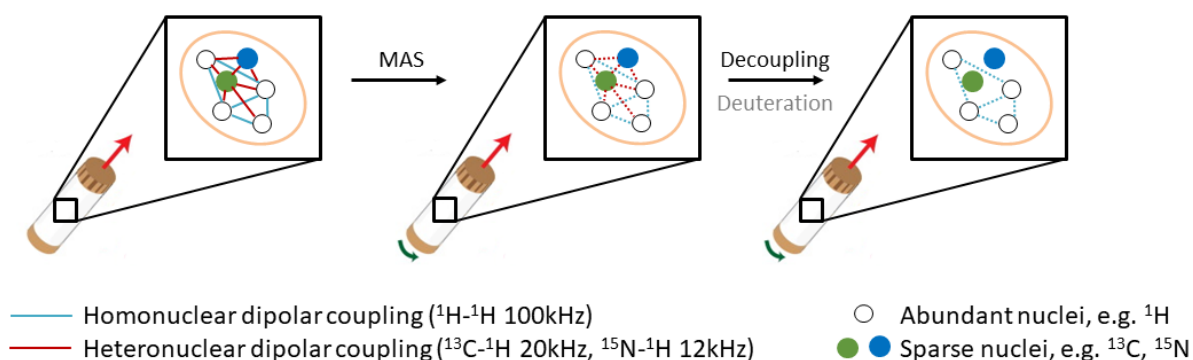


Figure 3: Scheme depicting methods to decrease the strength of dipolar couplings.

In the solid state, experimental setups can differ from the equivalent experiment in solution. For example, INEPT modules are a principle building block in solution NMR experiments to transfer magnetization by generating anti-phase magnetization for J -coupled spins. Cross polarization (CP) can achieve more efficient magnetization transfers via dipolar couplings^[9]. Using these transfers, backbone assignment experiments like in solution are possible. In analogy to solution NMR experiments, deuteration, ^1H -detected experiments compared to ^{13}C -detected experiments, and higher B_0 can increase the experimental sensitivity.

To elucidate contacts between residues which are spatially close but sequentially apart, radio-frequency-driven recoupling (RFDR) experiments were implemented, which gain similar

information as the NOESY experiment in solution. Here, dipolar couplings are reintroduced by rotor-synchronized pulsing. The resulting spectra contain cross-peaks of different intensities, which can reflect the distance between the interacting spins. However, phenomena like dipolar truncation, which describes the preference for short-distance compared to long-distance dipolar magnetization transfers during recoupling, decrease the number of cross peaks in RFDR spectra^[10]. Therefore, structure calculations and peak assignments via RFDR experiments are more challenging using solid-state rather than solution NMR spectra. Moreover, solution and solid-state relaxation experiments detect motion at different timescales and can complement each other.

In summary, solid-state NMR is a viable technique to study especially large molecules and complexes. While it shares similarities with solution NMR, solid-state NMR experiments are not yet as easy to establish and interpret as solution NMR experiments and still offer a lot of potential for developing new methods.

1.1.3 The contribution of NMR spectroscopy to the field of structural biology

NMR spectroscopy is widely used as a method to identify small molecules. In the last two decades, it has also served well for analyzing biochemical samples including more complex molecules. The milestones in the development of NMR spectroscopy can be easily visualized looking at the Nobel prizes awarded to the field. NMR as a method was established by Felix Bloch and Edward M. Purcell, reporting the first NMR measurements in bulk condensed phase and being awarded with the Nobel prize in 1952^[11]. A few years before, Isidor I. Rabi successfully analyzed the magnetic properties of nuclei^[11]. Roughly 40 years later, the Nobel prize was given to Richard R. Ernst in 1991, who advanced the field developing new pulse sequences and implementing two-dimensional experiments^[12]. Kurt Wüthrich was awarded for his contributions to calculating three-dimensional protein structures with solution NMR experiments in 2002, signifying the potential for structural biology^[13]. One year later, Paul C. Lauterbur and Sir Peter Mansfield shared the Nobel prize for developing magnetic resonance imaging, expanding the applicability of NMR-related experiments from *in vitro* samples to human patients^[14].

Today, both solution and solid-state NMR can provide information on protein structures like secondary structure elements via dihedral angles, 3D structures based on dipolar couplings, and physical parameters like folding, hydration, and protonation states (Figure 4). For solid-

state samples, no long-range order is needed as for other methods analyzing, for example, crystals^[15]. It can equally contribute to analyzing protein dynamics as a broad selection of experiments is available which can detect motion on nearly all regimes important for proteins, ranging from the picoseconds to the seconds timescales (Figure 6)^[15]. As both structural and motion-related features can be assessed, NMR can distinguish between static and motional disorder. In terms of molecular interactions, exchange regimes, binding affinities, and chemical shift perturbations (CSPs) can be quantified, identifying binding pockets with atomic resolution. These possibilities, however, can only be explored once the individual protein residues are assigned to peaks in the NMR spectra. As mentioned before, assignments can be achieved via the classical backbone and sidechain assignment approaches using scalar coupling-based experiments or via distance-based assignments utilizing dipolar couplings. Alternatively, NMR samples can also be biochemically modified to distinguish peaks and residues from each other via isotopic labelling or mutations, for example.

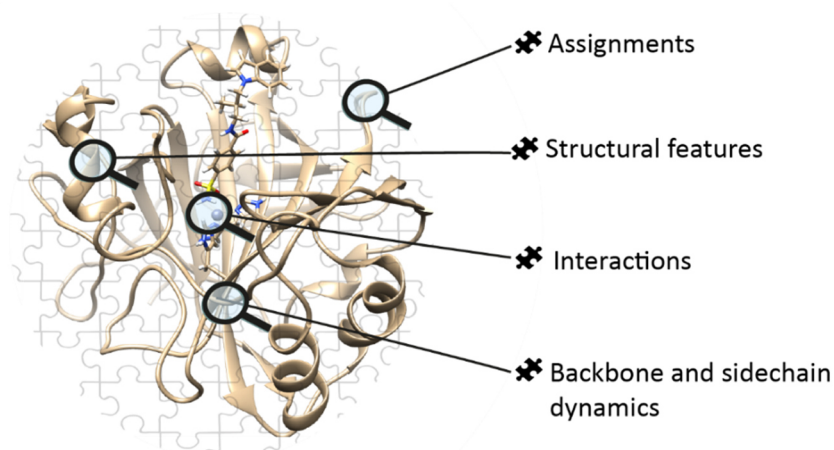


Figure 4: Scheme depicting the applications of NMR spectroscopy for characterizing proteins, here shown exemplarily for the X-ray structure of the hCAII-SBR complex (PDB: 1IF7^[16]).

In structural biology, various other techniques are available to characterize protein structures and motion, including X-ray and neutron diffraction, electron crystallography, electron microscopy (EM), electron paramagnetic resonance spectroscopy (EPR), X-ray or neutron solution scattering, and fiber or powder diffraction, which, together with NMR, account for most of the protein structures submitted in the Protein Database (PDB)^[17]. The most popular technique is X-ray diffraction because it can deliver high-resolution structures in a high throughput fashion. Electron microscopy has gained popularity since newly developed detectors and software deliver structures with down to 3 Å resolution^[18]. Recently,

AlphaFold^[19] has emerged predicting protein structures with astonishing accuracy based on known protein homologues and their structures.

In terms of dynamics, X-ray diffraction and cryo-EM can resolve motion by freezing samples at different time points to visualize different conformations. Here, a mostly qualitative assessment of motion has been possible, which comes with a certain bias towards the conformational ensemble. X-ray diffraction often depicts energetically favourable conformations due to crystallization and measurements at low temperatures. Recently, X-ray-free electron lasers have delivered promising results on reactions triggered by light or chemical diffusion, revealing also excited states at physiological temperature^[20]. Cryo-EM can image snapshots of several conformational states. However, it is also limited by how fast the protein solution can be vitrified. This process itself can introduce a bias for conformational states, for example, via the interaction of proteins with the frozen sample surface. In contrast, solution NMR impresses by adding detailed atom-specific information on both structures and dynamics of small proteins (<30 kDa), although it is less frequently used as a structure determination tool. Solid-state NMR has been successful in producing reasonably resolved protein structures in recent years and is less limited by size than solution NMR. However, for crystalline solid-state NMR and X-ray diffraction samples, a certain bias can be assumed as major rearrangements are unlikely to be feasible within the crystal lattice.

Protein dynamics can also be estimated by single-molecule super-resolution techniques like fluorescence-resonance energy transfer (FRET) and computational methods like molecular dynamics (MD) simulations. Super-resolution microscopy has developed single-molecule imaging methods which allow to study protein interactions in cells and conformational dynamics of proteins. Here, protein motion can be observed directly, and timescales can be resolved quantitatively. However, the overall resolution does not reach the atomic level and the visualized timescale is limited to slow motion by the detector^[21]. MD simulations can predict intramolecular protein motion and intermolecular interactions with atomic resolution on the other side of the scale, in the fast motion regime. Recently, very elaborate MD simulations have reached the fast millisecond timescale^[22]. Although MD models have been shown to predict dynamics increasingly accurately, they should be complemented with experimental data.

Considering the range of experiments available for NMR spectroscopy, gaps in structure- and dynamics-related characterizations by other methods in structural biology can be filled.

Especially microsecond motion is a time regime which is challenging to predict with MD and not yet detectable with super-resolution microscopy. Therefore, an integrative structural-biology approach has the highest potential to answer especially dynamics-related questions for different protein systems on nearly any timescale.

1.1.4 The characterization of protein-ligand complexes

One of the most common ways to manipulate protein function is by regulating protein interactions. These result from atomic interactions involving, for example, hydrogen bonds, ionic interactions, π -stacking, and Van-der-Waals interactions. One of the most important parameters is the binding affinity of the interaction partners, often described with the dissociation constant K_D for any kind of complex or the inhibitory constant K_I , which is mostly used to describe enzyme-inhibitor complexes. Overall, an interaction is modulated by the concentrations of the binding partners and the rates at which the complex associates and dissociates (equation (5)). The k_{on} rate constant reflects how often the interaction partners associate at a given concentration, which is further modulated by the concentrations of the protein $[P]_{free}$ and the ligand $[L]_{free}$. Very high k_{on} rates are finally limited by the number of collisions with the required orientation over a certain time, which was estimated to be $10^9 \text{ s}^{-1} \text{ M}^{-1}$ [23]. The k_{off} rate constant often defines the interaction strength more directly, determining how long the complex with the concentration $[PL]$ stays intact. Weak binders usually have K_D values around 1 mM while the K_D values of tight binders can reside in the nM to pM range.

$$K_D = \frac{[P]_{free}[L]_{free}}{[PL]} = \frac{k_{off}}{k_{on}} \quad (5)$$

The early understanding of protein interactions was based on the lock-and-key principle, assuming that a binding event takes place due to matching static three-dimensional surfaces[24]. However, it does not serve as a model for multi-step binding processes and does not account for different molecular conformations. The concepts of induced fit and conformational selection have been developed to model the complexity of molecular interactions more closely. While the induced fit relies on a conformational change of the binding partner due to the binding event, conformational selection has been proposed for cases where the molecule binds only one conformational state of its binding partner[24]. These concepts are based on the energy landscape of proteins, describing the different

populations and fluctuations of molecules which possess various distinct energy levels^[25]. Alongside many other biophysical methods, NMR spectroscopy can report on the affinity, residue-specific changes in the chemical environment, and populations of different conformations^[23].

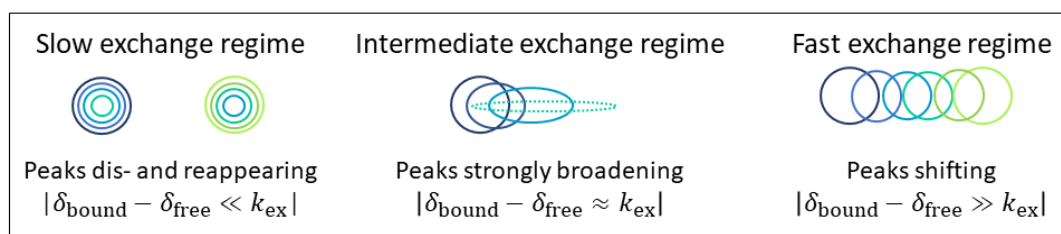


Figure 5: Overview on the three different chemical exchange regimes in solution NMR spectroscopy. The coloured circles represent the intensity contours of a peak depicting changes in shape and position during ligand titrations, assuming different chemical exchange regimes. Below, the mathematical relationship between the chemical shifts of the apo protein δ_{free} , the protein in ligand-bound state δ_{bound} , as well as the exchange rate k_{ex} are presented^[26].

In Figure 5, the different exchange regimes are summarized. In the fast-exchange regime, interactions with K_D values between 10 mM and 3 μM can be quantified^[26]. Here, the chemical shift difference between apo and ligand-bound state in Hz is much smaller than the exchange rate. For such protein-ligand complexes, the chemical shift is averaged over the different population ratios between apo and occupied protein during the ligand titration. The progression of the peak position can be followed comparing the different spectra.

For intermediate-exchange interactions, on the other hand, the chemical shift differences are in the same regime as the exchange rate. This results in broadened peaks with increasing ligand concentrations, lowering the signal-to-noise ratio sometimes even beyond the detection limit. To prevent titrations with this exchange regime, for example, the temperature, viscosity, and magnetic field strength can be adapted to speed up or slow down the exchange.

For slow exchange, the exchange rate is lower than the chemical shift difference between apo and bound state. The peak intensities in the apo state decrease while new peaks corresponding to the bound state appear and increase in intensity until, ideally, a fully saturated protein complex is observed. As the peaks cannot be tracked easily like for fast exchange, the peaks in the ligand-bound state might need to be assigned anew. If this appears not to be feasible, the closest peak with the smallest CSP is usually assumed to correspond to the peak undergoing slow exchange^[26]. With this method, however, CSPs can be underestimated, and wrong assignments can be obtained. For fast and slow exchange, the

populations of apo and ligand-bound protein can be determined comparing the averaged chemical shift to the chemical shifts of the apo or fully saturated protein and comparing the peak intensities of the peaks corresponding to the same residue, respectively. It should be stressed that additional parameters like pH, temperature, ligand solubility, and solvents can impede the titration experiment. Mapping the residues with changed chemical shifts on a known protein structure allows to identify the putative binding pocket but also residues indirectly affected by ligand binding.

1.1.5 NMR spectroscopy as a reporter on dynamics

For many decades, researchers visualized proteins as rather static objects. This was mainly the case because the available tools elucidating proteins on a molecular level resulted in snapshots of protein structures. In the last two decades, however, techniques have been developed and improved which combine information on the various timescales of dynamics that proteins undergo^[27].

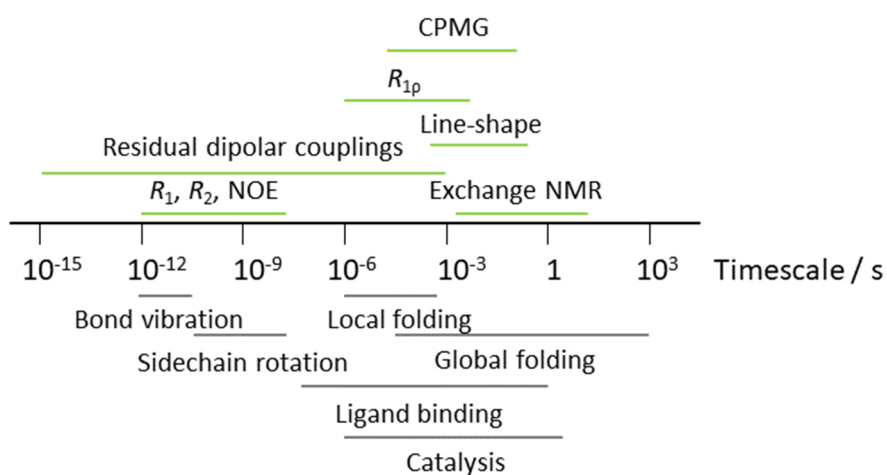


Figure 6: Solution NMR experiments detecting different timescales of dynamics for proteins.

The different timescales of motion include bond vibrations, rotations of, for example, sidechains, local and global folding, ligand association and dissociation, catalysis, and the overall molecular tumbling and diffusion, mostly dependent on the hydrodynamic radius of the molecule (Figure 6). NMR spectroscopy can detect motion qualitatively and quantitatively at thermal equilibrium with atomic resolution, depending on the available timescale-specific experiment^[28]. For function, the most important conformational changes of proteins usually take place on the micro- to second timescale, which underlines the importance of lineshape analysis, exchange NMR, and relaxation dispersion experiments. Dynamics on this timescale

are also known as chemical exchange (chapter 1.1.5), which describes fluctuations between two or more conformations^[29].

In solution, pico- to nanosecond motion is characterized by a standard relaxation data set of R_1 , R_2 , and NOE relaxation experiments for ^{15}N spins of backbone amides. These experiments are powerful tools to study local and sidechain dynamics. As mentioned in chapter 1.1.1, R_1 and R_2 relaxation rates are mainly driven by dipole-dipole interactions and CSA^[2b, 29]. These rates depend on the speed of the molecular tumbling. With slower molecular tumbling, low temperature or viscosity and strong anisotropic contributions, R_2 relaxation rates increase, which can diminish the sensitivity of especially NMR experiments with long pulse sequences. R_1 rates, on the other hand, are highest for small macromolecules and slow down the slower or faster molecular tumbling gets. In the solid state, slow R_1 rates can require very long recycle delays between scans, which can be sped up by doping samples with paramagnetic chemicals^[30].

On the opposite side of the detectable timescales, exchange spectroscopy (EXSY) and chemical-exchange saturation transfer (CEST) experiments allow to elucidate the contribution of exchange to relaxation on the seconds to tens of milliseconds timescale via differences in the isotropic chemical shifts^[31]. With EXSY, the diagonal and cross-peak intensities and peak shapes in a 2D-spectrum can be studied employing a defined mixing period with the spins along the z-axis^[32]. CEST can be utilized to detect otherwise “invisible” peaks of strongly skewed populations^[33]. It is sensitive to the same timescale as EXSY and therefore can be used to complement Carr-Purcell-Meiboom-Gill relaxation dispersion (CPMG) experiments, which probe exchange dynamics on the microsecond timescale. During CEST, a weak B_1 field is applied at different chemical shift offsets in the spectrum followed by a specified time for longitudinal relaxation to occur, modulating the intensity of the peak of interest^[33]. When the resonance condition with the major or minor state is approached and hit, the peak intensities decrease. CEST can determine the populations, the lifetimes of the short-lived conformations, its chemical shifts, and the line widths by perturbing/saturating the exchanging peaks with weak fields^[33].

Relaxation dispersion experiments comprise CPMG and $R_{1\rho}$ experiments which can quantify exchange contributions on the micro- to milliseconds timescale to transverse relaxation via oscillating isotropic chemical shifts^[31]. As described in the reviews by Ishima and Bagby^[28] and by Palmer^[29], CPMG experiments are sensitive towards micro- to millisecond dynamics. CPMG

pulse trains are employed with the carrier on resonance with the major conformational states^[33]. The R_2 rates are measured varying the delays between the CPMG pulses, which translate into the effective CPMG frequencies ν_{eff} . While the magnetization of the non-exchanging nuclei is always perfectly refocused, nuclei in residues undergoing conformational exchange experience different degrees of non-refocused chemical shift evolution, leading to dephasing, i.e. R_2 relaxation^[33].

With larger exchange rates R_{ex} than relaxation rates, the contributions can be simplified to:

$$R_2^i(\nu_{\text{eff}}) = R_{2,0} + R_{\text{ex}}(\nu_{\text{eff}}^i) \quad (6)$$

Instead of recording several relaxation delay points, only two points at time zero and a specified constant time are sufficient to calculate each R_2 value per ν_{eff} condition. When motion occurs at a similar regime as ν_{eff} , $R_2(\nu_{\text{eff}})$ decreases with increasing ν_{eff} . Several optimizations of CPMG experiments have been implemented to, for example, prevent J -coupling evolution during the delay times of the CPMG train^[34].

$R_{1\rho}$ relaxation experiments are sensitive to dynamics in the microsecond range and were implemented both in solution and the solid state. Bloch-McConnell relaxation dispersion (BMRD) in solution and solids and near-rotary resonance relaxation dispersion (NERRD) in solids will be discussed here. In both experiments, the spins of interest are locked at different effective fields at the angle θ ^[34]:

$$R_{1\rho} = R_1 \cos^2 \theta + (R_2^0 + R_{\text{ex}}) \sin^2 \theta \quad \text{with} \quad \tan \theta = \omega_1 / \Omega \quad (7)$$

The amplitude of the spin-lock field is defined as ω_1 , and Ω is the population-average resonance offset frequency. R_2^0 is estimated to reveal the exchange contribution. Several experiments have been developed to account for evolving J -couplings, which impact the T_2 relaxation. For example, ^1H and ^{15}N spins can be decoupled applying 180° pulses on ^1H to prevent ^{15}N CSA and ^1H - ^{15}N dipolar couplings^[2a, 34]. Both on-resonance and off-resonance versions of this experiment are available, defined by the position of the carrier and thus requiring different data processing^[2a]. While the experimental setups of BMRD and NERRD experiments are very similar, the different field strengths, in relation to the rotor frequency, allow BMRD to be sensitive for changes of the isotropic chemical shift while NERRD detects reorientations of bond angles^[15]. BMRD experiments can be measured both in solution and solids. NERRD is exclusively available for solid-state NMR.

The BMRD experiment is usually used to detect ^{15}N -backbone dynamics of proteins on the fast millisecond to slow microsecond timescale via fluctuations of the isotropic chemical shift^[35]. The equation describing two-site exchange with skewed populations p_1 and p_2 , the chemical shift difference $\Delta\omega_{12}$, the effective radio frequency field strength ω_e , the timescale $\tau_{ex} = \frac{1}{k_{ex}}$, and the baseline relaxation rates R_1 and $R_{1\rho}^0$ is shown in equation (8)^[36].

$$R_{1\rho} = R_1 \cos^2 \theta + R_{1\rho}^0 \sin^2 \theta + \frac{\sin^2 \theta p_1 p_2 \Delta\omega_{12}^2 k_{ex}}{k_{ex}^2 + \omega_e^2} \quad (8)$$

With magic-angle spinning (MAS) solid-state NMR, a larger range of motions can be detected compared to solution NMR due to the lack of molecular tumbling and different hardware requirements (chapter 1.1.2)^[15]. Various techniques from reintroducing dipolar couplings, relaxation experiments adapted from solution NMR, and quadrupolar interactions via deuterium spins are available to detect motion^[15]. The interpretation of relaxation data tends to be more complicated than in solution due to anisotropic contributions, which can dominate the relaxation-induced signal decay compared to dynamics-induced contributions and can only be partially alleviated by deuteration^[31].

The NERRD experiment can measure microsecond dynamics of $\tau_c \sim \omega_e^{-1}$ by reintroducing dipolar spin couplings when the spin-lock fields approach the half- or rotary-resonance conditions^[15, 31, 36]. During the spin lock, the spins are in the transversal plane, described by the tilt angle $\beta_e = \pi/2$ ^[37]. While ^{15}N -NERRD experiments allow to quantify dynamics, the anisotropic interactions for protons are difficult to quantify even with deuteration and 100 kHz MAS, thus interfering with the interpretation of the relaxation data and only allowing a qualitative assessment of e.g. ^1H -NERRD data^[36]. The ^1H -NERRD experiments can be described by equation (9) for on-resonance conditions with $\beta_e = (\pi/2)$, displaying the impact of the homonuclear dipolar coupling d_{II} , and the relation between the spinning speed ω_r and the effective radio frequency field strength ω_e as a function of spectral density $J(\omega)$.

$$R_{1\rho}^{AB}(\pi/2) = \frac{1}{4} d_{II}^2 \left\{ \frac{1}{6} J(\omega_r) + \frac{1}{12} J(2\omega_r) + \frac{9}{2} J(\omega_I) + 3J(2\omega_I) + \frac{3}{8} J(2\omega_e - 2\omega_r) + \frac{3}{4} J(2\omega_e - \omega_r) + \frac{3}{4} J(2\omega_e + \omega_r) + \frac{3}{8} J(2\omega_e + 2\omega_r) \right\} \quad (9)$$

The spectral density function derived from the extended model-free approach (10) is dependent on the order parameter S^2 , which here describes the amplitude of motion with many degrees of freedom for $S^2 = 0$ and rigid structures for $S^2 = 1$, and the timescale τ_c of the bond vector motion, assuming the motion is dominated by only a single timescale^[15].

$$J(\omega) = \frac{2}{5} \frac{(1-s^2)\tau_c}{1+(\omega\tau_c)^2} \quad (10)$$

Overall, different experiments in solution and in the solid state complement each other in terms of timescales and the properties reporting on motion, for example, the chemical shift or bond fluctuations. All these experiments also come in different flavors allowing to choose between numerous nuclei as reporters on relaxation and various versions, which, for example, can suppress solvent signals more efficiently or are better suited to obtain dynamics for large molecules.

1.1.6 Isotopic labelling and selective methyl labelling

Parts of this chapter were adapted from^[38]: J. Kotschy, R. Linser, 2020, “Proton-detected solid-state NMR and its applications to membrane proteins” in “Solid-state NMR Applications in biomembrane structure”, edited by Frances Separovic and Marc-Antoine Sani, pages 15-1 to 15-35, IOP Publishing, DOI: 10.1088/978-0-7503-2532-5ch15.

The natural abundance of the most common nuclei in biomolecules is low, except for protons (Table 2). Especially for higher-dimensionality experiments, like backbone assignments of proteins, uniform ¹³C-, ¹⁵N-labelling is required. Historically, the most popular approach has been to express biomolecules like proteins in *Escherichia coli* (*E. coli*) cell cultures in isotopically uniformly labelled media. This method can be applied for a broad range of proteins and is easily established. Alternatively, cell-free protein expression systems have been implemented as well, which can obtain complex labelling schemes at lower expenses^[39]. In addition, single amino acids can be labelled by amber stop codons or specifically charged tRNA molecules^[39]. Post-translations modifications can be introduced both via cell-free expression and uniform labelling in insect or mammalian cells. The correct folding of complex or highly amphipathic proteins seems usually more feasible via cellular expression due to conditions closer to the native environment and chaperones aiding the folding process.

Moreover, deuteration is key to study large proteins and to obtain reasonable solution or solid-state NMR spectra. At longer correlation times in solution and solid-state experiments with MAS lower than 100 kHz, dipolar couplings strongly accelerate T_2 relaxation (chapter 1.1.2). These rates can be lowered diluting the protons in the sample by deuteration. Moreover, highly resolved proton-detected spectra can be already obtained without deuteration at MAS frequencies as high as 100 kHz in solids^[40]. Of course, an overall lower proton abundance upon deuteration can negatively impact the feasibility of pulse sequences

involving protons. Different labelling techniques have been established to incorporate protons only at the desired sites. For deuteration via cell expression, the cells need to be adapted to grow in media with deuterated carbon sources in D₂O. Various protocols are available for this purpose, including a stepwise adaptation to deuterated media, or growing the cells in usual LB medium to high density to transfer them to isotopically labelled medium afterwards (chapter 4.9.1-2). Perdeuterated samples are defined as deuterated protein samples with protonated amide groups, which has been a popular labelling strategy as it is a simple procedure and allows backbone assignments. These amide protons are re-introduced via H/D exchange after expression during protein purification in H₂O-based buffers. Fully deuterated protein samples can be achieved by purifying the proteins in D₂O-based buffers or incubating the final protein sample in D₂O-based buffers. Solvent-inaccessible residues in hydrophobic pockets or amides involved in strong hydrogen bonds can slow down or prevent H/D exchange, which possibly can be alleviated by un- and re-folding the protein.

Moreover, several other techniques balancing the H/D ratio have been established. Fractional perdeuteration combines a deuterated or protonated carbon-source with H₂O- or D₂O-based media, called inverse and coined fractional deuteration, respectively^[41]. While perdeuterated samples reach high deuteration percentages for all sites except amides, fractionally deuterated samples have an overall lower deuteration level producing different isotopomers for all amino acids. Looking at aromatic amino acids, protonated carbon sources can deliver proton reporters at the sidechains while reducing the overall proton density^[42]. For example, Mance *et al.* applied coined fractional deuteration for the membrane protein KcsA for the solid state in 2015^[41a]. They reported a high deuteration level at C_α atoms and low deuteration for side-chain carbons and amides. They stated that perdeuteration and fractional deuteration achieved similar H^N peak resolution at 55 kHz MAS^[41a]. Similarly, the reduced adjoining protonation (RAP) method involves a deuterated carbon-source and a specific ratio of D₂O to H₂O in the expression medium. This method enables access to otherwise non-exchangeable proton sites, e.g. methyl groups^[43]. Moreover, specific types of amino acids can be protonated via proton-cloud labelling providing protonated amino acids in addition to the deuterated expression medium. Stereo-array isotope labelling (SAIL) works similarly in the context of cell-free protein expression^[44]. Common challenges associated with these techniques are slow or even arrested cell growth in deuterated media as well as scrambling. Due to the additional neutron in deuterium compared to hydrogen nuclei, enzyme kinetics

and interactions may deviate considering, for example, hydrogen bonds and pH values. Scrambling describes the metabolization and use of precursors beyond the intended scope. The low experimental sensitivity for large slowly tumbling molecules in solution NMR, mentioned in chapter 1.1.1, can also be addressed using methyl groups as reporters instead of e.g. amides of the protein backbone. For this purpose, perdeuterated proteins with protonated methyl groups of selected amino acids can be prepared. Only six amino acids from the proteinogenic repertoire bear methyl groups, namely Ala, Ile, Leu, Val, Thr, and Met, which each can be labelled specifically using various precursors and protocols. Consequently, the number of peaks in the methyl spectra can be defined as desired and spectral crowding can be reduced. Methyl groups are also less prone to fast T_2 relaxation compared to amides at the backbone due to their location at the relatively flexible amino acid side chains, reducing line broadening and providing reasonable sensitivity for proteins of >100 kDa size in solution^[45]. Moreover, methyl groups include three protons while the amide nitrogen is bound to only one proton. Theoretically, this should lead to higher sensitivity for solution NMR experiments. In solids, it should be considered to substitute one or two methyl-protons by deuterium as the dipole-dipole couplings are very high at such low proton-proton distances^[46]. While high-speed MAS of up to 110 kHz can alleviate these interactions to some extent, maximum sensitivity for rigid methyl groups was predicted to be achieved only at >300 kHz spinning speed, which is not yet feasible^[47]. In solution, samples containing H_3C -groups produce higher sensitivity and similar peak widths compared to HD_2C - or H_2DC -groups^[48].

To specifically protonate only the methyl groups of some amino acids with an otherwise deuterated background, precursors with low scrambling are required^[45d]. In contrast, glucose or pyruvate are suited to promiscuously label amino acids. For example, ILV-methyl labelling can be specifically obtained using 2-acetolactate and α -ketobutyrate for valine and leucine, and isoleucine labelling, respectively (Figure 7). To avoid carbon-carbon scalar couplings, only the C_γ - or C_δ -methyl groups should be ^{13}C -labelled. Together with specific protonation, all other carbon atoms in the amino acids would be unlabelled and bound to deuterium instead of hydrogen atoms. Moreover, the methyl groups for valine and leucine can be labelled stereospecifically, meaning that only one of the geminal groups, either the pro(*R*)- or pro(*S*)-methyl group, is isotopically labelled. For the solid state, the equivalent sample should contain

HD₂-¹³C-methyl groups to account for the dipolar couplings between the protons of the same methyl group, which are negligible in solution. Inter-residual dipolar couplings between methyl protons in solids remain.

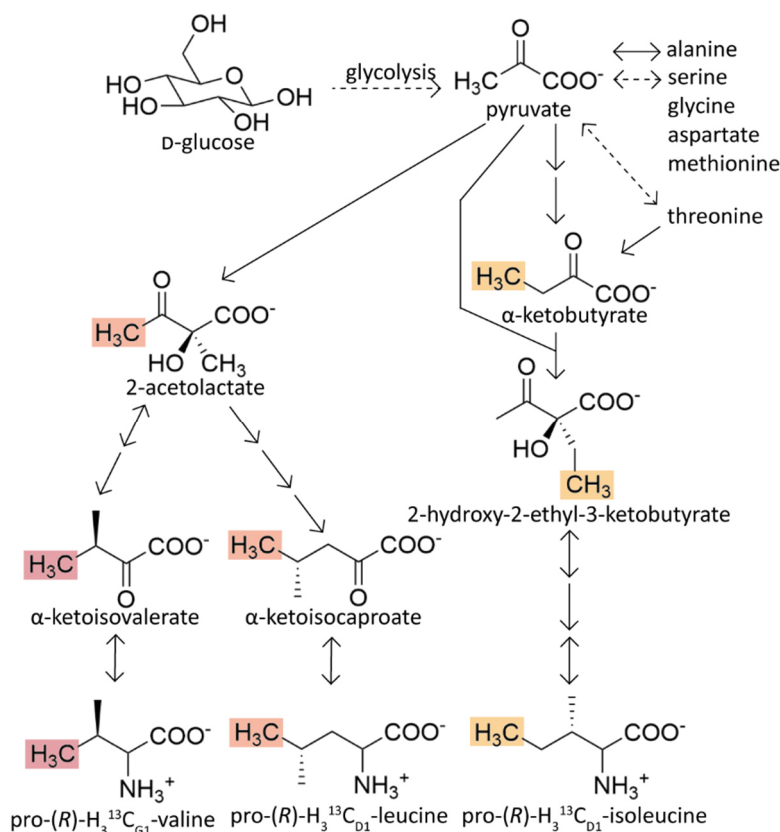


Figure 7: Scheme of amino acid synthesis pathways relevant for stereospecific labelling of pro-(R)-methyl groups of the residues valine, leucine, and isoleucine in *E. coli*^[45d]. The arrows indicate enzymatic conversions of molecules in the cells. Arrows pointing in two directions represent both anabolic and catabolic reactions. The dotted arrows describe several required conversion steps. The labelled methyl groups were shaded in red for valine, orange for leucine, and yellow for isoleucine.

All in all, a large variety of labelling schemes is available, which should be chosen based on the individual research question at hand and the practically feasible time- and cost-related investments. A universal tool especially for large proteins, however, is deuteration, which is often key to observe well-resolved NMR spectra both in solution and the solid state.

2. The Zn²⁺-complexing drug target hCAII

2.1 Introduction

The published X-ray structures of hCAII (for example, PDB 1IF7^[16]) refer to residue K127 instead of K126, which follows in the sequence after T125. This mismatch was eliminated for the X-ray structure determined here (chapter 2.2.1). For clarity, all residue numbers >125 reported in the literature were adapted to the numbering used here (for example, T200 will be called T199 in the following).

2.1.1 Known characteristics of hCAII

The enzyme hCAII catalyzes the reversible hydration of carbon dioxide to form hydrogen carbonate (Figure 9) at an impressive rate of one million conversions per second at pH 9 and 25 °C^[49]. hCAII has therefore one of the fastest catalytic rates known for proteins. The native hCAII protein consists of 260 amino acids (roughly 29 kDa) and was first identified in mammalian erythrocytes by Meldrum and Roughton in 1933^[50]. Since then, it has been revealed that it belongs to the family of α -type CA whose members include all CA enzymes present in animals^[51]. The other two families of β - and γ -CA evolved independently and represent mainly CA in plants and methane-producing bacteria, respectively. The α -CA enzymes have several isozymes: In humans, 15 isozymes have been found^[52]. While hCAI and hCAII are soluble proteins, for example, hCAIX and hCAXII are associated with the membrane. Several post-translational modifications were identified for the different isozymes and might contribute to the regulation of the activity of carbonic anhydrases^[53]. The importance of understanding how carbonic anhydrases work is underlined when considering that the isozymes of hCA are involved in a myriad of disease conditions e.g. glaucoma^[54], epilepsy^[55], obesity^[56], and cancer^[57]. Potentially, carbonic anhydrases of bacteria, fungi, or pathogens could also be targeted to treat infectious diseases^[57a].

Especially hCAII is widely distributed over different cell types and organs in humans^[58]. It localizes to the cellular cytoplasm and regulates the pH and osmotic pressure, and is indirectly involved in respiration^[51a], bone development^[59], fluid secretion in the anterior chamber of the eye, and ion homeostasis in duodenum, kidney, and pancreas^[60]. For example, it was shown that hCAII interacts with SLC26A6, which exchanges Cl⁻ and HCO₃⁻ at the plasma membrane^[60a]. Mutating hCAII resulted in reduced SLC26A6 activity^[60a] but activation of

SLC26A6 by PKC caused hCAII to be delocalized from the membrane^[60a]. There are around twelve approved drugs available to address intraocular pressure, glaucoma, insomnia, hypertension, epilepsy, osteopetrosis, and Ménière's disease targeting hCAII^[60b].

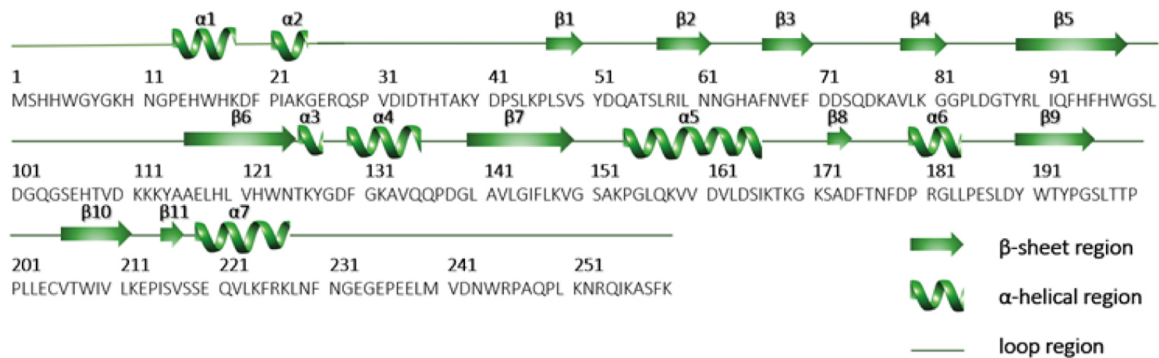


Figure 8: Designated secondary structure elements aligned with the sequence of hCAII.

The structure of hCAII was solved via X-ray crystallography by Liljas *et al.* in 1972^[61] and since, >100 structures of mainly hCAII-ligand complexes have followed^[17]. The secondary structure of hCAII consists of mainly β -sheets, few α -helices, and few β -turns which together form one, mainly spherical molecule in which only the N-terminus is more loosely associated with the rest of the enzyme (Figure 8). Three histidine residues, H94, H96, and H119, complex Zn^{2+} at the catalytic center of the enzyme. Zn^{2+} ions form tetrahedral complexes with water or hydroxide as the fourth coordination partner. It has been shown that the pK_a of water bound to Zn^{2+} decreased from fourteen to seven resulting in a drastic increase of acidity^[62]. The mechanism of hCAII converting water and carbon dioxide to bicarbonate is schematically shown in Figure 9. It was proposed that the first 23 residues of the enzyme can be removed without diminishing stability or activity^[63].

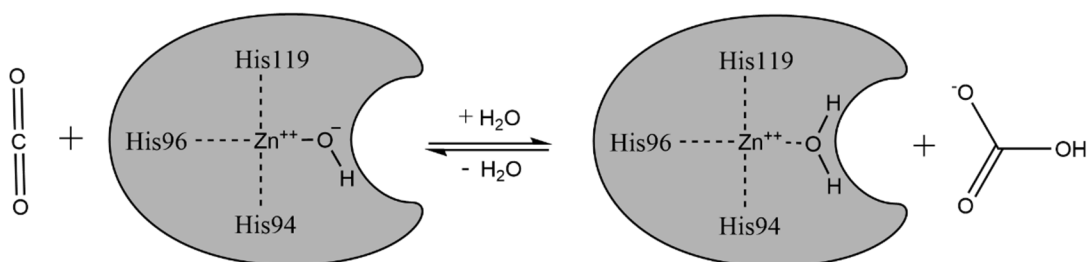


Figure 9: Scheme displaying the catalytic mechanism of hCAII converting CO_2 and water into bicarbonate^[58].

Close to the catalytic site, H64 was identified as a proton shuttle, which intercepts the protons released from the water molecule when complexed by the Zn^{2+} ion^[58]. Mutating H64 resulted in a depletion of carbon dioxide hydration, which could be partially recovered by addition of

e.g. exogenous proton donors^[58, 64]. Mutating the adjacent N62 also results in reduced activity of hCAII due to the loss of a hydrogen bonding partner of H64^[65].

In addition, various other amino acids close to the catalytic site indirectly contribute to the enzymatic activity of hCAII. T198 is evolutionary conserved in all carbonic anhydrase isoforms, which is also reflected by a ~100% reduction of enzymatic activity when it is mutated to alanine^[66]. The hydrogen bond network between Zn^{2+} , HO^- , E106, and T198 is conserved in all non-plant carbonic anhydrases^[67]. The network arranges the active site in a way that the bound CO_2 molecule is susceptible for the nucleophilic attack of the OH^- ion exposing a lone pair towards the CO_2 molecule^[58, 66b, 68].

Various residues of hCAII form a cone-shaped binding pocket next to the catalytic site towards the protein surface with partly hydrophilic and hydrophobic surfaces. W5, W7, W16, F20, F130, L197, T198, and T199 contribute to the “hydrophobic wall” in the binding pocket next to the catalytic site and presumably are responsible for anchoring the N-terminus to the rest of the protein. Instead, residues H3, H4, R58, H64, E69, D71, D72, H94, H96, E106, H119, D129, K132, and K169 add hydrophilicity to the binding pocket. Apart from the catalytic pocket, hCAII contains a hydrophobic core with F66, F70, F93, F95, W97, F175, F178, and F225, which is conserved in hCAI and hCAII. In addition, there are many conserved water molecules that were found in X-ray diffraction or NMR protein measurements^[51a, 66a]. An evolutionarily conserved water network has been identified with the time for geometric reorganization matching the residence time of Zn^{2+} -bound water or hydroxide on the fast microsecond timescale^[66a]. Inhibitors have been shown to interact with such “deep water” molecules, for example, the water molecule interacting with the amide group of T198, and to partially displace other water molecules in the water network^[66a].

Recently, R_1 and $R_{1\rho}$ relaxation rates derived from solid-state NMR showed that the unoccupied hCAII binding pocket is mostly rigid with picosecond to nanosecond motion being mainly restricted to the loops connecting secondary structures and microsecond to millisecond motion primarily occurring for residues at the protein surface^[69]. However, solution NMR R_2 relaxation and R_{ex} rates revealed motion on the microsecond timescale for some residues of the active site loop around S196-E204, which was suggested to correlate with “open” and “closed” states of hCAII^[70]. In addition, R_{ex} rates for the residues around the H64 proton shuttle and G6 were elevated^[70].

In summary, the kinetics of the reaction and the dynamics of hCAII within the complex have been studied extensively as mentioned above, identifying, for example, H64 as proton shuttle, the water network in the active site, and different states of the active site loop of hCAII. To untangle the different contributions, inhibitors can be used to “freeze” proteins in a similar state as when substrate is bound. Drug discovery campaigns benefit from such knowledge on the protein binding site to design tightly binding lead compounds.

2.1.2 The binding site of hCAII and binding inhibitors

hCAII has one main binding site with a cone-like shape and ~ 15 Å width. It stretches from the hCAII surface to the tetrahedral coordination site of Zn^{2+} , which is complexed by H94, H96, and H119 with an affinity of ~ 4 pM^[68A, 71]. The three histidine residues are relatively rigid and stabilized by i.a. hydrogen bonding with T199, E117, Q92, and N243^[68a] (Figure 8). The fourth coordination site in the Zn^{2+} complex is occupied by water, hydroxide, ligands, or inhibitors. The CO_2 molecule is assumed to be guided to a position adjacent to the water or hydroxide molecule via Q135^[72]. In addition, there is a low-affinity binding site at the N-terminus involving residues 2-19^[73]. Besides catalyzing the hydration of CO_2 , it was also shown that hCAII has a weak esterase activity^[74]. Moreover, hCAII can also convert cyanamide to urea^[75]. Unfortunately, once urea is formed and bound at the catalytic center, it is not released and therefore depletes protein activity^[75]. The activity of hCAII is also decreased replacing the Zn^{2+} ion with a Co^{2+} ion^[76].

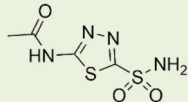
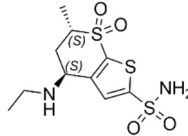
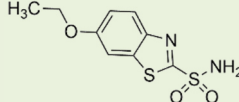
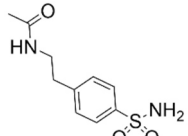
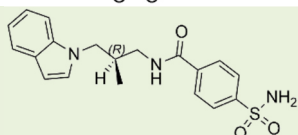
The wide binding site of hCAII provides a large space for interaction partners. The chemical space of potential ligands has been studied extensively, resulting so far in >100 ligands binding to hCAII^[77]. Interestingly, various modes of inhibition have been identified for hCAII inhibitors summarized in a review by Supuran, C. T., in 2016^[78]. These include other divalent ions like Cu^{2+} or Ni^{2+} displacing the Zn^{2+} ion, molecules like sulfonamides and monovalent anions coordinating Zn^{2+} , molecules like phenols and carboxylates binding to the water or the hydroxide ion, molecules like coumarins blocking the active site, and molecules interacting via a different binding site. For some inhibitors like secondary or tertiary sulfonamides or imatinib, the mechanism of inhibition is still unknown. The inhibitors binding to hCAII via Zn^{2+} coordination can be further divided into three groups: There are anions containing protonated ligand atoms like HSO_3^- and HS^- , and anions without such a protonated ligand atom^[79]. The latter includes anions which change the coordination site of Zn^{2+} like I^- , Br^- , or

N_3^- ^[80], or anions which do not distort the tetrahedral Zn^{2+} coordination site as they form hydrogen bonds with T198 like SCN^- or CH_3COO^- ^[79a, 81]. Anionic ligands with sulfonamide warheads represent inhibitors with some of the highest binding affinities to hCAII, like acetazolamide and dorzolamide.

In addition to sulfonamide warheads, the Klebe group reported that a central aromatic ring enabling CH- π -interactions with side chain of L197 was often found to be associated with high affinity to hCAII, which was supposed to resemble the intermediate structures of CO_2 hydration^[82]. The sulfonamide ligands with a benzene ring were also assumed to either be already charged upon binding or to be deprotonated after binding^[82]. At least one intermediate state during the binding process was hypothesized^[82-83]. In addition, k_{on} was reported to increase with longer alkyl chains attached to the benzene ring in para position^[82-83]. Acetazolamide, dorzolamide, ethoxzolamide, M25, and SBR are ligands fitting these criteria and have binding affinities in the nano- to picomolar range (Table 3). Out of these four ligands, dorzolamide and SBR bear the highest number of non-exchangeable protons, which is favourable for any proton-detected experiments. Interestingly, SBR shows a low electron density at the tail region in the X-ray structure suggesting heterogeneous conformations or dynamics while bound to hCAII^[82, 84]. Notably, acetazolamide and M25 occupy also a second binding site at the N-terminus of hCAII.

Especially for treatments with sulfonamides, which was the first group of hCA-targeting drugs, side effects often occurred due to a lack of specificity towards the different isozymes of the CAs^[85]. For CAs in the extracellular space, positively charged inhibitors improved the selectivity for these CAs^[86]. T199 was suggested to play a critical role in distinguishing ligands selecting hCAI or hCAII^[58]. To prevent interactions with hCAII, inhibitors sterically clashing with A65 have been designed as the occurrence of alanine at this position is exclusive for hCAII^[87]. For inhibitors selective for CA IX and XII, prodrugs were developed which are only activated in hypoxic tumor tissue^[88].

Table 3: Structures, affinities, and the number of non-exchanging protons (non-ex. protons) of several sulfonamide inhibitors binding to hCAII.

Ligand structure	Ligand abbreviation/PDB ligand name and IUPAC name	K_D / nM	PDB entries	Non-ex. protons
	Acetazolamide <i>N</i> -(5-sulfamoyl-1,3,4-thiadiazol-2-yl)acetamide	15 ^[89]	3HS4	3
	Dorzolamide (4 <i>S</i> ,6 <i>S</i>)-4-(ethylamino)-6-methyl-7,7-dioxo-5,6-dihydro-4 <i>H</i> -thieno[2,3- <i>b</i>]thiopyran-2-sulfonamide	$K_i = 9$ ^[90] $K_D = 43$ ^[91]	4M2U	13
	Ethoxzolamide 6-ethoxy-1,3-benzothiazole-2-sulfonamide	10 ^[89]	3CAJ	8
	M25 <i>N</i> -[2-(4-sulfamoylphenyl)ethyl]acetamide	900 ^[73]	2NNS	11
	SBR <i>N</i> -[(2 <i>R</i>)-3-indol-1-yl-2-methylpropyl]-4-sulfamoylbenzamide	0.03 ^[16]	1IF7, 6SDL	18

Although hCAII and its ligands have been characterized thoroughly, there are only few pieces of information available on the dynamics of the hCAII-inhibitor complexes. Mostly, hCAII complexes were assumed to be structurally rigid^[92]. One of the reasons is that classical tools in structural biology like X-ray crystallography and cryo-EM are limited in revealing dynamics, as discussed in chapter 1.1.3. However, ligand dynamics could be relevant for drug development, biotechnological applications for CO₂-fixation^[93], and directed evolution for hCAII and other biocatalysts to accept a wider range of substrates^[94].

2.2 Results and discussion

Parts of this chapter have been published in^[95]: J. Kotschy, B. Söldner, H. Singh, S. K. Vasa, R. Linser, *Angew Chem Int Ed* **2023**, e202313947.

Testing where and how specific small molecules bind to proteins can contribute to drug development. X-ray crystallography is one of the most popular techniques to study such protein-ligand complexes. It can uncover, for example, the interaction surface by resolving the ligand location, the occupancy, and the conformational heterogeneity of protein-ligand complexes. The classical method using one protein crystal is limited, however, in its capability to distinguish static heterogeneity and dynamics using B factors. At most, it can resolve, for example, two distinct amino acid sidechain conformations when two states are roughly equally populated. A single crystal structure also cannot inform on any time regime of dynamics. NMR spectroscopy is a valuable tool to identify protein or ligand conformations and dynamics at certain timescales in a qualitative or quantitative manner. Therefore, X-ray crystallography and NMR spectroscopy can complement each other. X-ray crystallography can serve for quick structure determination (chapter 2.2.1) while NMR spectroscopy can sample the dynamics of moieties, distinguishing motion from static disorder (chapters 2.2.2-2.2.4).

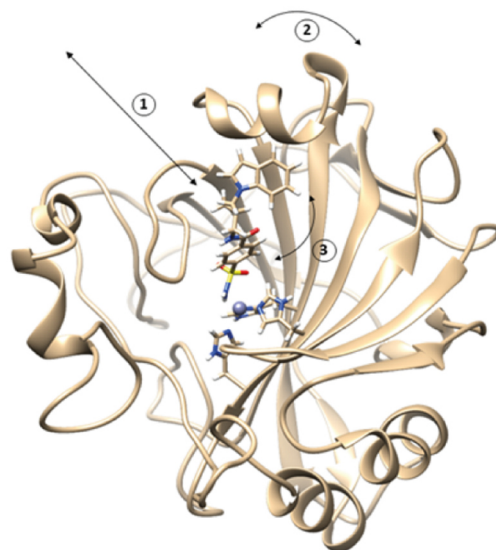


Figure 10: X-ray structure of the hCAII-SBR complex determined here (chapter 2.2.1) with arrows indicating association and dissociation of the complex (1), hCAII dynamics (2), and ligand dynamics (3).

Here, the interactions and dynamics of the enzyme-ligand complex formed by hCAII and the ligand SBR with a 30 μM affinity was studied (Figure 10). Despite the extremely high affinity, the X-ray structures showed that the ligand SBR seemed to be either dynamic or to adapt different conformations with the lowest B factors at the Zn^{2+} binding site and the highest B

factors at the indole ring. In addition, Steffen Glöckner from the Klebe group reported unusual isothermal calorimetry curves in his dissertation, which led to the suggestion that several binding events occur in the same binding pocket. Here, X-ray crystallography, and solution and solid-state NMR spectroscopy were used to probe ligand dynamics.

To detect ligand dynamics within the hCAII-SBR complex via NMR, several critical sample requirements needed to be met. First, the nucleus reporting on dynamics had to be chosen. Labelling ligands isotopically is often not feasible due to both high costs and arduous synthesis based on the available, isotopically labelled precursors. Thus, proton-detected relaxation experiments were selected. The NMR-active hydrogen atoms have nearly 100% natural abundance (Table 2) but come with challenges related to the ligand-selective NMR relaxation experiments. Consequently, a ligand with many non-exchanging protons was chosen, which is called SBR. However, not only the ligand SBR carries protons in the sample. It is outnumbered by the protons of hCAII by a factor of roughly 100. Protein deuteration can limit the number of protons of proteins in a sample, which was applied here yielding a perdeuterated hCAII sample (chapter 1.1.6). In addition, hCAII was ^{13}C -, ^{15}N -labelled so any remaining protons, namely H^{N} and H^{C} , were covalently bound to these nuclei and thus could be distinguished from protons bound to ^{12}C and ^{14}N atoms, which were exclusively present in the ligand structure. Being able to distinguish protein and ligand protons allowed to cancel the magnetization transfer pathways belonging to these undesired protein protons. To further reduce the background of undesired proton signals, deuterated buffer components were used and, when possible, dissolved in D_2O or $\text{d}_6\text{-DMSO}$. Residual solvent signal was removed by phase cycling in the chosen pulse sequence. Finally, pulse sequences both for solution and solid-state NMR needed to be designed to sample ligand dynamics at the expected time scale (chapter 2.2.3). Catalysis and conformational changes often take place on the microsecond to second timescale thus relaxation experiments focussing on dynamics in the microsecond regime were chosen. Also, the hCAII conversion rate of max. $10^6/\text{s}$ indicates that catalysis-related processes of the protein likely take place on the fast microsecond timescale^[49]. However, it should be noted that ligand dynamics can take place irrespective of protein dynamics and might not be interlinked.

Most importantly, the hCAII occupancy with SBR would need to be maximized and the amount of free SBR correspondingly minimized. In solution, the K_{D} value allows to estimate the ideal ligand-protein ratio. If not chosen properly, free SBR might dominate the overall ligand signal

as the faster tumbling speed of free ligand compared to bound ligand comes with slower T_2 relaxation. Overall, the sample preparation for solid-state required more optimization, because the success of growing protein crystals and the introduction of the ligand were critical steps.

The hCAII-SBR sample optimization for solids (chapter 2.2.1), the interaction surface between hCAII and ligands (chapter 2.2.2), the SBR proton dynamics in solution and solids (chapter 2.2.3), and hCAII dynamics in the complex (chapter 2.2.4) will be addressed in the following chapters. Besides existing hCAII amide assignments (chapter 2.2.2), new methyl sidechain assignments (chapter 2.2.5) were obtained, representing additional reporters for future hCAII studies.

2.2.1 Screening high-occupancy conditions for hCAII-SBR crystals via X-ray crystallography

To detect SBR ligand protons while bound to hCAII crystals, the overall protein yield after expression and purification, the protein crystal yield, the ligand occupancy in the complex, and protein and solvent deuteration levels needed to be optimized. These challenges were tackled consecutively starting with the expression and purification of hCAII.

First, the protein expression with different *E. coli* cell lines and protocols was tested to maximize the yield of D-, ^{13}C -, ^{15}N -labelled hCAII. The main bottleneck was adapting the transformed cells to grow in deuterated media. Two protocols and two cell strains, namely BL21 Codon Plus (DE3)-RIL (purchased from Agilent) and Rosetta 2 (DE3) cells (purchased from Novagen), were used. While both cell strains are suited for T7 expression vectors, enabling IPTG-induced expression, the Rosetta 2 (DE3) cells provide additional tRNAs matching codons of eukaryotic genes which rarely occur in *E. coli*, including codons CCC for proline and GGA for glycine. In total, these codons occur fifteen times in the hCAII sequence and its N-terminally attached Glutathione S-transferase-(GST)-tag sequence. Providing more tRNA molecules for translating the required codons of the GST-hCAII construct likely sped up the production of hCAII. Looking at Figure 11 panel A, this can explain why higher expression yields were obtained for hCAII in Rosetta 2 (DE3) cells.

Next, two different expression protocols were assessed. Each included several incubation steps in complex cell culture media, successively increasing the D_2O content (Figure 51). After transformation, all large colonies were combined and used for the D_2O adaptation protocol. This seemed to decrease the risk of arrested growth during the adaptation procedure. The

most critical step was the final transfer to the medium dissolved in 100% D₂O. Here, the two protocols differed in the cell densities chosen for the last adaptation step. Either the 100% D₂O culture was inoculated with a cell suspension aliquot resulting in an OD₆₀₀ of 0.1 or with the whole cell pellet from the 90% culture. In panel A in Figure 11, the lysates of expressions using either protocol are shown after separating the protein with an SDS-PAGE. Transferring the cells at an OD₆₀₀ of 0.1 resulted in higher overall hCAII yields.

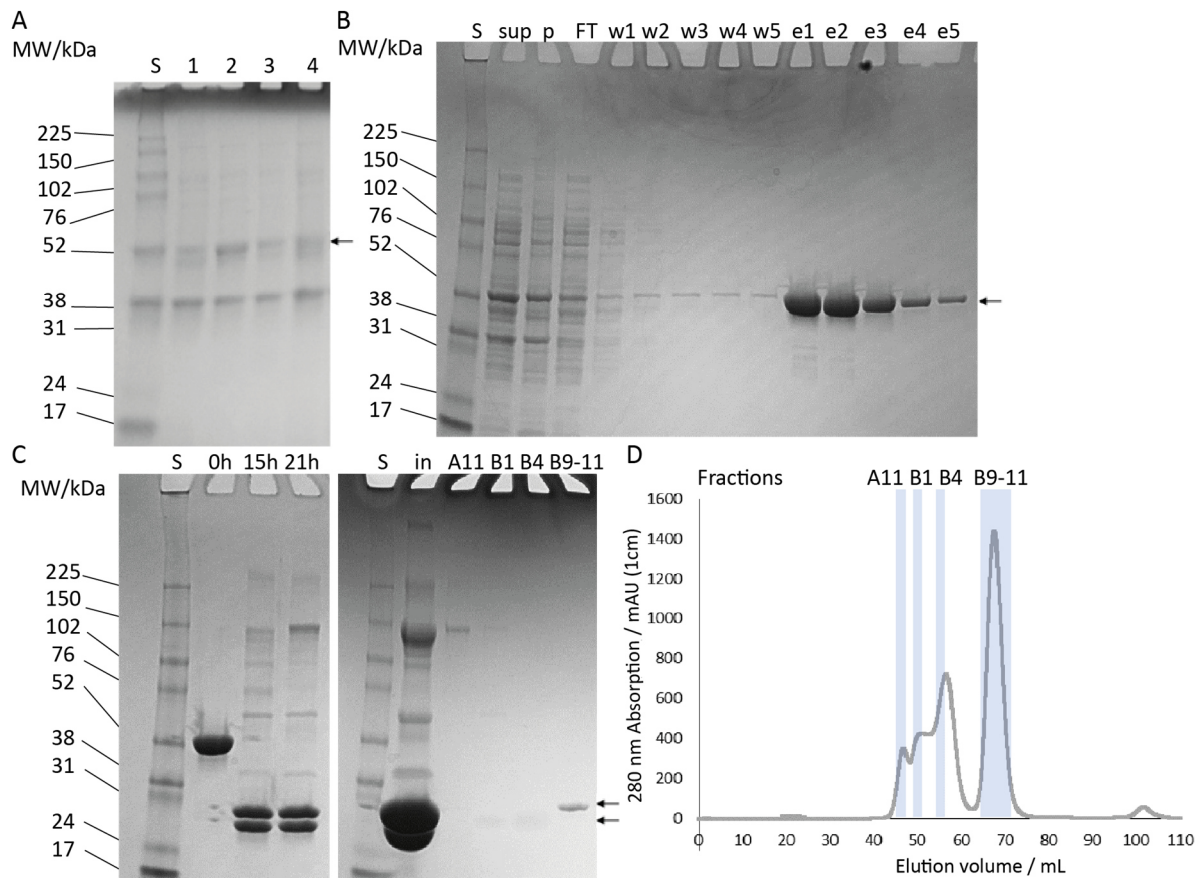


Figure 11: Expression and purification of hCAII. The GST-hCAII construct has a size of 55.9 kDa, which results in a 29.8 kDa construct after cleavage. S: Amersham ECL Rainbow Marker. **A.** SDS-PAGE with hCAII lysates after expression in Rosetta 2 (DE3) and BL21 Codon Plus (DE3)-RIL cells using either the whole cell pellet of the 90% adaptation culture to inoculate the final 100% D₂O culture (1, 3) or inoculating the cell culture at OD₆₀₀ 0.1 every time (2, 4), respectively. The arrow identifies the band corresponding to hCAII. **B, C.** SDS-PAGE of purification samples with D-, ¹³C-, ¹⁵N-labelled hCAII (2 L expression with Rosetta 2 (DE3) cells). Sup: supernatant after 2nd centrifugation; p: pellet after 2nd centrifugation; FT: flow through, w1-5: wash fractions; e1-5: elution fractions; 0h, 15h, 21h: incubation time for thrombin cleavage at room temperature; SEC: Size Exclusion Chromatography; in: input; A11, B1, B4, B9-11: SEC fractions as shown in panel D. The arrows identify the bands corresponding to the thrombin cleavage products. **D.** SEC chromatogram corresponding to panel C, displaying the absorption of 280 nm UV-light of different fractions (milli-absorption units (mAU) at 1 cm sample diameter).

The purification protocol had been established before^[69] and was only slightly optimized, for example, by adding the thrombin protease powder directly to the sample to cleave off the GST-tag. The activity of hCAII after expression and purification could be confirmed with an activity assay but was not further quantified (Carbonic Anhydrase Activity Assay Kit, data not

shown). Overall, the hCAII expression protocol with Rosetta 2 cells and subsequent purification yielded up to 15 mg protein per litre expression medium, which was either unlabelled or ^{13}C -, ^{15}N -labelled M9 medium. Upon deuteration, yields dropped to ~ 2 mg/L but could be boosted to reach 5 mg/L upon addition of isotopically labelled amino acids, for example, used for methyl labelling. Evidently, expression yields significantly decreased in deuterated media, while ^{13}C -, ^{15}N -labelling had no measurable effect compared to LB media. Concerning the crystallization of hCAII, several conditions were screened to improve ligand occupancy while maintaining crystal integrity and homogeneity. Previously, a manual protocol had been implemented for the preparation of solid-state NMR samples^[69]. The crystallization conditions were optimized in a partly automated manner using pipetting robots to save time while concomitantly increasing robustness and improving reproducibility. As a read-out system for screening different crystallization conditions, solid-state NMR-related rotor filling and measurements take time and effort, thus X-ray crystallography was chosen. Many X-ray structures and crystallization protocols of hCAII-ligand complexes, including the hCAII-SBR complex, are available^[16, 69, 82]. While a few large crystals of ideally cubic shape are required for X-ray crystallography, many small crystals are needed for a single solid-state NMR sample as rotors with diameters of 0.7-1.3 mm need to be filled with crystals. Starting out with the published setups and conditions, the crystallization of 10 mg/mL unlabelled apo-hCAII was most successful at 2.6-3.0 M $(\text{NH}_4)_2\text{SO}_4$, pH 7.0-8.0, and 4 °C in sitting-drop plates (Figure 12). Fewer crystals were obtained at pH values between 6.0 and 7.0, incubation at 20 °C, and lower protein concentrations. To harvest the crystals and to decrease crystal size, a pipette tip was used to stir the suspension and to gently detach the crystals from the well surface. The crystallization could be scaled up to a 24-well plate which did not interfere with crystal formation along with isotopic labelling of hCAII (Figure 12, panels A-D). For D-, ^{13}C -, ^{15}N -labelled hCAII, the overall yields and the crystal size decreased. Upon addition of methyl labelled amino acids during expression, the crystal shape changed, and spiky ball-shaped crystals were obtained (Figure 12, panels E-F). While such crystals would not be ideal for X-ray diffraction, they can be used for solid-state NMR experiments. Furthermore, it was tested whether apo-hCAII can also be crystallized in deuterated buffer, further reducing the number of protons in the sample. For an ideal peak resolution and experimental sensitivity in the solid state, only ligand protons would be present in the sample.

Three different strategies were employed: The protein hCAII was directly crystallized in D₂O-based precipitant buffers, seeds of labelled hCAII crystals were added to boost hCAII crystallization in D₂O, and the already grown hCAII crystals were incubated in D₂O. Unfortunately, no hCAII crystals grew in D₂O-based precipitant with and without seeds over the period of two months. The crystal seeds dissolved after adding them to the crystallization buffers. The seeding was performed according to the instructions provided for the Seed Bead Kit diluting the seed stock in the hCAII-precipitant mixture by 1:1, 1:10, or 1:100. Only after roughly one year, very few small crystals were observed in the crystallization batch to which seeds have been added. These were not sufficient for a solid-state sample. For the third approach, the prepared hCAII crystals were transferred to a reaction tube, most of the supernatant was removed, and the supernatant was lyophilized overnight. The obtained powder was then dissolved in D₂O and added to the crystal suspension. Incubating the mixture overnight at 4 °C, visual inspection revealed that the number of crystals declined drastically. Under a brightfield microscope, it could be confirmed that only few crystals were left and most of the particles in the suspension corresponded to precipitant. However, it was not tested whether the crystals also dissolved when the powder obtained by lyophilization was dissolved in water. Nevertheless, the remaining crystals were filled into a solid-state NMR rotor. The resulting ¹H-¹⁵N spectrum unfortunately had insufficient quality, likely due to high sample heterogeneity. It can be concluded that deuteration either impaired protein folding or drastically slowed down the crystallization process which can be attributed to solvent isotope effects^[96].

The crystalline hCAII-SBR complex was prepared via two different approaches: Soaking the ligand SBR in prepared apo-hCAII crystals and co-crystallizing hCAII in presence of SBR (Figure 52). It should be noted that most published complex structures of hCAII have been determined via soaking^[82]. For the soaking procedure, the hCAII crystals were transferred to a reaction tube and ligand was added to the supernatant of the crystal suspension. The ligand was added as a powder or from a 100 mM ligand solution in d₆-DMSO in a stepwise fashion, ultimately reaching a ~1:2 protein-ligand ratio. Ideally, the ligand concentrations were chosen below ~200 μM, which is the solubility limit of SBR in aqueous solutions, to avoid ligand precipitation. In between two titration steps, the suspension was mixed thoroughly and incubated on ice for 10 min. The crystals withstood these conditions, and no destabilization of crystals or increased precipitation was visible via brightfield microscopy. Moreover, soaking

was also tried adding the ligand powder directly without DMSO. For co-crystallization, the desired amount of ligand was added from a 100 mM d_6 -DMSO stock to the final protein-precipitant solution when setting up the crystallization with a protein-ligand ratio of 1:2. The incubation time needed for crystal growth and the crystal shape were not hampered noticeably by adding the ligand. Visual inspection indicated that the overall yield of crystals was slightly reduced. Subsequently, one crystal per condition was picked for X-ray diffraction measurements to determine the ligand occupancy.

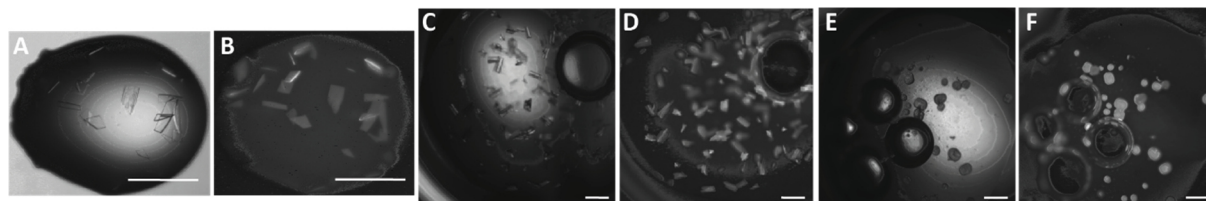


Figure 12: Brightfield and UV absorption images of apo-hCAII crystals taken with a JANSi UVEXm microscope. All scales correspond to a length of 150 μm . Unlabelled hCAII crystals in a 96-well (A, B), D-, ^{13}C -, ^{15}N -labelled hCAII crystals in a 24-well plate (C, D), and D-, ^{13}C -, ^{15}N - and ILV-methyl labelled hCAII crystals in a 24-well plate (E, F) visualized via brightfield and UV absorption, respectively.

In panels A and B of Figure 13, the active site of the calculated hCAII-SBR complex X-ray structures and the corresponding electron densities are displayed. The soaked and co-crystallized structures showed a high overall similarity. For the co-crystallized structure (PDB: 8R1I), a resolution of 1.46 \AA was achieved with well-defined electron densities for the ligand SBR (Figure 13). A slightly lower resolution of 1.48 \AA was obtained for the soaked sample (Table 23). With $R_{\text{free}}=0.19$, 0.39% Ramachandran outliers, and a low clash score, the parameters for the co-crystallized structure lay within the recommended limits (Figure 13, panels C and D). The soaked structure performed slightly worse with $R_{\text{free}}=0.20$, 0.00% Ramachandran outliers as well as a two-fold higher clash score than the co-crystallized structure. It should be noted that Ramachandran outliers are more likely to occur for less well-defined electron densities due to unfolded and/or flexible structures, which is partly the case for the N-terminus. In contrast, no electron density for the ligand was observed when soaking was performed by directly adding ligand powder. Most likely, the ligand powder was not sufficiently dissolved in the precipitant buffer leaving the hCAII crystals in apo state. Preparing the ligand in a DMSO stock first was essential to dissolve the ligand. Overall, the co-crystallized structure indicated that co-crystallization is better suited for ligand proton-relaxation studies via solid-state NMR than soaking due to the higher ligand occupancy.

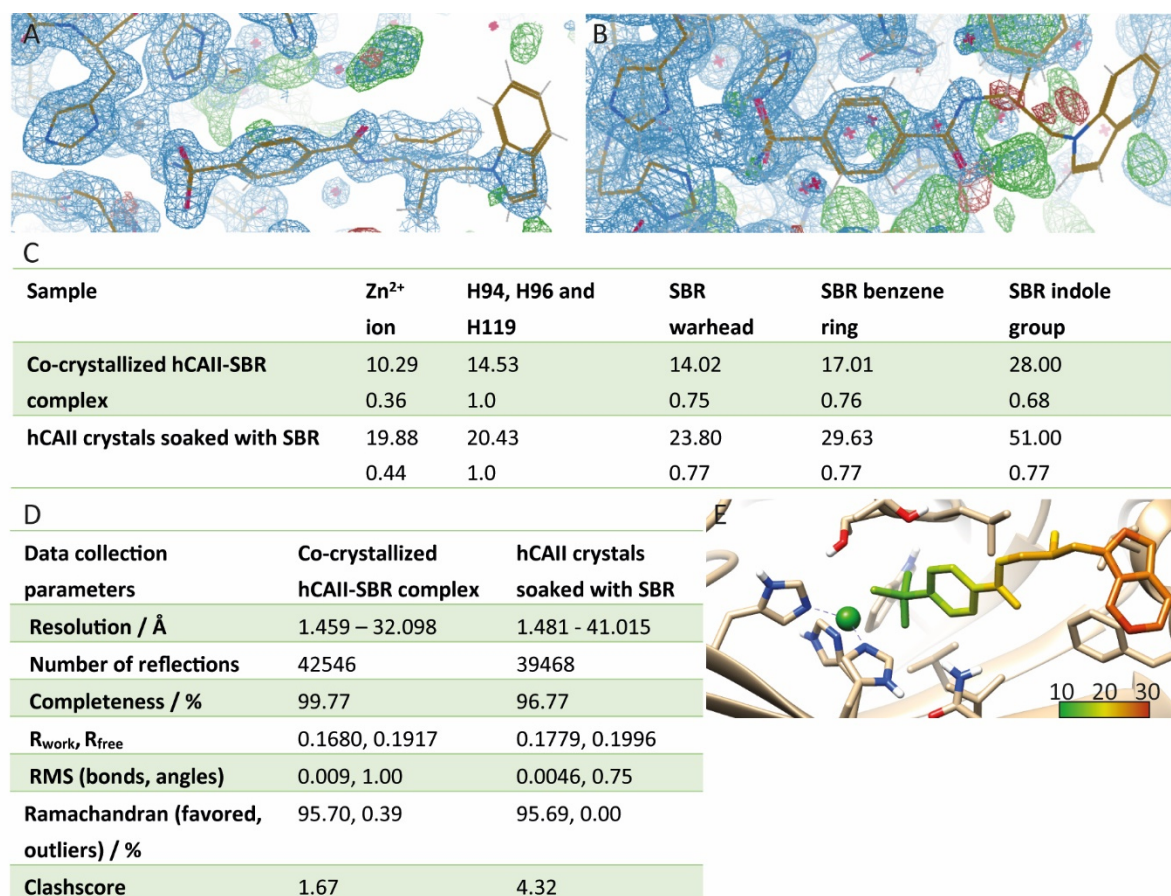


Figure 13: X-ray diffraction data. **A, B.** Processed X-ray diffraction data displaying the refined hCAII-SBR complex structures and corresponding electron density maps. The co-crystallized hCAII-SBR sample refers to the structure shown in panel **A**, while the sample with hCAII crystals soaked with SBR from a 100 mM ligand solution in *d*₆-DMSO is represented by panel **B**. **C-E.** Processing parameters and B factors derived from the X-ray structures shown in panels **A** and **B**. In panel **C**, the (averaged) B factors are shown above, and the occupancy given by the software Phenix^[97] is shown below in each table cell. For histidine residues H94, H96, and H119, which complex the Zn²⁺ ion, the B factors of both side chain nitrogen atoms were averaged. The averaged B factors of the SBR warhead were derived from the two oxygen and the nitrogen atoms of the sulfonamide group. For the SBR indole group, the B factor belongs to the nitrogen atom. **D.** Summary of the main X-ray diffraction parameters for the hCAII-SBR crystals achieved via co-crystallization or soaking. The data sets were recorded at 100 K (Table 23). **E.** B factors plotted on the SBR structure and Zn²⁺ within the co-crystallized hCAII-SBR complex.

The calculated structures were compared to the two X-ray structures of the hCAII-SBR complex published by Glöckner, S., *et al.* and Grzybowski, B. A., *et al.* with each 1.08 and 1.98 Å resolution, respectively (Figure 14, panel A). Evidently, the overall fold of hCAII is very similar. The electron densities for the ligand, however, matched the ligand structure to a lesser extent than in the structure calculated here. The structure from Grzybowski, B. A., *et al.* showed bulky, less defined electron density clouds for SBR in the active site of hCAII (taken from a figure in the paper, no electron density map submitted to the PDB database). In the structure by Glöckner, S., *et al.*, the local electron density was too low to accommodate the indole group of the ligand. Both structures calculated and presented here, soaking the crystals

with SBR from a DMSO stock solution and co-crystallizing the hCAII-SBR complex, ameliorated the details on where and how the ligand is placed in the complex structure with well-defined electron densities matching the ligand structure (Figure 13).

The electron densities represented by B factors were used to compare and estimate ligand occupancies for the calculated X-ray structures (Figure 13, panels C-E). B factors serve as qualitative measure for atomic displacement in X-ray structures with up to 15% accuracy^[98]. More rigid structures are represented by low B factors while high B factors hint towards high flexibility, resulting from e.g. vibration and static disorder^[98a]. Such parameters, in addition to electron densities in cryo-EM structures, have been used as indicators for flexibility to subsequently study these putative motions via NMR spectroscopy^[37a]. Different conformations per residue can manifest in several electron densities due to long lifetimes with moderate B factors, also called static disorder. More heterogeneous structures with fast-exchanging conformations can result in higher B factors, while randomized fast motion can lead to “invisible” structures with dramatically increased B factors. The stability of the conformations, tantamount to the exchange rates, thus strongly affect the electron densities. Here, the software Coot^[99] was used for X-ray structure refinement and determination of the relative occupancies of hCAII with Zn²⁺ and SBR based on the provided electron densities. The occupancies of SBR and Zn²⁺ were determined to be ~75% and 36% for the co-crystallized structure (Figure 13, panel C). The B factors in the co-crystallized structure for the indole ring of SBR are roughly twice as large as the B factors of the warhead of SBR, Zn²⁺, and exemplarily chosen hCAII atoms. In contrast, the soaked sample shows nearly twice as large B factors for both the warhead and the indole ring compared to the co-crystallized structure. The increasing B factors of ligand relative to the protein atoms reflect lower occupancy of hCAII with the ligand SBR.

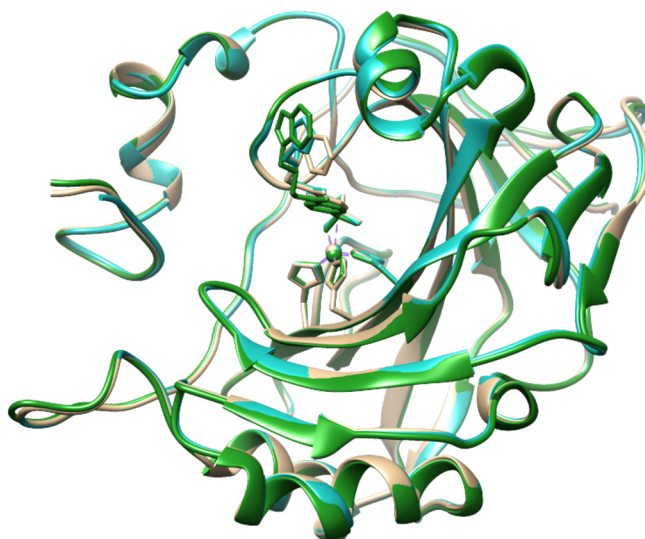


Figure 14: Overlay of hCAII-SBR complex structures. The co-crystallized X-ray structure is coloured in beige (PDB: 8R1I), the 1F7^[16] PDB structure is depicted in green, and the 6SDL PDB structure is shown in turquoise. The sidechains of H94, H96, and H119 are shown as well. No significant differences in the folding of hCAII can be seen. The positions of the Zn²⁺ ion, the ligand SBR, and some loops differ slightly.

These data suggest an overall higher ligand occupancy in the co-crystallized sample. Therefore, a lower performance of the soaking procedure in terms of homogeneous ligand occupancy can be concluded. Both B values of the co-crystallized and soaked structures indicate static disorder or flexibility for the indole group compared to the warhead group. These findings indicate that the ligand SBR adapts various conformations while bound to hCAII which cannot be explained by two-site exchange on the slow-exchange timescale.

2.2.2 Chemical shift perturbations for hCAII induced by ligands in solution and solids

For ligand titration studies, amide chemical shifts are well-suited to report on local changes of the chemical environment of the protein backbone upon ligand binding. The assignments of hCAII amide peaks were readily available^[1] and further improved by adding amide sidechain assignments from literature^[100]. The hCAII amide peak assignments were transferred using 2D ¹H-¹⁵N and 3D HNCA spectra, which were useful especially for crowded peak regions in solid-state spectra. Assignments and spectra in solution and solid-state were handled independently, although most chemical shifts are similar (Figure 15). The hCAII construct consisted of 260 amino acids plus five amino acids, which remain after cleaving off the GST-tag with thrombin. These five residues were not assigned. In solution, 87% of the backbone residues were assigned successfully (Table 30). In addition, 22 out of 28 amide sidechain assignments of asparagine, glutamine or tryptophane could be transferred^[100]. In

the solid state, the backbone amide assignments covered 90% of the sequence plus six out of seven tryptophane sidechain assignments (Table 30).

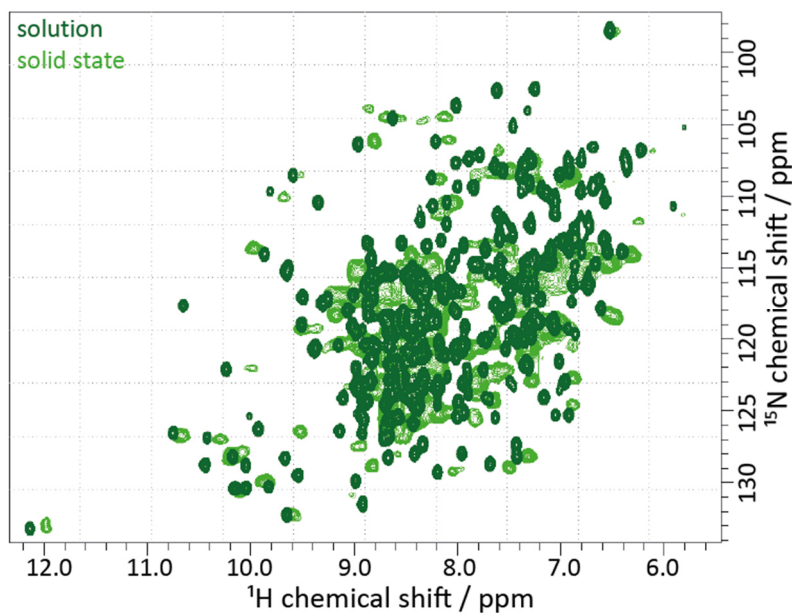


Figure 15: Overlay of 2D ^1H - ^{15}N spectra comparing apo-hCAII in solution (dark green) and the solid state (light green). The $720\ \mu\text{M}$ ^{13}C -, ^{15}N -labelled hCAII sample in solution was kept in 80 mM phosphate and 20 mM NaCl at pH 7.4 with 10% D_2O . The spectrum was recorded at 37 °C on an 800 MHz spectrometer for 20 min. The crystalline solid-state sample was prepared in 100 mM TRIS buffer, $\sim 2.8\ \text{M}$ $(\text{NH}_4)_2\text{SO}_4$ at pH ~ 7.9 and filled into a 1.3 mm rotor. The spectrum was recorded at $\sim 25\ ^\circ\text{C}$ and 55.56 kHz MAS on a 700 MHz spectrometer for two hours.

Both solution and solid-state NMR experiments are viable tools to study ligand interactions, protein structures, and dynamics with several advantages and disadvantages each. One of the challenges for solid-state NMR experiments are broader peaks, especially for protons, which increase the chances for peaks to overlap. Exemplarily, the linewidths of T168 in a ^1H - ^{15}N 2D spectrum of apo-hCAII were determined in solution and solids. It was shown that in the ^{15}N dimension, the peak widths with 0.069 ppm in solution and 0.073 ppm in solids at 55.56 kHz MAS were similar. In the ^1H dimension instead, 0.047 ppm in solution and 0.13 ppm in solids at 55 kHz MAS revealed strong line broadening in the solid state. This is not surprising as one of the largest differences between solids and solution can be attributed to increased proton dipolar couplings in solids, as mentioned in the introduction (chapter 1.1.2).

With the amide assignments readily available, chemical shift changes of hCAII residues upon ligand binding directly characterize where and how interaction partners bind in solution and in the crystal. Two ligands were investigated, namely SBR and M25. First, the proton signals of both ligands were assigned mainly using HMBC, HSQC, and TOCSY spectra. The SBR stock solutions were prepared in DMSO as its solubility limit in water is low with $\sim 200\ \mu\text{M}$. The

proton chemical shifts of SBR in DMSO compared to D₂O slightly changed (compare Figure 54 and Figure 55). For M25, all stocks and thus NMR spectra were prepared and acquired in D₂O (compare Figure 56).

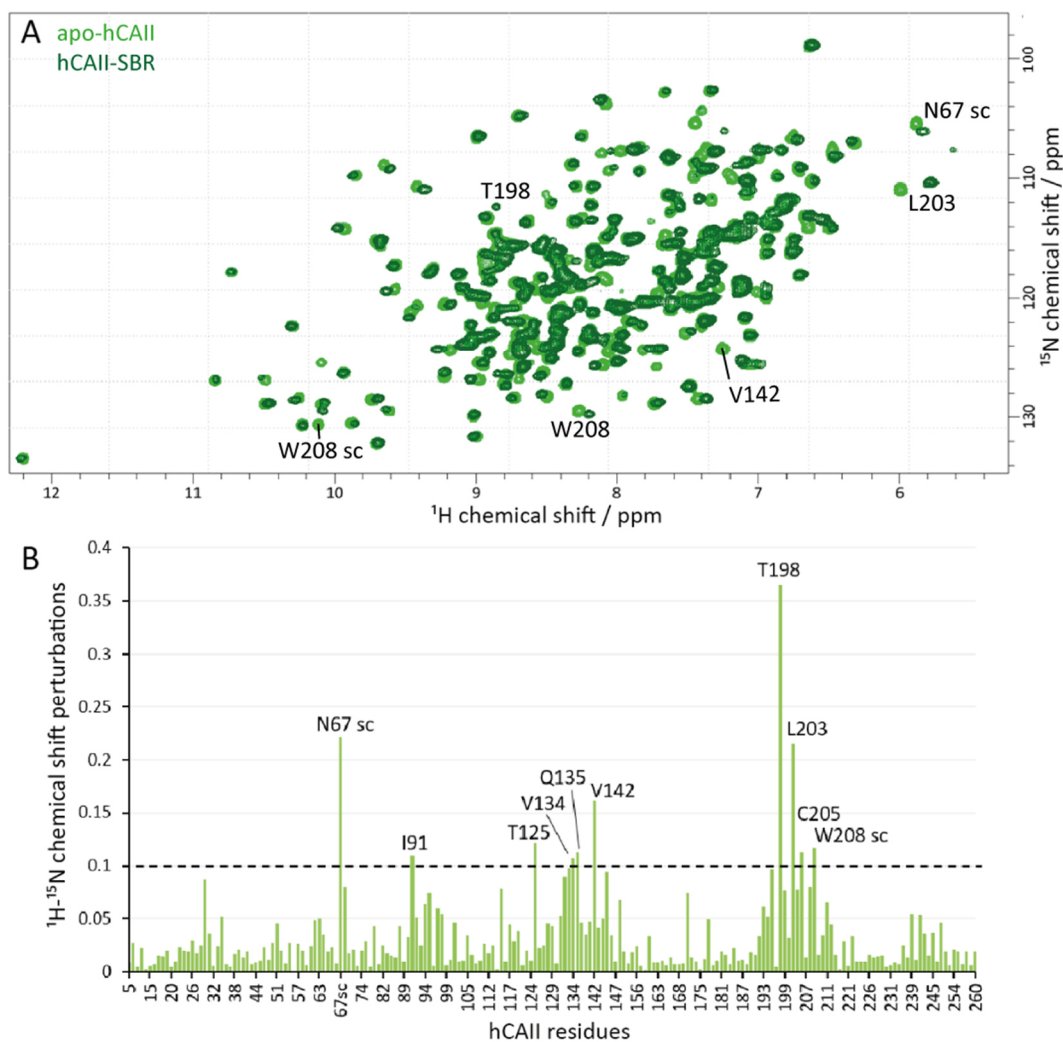


Figure 16: hCAII-SBR titration comparing apo-hCAII and the hCAII-SBR complex at 1:1 molar ratio in 80 mM phosphate and 20 mM NaCl at pH 7.4 with 10% D₂O. The spectra were recorded in solution at 37 °C on an 800 MHz spectrometer for each ~40 min. **A.** Overlay of two 2D ¹H-¹⁵N HSQC spectra of apo-hCAII (light green) and hCAII-SBR (dark green). **B.** CSP plot of H^N and ¹⁵N peaks. Sc: amino acid sidechain.

Next, a series of ligand titration steps was performed to uncover the exchange rates of ligand binding and the chemical shift (δ) perturbations (CSPs) between free and bound protein via solution NMR experiments (Figure 16). In solids, spectra of apo-hCAII and hCAII-SBR crystals were compared (Figure 17). Both in solution and solids, the nearest-neighbour-approach was utilized to assign peaks after adding ligand. To quantify the CSPs, the differences between apo-hCAII and hCAII-complex states were calculated using equation (11).

$$\text{CSP} = \sqrt{(\delta_{\text{H,apo}} - \delta_{\text{H,complex}})^2 + \frac{1}{100} * (\delta_{\text{N,apo}} - \delta_{\text{N,complex}})^2} \quad (11)$$

The solution NMR titration with SBR revealed that SBR binds to the catalytic pocket of hCAII, as expected, with slow exchange (Figure 16). Several clusters of residues with strong perturbations were observed (secondary structures and sequence are correlated in Figure 8). All residues were localized adjacent or at the catalytic site of hCAII, namely around amino acids I91 ($\beta 5$), T125 to V142 (located at $\alpha 3$, $\alpha 4$ and $\beta 7$), and T198 to W208 (located at $\beta 10$ and the prior loop). The sidechain of N67 ($\beta 3$) also showed perturbations but was not part of a cluster. Both the exchange regime and the clusters of amino acids with CSPs were similar to the titrations with other ligands, e.g. acetazolamide^[40a] and dorzolamide^[70].

In the solid state, the chemical shifts of amino acids Y7, G102, E106, G170, F175, N177, W208, K227, and D242 differed between apo-hCAII and hCAII crystals soaked with SBR (Figure 17). Compared to the CSPs in solution, the CSPs found for G102 and E106 correspond to the same cluster as I91 in solution. W208 showed CSPs both in solution and solids as well. These residues confirmed changes close to the binding pocket. For the second cluster, CSPs were detected in the solid state but were overall lower than in solution. The amino acids G170, F175, N177, and K227 are placed at the surface of the hCAII molecule, relatively far away from the catalytic pocket. These residues likely indicated larger CSPs due to changes in the crystal-crystal contacts between hCAII monomers which might have changed when SBR (and d_6 -DMSO) was soaked into the crystals. Y7 and D242 are part of unfolded loops at the N- and C-termini, respectively, and closer to the catalytic site of hCAII. It can be rationalized that these regions are more flexible in solution and therefore might have only shown CSPs due to ligand binding once they were restricted by the crystal lattice in solids. Overall, the signal-to-noise ratio of the CSP plots is lower than in solution, likely due to contributions of crystal-crystal contacts and lower ligand occupancy.

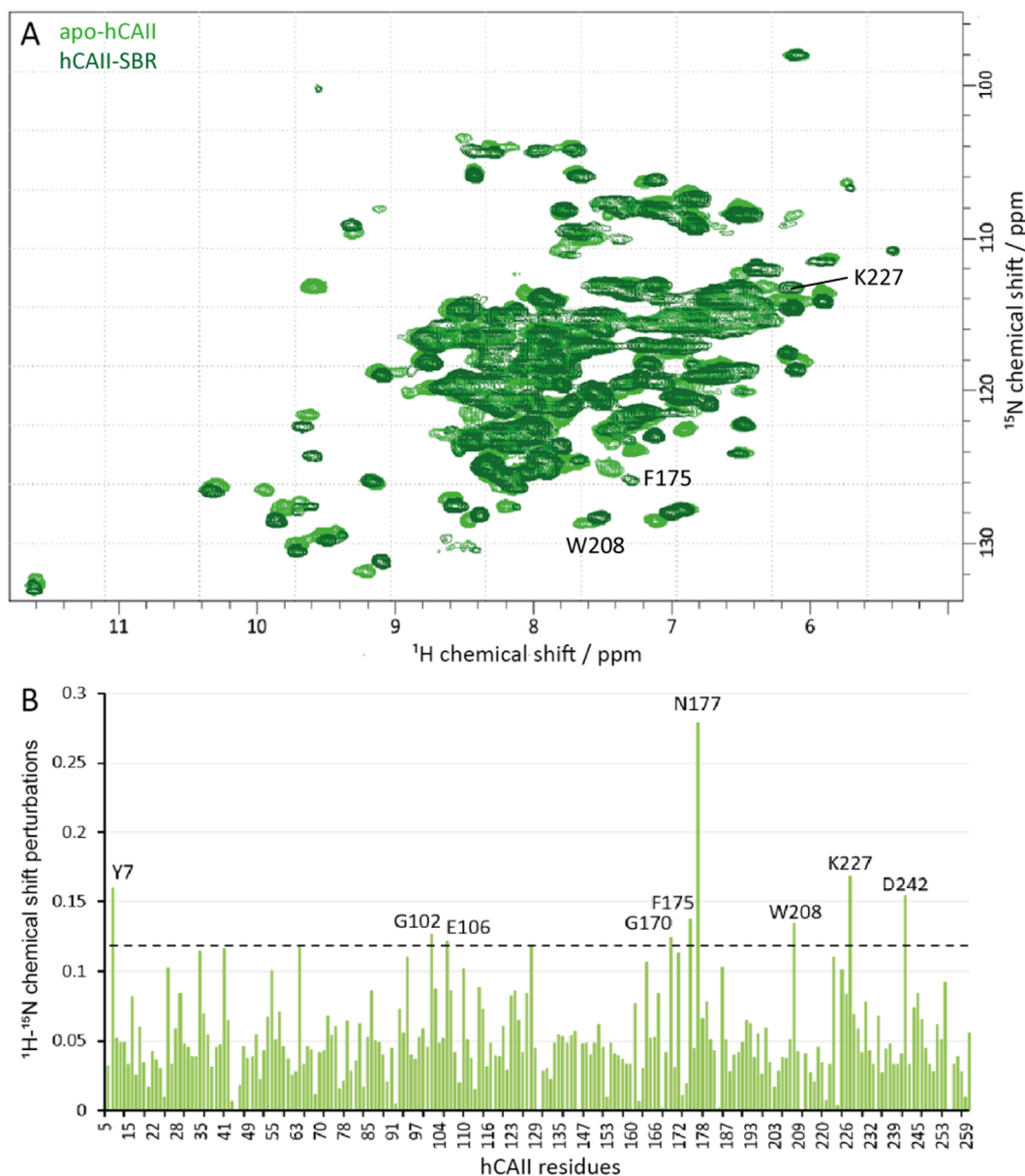


Figure 17: Comparing the solid-state NMR spectra of crystalline apo-hCAII and the hCAII-SBR complex in 100 mM (d_5 -)TRIS buffer, ~ 2.8 M $(\text{NH}_4)_2\text{SO}_4$ at pH ~ 7.9 , each filled into 1.3 mm rotors. The complex sample was prepared by soaking the crystals with ligand in 1:2 molar ratio. The spectra were recorded at ~ 25 $^\circ\text{C}$ and 55.56 kHz MAS on a 700 MHz spectrometer. Overlay of two 2D ^1H - ^{15}N spectra (A) and the corresponding CSP plot of $^1\text{H}^{\text{N}}$ and ^{15}N peaks (B) comparing apo-hCAII and the hCAII-SBR complex.

In addition, HNH-RFDR solid-state spectra were acquired for roughly one week on a 700 MHz spectrometer to calculate the structure of the hCAII-SBR complex. The sample was obtained by soaking the ligand into prepared hCAII crystals and filling those into a 1.3 mm rotor. The software Aria^[101] was supplied with an X-ray structure, the RFDR and HSQC peak lists, as well as assignments previously obtained for hCAII. Unfortunately, Aria was not able to add any assignments to the already provided assignments and found only ~ 177 distance restraints. The low number of restraints resulted from a lack of cross peaks, which then led to poor

convergence of the calculated structures. Neither these solid-state RFDR spectra, nor similar solution NOESY experiments revealed intermolecular cross peaks between hCAII and SBR. Long intermolecular distances, motion, and exchange processes on disadvantageous timescales might explain why no intermolecular cross peaks were observed.

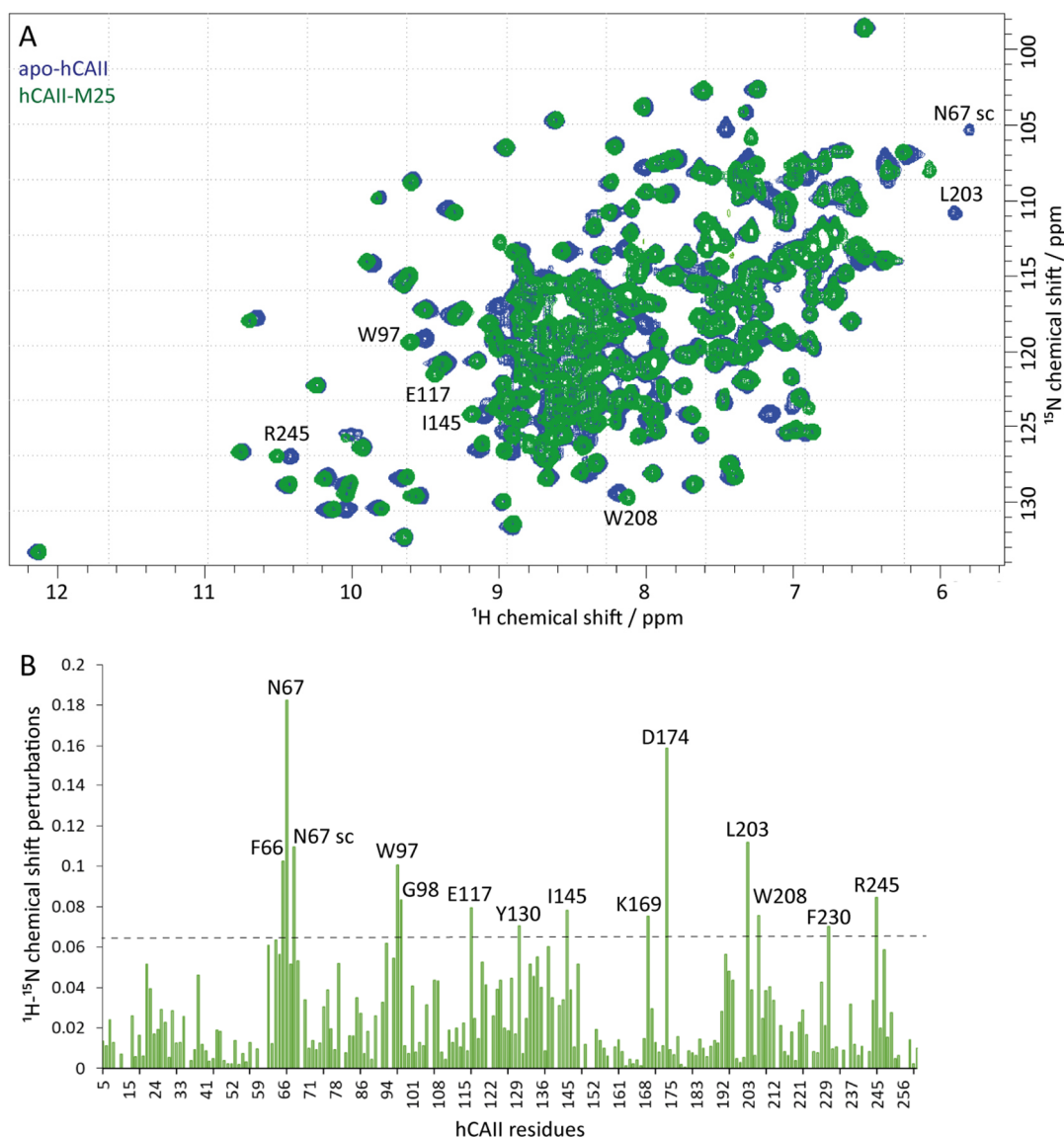


Figure 18: hCAII-M25 titration in solution with $720 \mu\text{M}$ ^{13}C -, ^{15}N -labelled hCAII with M25 in 1:1 molar ratio in 80 mM phosphate and 20 mM NaCl at pH 7.4 with 10% D_2O . **A.** Overlay of two 2D ^1H - ^{15}N HSQC spectra of apo-hCAII (blue) and hCAII-M25 (dark green) recorded at 37 °C on an 800 MHz spectrometer for 20 min. **B.** CSP plot of ^1H and ^{15}N peaks comparing apo-hCAII with the hCAII-M25 complex.

As alternative to the ligand SBR for ligand dynamics studies, the ligand M25 was titrated to hCAII in solution as well. Based on a known X-ray structure (PDB: 2NNS^[73]), two binding pockets were expected, one at the catalytic site of hCAII and the other one interacting with the N-terminus of hCAII (amino acids ~1-20). CSPs of residues W97, G98, E117, F130, I145, L203, and W208 roughly match the clusters observed for ligand SBR in solution and solids,

which consisted of I91-E106, T125-V142 and T198-W208. In addition, the peaks corresponding to N66, N67, and the N67 sidechain show perturbed chemical shifts in the presence of M25, like the changes for SBR in solution. The CSPs introduced by M25 seem to be more pronounced than for SBR. In contrast, CSPs by M25 for residues F230 and R245 coincide with perturbations for hCAII residues K227 and D242, which were only introduced in solids upon addition of SBR. Moreover, residues K169 and D174, which were perturbed by M25, are not close to the catalytic site of SBR and rather on the surface of hCAII. These residues might have been indirectly affected by M25 binding. On the other hand, it can be hypothesized that the region around K169, D174 and F230 might undergo unspecific transient interactions with free M25 molecules, while no strong chemical shift changes were detected for the putative second binding site according to the X-ray structure.

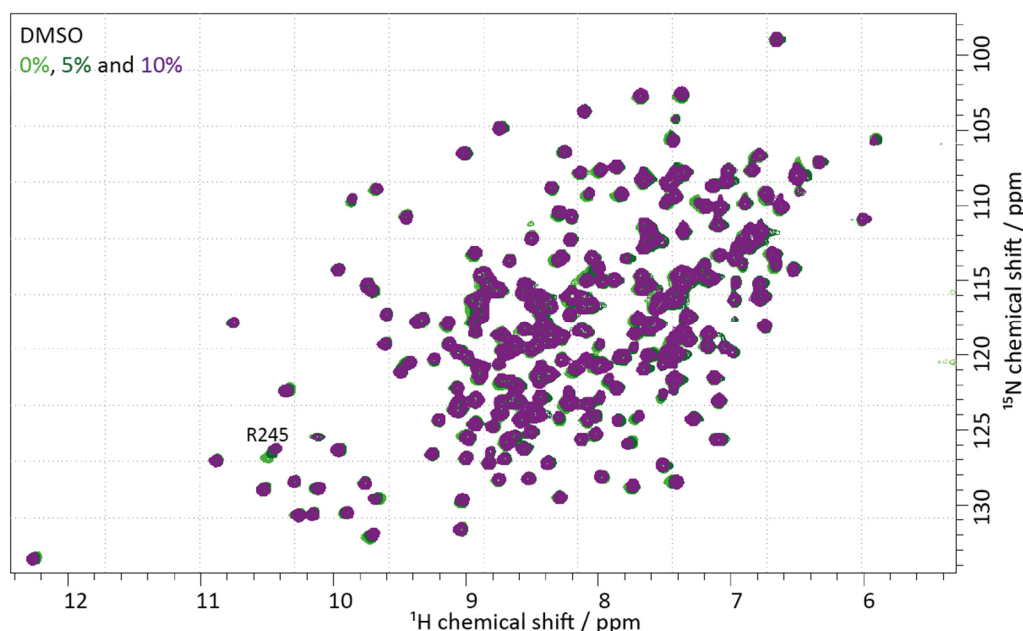


Figure 19: Overlay of three 2D ^1H - ^{15}N HSQC spectra comparing apo-hCAII and hCAII in the presence of up to 10% d_6 -DMSO. The sample contained $200 \mu\text{M}$ ^{13}C -, ^{15}N -labelled hCAII in 80 mM phosphate and 20 mM NaCl at pH 7.4 with 0% (light green), 5% (dark green), and 10% (purple) d_6 -DMSO. The spectra were recorded at 37 °C on an 800 MHz spectrometer for 45 min.

For protein-ligand titrations, usually ligand stocks of high concentration are used with solvents like DMSO. While SBR was dissolved in d_6 -DMSO, M25 was directly dissolved in water, and therefore no additional solvent was introduced obtaining the hCAII-M25 complex. To disentangle binding of DMSO solvent molecules from binding of the ligand SBR, titrations with DMSO were performed. In Figure 19, the spectra of hCAII with different amounts of DMSO were overlaid. The fingerprint of hCAII overall only slightly changed upon addition of

DMSO. For some residues, small perturbations can be seen which hint towards the fast-exchange regime. This is most pronounced for residue R245 which is located at C-terminus of hCAII (Figure 14 and Figure 19). For this residue, only small CSPs were detected during the hCAII-SBR titration. It can be deduced that the CSPs detected in Figure 16 and Figure 17 originate from the ligand SBR and that the additional amounts of DMSO in the sample can be neglected.

Finally, the ligands SBR and M25 bind to hCAII in solution according to the expectations based on the corresponding X-ray structures. Not only apo-hCAII but also hCAII in complex with SBR seems to experience similar chemical environments in the solution and solid-state samples according to the NMR spectra. Although d_6 -DMSO was introduced upon addition of SBR, its effect on the chemical environment of the backbone amides of hCAII is negligible.

2.2.3 Ligand motion in hCAII-bound state detected in solution and solids

The following chapter focuses on SBR ligand dynamics in the hCAII-bound state detected via solution and solid-state NMR. The X-ray structure of the complex (chapter 2.2.1) showed higher B factors especially for the indole moiety of SBR, indicating increased static disorder or flexibility. Out of the various timescales of motion, which can be detected via relaxation NMR experiments (chapter 1.1.5), the ligand SBR is likely to move on a fast rather than a slow timescale. Different methods were utilized to determine ligand dynamics in hCAII-bound state. In addition, the chemical environments of the ligand in apo and hCAII-bound state were compared.

In solution, second to slow millisecond dynamics are imprinted in the peak widths in simple 1D or 2D NMR spectra. Comparing the proton line shapes of free and hCAII-bound ligand, the refocused INEPT NMR spectra in solution displayed different peak widths of 0.034 and 0.044 ppm, respectively, for the cross peaks between spins 2 and 3 (Figure 20). This small increase in linewidth can be attributed to slower tumbling of SBR in the complex (~30 kDa) compared to free ligand (0.37 kDa). For other peaks, especially diagonal peaks, the broader lines led to peak overlap. Consequently, it was assumed that free and hCAII-bound SBR did not undergo dynamics or chemical exchange on the seconds to slow milliseconds timescale. In addition, the signals of the free and protein-bound ligand resembled each other well and did not report on strong changes in the chemical environment when bound to the protein in solution. Additional peaks, potentially due to multiple conformations, could not be assigned.

As elaborated in chapter 2.2.3, the ^1H -BMRD and ^1H -NERRD NMR relaxation experiments^[36] were chosen to detect microsecond dynamics of the ligand SBR in solution and the solid state, respectively (chapter 1.1.5). Acquiring and analyzing ^1H -relaxation data is a challenging task, as the anisotropic contributions, especially in the solid state, cannot be fully accounted for by high MAS, high radiofrequency field strengths, and sample deuteration^[15]. As for all homonuclear 2D experiments, both chemical shift and scalar couplings evolve during the indirect chemical shift evolution period t_1 . The relaxation data were processed by Suresh Vasa, accounting for off-resonance effects during the spin lock in solution and the solid state, fitting the data, and calculating the associated errors with the Monte-Carlo procedure^[36].

To detect ligand protons in hCAII-bound state, all protein and water protons needed to be removed from the spectrum. As all protein protons are bound to ^{13}C or ^{15}N atoms, their scalar coupling was utilized to remove all remaining proton signals of hCAII. Most ligand protons were bound to the NMR-inactive ^{12}C and ^{14}N atoms and thus did not evolve scalar couplings. All buffers and solvents were used in a deuterated fashion.

In solution, a ligand-specific ^1H -BMRD experiment was implemented (Figure 89). The refocused INEPT was selected to obtain ^1H - ^1H cross peaks, which provide higher resolution than the diagonal peaks, and to prevent additional high-power pulses, as in e.g. the TOCSY experiment. The probe as part of the spectrometer hardware is sensitive towards high power pulses, which are already imposed with the spin lock. In both refocused INEPT and TOCSY, homonuclear J -couplings evolve while the chemical shift is refocused. The ^{13}C -, ^{15}N -filters^[35] dephase proton magnetization bound to ^{13}C - and ^{15}N -nuclei while the magnetization of protons which are bound to ^{12}C and ^{14}N nuclei is refocused. In total, four ^{13}C -, ^{15}N -filters were implemented to decrease the contributions of protein protons to the signal as much as possible. During the spin lock, the proton magnetization was kept at a 35.3° angle relative to the z -axis at which NOESY and ROESY contributions cancel^[102]. High-power pulses like for the spin lock heat up the sample over the course of the experiment. To prevent unequal temperature changes, heat compensation was scheduled at the beginning of the pulse sequence with a spin lock pulse accounting for the difference to the condition with the highest spin lock. In addition, the different relaxation delays were balanced by adding a compensating delay after the spinlock module.

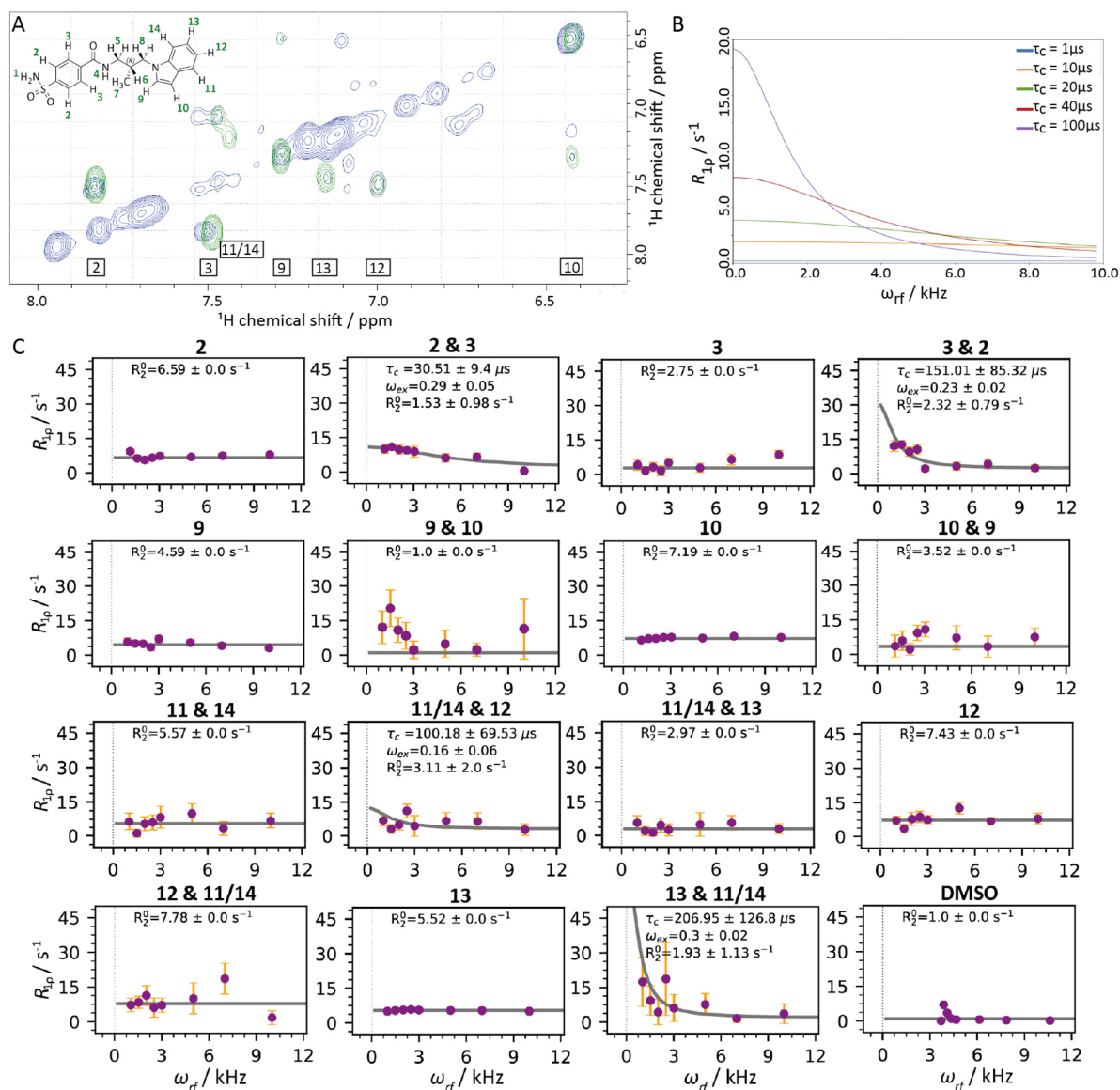


Figure 20: $R_{1\rho}$ relaxation dispersion experiments in solution (^1H -BMRD). **A.** ^1H $R_{1\rho}$ relaxation spectra comparing the aromatic protons of free SBR in D_2O (green) or SBR bound to $483 \mu\text{M}$ ^{13}C -, ^{15}N -labelled hCAII at equimolar ratio in 80 mM phosphate buffer, 20 mM NaCl, and <10% d_6 -DMSO at pH 8 (blue). The shown spectra were recorded at 25 °C with a 1 kHz spin lock and 10 ms mixing time. The whole dataset for SBR was recorded for 17 hours, while the dataset for the complex was acquired for 3.5 days. **B.** Simulated relaxation dispersion spectra in solution assuming an excited population of $p_E = 1\%$ and a chemical shift difference between excited and ground state of $\Delta\omega_{ex} = 1\text{ppm}$, kindly prepared by Suresh Vasa. **C.** $R_{1\rho}$ relaxation dispersion plots with fitted curves (grey) corresponding to the data points (purple) of the ^1H -BMRD experiments for SBR bound to hCAII with error bars (orange) derived from the Monte Carlo error estimation. The numbers of the spins correspond to panel A. Data plots prepared by Suresh Vasa.

The BMRD relaxation data for SBR in complex with hCAII are shown in Figure 20. For comparison, relaxation dispersion curves were simulated for different regimes of motion (Figure 20, panel B). The faster the motion, the lower the $R_{1\rho}$ value at 0 kHz spin lock and the flatter the decrease of the relaxation dispersion curve towards higher spin locks. Completely flat curves indicate no measurable microsecond motion based on chemical shift fluctuations.

The peak corresponding to DMSO, which was introduced as a solvent from the SBR stock solution, was treated as a reference peak which, indeed, shows no microsecond motion and low error bars (Figure 20, panel C, bottom right). Relaxation dispersion plots were only prepared for assigned and sufficiently separated peaks, which was the case for the aromatic protons. The resulting relaxation dispersion plots are mostly flat indicating no microsecond motion. Only for the cross peaks between spins 2 and 3, spins 11/14 and 12, and spins 13 and 11/14, the relaxation dispersion curves revealed potential microsecond motion. The data points for the cross peaks between spins 11/14 and 12, and 13 and 11/14, however, show that the changes lie within the margin of the error bars. Consequently, these fits do not indicate dynamics. Some of these data sets are also relatively noisy with mostly high error bars, for example, for the cross peaks of spins 12 and 11/14, and spins 9 and 10. The scattered data points can be attributed to intra-ligand ^1H - ^1H scalar couplings which might have evolved. For the cross peaks between spins 2 and 3, microsecond dynamics were detected. These spins are located at the benzene ring of the ligand SBR (Figure 20, panel A). The timescale of the dynamics with $151 \pm 85 \mu\text{s}$ and $31 \pm 9 \mu\text{s}$ deviate from each other, although the same timescale of dynamics should underlie. Also, no dynamics were detected for the diagonal peaks, although the dynamics of the cross peaks should have been imprinted by the spin in the direct dimension (Figure 20). It remains unclear why the derived dynamics from diagonal and cross peaks differ. Next, the microsecond dynamics detected for spins 2 and 3 at the benzene ring of SBR were probed via solid-state NMR spectroscopy and MD simulations. In solids, a 2D ^1H - ^1H NERRD experiment was established to detect motion on a similar timescale as the ^1H - ^1H BMRD experiment in solution. As mentioned in chapter 1.1.5, the NERRD experiment can identify microsecond motion. The fluctuations of the bond vectors are detected by analyzing the relaxation dispersion plots depicting $R_{2,\text{eff}}$ at different spin lock fields. Close to the half-, single-, or double-rotary-resonance condition, $R_{2,\text{eff}}$ is elevated in case microsecond fluctuations are present (Figure 22, panel B). Very close to the resonance conditions, the measured $R_{2,\text{eff}}$ values are less reliable, therefore only data points ≥ 3 kHz away from the resonance condition were measured. Similar to the BMRD fits, steeper curves correspond to slower and the flatter curves correspond to faster motion^[15]. Like the ^1H -BMRD experiment, the ^1H -NERRD experiment included eight different spin lock fields, water suppression, temperature compensation, and constant time compensation (Figure 21). To specifically select ligand protons, ^{13}C -, ^{15}N - and ^1H -REDOR filters were both tested. However,

only the ^{13}C -, ^{15}N -filters delivered sufficient suppression of undesired proton signals, like protons of hCAII. Proton buffer signals were further reduced by using deuterated chemicals. In solids, an on-resonance spin lock was applied. Unexpectedly, J -coupling-based experiments like a refocused INEPT did not achieve cross peaks. Dipolar coupling-based experiments with NOE or RFDR mixing also failed to produce cross peaks. Most likely, the cross peak intensities were low and did not appear at a reasonable signal-to-noise ratio. Finally, only diagonal peaks were used implementing a NOESY module without any NOESY mixing time. Any other 2D ^1H - ^1H experiment would have served this purpose as well, essentially delivering 1D resolution.

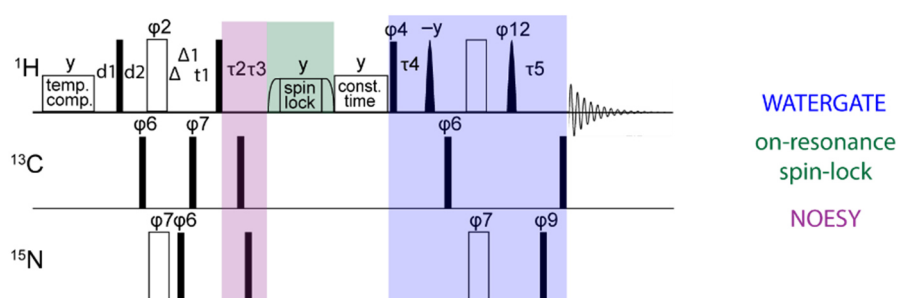


Figure 21: Ligand-selective ^1H - ^1H NERRD pulse sequence developed by Suresh Vasa. Even though the sequence potentially grants 2D resolution, only the diagonal peaks were finally used yielding 1D resolution. The pulse sequence included temperature compensation (temp. comp.), a ^{13}C -, ^{15}N -INEPT-filter, the evolution time for the indirect dimension t_1 , a NOESY sequence (magenta), an on-resonance spin lock (green), constant time compensation (const. time), and water suppression (blue). The spin lock times amounted each to 2, 5, 10, 20 and 40 ms. The delays were defined as the following: $d_1=1$ s, $d_2=3.57$ ms, $\Delta=1.97$ ms, $\Delta_1=1.59$ ms, $\tau_2=2.5$ ms, $\tau_3=2.49$ ms, $\tau_4=2.54$ ms, and $\tau_5=942$ μs . The phase cycling was performed with $\phi_2=0\ 1\ 0\ 1\ 2\ 3\ 2\ 3$, $\phi_4=0\ 0\ 1\ 1\ 2\ 2\ 3\ 3$, $\phi_6=0\ 0\ 0\ 0\ 0\ 0\ 0\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$, $\phi_7=0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$, $\phi_9=0\ 0\ 0\ 2\ 2\ 2\ 2$, and $\phi_{12}=2\ 2\ 3\ 3\ 0\ 0\ 1\ 1$.

Here, a co-crystallized hCAII-SBR sample in a 1.3 mm rotor was prepared to increase the sensitivity of the experiment compared to a sample in a 0.7 mm rotor. It was assumed that the anisotropic contributions to proton relaxation, which can be removed more efficiently in 0.7 mm rotors due to fast spinning at 100 kHz MAS, were of less importance for high sensitivity than the number of spins available during the experiment. The 1.3 mm rotor has a five to six-fold larger volume, accommodating roughly one milligram of protein^[103].

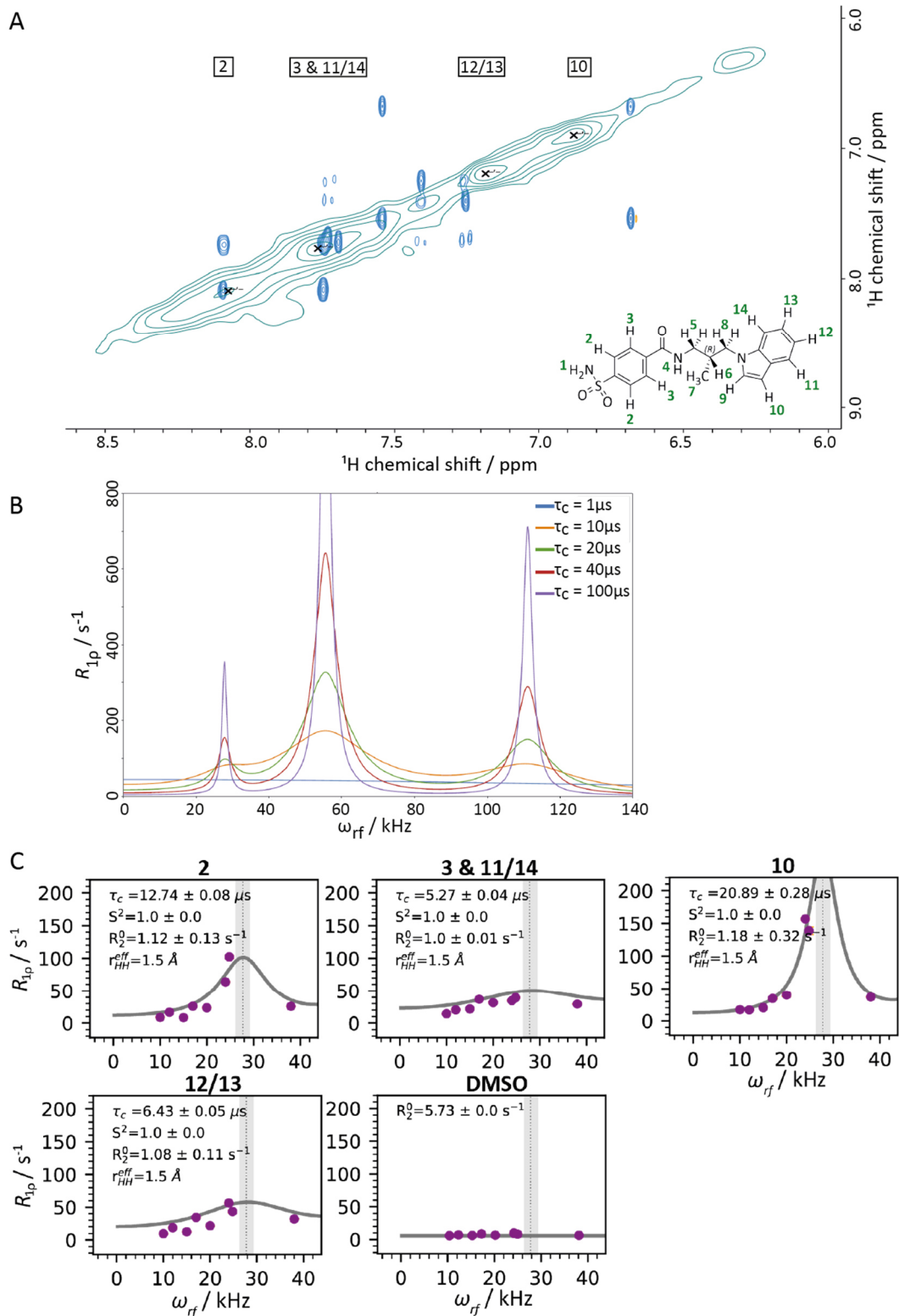


Figure 22: R_{1p} relaxation dispersion experiments in the solid state (^1H - ^1H -NERRD) at 55.56 kHz MAS and 298 K using ~ 1 mg co-crystallized D-, ^{13}C -, ^{15}N -labelled hCAII-SBR complex in 2.9 M $(\text{NH}_4)_2\text{SO}_4$, 50 mM d_5 -TRIS buffer, and $<10\%$ d_6 -DMSO at pH 7.9 in a 1.3 mm rotor, recorded for 2 days and 10 hours. **A.** Overlay of 2D ^1H - ^1H NOESY spectrum in solids (green) and a ^1H - ^1H TOCSY of free SBR in solution (blue). **B.** Simulated relaxation dispersion profiles assuming $S^2=0.97$, $R_2^0=0$ s^{-1} , and $r_{\text{HH}}^{\text{eff}}=3.0$ \AA , kindly provided by Suresh Vasa. **C.** R_{1p} relaxation dispersion data (purple) fitted with curves (grey) with very small error bars (orange), which were derived from the Monte Carlo error estimation and provided by Suresh Vasa. The number of the spins are the same as in panel A.

Using the tailored pulse sequence to select ligand proton signals, peaks matching the expected chemical shifts for SBR protons were detected in the solid state (Figure 22, panel A, green spectrum). Due to heterogeneity and strong anisotropic interactions between protons in solids, the peaks are broadened and overlap especially for the aliphatic and a few aromatic diagonal peaks. Strikingly, the extracted relaxation dispersion profiles matched the expected relaxation dispersion curves very well and indicate microsecond motion for several ligand protons. Microsecond motion was obtained for the diagonal peaks corresponding to spin 2, spins 3 & 11/14, spin 10, and spins 12/13, which are part of the benzene and indole moieties of SBR. The peaks for spins 3, 11, and 14 as well as for spins 12 and 13 overlap and thus cannot be analysed separately. A quantitative assessment of the amplitude of motion was not possible due to its convolution with site-specific effective ^1H - ^1H distances, which directly modulate the extent of $R_{1\rho}$ relaxation.

To understand why the BMRD data only indicate microsecond motion for proton spins 2 and 3 of the benzene ring of SBR and why the NERRD data reveal microsecond motion for proton spins 2, 3, 10, and 11-14, MD simulations mimicking the state of the hCAII-SBR complex in solution were performed for 4 μs by Benedikt Söldner. It should be noted that the MD simulations were set up at room temperature, while the X-ray data were acquired at -173°C , which stabilizes any possibly flexible moieties of the SBR in the thermodynamically most favourable conformation. The MD data revealed ring flips of nearly 180° for the benzene ring of SBR on the microsecond timescale and much faster motion for the indole ring. In addition, short-lived π -stacking interactions were observed between the indole ring and the F130 sidechain of hCAII. Otherwise, the indole moiety of SBR moved rather randomly at a speed beyond the timescales detected by BMRD and NERRD experiments. Hence, the motion revealed in both experiments must result from the benzene ring flips as only these dynamics occur on the proper timescale.

One main difference between the NERRD and the BMRD data lies in the physical properties they detect. NERRD detects fluctuations of the bond vector orientation even with regard to distant protons which are still part of the dipolar coupling network. The chemical shift changes detected by BMRD experiments reflect the chemical environments of the spins. If these changes are rather local, other spins might not sense these fluctuations. Such changes were detected with the ^1H - ^1H BMRD experiment for the benzene ring due to chemical shift

differences before and after a ring flip. In the ^1H - ^1H NERRD experiment, the H-H bond vector fluctuations of the protons in the benzene ring modulated the proton dipolar coupling network which was sensed by the proton dipolar coupling network of protons of the indole moiety of the ligand SBR despite a $\sim 8 - 12$ Å distance in a stretched ligand conformation. MD simulations indicated that this distance is decreased at certain time points due to the flexibility of the indole group. Microsecond dynamics of the benzene ring were therefore detected via BMRD and NERRD experiments, although they are sensitive towards different timescales with slow microsecond to fast millisecond motion and microsecond motion, respectively^[15]. Alternatively, a slower dynamics component of the indole group itself could have been overlooked when the MD simulation data were analysed or not detected as the 4 μs MD simulation could have been too short. Moreover, the dynamics of SBR bound to hCAII in soluble and crystalline state could deviate due to the different degrees of freedom within the complex.

In summary, the hypothesized microsecond dynamics of the ligand SBR in hCAII-bound state were detected via solution and solid-state relaxation dispersion experiments. These microsecond dynamics were attributed to the benzene ring of the ligand and not to the indole moiety, which moves on a faster timescale, according to the MD results. The solution ^1H - ^1H BMRD experiment indicated microsecond dynamics for the benzene ring, but only with a small decrease of the fitted relaxation dispersion curve. The established solid-state ^1H - ^1H NERRD experiment detected the motion more reliably. However, many protons in proximity to the flexible benzene ring sensed the microsecond dynamics motion. As a result, MD simulations were necessary to assign the dynamics to the protons with atomic resolution. Modifying the benzene ring at the ortho position could prevent the flipping motion and might further increase the affinity of the ligand SBR. Moreover, modifications decreasing the flexibility of the indole ring could lead to higher affinity as well. Here, the overall solubility of the ligand should be monitored during optimizations and improved, if possible.

2.2.4 hCAII backbone dynamics in the hCAII-SBR complex in solution

It was hypothesized that motion in the hCAII-SBR complex could be a collective movement pattern, involving both ligand and protein residues. Especially residues like F130 (chapter 2.2.3) interacting with SBR might be similarly flexible as the ligand. For this purpose, ^{15}N -CPMG experiments were performed in solution probing slow micro- to fast millisecond

dynamics of the hCAII backbone amides. The data points accounting for different CPMG frequencies were fitted assuming no or two-site chemical exchange^[28-29].

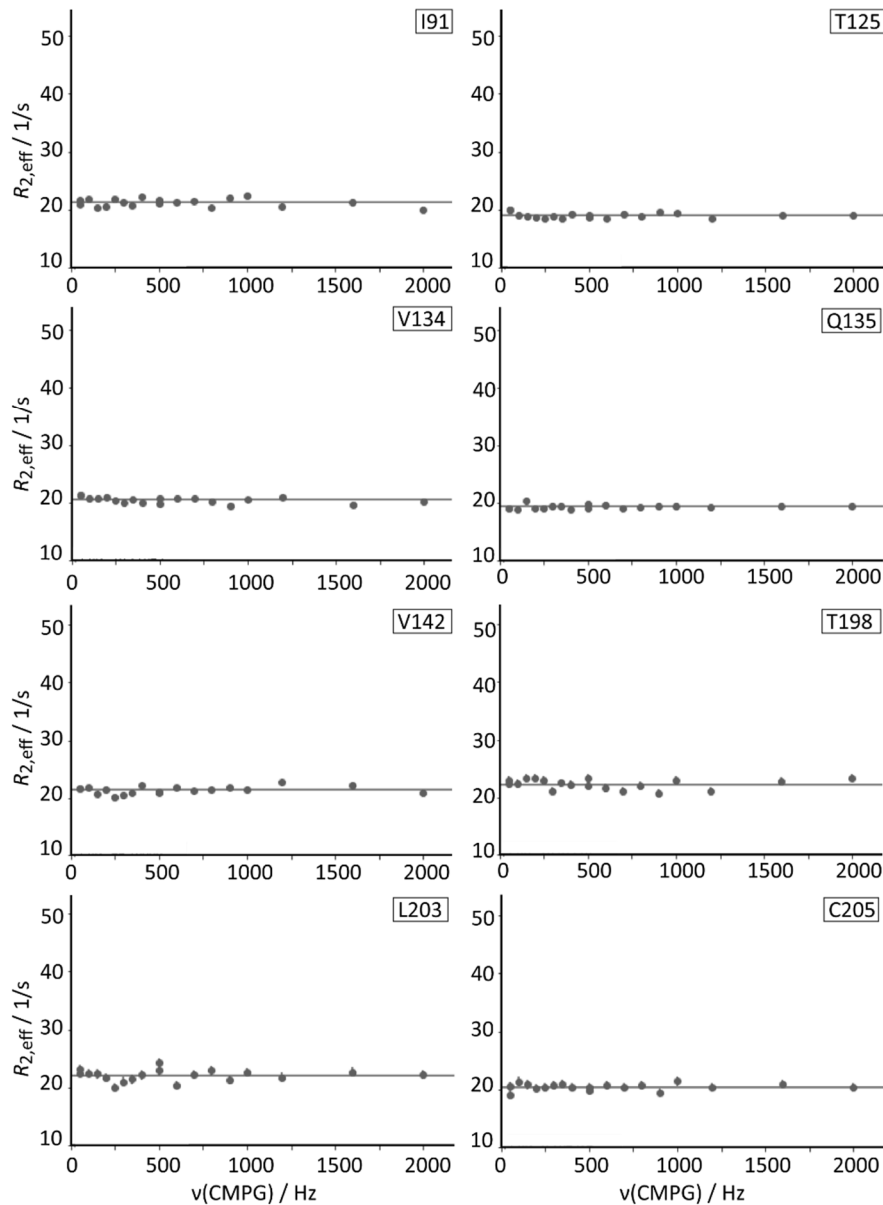


Figure 23: Selection of ^{15}N -CPMG relaxation dispersion plots probing hCAII backbone dynamics in the hCAII-SBR complex in solution at 0-200 Hz CPMG frequencies. The corresponding hCAII residues are specified in the right top corner of each plot. The data were acquired on a 600 MHz spectrometer equipped with a cryo-probe at Forschungszentrum Jülich.

Based on previous measurements^[69], it has been established that the loop regions of the apo-hCAII backbone include several residues which undergo slow microsecond motion, like W208 or N124. Adding ligands such as acetazolamide or dorzolamide, slow microsecond dynamics were dampened to some extent^[40a, 70]. Similarly, no dynamics were detected for the hCAII backbone in complex with SBR. In Figure 23, some exemplary relaxation dispersion profiles for hCAII residues are shown, all being flat. It seems SBR binding does not only happen with

high affinity, it also strongly stabilizes the overall backbone of hCAII on the millisecond timescale. Compared to acetazolamide and dorzolamide, SBR has a larger tail resulting in a larger interaction surface with hCAII and a smaller percentage of unoccupied space in the binding site. Backbone dynamics potentially induced by motions of SBR might be below the detection limit of the ^{15}N -CPMG experiment. Moreover, it remains unclear if the hCAII sidechains could be involved in motion like the ligand SBR (chapter 2.2.3). Unfortunately, only few sidechain assignments have been published until now^[66a].

2.2.5 Methyl chemical shift assignments for the hCAII sidechains

While amide assignments can report on structures and dynamics of the backbone and hydrophilic amide sidechains, methyl sidechains can reveal properties of hydrophobic regions in protein structures. The active site of hCAII contains several methyl-bearing residues including L197 as part of the “hydrophobic wall”. It was hypothesized that the ligand flexibility of SBR, as described in chapter 2.2.3, might also influence methyl sidechain structures and dynamics. Therefore, methyl sidechains were assigned *de novo* for amino acids isoleucine, leucine, and valine (ILV). Here, an experimental approach was utilized with time-shared 3D XXH NOESY and RFDR experiments in solution and the solid state. The number of peaks in methyl spectra can be adapted easily by choosing an appropriate labelling approach (chapter 1.1.6). On one hand, too large distances between methyl groups complicate their identification based on distance restraints due to a lower number of cross peaks. Too many peaks, on the other hand, lead to peak overlap in the 2D ^1H - ^{13}C experiment. To reduce the total number of peaks avoiding peak overlap, no alanine or threonine methyl groups were labelled. Methionine only occurs two times in the amino acid sequence with very few contacts and therefore would have had a low benefit for the assignment process. For isoleucine, valine, and leucine in the solid-state NMR samples, only one proton per methyl group was labelled to prevent line broadening due to slightly heterogeneous chemical shifts of the three protons in the solid state (chapter 1.1.6). For valine and leucine, no stereospecific precursors for HD_2^{13}C -methyl labelling were available thus both methyl groups were each labelled. For comparability, the same labelling scheme was used for the solution NMR studies.

In the ^1H - ^{13}C 2D solution and solid-state spectra, 86 peaks were detected (Figure 24, panel A). Based on the labelling scheme and the hCAII sequence, 95 peaks originating from 52 residues were expected in an ^1H - ^{13}C spectrum consisting of nine isoleucine, 26 leucine, and 17 valine

residues with each twice the peak number for leucine and valine residues. The nine peaks difference to the expected number of peaks could either be explained by the same chemical shift for geminal valine and leucine sidechains, intermediate-exchange dynamics, or a mixture of both. Isoleucine peaks were clearly separated from valine and leucine peaks with eight peaks with upfield-shifted ^{13}C chemical shifts below 12 ppm.

In Figure 24 panel B, methyl groups with contacts to water as well as valine and leucine residues were discerned to sort the methyl peaks into groups. Distinguishing the corresponding residue types and the hydrophilicity of the environment can support both the manual and automated assignment processes as it reduces the number of possible candidates per peak. An IV-methyl labelled sample was prepared upon addition of 15 mg unlabelled leucine to one litre of expression culture, which had already been supplemented with ILV-methyl labelling precursors. The difference between the resulting spectrum and a spectrum corresponding to an ILV-methyl labelled sample should identify the peaks originating from leucine residues. Notably, the peaks were not absent in the IV-methyl labelled spectrum but had lower peak intensities. Apparently, some leucine residues were labelled, likely because the excess of unlabelled leucine was insufficient.

Moreover, valine peaks also tend to be shifted slightly upfield compared to the average ^{13}C chemical shift of leucine residues. Apart from eight clearly identified isoleucine peaks, three additional peaks had small ^{13}C chemical shifts between 12 and 14 ppm. These could be assumed to correspond to three valine or two valine peaks and one isoleucine peak. In total, 46 leucine and ~32 valine peaks were identified with a small degree of uncertainty for peaks with medium peak height (Figure 24, panel B).

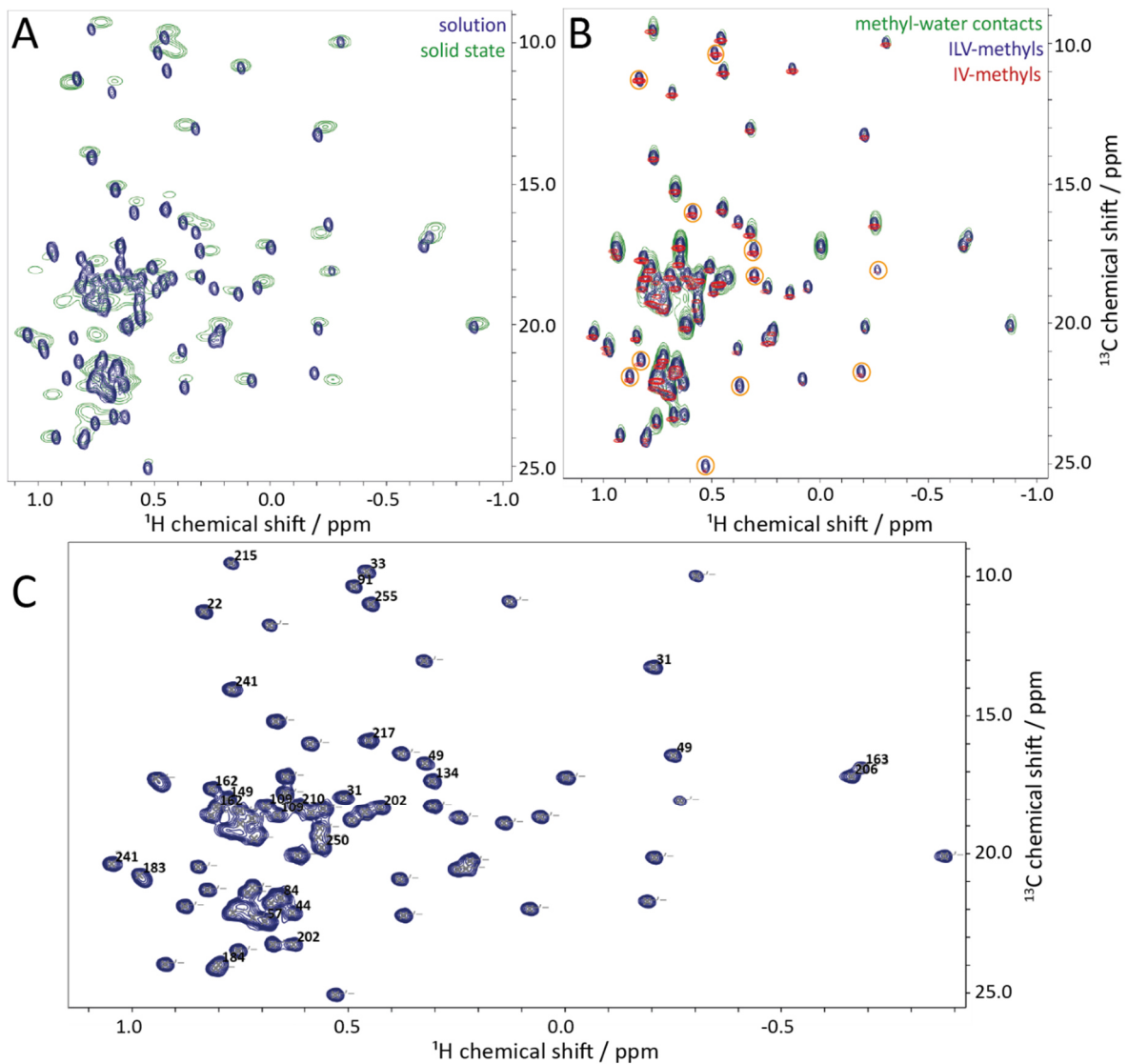


Figure 24: Overlays of ^1H - ^{13}C HMQC NMR spectra of apo-hCAII. **A.** Spectra recorded on a $\sim 300\ \mu\text{m}$ hCAII sample with DCN- and HD_2^{13}C -ILV-methyl labelling in $80\ \mu\text{M}$ phosphate buffer and $20\ \mu\text{M}$ NaCl at pH 7.6 on an 800 MHz spectrometer at $25\ ^\circ\text{C}$ in solution for 16 min (blue) and a 1.3 mm crystallized solid-state hCAII sample in $2.9\ \text{M}$ $(\text{NH}_4)_2\text{SO}_4$ and $50\ \text{mM}$ d_5 -TRIS buffer at pH 7.9 at 55 kHz on a 700 MHz spectrometer at $23\ ^\circ\text{C}$ for one hour (green). **B.** Solution NMR spectra of ILV- (blue, same as in A) and IV-methyl labelled (red) hCAII samples. The red spectrum refers to $\sim 40\ \mu\text{M}$ hCAII with DCN- and HD_2^{13}C -IV-methyl labelling, measured on a 700 MHz spectrometer at $23\ ^\circ\text{C}$ for 9 h. The green spectrum corresponds to a plane at the ^1H chemical shift of water (4.7 ppm) of a 3D HCH-NOESY spectrum recorded at the same conditions as the red spectrum but for one day and 22 hours. Residues circled in orange highlight peaks with a weak or no cross peak to water. **C.** Compiled manual and FLYA-based solution and solid-state methyl assignments (Table 4).

Moreover, water contacts of the methyl sidechains of isoleucine, leucine, and valine were determined with 3D HCH solution spectra (Figure 24, panel B). The peaks circled in orange had drastically reduced water contacts compared to the other peaks. This was the case for two isoleucine, two valine, and seven leucine peaks. In solids, overall fewer cross peaks between methyl sidechains and water were detected (Figure 58).

To acquire the 3D XXH NOESY and RFDR experiments efficiently, time-shared experiments^[104] were implemented in solution^[105] (Figure 90) and the solid state by Dr. Suresh Vasa (Figure 91). Each time-shared experiment yielded the information context commensurate to four individual experiments, namely CCH, CNH, NCH, and NNH experiments, which were recorded simultaneously by accommodating both C and N chemical shift evolution in one indirect evolution period. Concomitant magnetization transfers were either based on the simultaneous evolution of *J*-couplings between HN and HC nuclei or simultaneous CP. The spectra were separated afterwards either using multiplexing or via nuclei-specific offsets in the proton and/or first heteronuclear dimension. The magnetization was transferred from one methyl or amide group to another methyl or amide group as described in panel A in Figure 25. Carbon and nitrogen nuclei serve well to disambiguate contacts between protons as they offer large dispersion, yielding less peak overlap compared to 3D HXH experiments, for example. Nevertheless, 3D HCH and HNH experiments were recorded to additionally support the assignment process. Moreover, 3D CCH experiments with short NOESY or RFDR mixing times were acquired to identify the short-distance contacts between geminal methyl groups of leucine or valine residues. In addition to the time-shared 3D experiments, a time-shared 4D HXXH NOESY experiment was recorded in solution. However, the resulting data sets could not be processed successfully, and the individual spectra could not be retrieved and used for assignments.

The number of cross peaks determines how many assignments can be successfully derived from the data. The signal-to-noise ratios of cross peaks are modulated by the distance between the nuclei, the mixing time, the sensitivity of the experiment, and the number of scans dedicated to acquiring high-quality spectra. The sample quality and concentration as well as line broadening due to heterogeneity or motion affect the peak intensities, too, but similarly apply for diagonal peaks. To assess the quality of the NOESY and RFDR spectra, the total number of expected peaks and the number of peaks detected in the spectra were estimated (Figure 25, panel C). For the estimated peak number, a distance limit needs to be chosen. The cross peak intensities in NOESY experiments are modulated by the internuclear distance factor of r^{-6} while the cross peak intensities in RFDR experiments are scaled by the r^{-3} term^[106]. It has been reported that, in ideal scenarios, RFDR experiments can achieve cross peaks for spins separated by up to 13 Å^[107] while NOESY experiments can reach up to

8 Å^[108]. Assuming distance limits of 5 and 7 Å in a CCH or HCH experiment, the number of expected inter-methyl cross peak pairs was doubled according to the distances derived from the calculated X-ray structure.

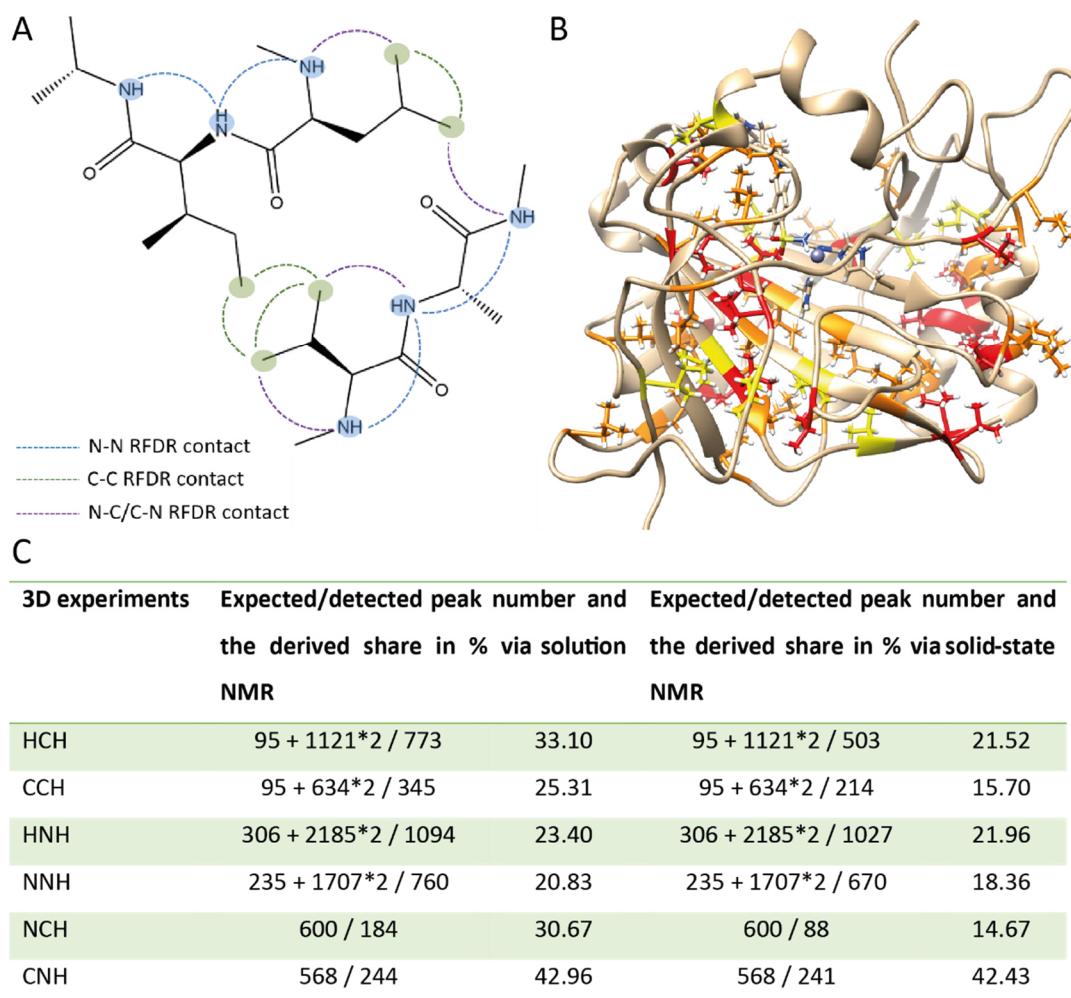


Figure 25: Methyl assignments. **A.** Exemplary scheme of methyl-methyl, methyl-amide, and amide-amide contacts detected in time-shared 3D XXH RFDR experiments. The atoms of the interacting nuclei are connected by differently coloured lines. **B.** X-ray structure of the hCAII-SBR complex (chapter 2.2.1) with labelled isoleucine (yellow), leucine (orange), and valine (red) methyl sidechains. **C.** Table displaying the expected and detected peak numbers for the individual experiments recorded in solution and in the solid state. A distance limit of 5 Å was assumed for RFDR contacts, which were estimated by FLYA and kindly provided by Dr. Alexander Klein. The sum of the expected peaks was divided each into the diagonal and the cross peaks (diagonal + cross peaks*2).

Here, a 5 Å distance limit was chosen to estimate the number of expected peaks, which were then compared with the number of peaks detected in the acquired spectra. 18-23% of the expected peaks were found in 3D HNH and NNH experiments in solution and the solid state. Out of all 3D spectra, the CNH spectra were most sensitive with roughly 42% of the expected peaks detected in solution and solids. Instead, the number of peaks for the 3D HCH, CCH, and NCH experiments deviated comparing solution and solid-state experiments. In solution, 25-33% of the expected peaks were identified in the spectra whereas in solids only 14-22% of

the expected peaks were visible. Interestingly, this was not only the case for the time-shared 3D CCH and NCH experiments but also for the separately acquired 3D HCH experiment. Thus, these differences do not seem to be related to the sensitivity of the time-shared experiment in solution and solids. To properly assess the performance of the time-shared experiments, all 3D experiments would have had to be acquired separately in addition, which was out of the scope of this project. Only the 3D CCH with long RFDR mixing time has been acquired both as a single and time-shared experiment in solids. Evidently, 132 and 130 peaks were picked in the single and time-shared 3D solid-state CCH spectra, respectively, which highlights the good performance of the time-shared experiment. It should be considered, however, that the mixing time in the single CCH was longer than in the time-shared CCH, with the numbers of RFDR blocks $l_{11}=640$ and $l_{11}=384$, respectively. It can be speculated that these different mixing times had a minor effect on the number of cross peaks.

Figure 26 points out the differences between the number of cross peaks for solution and solid-state spectra, exemplarily shown for 3D CCH spectra. Overall, the time-shared experiment in solution performed better than the corresponding experiment in the solid state, as elaborated above. For the time-shared experiment, less measurement time of roughly three and a half days was needed in solution compared to ten days in the solid state. The signal-to-noise ratios in the solution spectra were often higher compared to the solid-state spectra. In panel A, six cross peaks are visible in the solution CCH spectrum, whereas only one cross peak appears in the solid-state spectrum indicating methyl-methyl contacts for a methyl group with ^1H and ^{13}C chemical shifts of 0.77 and 10 ppm, respectively. In panels B and C, cross peaks between two geminal valine methyl groups are depicted, which presumably correspond to V241. The peaks can be identified to belong to the same residue, because the cross peak appears not only in the spectrum with long but also the spectrum with short mixing time in solution and solids. During assignments, reduced cross peak numbers limit the assignment success.

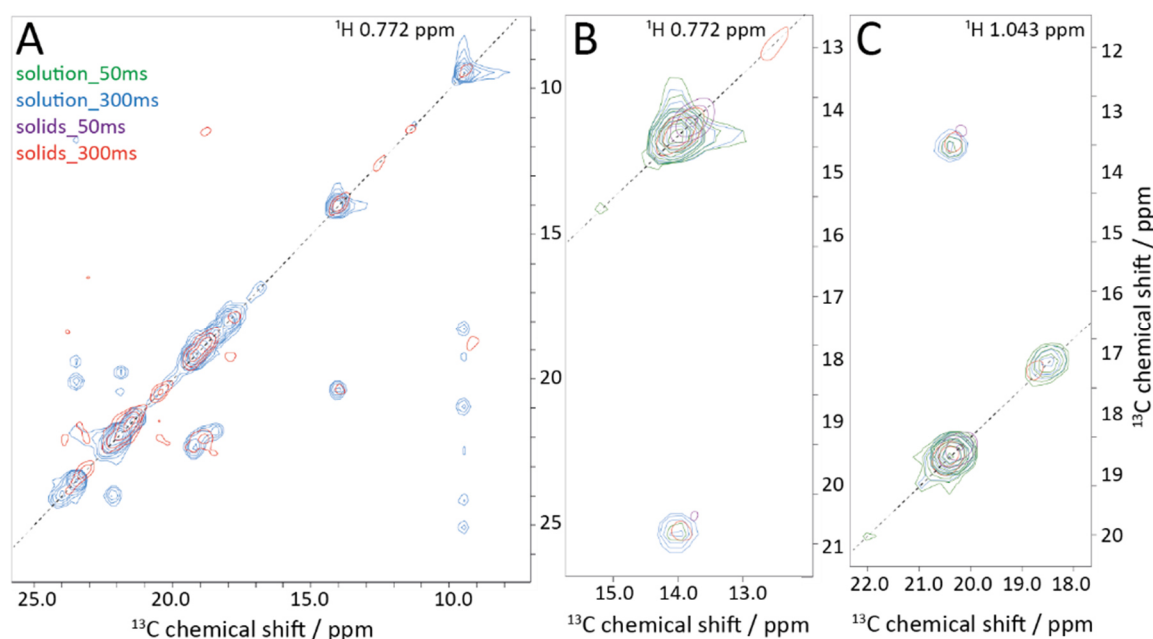


Figure 26: Overlays of CCH planes at different ^1H chemical shifts for solution and solid-state (A, B, C). In addition to the time-shared experiments, RFDR spectra serve to distinguish short- and long-distance contacts with mixing times of 7 ms ($I_{11}=384$) and 11.5 ms ($I_{11}=640$) in solids, respectively. The solution CCH NOESY spectra were both derived from time-shared experiments with 50 and 300 ms mixing times. Only positive peak contours are shown. The diagonal is indicated by a dotted line. Phase distortions occurred for high-intensity solution NMR peaks close to the spectrum edges. The solution data set was recorded for a $\sim 300\ \mu\text{M}$ hCAII sample with DCN- and HD_2^{13}C -ILV-methyl labelling in $80\ \mu\text{M}$ phosphate buffer and $20\ \mu\text{M}$ NaCl at pH 7.6 on an 800 MHz spectrometer at $25\ ^\circ\text{C}$ for five days and 16 hours. The solid-state spectra originate from DCN- and HD_2^{13}C -ILV-methyl labelled hCAII crystals in a 1.3 mm rotor measured at 55 kHz and $23\ ^\circ\text{C}$ on a 700 MHz spectrometer for roughly ten days.

Manual assignments were performed correlating assigned amide peaks with unassigned methyl peaks. Most assignments in solution were based on the 3D NCH and CNH spectra. However, the 3D NCH spectrum recorded in solids had insufficient quality and therefore could not be used in the assignment process. Consequently, the solid-state assignments were mainly based on CNH, HNH, HCH, and CCH spectra. All assignments were verified using several approaches and distinguished by assignment certainty. Assignments were considered unambiguous if only one residue matched a peak or the same peak was assigned via different strategies, like finding cross peaks both in the NCH and the CNH spectra. In addition, an IV-methyl labelled sample helped to distinguish leucine and valine methyl peaks. Methyl-water contacts extracted from the 3D HCH spectra were utilized to support the assignments. The 3D spectra allowed to deconvolute broad 2D peaks due to better resolution. Moreover, the methyl-methyl cross peaks were only considered unambiguous if they were obtained for both possible magnetization transfer pathways in the 3D CCH spectrum. X-ray structure-derived distance restraints were utilized to assign the methyl peaks with different intensities (chapter 2.2.1). For backbone-sidechain magnetization transfers via dipolar couplings, the amide-

methyl cross peak with the highest intensity does not necessarily belong to the same amino acid, because it depends on the sidechain length and conformation. The peaks in crowded regions in the spectra and hCAII regions with many methyl groups were most challenging to assign. Combining the NCH and CNH as well as the CCH and HCH spectra complementarily proved to be a robust and straight-forward approach, especially for crowded spectral regions. Assignment candidates with similar distance restraints for one cross peak were considered equally. Of course, the assignments got easier as soon as more peaks had been assigned and the pool of possible candidates shrunk, assigning the remaining peaks. It should be noted that this approach introduces a certain degree of bias. In addition, inconsistencies between expected and detected cross peaks in the spectra complicated the assignment process.

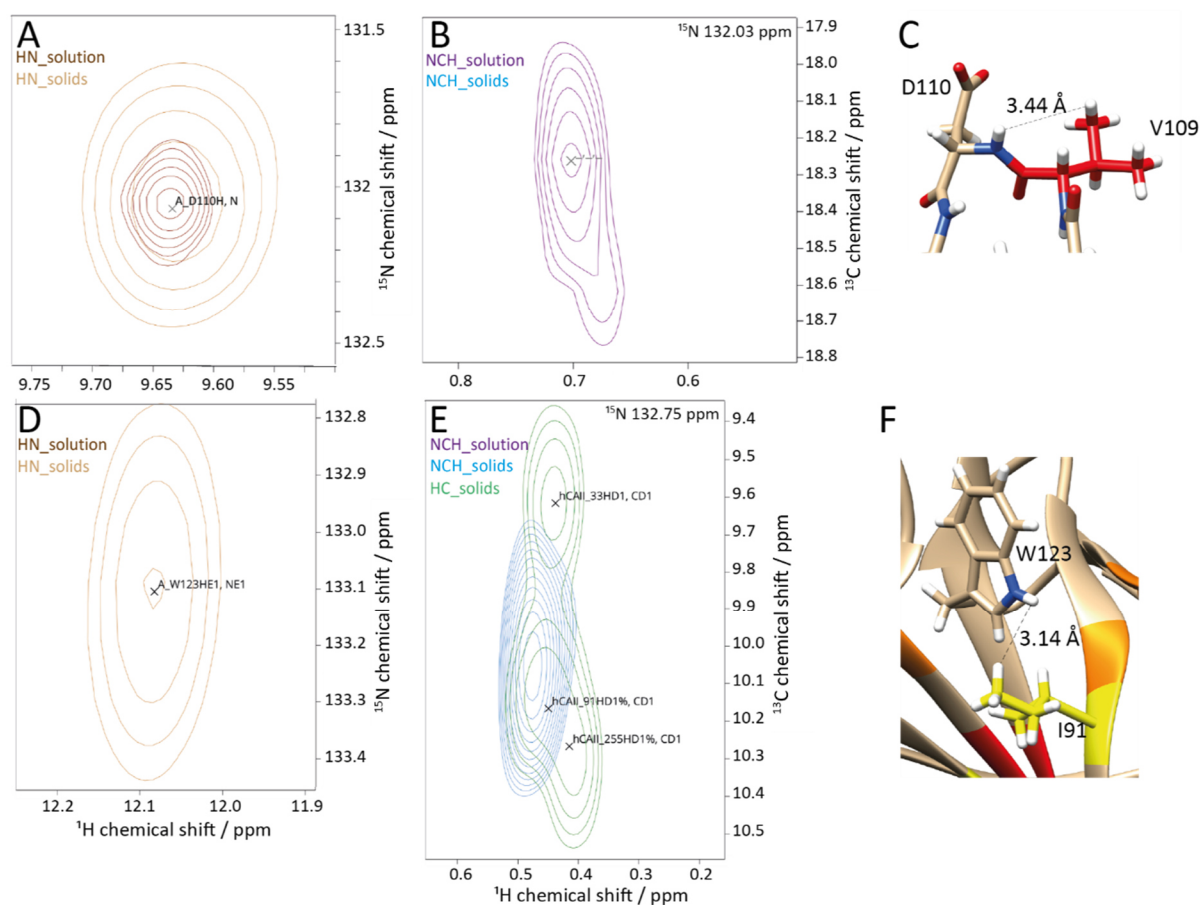


Figure 27: Exemplary assignments resulting from 3D NCH spectra, which were correlated with 2D HN spectra recorded in solution and the solid state corresponding to Figure 26. **A, D.** ^1H - ^{15}N HSQC spectra showing the peaks assigned to the backbone amide of D110 and the sidechain amide of W123, respectively, recorded for one hour in solids and 16 min in solution. **B, E.** The NCH planes corresponding to panels A and B, respectively. **C, F.** Distances between atoms according to the X-ray structure (chapter 2.2.1). Carbon sidechain atoms were coloured in yellow, orange, or red for isoleucine, leucine, or valine residues.

In Figure 27, assignments are shown exemplarily using 2D hNH and 3D CNH spectra of solution and solid-state apo-hCAII samples. Here, two amide peaks, well-separated from other peaks

in the ^{15}N dimension, were chosen and matched with peaks in the NCH spectra. A peak matching the ^{15}N chemical shift of D110 was only detected in the NCH solution spectrum, while no peak was detected in the solid-state spectrum (Figure 27, panel B). The crystal structure revealed that the only methyl-bearing residue in the vicinity of D110 is V109 (Figure 27, panel C). Here, the peak shape of the cross peak in panel B indicates that it likely represents two peaks. These two peaks should correspond to the two geminal groups of V109. The missing cross peak in the solid-state spectrum can possibly indicate a lower signal-to-noise ratio than in the solution NMR spectrum, heterogeneous conformations for the V109 sidechain, or motion for the sidechain present in solids. In the second example (Figure 27, panels D-F), the sidechain amide peak for W123 was only detected in the solid-state spectrum and accordingly, only the solid-state NCH spectrum shows a corresponding methyl-amide cross peak (Figure 27, panel E). Again, the estimated distance restraints were searched for possible interaction partners for W123 (Figure 27, panel F). I91 is likeliest to match the ^1H - and ^{13}C chemical shift revealed in the NCH solid-state spectrum. The absence of the W123 sidechain amide peak in the solution spectrum can be explained by the different J -couplings and therefore required INEPT delays for tryptophane sidechain and backbone amides, leading to low signal for tryptophane sidechain amides.

The final manual methyl assignments for hCAII in solution and the solid state are depicted in Table 4. Taken together, 24 unambiguous methyl assignments were achieved manually based on solution and solid-state spectra. This corresponds to 25% of all expected methyl groups for hCAII. For ~42% of the 52 residues at least one assignment was obtained in solution and the solid state via manual assignments. The partition of unambiguous assignments of all assignments in solution and solids is similar with 42% and 50%, respectively.

Out of the 14 assignments for the same nuclei in solution and solids, several of the corresponding peak positions also match in solution and the solid state. The assignments of isoleucine methyl groups in solution and solids strongly resemble each other for I22, I33, and I91 considering a 0.03 and 0.25 ppm tolerance for ^1H and ^{13}C chemical shifts, respectively. This is also the case for one methyl group of L250. The peaks corresponding to V149, V162, and V222 are located in crowded peak areas in the spectrum, which might explain the deviations of the chemical shift assignments for these peaks. Moreover, the assignments for peaks L100, V109, L156, and V206 differ comparing solution and solid-state assignments. Deviations of chemical shifts obtained in solution and the solid state can be expected as the

protein samples were prepared differently involving different buffers. Exemplarily, this is the case for I255 whose ^{13}C chemical shift in solution deviates from the ^{13}C chemical shift in solids. Considering the peak position relative to the other isoleucine peaks, it seems I255 was the only isoleucine whose chemical environment and/or dynamics changed significantly between solution and solids. This hypothesis can be supported by the exposed location of I255, which is close to the C-terminus on the surface of hCAII. The same is true for L100. For V109, however, it can be concluded that the weak cross peak observed in solids to D110 most probably is an artifact as it differs from the results obtained in solution. It is worth noticing that the assignment for L156 in solution was equally likely to match L183. The same assignment uncertainty was revealed for the peaks corresponding to L147 and V217 in solution. Assuming the solid-state peak assignment for L156 is correct, the peak assigned in solution should consequently correspond to L183. The assignment for V206 probably differs as it has many other valine residues in spatial proximity and could have been wrongly assigned either in solution or solids.

Table 4: Methyl ^1H and ^{13}C chemical shift assignments for hCAII. Unambiguous and ambiguous assignments are shown in black and grey, respectively. Geminal groups were treated separately but not assigned stereospecifically. FLYA assignments (kindly shared by Dr. Alexander Klein) were based on the same data sets (chapter 5.8). Manual and FLYA-based ^1H or ^{13}C chemical shift assignments with 0.03 and 0.25 ppm tolerances were matched and marked with an asterisk.

Assigned methyl groups	Manually assigned peaks				Peaks assigned by FLYA			
	Solution		Solid state		Solution		Solid state	
	^1H / ppm	^{13}C / ppm	^1H / ppm	^{13}C / ppm	^1H / ppm	^{13}C / ppm	^1H / ppm	^{13}C / ppm
V31-H _{G1} C _{G1}	0.511	17.942	0.728	21.730	0.511*	17.932*	0.535	18.314
V31-H _{G2} C _{G2}	-0.204	13.245	-	-	-0.202*	13.209*	0.375	15.925
L44-H _{D1} C _{D1}	0.630	22.126	-	-	0.621	22.023*	0.688	21.897
L47-H _{D1} C _{D1}	-	-	0.657	18.573	0.631	21.788	0.719	21.699
V49-H _{G1} C _{G1}	-0.250	16.415	-	-	-0.250*	16.415*	-0.215	16.555
V49-H _{G2} C _{G2}	0.323	16.700	-	-	0.325*	16.601*	0.358	16.379
L57-H _{D1} C _{D1}	0.695	22.425	-	-	0.643	21.372	0.694	22.130
L84-H _{D1} C _{D1}	-	-	0.636	21.957	0.689	23.936	0.744	21.629
L90-H _{D1} C _{D1}	-	-	0.620	21.862	0.861	20.366	0.725	19.901
L100-H _{D1} C _{D1}	0.744	21.908	0.751	20.360	0.757*	21.972*	0.728*	19.227
L100-H _{D2} C _{D2}	0.847	20.471	-	-	0.591	19.606	0.749	23.471
V109-H _{G1} C _{G1}	0.665	18.609	0.526	24.839	0.597	18.416	0.640	18.309
V109-H _{G2} C _{G2}	0.691	18.259	-	-	0.357	16.221	0.342	17.134
L118-H _{D1} C _{D1}	-	-	0.984	20.607	0.533	25.073	0.749	21.737
L118-H _{D2} C _{D2}	-	-	-	-	-0.265	18.057	0.775	19.963
L120-H _{D1} C _{D1}	-	-	-	-	0.374	19.195	0.722	21.743
V121-H _{G1} C _{G1}	-	-	0.738	19.130	1.038	20.366	0.722*	21.743

V134-H _{G1} C _{G1}	0.307	17.335	-	-	0.299*	17.363*	0.76	17.391
V134-H _{G2} C _{G2}	0.945	17.246	-	-	0.569	15.928	0.792	18.694
V142-H _{G1} C _{G1}	-	-	-	-	0.667	15.166	0.701	15.046
V142-H _{G2} C _{G2}	-	-	-	-	-	18.304	0.572	18.347
L143-H _{D1} C _{D1}	-	-	0.427	21.090	0.750	-	0.724	21.700
L143-H _{D2} C _{D2}	-	-	0.214	19.991	-	-	0.749	21.800
L147-H _{D1} C _{D1}	0.977 (or V217)	20.901 (or V217)	-	-	0.973*	21.039*	0.742	20.017
V149-H _{G1} C _{G1}	0.782	17.994	0.704	18.303	0.616	18.387	1.064	20.226
V149-H _{G2} C _{G2}	0.647	17.772	-	-	0.411	16.288	0.788	13.710
L156-H _{D1} C _{D1}	0.983 (or L183)	20.778 (or L183)	0.687	22.276	0.981*	-	0.754	18.628
L156-H _{D1} C _{D1}	0.713	21.572	-	-	0.787	20.862	0.716	21.891
V159-H _{G1} C _{G1}	-	-	-	-	-	18.599	0.792	18.651
V160-H _{G1} C _{G1}	-	-	-	-	-	18.436	0.772	18.448
V162-H _{G1} C _{G1}	0.806	18.308	0.788	19.202	0.264	18.567	0.801*	18.733
V162-H _{G2} C _{G2}	0.816	17.614	-	-	0.251	20.274	0.756	17.430
L163-H _{D1} C _{D1}	-0.687	16.851	-	-	-0.204	20.093	0.702	21.983
L183-H _{D1} C _{D1}	0.983 (or L156)	20.778 (or L156)	-	-	0.971*	20.842*	0.725	21.748
L184-H _{D1} C _{D1}	-	-	0.810	24.017	0.370	22.170	0.756	23.545
L197-H _{D1} C _{D1}	-0.003	17.211	-	-	0.745	18.534	0.739	18.790
L202-H _{D1} C _{D1}	0.627	23.258	-	-	0.657*	23.270*	0.746	23.605
L202-H _{D2} C _{D2}	0.426	18.318	-	-	0.443*	18.394*	0.779	20.287
L203-H _{D1} C _{D1}	-	-	0.795	18.737	0.726	18.544	0.736	18.55*
V206-H _{G1} C _{G1}	-0.662	17.168	0.555	18.291	-	18.431	0.333	17.229
V206-H _{G2} C _{G2}	0.617	20.059	-	-	-	18.628	0.741	18.263
V210-H _{G1} C _{G1}	-	-	0.620	18.531	0.645	18.366	0.547	18.424*
V217-H _{G1} C _{G1}	0.977 (or L147)	20.901 (or L147)	-	-	0.627	18.353	0.526	18.417
V217-H _{G2} C _{G2}	0.453	15.877	-	-	0.453*	15.870*	0.401	12.887
V222-H _{G1} C _{G1}	0.646	17.172	0.672	22.022	0.819	17.338	0.793	20.271
V222-H _{G2} C _{G2}	0.462	18.459	-	-	0.31	18.404*	1.065	13.706
L223-H _{D1} C _{D1}	0.751	18.458	-	-	0.756*	21.967	0.722	18.673
L223-H _{D2} C _{D2}	-	-	-	-	0.927	24.019	0.758	23.292
L239-H _{D1} C _{D1}	-	-	0.113	21.919	0.705	21.923	0.742	20.046
V241-H _{G1} C _{G1}	-	-	1.056	20.299	1.025	18.552	1.070*	20.280*
V241-H _{G2} C _{G2}	-	-	0.780	13.869	-0.664	17.166	0.797*	13.740*
L250-H _{D1} C _{D1}	0.758	21.894	0.897	19.963	0.751*	21.949*	0.738	21.882
L250-H _{D2} C _{D2}	0.561	19.776	0.555	19.744	0.565*	19.708*	0.703	18.571
I22-H _{D1} C _{D1}	0.834	11.264	0.860	11.403	0.841*	11.209*	0.769	11.208
I33-H _{D1} C _{D1}	0.457	9.820	0.439	9.616	-	9.874*	0.479	9.546*
I59-H _{D1} C _{D1}	-	-	-	-	-	9.935	0.816	11.263
I91-H _{D1} C _{D1}	0.487	10.355	0.451	10.166	0.492*	11.311*	0.776	9.703

I145-H _{D1} C _{D1}	-	-	-	-	0.677	11.696	0.160	10.481
I166-H _{D1} C _{D1}	-	-	-	-	-	-	0.783	9.731
I209-H _{D1} C _{D1}	-	-	0.127	10.808	0.488	10.700	0.155*	10.526
I215-H _{D1} C _{D1}	0.772	9.516	-	-	0.778*	9.466*	0.469	10.025
I255-H _{D1} C _{D1}	0.448	10.985	0.416	10.266	0.443*	10.975*	0.475	9.798
	Manual assignments				FLYA assignments			
	Solution		Solid state		Solution		Solid state	
Assigned methyl groups	¹H and ¹³C / ppm		¹H and ¹³C / ppm		¹H / ppm	¹³C / ppm	¹H / ppm	¹³C / ppm
All assigned peaks	38		28		53	58	62	62
All assigned residues	26		25		38	42	44	44
Unambiguously assigned peaks	16		14		22	26	2	2
Unambiguously assigned residues	14		14		15	17	2	2
All assignments confirmed by FLYA	22		9		-	-	-	-
Ambiguous assignments confirmed by FLYA	13		5		-	-	-	-

Looking back at the water-methyl NOESY contacts depicted in panel B in Figure 24, the residues identified for no or low-intensity cross peaks should be mostly located close to hydrophobic patches. This statement can be confirmed for I91 and L134, which are close to the hydrophobic part of the binding pocket, which is in the vicinity of the indole moiety of ligand SBR in the complex structure. The other peaks with weak water contacts were mostly not assigned, except for I22. Surprisingly, I22 showed only a low-intensity water peak, although it is located at the hCAII protein surface. It can be speculated that other contributions like sidechain flexibility interfered with the magnetization transfer for this specific cross peak, as I22 is part of the N-terminus of hCAII.

Moreover, automated FLYA^[7-8] assignments were obtained by Dr. Alexander Klein, which were compared to the manual assignments (Table 4). FLYA has been developed mainly for small proteins and solution NMR data sets. For larger protein targets, usually additional data like NMR spectra of mutants have been utilized. In addition, peak linewidths are much broader in solids compared to solution, adding to the deviations of the peak centres. Therefore, the FLYA algorithm parameters were edited and implemented for this purpose. The BMRB database was used as reference for chemical shifts. Strong assignments by FLYA were considered unambiguous.

In total, 29 unambiguous FLYA assignments in solution and only two unambiguous assignments in solids were achieved, considering either both or one of the predicted ^1H and ^{13}C chemical shifts per methyl group. Evidently, the FLYA assignments were less successful for the solid-state spectra. The reason was probably the lower number of cross peaks in the 3D NCH, CCH, and HCH spectra in addition to the overall broader peaks and higher chemical shift deviations in the solid-state spectra. Matching all manual and FLYA-based assignments, 59% of the ambiguous manual solution assignments and 36% of the ambiguous manual solid-state assignments could be confirmed by FLYA, considering only methyl groups for which both ^1H and ^{13}C chemical shifts have been predicted by FLYA. Consequently, 13 ambiguous assignments in solution and five ambiguous assignment in solids were supported by matching FLYA assignments and thus were considered unambiguous. Among those was the solution NMR peak which was either attributed to L156 or L183 based on manual assignments. FLYA predicted a similar assignment for L183, thus this peak likely does not correspond to L156. The same has been indicated by the manual solid-state assignment for L156, as mentioned above.

In summary, both the manual and FLYA-based methyl assignments were more successful based on solution compared to solid-state NMR data sets. In the solid state, anisotropic contributions like ^1H - ^1H dipolar couplings and the associated dipolar truncation broaden and reduce the number of cross peaks (chapter 1.1.2). Unfortunately, long-range cross peaks are mostly affected, which are especially valuable during assignments. While slightly more unambiguous assignments were obtained by FLYA than via manual assignment in solution, FLYA struggled with extracting unambiguous assignments from the solid-state spectra. For these more challenging data sets, handling the data manually achieved more unambiguous assignments. However, any assignments based on more noisy data bear a higher probability for false assignments.

Comparing these assignments to the published ^{13}C -methyl sidechain chemical shifts^[70], roughly 50% of the ^{13}C assignments match assuming a tolerance of 0.5 ppm (Table 29). Unexpectedly, no isoleucine ^{13}C chemical shifts match, although the isoleucine assignments in solution obtained here are considered unambiguous. Most likely, the other 50% of assignments deviate due to different sample conditions.

At the end, 52 out of 86 methyl peaks of the hCAII sidechain methyl groups have been at least partially assigned, including residues I91, V121, V134, L197, and L203, which are close to the

ligand binding site, assuming similar chemical shifts in solution and the solid state. Especially in solution, the time-shared 3D XXH experiments delivered optimal data for methyl assignments, reflected both by manual and automated assignments. Furthermore, manual and automated approaches served each as verification tools for each other as the potential bias by the user or the algorithm should differ. Of course, peak picking represents another source of errors, which equally influences manual and FLYA-based assignments. Overall, ^1H - and ^{13}C -peak assignment pairs for valine and leucine residues as well as for the $\text{H}_\text{D}\text{C}_\text{D}$ -methyl groups of isoleucine residues have been successfully achieved, therefore these new assignments promise to serve well as reporters for structural and dynamics-related investigations towards protein methyl sidechains.

2.3 Conclusion and outlook

In summary, several functionally relevant features of hCAII-ligand interactions were revealed in addition to improved NMR sample preparation strategies and establishing new solid-state NMR experiments for detecting ligand dynamics and assigning methyl sidechains.

While ligand soaking approaches are often used to obtain X-ray structures, co-crystallization for the hCAII-SBR complex led to higher ligand occupancies (PDB: 8R1I) required for NMR studies of the ligand in protein-bound state. These optimizations were essential to detect ligand signal and to successfully establish the ^1H - ^1H NERRD experiment to site-specifically identify ligand dynamics within the binding pockets of a protein. Prerequisites are that the protein is isotopically labelled, and that the ligand has non-exchangeable protons. Here, the benzene ring flip of the ligand SBR while bound to hCAII was sensed by several spins in the NERRD experiment, revealing the extent of the intra-ligand proton dipolar coupling network. The dynamics could only be assigned site-specifically with the help of MD simulations. In addition to the microsecond dynamics of the benzene ring, the indole group of SBR was revealed to undergo faster motion while bound to hCAII. Experimentally, the challenges associated with the quantification of proton relaxation still have not been solved and remain a limitation of this method. It would be necessary to selectively disrupt the proton network. The currently available techniques for this purpose mainly comprise of deuteration and MAS. Homonuclear couplings unfortunately cannot be disrupted via pulsing. Also, it remains questionable if this technique can be developed further to serve drug binding studies on a high-throughput scale.

Furthermore, the question remains how high-affinity interactions and ligand flexibility can be present at the same time. The high affinity of SBR was mainly assigned to its larger interaction surface compared to e.g. acetazolamide. As expected for a rigid complex, micro- and millisecond dynamics present for the backbone of apo-hCAII backbone are abolished upon ligand binding. The binding of a flexible ligand seems to represent an energetic gain high enough for near-covalent binding conditions with 30 pM affinity^[16]. It can be suggested that internal dynamics in the ligand can be tolerated or even be advantageous. Previous research has shown that remaining ligand dynamics reduce enthalpic contributions but increase binding affinity from the viewpoint of entropy^[109]. This apparent contradiction calls for further experiments. Isothermal calorimetric investigation of ligand binding can identify the individual enthalpic and entropic contributions, for example, owing to displaced water

molecules or interactions with F130 via π -stacking, which might result in lower SBR affinity when mutated. Studying sidechains close to the ligand binding site could complement the ligand relaxation data and uncover how high affinity and dynamics can fit into the same picture. At the binding site, the methyl sidechains of one isoleucine (I91), four valine (V121, V134, V142, V206), and three leucine (L140, L197, L203) residues face the ligand SBR, according to the X-ray structures. Also, faster backbone and water network dynamics could occur, which are, however, difficult to capture via NMR spectroscopy.

Here, 52 ILV-methyl peaks have been assigned, at least partially, based on manual and FLYA-assignments derived from time-shared 3D XXH NOESY and RFDR experiments. Sample preparation, measurement time, and high spectral quality were more favourable in solution compared to the equivalent time-shared experiment in the solid state. Consequently, the time-shared solid-state experiment established here is only recommended for proteins for which solution NMR studies are not a viable option. As the solid-state data were of lower quality, the automated FLYA-based assignments were less successful in the solid state compared to solution or the manual assignments. Alternatively, sidechain-to-backbone experiments correlating intra-residual amide and methyl groups could be used to deliver higher assignment success in the solid state^[110]. These experiments, however, require ¹³C-labelling of the whole sidechain, which are not well-suited for subsequent ¹³C-methyl relaxation experiments, for example.

Overall, the understanding of ligand binding and ligand dynamics in the hCAII-SBR complex in solution and crystals as well as hCAII methyl assignments could be advanced. Moreover, methods to detect proton relaxation of a ligand in protein-bound state and methyl assignments in the solid state have been developed and established. It should be emphasized that several combinations of methods were used to clarify how to interpret the data and to validate new methods. Especially X-ray crystallography and solid-state NMR, NMR relaxation and MD simulations, and manual and automated peak assignments complemented each other well. Further research on hCAII using the ligand dynamics tool established here could focus on how ligand dynamics are affected by post-translational modifications, mutations, other CA isozymes, or promiscuous hCAII variants catalysing reactions relevant for the chemical industry. In addition, the newly obtained assignments of hCAII methyl groups could be used as reporters on sidechain dynamics in the binding pocket.

3. The intramembrane protease GlpG

3.1 Introduction

3.1.1 Membrane proteins and their isolation from the native environment

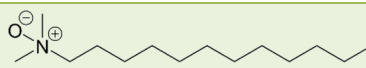
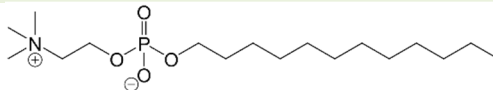
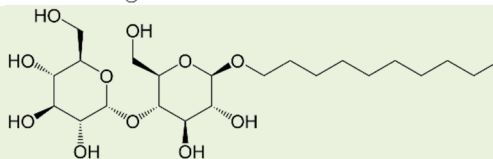
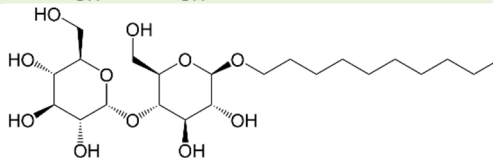
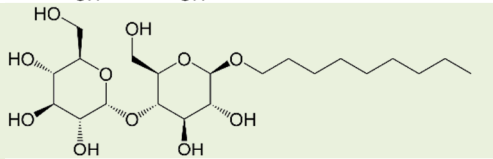
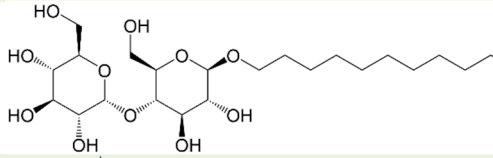
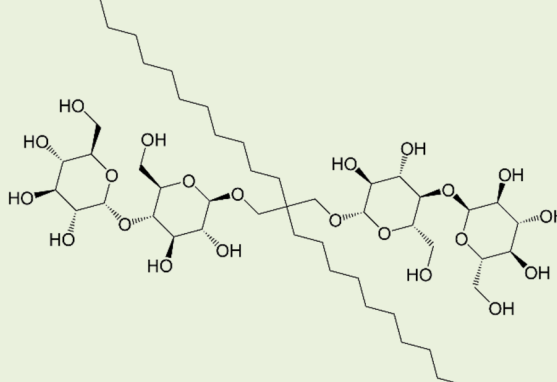
Membrane proteins represent diverse types of proteins associated with, embedded in, or anchored to lipid bilayers, which include, for example, the plasma membrane, membranes of organelles, and vesicles. Within these lipid bilayers, membrane proteins account for roughly 50% of the cellular plasma membrane mass^[111]. The importance of membrane proteins is reflected by their functional contributions in living cells and organisms, including adhesion, signalling, and transport via membranes^[112]. Therefore, roughly 26% of the protein-encoding genes in humans serve for membrane protein production^[113]. From a pharmaceutical point of view, membrane proteins are of interest, too, as they account for ~40% of all FDA-approved drugs targets^[113b]. For example, G-protein coupled receptors like adenosine receptors have already been used or are currently investigated to treat various diseases from asthma to Parkinson's disease^[114]. Targeting ligand-gated ion channels like inhibitory type A γ -aminobutyric acid (GABA_A) receptors can induce anaesthesia^[115] while e.g. voltage-gated calcium channels are involved in conditions like hypertension^[115].

To understand the various functions of membrane proteins and to potentially target them to treat diseases, they need to be isolated for structural and functional studies. The PDB database provides access to published protein structures and, in 2023, only identified ~6% of all published protein structures as entries annotated as membrane protein structures^[17]. To isolate proteins, they need to be produced at larger scale first. Overexpression is a commonly used tool to increase the yields of proteins when their natural abundance is low. Unfortunately, especially for membrane proteins, it can induce cellular stress responses due to, for example, protein aggregation or toxicity related to their native function. Membrane proteins are mostly extracted from their native membranes via solubilization with detergents. The native environment of transmembrane proteins includes the membrane, surrounding the transmembrane domain with a belt of lipids, and two hydrophilic environments with different ionic strengths, pH values, and viscosities, i.e. the extracellular space and the cytoplasm. In the lipid bilayer, membrane proteins are tightly engaged with certain lipids and other membrane proteins. In 1972, Singer and Nicolson described the random motion of membrane

proteins in the two-dimensional space of the lipid bilayer as fluid mosaic model^[116]. During the following years, it was discovered that the lateral diffusion of membrane proteins can be restricted or facilitated by various interaction partners like lipids, membrane or soluble proteins, or the cytoskeleton^[117]. For example, lipids in eukaryotic membranes were found to assemble into rafts, which many membrane proteins seem to interact with^[118]. Also bacterial membranes were hypothesized to form microdomains with clusters of lipids in specific regions^[119]. In addition, the constitution of membranes differs between the two layers of the bilayer. In Gram-negative bacteria like *E. coli*, the cellular envelope is even more diversified including two separate bilayers, the inner and the outer membrane, enclosing the periplasmic space and the peptidoglycan layer. The outer membrane is asymmetric and contains mainly phosphatidylethanolamine, lipopolysaccharides, and proteins^[120]. The inner membrane, however, is symmetric. Its lipid composition includes 75% phosphatidylethanolamine, 20% phosphatidylglycerol, and 5% cardiolipin^[121].

Solubilisation strategies and vehicle systems, in which the membrane proteins are kept during purification and measurements, can be chosen from detergent micelles/bicelles/liposomes, amphipols, styrene and maleic acid copolymers, and lipid-protein nanodiscs^[122], which shield the hydrophobic transmembrane patches from the surrounding water molecules. Amphipols are amphipathic polymers wrapping around the hydrophobic transmembrane regions of the membrane protein. Styrene-maleic acid copolymer-lipid and protein-lipid nanodiscs consist of lipids sandwiched between the membrane protein and a belt of polymers or proteins, respectively, mainly interacting with the hydrophobic lipid tails^[122-123]. The most commonly used vehicles are detergent-based^[122]. In micelles, detergent monomers form a sphere-like structure around the hydrophobic region of the proteins. Bicelles form a membrane-like structure around the hydrophobic transmembrane region of the protein while cone-shaped detergents cover the edges of the bilayer in a half-ball shape. Liposomes embed one or several proteins in lipid bilayers which are slightly bent and form overall a ball-shaped complex. Evidently, membrane proteins experience a more native-like environment when they are surrounded by lipids instead of detergents, decreasing the likelihood for them to aggregate, to degenerate, or to lose their native function. However, using polymers or nanodiscs does not guarantee the native state of proteins. For example, they can wrap around membrane proteins too tightly and alter their folding or flexibility.

Table 5: Selection of different detergents sorted by the molecular weight of the micelles. The detergent name, structure, Critical Micelle Concentration (CMC), and Molecular Weight (MW) of the micelle are shown.

Detergent name (abbreviation)	Detergent structure	CMC / % (w/v)	MW / kDa
<i>n</i> -Dodecyl- <i>N,N</i> -dimethylamine- <i>N</i> -oxide (LDAO)		0.023	17
<i>n</i> -Dodecylphosphocholine (FC-12)		0.047	19
<i>n</i> -Octyl- β -D-maltopyranoside (OG)		0.53	8 - 29
<i>n</i> -Decyl- β -D-maltopyranoside (DM)		0.087	33
<i>n</i> -Nonyl- β -D-maltopyranoside (NG)		0.2	41
<i>n</i> -Dodecyl- β -D-maltopyranoside (DDM)		0.0087	4 - 76
2,2-Didecylpropane-1,3-bis- β -maltopyranoside (LMNG)		0.001	Large

Detergents can be characterized by their structural features, harshness, critical micelle concentration (CMC) and micelle size. The CMC is the concentration above which the detergent monomers start to aggregate and form micelles. Ionic detergents with a non-zero net charge like SDS are considered harsh, zwitterionic detergents with a zero net charge like LDAO are characterized as less harsh, while nonionic detergents like DM are viewed as mild detergents. A selection of popular detergents is shown in Table 5. For example, γ -secretase was shown to be active when solubilized with the detergent CHAPSO while it was inactive in triton X-100^[124] or DDM^[125]. Rhomboid YqgP from *Bacillus subtilis* was observed to be active

in DDM, digitonin, and NG and was reported to be inactive in CHAPS or CHAPSO micelles^[126]. It is thus advisable to screen various detergents for solubilization to determine whether the membrane protein retains its activity.

Understanding the highly specialized native environment of membrane proteins helps to answer why their functions and native structures can change when they are removed from this environment. As mentioned before, it is unavoidable to transfer membrane proteins to a vehicle for purification. Meanwhile, the overall aim is to study the membrane proteins under conditions which retain the correct fold and activity and mimic the native environment. Measures to evaluate the state of the protein are, for example, activity or interaction assays, and the dispersion and peak linewidths in NMR spectra.

3.1.2 The family of rhomboid-like proteases

Proteases are proteins which cleave other proteins and therefore can indirectly affect essential cellular functions like cellular metabolism, cell division, and cell death. Rhomboid-like proteases are membrane proteins and have the powerful ability to irreversibly activate membrane-bound proteins by releasing them to e.g. the cytoplasm^[127]. In addition, they are ubiquitously present in all orders of life^[128]. The rhomboid-like family comprises active proteases called rhomboids, inactive rhomboid proteins (iRhoms), and another group of non-associated, inactive rhomboid-like proteins^[129]. In mammals, Rhomboid-related protein (RHBDL)1-4 and Presenilin-associated rhomboid-like protein (PARL) are the only active rhomboids which have been discovered so far^[130]. Studies suggest that RHBDL2 is involved in wound healing and impacts tumour proliferation, and RHBDL4 is associated with colorectal cancer and involved in amyloid precursor protein fission associated with Alzheimer's disease^[1]. Many inactive rhomboid-like pseudoproteases were discovered as well like the protein Derlin, which is involved in the endoplasmic reticulum-associated degradation pathway^[131]. In *Drosophila*, the first member of the rhomboid-like family was identified by random mutations leading to a phenotypic change of the embryos when the sequence of Rhomboid-1 was altered^[132]. Subsequently, it was demonstrated that Rhomboid-1 releases the soluble domain of the protein Spitz from its membrane anchor and activates the Epidermal Growth Factor Receptor^[133]. Being involved in such a meticulously orchestrated process like embryogenesis, the activity of Rhomboid-1 demands to be highly regulated with precise temporal and spatial resolution. Rhomboid-7 was found to influence mitochondrial

fusion events in *Drosophila*^[134]. Interestingly, rhomboid proteases from malaria parasite *Plasmodium falciparum* enable the cell entry and hence could potentially be targeted to treat infectious diseases like malaria^[135].

Structurally, rhomboid-like proteases are characterized by a catalytic dyad formed by S201 and H254 in the transmembrane region^[130]. Rhomboid-like serine proteases are not evolutionally related to soluble serine proteases and have a high sequential variety within their family^[129]. GlpG from *E. coli* shares roughly 39% of its protein sequence with GlpG from *Haemophilus influenzae*, while it is 26% with Rhomboid isoform A from *Drosophila melanogaster*, 22% with rhomboid protease 1 from *Plasmodium falciparum* NF54, and 25% with RHBDL4 isoform X1 from *Homo sapiens*, according to the protein database of the National Library of Medicine^[17, 136]. The transmembrane domain of the members of the rhomboid-like family consists of at least six transmembrane α -helices (TM1-6, see Figure 28). Interestingly, residues from TM2, TM4, and TM6 including residues S201 and H254 are conserved among several rhomboid-like proteases, including GlpG from *E. coli*^[129].

Overall, rhomboid-like proteases are ubiquitously present and are implicated in several important cellular functions. While it might be difficult to generalize structural and functional characteristics, given the large diversity among the family members, they share certain similarities. Understanding the catalytic mechanism and interactions of rhomboid-like family members is not only essential to identify in which signalling pathways they are involved but it also provides leverage to regulate disease-associated processes.

3.1.3 The structure and proteolytic activity of GlpG

GlpG is an active rhomboid-like serine protease from *E. coli* with a catalytic dyad involving S201 and H254. While part of the cleavage mechanism of GlpG is understood, native substrates remain elusive. GlpG is often used as model protein representing the rhomboid-like family due to its robust activity and overall stability. GlpG can withstand up to 60 °C, a pH range of 5.5 to 8.5, an ionic strength between 50 and 250 mM, and 0-10 mM concentrations of divalent ions. Like other membrane proteins, GlpG cannot tolerate low salt concentrations beyond 25 mM and, of course, detergent concentrations below the CMC^[137].

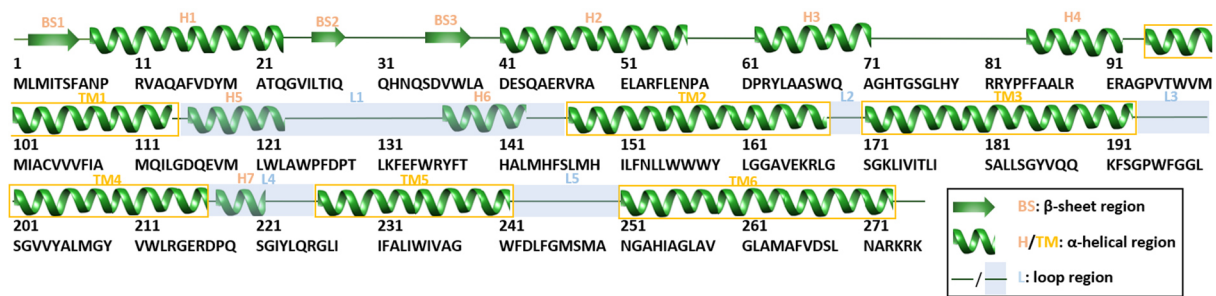


Figure 28: Predicted secondary structure elements aligned with the sequence of GlpG (AlphaFold: P09391^[19]).

Several structures of GlpG constructs and protein-ligand complexes have been published with the first structure solved by Wang *et al.* via X-ray crystallography in 2006^[138]. Structurally, GlpG consists of a soluble N-terminal domain (amino acids 1-86) and the GlpG core with TM1-6 (amino acids 87-276) (Figure 28). While the N-terminus, C-terminus, L2, and L4 face the extracellular space, the loops L1, L3, and L5 are exposed to the periplasmic space in *E. coli* membranes. In most crystal structures, the N-terminal domain of GlpG was cleaved off as no full-length structure could be crystallized up to now. In addition, the sequence of the N-terminal domain is not conserved and its absence does not abrogate GlpG activity^[139]. These reasons contributed to the N-terminal domain often being neglected. On the other hand, the N-terminal domain is supposed to drive dimerization via domain swapping^[140]. The relevance of dimer formation is given by the *in vivo* occurrence of GlpG dimers and supposedly improved activity of the dimer in DDM over the monomer in DM micelles^[141]. While no structure of full-length GlpG was determined experimentally, AlphaFold predicted a model structure which aligns very well with the structures of the N-terminus domain^[142] and GlpG core^[19, 138a].

Overall, X-ray crystallography structures of GlpG core revealed two different conformations. Upon catalysis, GlpG was suggested to partially unfold the substrate to allow access to the active site of GlpG^[143]. This active conformation was proposed to have L5 lifted outwards and TM5 in a bent conformation exposing the active site^[139, 144]. In the apo state, a closed conformation supposedly shields the active site^[138a]. Structurally, this was described by L5 being folded inwards and a straightened TM5^[139, 144]. The active site is built around the catalytic dyad with S201 and H254, which are part of TM4 and TM6, respectively. Mutating S201 of GlpG to alanine, cysteine, or threonine results in a loss of function and a slightly decreased melting temperature^[137]. Similarly low activity levels but less changes in the melting temperature were reported for mutants H254A and H254K^[137]. Notably, mutations in TM5 enhanced the activity of GlpG by up to ten-fold in contrast to mutations in L5, which did

not increase the activity of GlpG^[9,10]. MD simulations also confirmed that TM5, L4, and L5 retain the highest structural flexibility compared to the other transmembrane helices and loops^[145]. The role of L4 remains largely elusive but mutations of hydrophobic amino acids in this loop decreased the activity of GlpG from *Haemophilus influenzae*^[146]. Anchoring TM2 to TM5 or disabling interactions between L1 and the core decreased GlpG activity^[147]. Mutations in L1 can diminish GlpG activity as well^[142, 148]. Based on the high degrees of flexibility of L5 and TM5, they both were suggested to regulate GlpG activity each or cumulatively. L1 is either affected by GlpG activity regulation or adds to it indirectly (Figure 31).

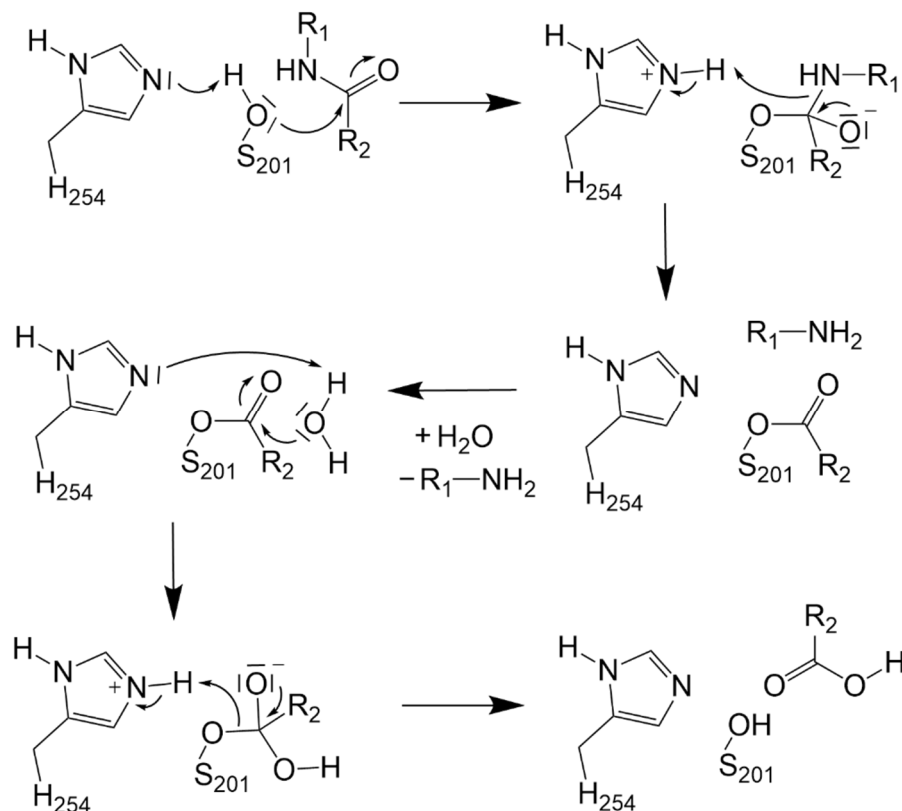


Figure 29: Schematic catalytic mechanism of rhomboid-like serine proteases, exemplified for GlpG from *E. coli*.

Another essential component for proteolysis is the access of water molecules to the active site. Other than the hydrophilic surface of the lipid bilayer, the inside of the bilayer and the transmembrane surface of GlpG are dominated by hydrophobic moieties. Therefore, the supply with water at the active site needs to be enabled via an intrinsic “channel” within GlpG. Some water molecules around the catalytic dyad of GlpG were resolved in the X-ray structures, which likely take part in the proteolysis reaction^[138c, 144]. In addition, MD simulations indicated that one internal water retention site seems to be conserved, which is enclosed by H141, S181, S185, Q189, G199, G202, and V203^[145].

Interestingly, GlpG was claimed to actively narrow its surrounding lipid bilayer^[149]. The lipid composition seems to play a role in membrane thinning as well^[149b]. A narrower lipid bilayer also increases the water accessibility, especially close to L1^[149a]. This coincides with the findings that GlpG activity is solely measured in detergent micelles with alkyl chains of 10-12 carbons, while it is abrogated in detergents with longer alkyl chains^[150]. The thickness of the hydrophobic transmembrane region of GlpG was estimated to be ~20-30 Å with the catalytic S201 residue buried a few Ångströms deep in the intramembrane region^[149, 151]. Bondar *et al.* furthermore proposed that the location of the catalytic residue S201 in the lipid bilayer complex changes over time and depends partly on the surrounding lipid species^[148b]. They found indications that L1 influences the angular orientation of GlpG within the bilayer, too^[148b]. The connection between membrane thinning and GlpG activity is not yet clear.

In terms of solubilization vehicles, GlpG retains a very similar conformation in micelles and DMPC/CHAPSO bicelles with small differences in the loop regions, e.g. L5, comparing various X-ray structures^[151]. It is worth noticing that some GlpG structures show that detergent and lipid monomers protrude into the GlpG transmembrane domain and seem to interfere with the active site^[151]. However, this does not seem to impede the activity of GlpG core in detergent micelles of DDM, DM, and FC-12 with alkyl chains of 10-12 carbons and could be an artifact^[150, 152]. In phospholipid bilayers, GlpG has lower activity^[150, 152].

Overall, many details on GlpG catalysis including protein dynamics remain to be clarified. The process of substrate recognition, bending and release, as well as the transport of water molecules to the active site need to be studied further^[127a].

3.1.4 Substrates and inhibitors interacting with GlpG

Although the native substrates of GlpG in *E. coli* are unknown, artificial single-pass helical transmembrane peptides have been identified, which GlpG can cleave^[153]. These artificial substrates were modified from the native substrates Spitz, TatA, and Gurken of other rhomboid-like proteases. The peptide Spitz from *Drosophila* was identified to be cleaved by Rhomboid-1 within the EGFR signalling cascade^[154]. TatA is involved in inter-cellular communication in *Providencia stuartii*, which requires its cleavage by AarA^[155]. In *Bacillus subtilis*, YqgP was found to cleave the substrate Gurken^[153]. In addition, the TM2 helix of the lactose permease LacY (LacYTM2) from *E. coli* was isolated to serve as GlpG substrate^[156]. Surprisingly, these substrates do not share the same amino acid sequence around the

cleavage site (Figure 30)^[153]. Instead, several TM-helix-destabilizing amino acids like proline, glycine, serine, and glutamine were found next to the cleavage site^[153]. Notably, all cleavage sites of the single-pass transmembrane substrates shown here can be found at the edge or outside of the transmembrane region (Figure 30).

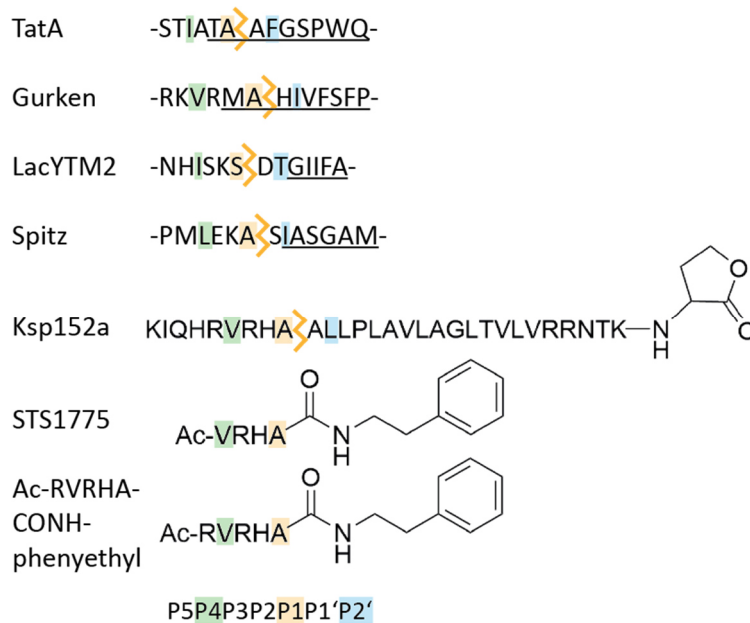


Figure 30: Structures of GlpG inhibitors and substrates^[153]. Positions P5-P2' are highlighted by coloured backgrounds. Cleavage sites are indicated by yellow lightning bolts (between P1 and P1'). Predicted transmembrane regions are underlined^[157].

Mutating the amino acids close to the cleavage site, several preferred characteristics for each position could be extracted^[153]. The residues flanking the cleavage site towards the N-terminus of the substrate are called positions 1-5 (P1-5). The residues towards the C-terminus of the substrate were dubbed 1'-5' (P1'-5'). Substituting residues at P1 for TatA, amino acids with small sidechains were preferred for TatA to be cleaved by AarA^[153]. It was also revealed that exchanging residues at P1, P4, and P2' had the highest impact on cleavage by AarA, favoring hydrophobic side chains at P2' and large hydrophobic sidechains at P4^[153]. The requirements for these three critical positions are in congruence with the residues at the corresponding positions in Gurken, LacY^{TM2}, and Spitz (Figure 30).

Small molecules and peptide-like inhibitors have been designed to react with the catalytic S201 mimicking an intermediate step of substrate cleavage at the binding site. Among these small molecules, phosphonofluoridates^[158], isocoumarins^[159], β -lactams^[160], and β -lactones^[161] have been designed to bind to GlpG. Moreover, inhibitors including small peptide sequences were created, including peptidyl chloromethylketones^[162] and peptidyl

aldehydes^[138b]. Unfortunately, they bind to GlpG rather weakly with low selectivity and potency^[163]. Instead, *N*-substituted peptidyl ketoamides can selectively bind rhomboids with nanomolar affinities in a slow non-competitive fashion^[163]. Despite the formation of a covalent bond with S201, the interaction is reversible^[163]. Several parameters like the modification at the ketoamide nitrogen, the length of the peptide sequence, and the warhead at the C-terminal alanine residue have been optimized. The inhibitor achieving the highest affinity and inhibition was a peptidyl ketoamide with a 4-phenyl-butyl modification at the ketoamide nitrogen and an acetylated peptide sequence of Ac-RVRHA^[163]. This compound selectively binds GlpG compared to rhomboid-like proteases from mice and flies^[163]. The slightly modified compounds STS1775 and Ac-RVRHA-phenylethyl (Figure 30) showed similar affinities and inhibition potencies for GlpG^[163].

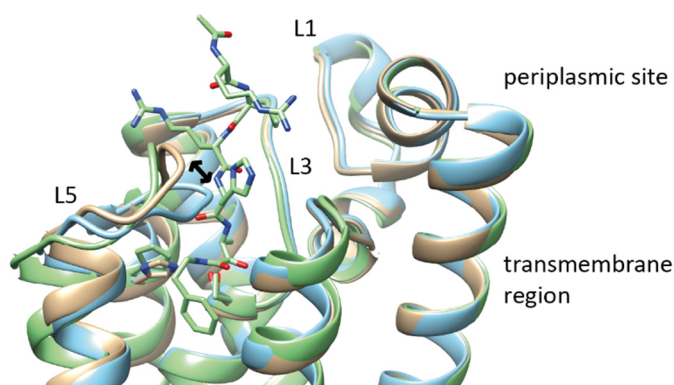


Figure 31: Comparison of GlpG conformations. Apo-GlpG structure in the active conformation (beige, AlphaFold: P09391^[19]), apo-GlpG in the inactive conformation (blue, PDB: 2IC8^[138a]), and GlpG in complex with a peptide-like inhibitor (green, PDB: 5MT6^[163]). The arrow indicates the different conformations of L5.

As mentioned in chapter 3.1.3, the substrate interaction with GlpG has been proposed to consist of multiple steps of recognizing, unfolding, and binding the substrates. This process is likely simplified for binding artificial inhibitors. Still, GlpG might need to undergo some conformational rearrangements like lifting L1 to bind them (Figure 31). To understand the different contributions, the non-helical inhibitor STS1775 and several helical substrate-like inhibitors were studied and will be discussed in the following chapters.

3.2 Results and discussion

The transmembrane protease GlpG is a representative of the family of rhomboid-like proteases whose human members are associated with neurodegenerative diseases and cancer^[1]. Up to now, the N-terminal domain of GlpG (residues 1-86) and GlpG core (residues 87-276) have been studied separately (Figure 32). Apo and inhibitor-bound states have been compared via NMR spectroscopy or X-ray crystallography solely for GlpG core (Figure 32), assuming the N-terminal domain to be negligible. Although cleaving off the N-terminal domain did not result in large chemical shift changes for the core in the solid state^[164], its absence was reported to affect the activity of GlpG^[142]. Studying the full-length construct should capture the enzyme in a more native-like state. To understand how dynamic these proteases are and how they bind substrates, the millisecond motion of full-length GlpG was investigated here in apo, inhibitor-, and substrate-bound form. In the following chapters, GlpG samples for solution NMR measurements were prepared and optimized (chapter 3.2.1) to assign the GlpG peaks (chapter 3.2.2), to study inhibitor binding (chapter 3.2.3), and to determine GlpG dynamics (chapter 3.2.4). As the sensitivity of solution NMR experiments suffers from slow tumbling speeds, GlpG was studied in detergents and not in liposomes or bilayers. In crystals, the full-length GlpG construct has never been crystallized in contrast to GlpG core. It was attempted to reproduce GlpG core crystallization to prepare solid-state NMR samples, but these screenings had not been successful (chapter 3.2.5).

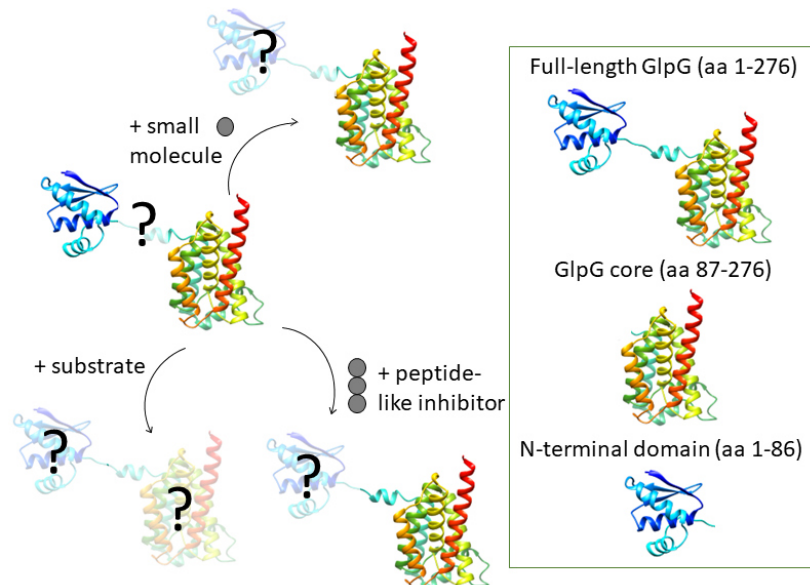


Figure 32: Scheme highlighting published GlpG constructs and complexes, based on the predicted AlphaFold structure for full-length GlpG (P09391^[19]). Question marks indicate missing information. aa: amino acids.

3.2.1 Optimizing the preparation of full-length GlpG NMR samples

Different *E. coli* strains and protocols were tested to optimize the expression of full-length GlpG wild type. Standard cell lines like BL21 cells but also C41 and C43 cells, optimized for the expression of toxic or membrane protein constructs, were considered. The usage of Lemo21 cells allows to limit, for example, protein aggregation by adding rhamnose, which regulates the T7 RNA polymerase activity^[165]. Moreover, LOBSTR cells are well-suited for His-tagged proteins as the native His-rich proteins ArnA and SlyD were modified in *E. coli* to decrease contaminations in His-affinity chromatography experiments^[166]. Testing these different cell lines, the expression yields for GlpG were highest in LOBSTR and C43 cells (Figure 33, panel A). In LOBSTR cells, fewer contaminating proteins were detected on the SDS-PAGE compared to the sample derived from C43 cells. Conversely, C41, BL21, and Lemo21 cells delivered lower GlpG yields (Figure 33, panel B). In Lemo21 cells, the GlpG yields were maximized when no rhamnose was added, which defeated the purpose of using this cell line (Figure 61). High ratios between the protein of interest and other proteins usually lead to higher yields and purity of the final samples. Therefore, LOBSTR cells were selected.

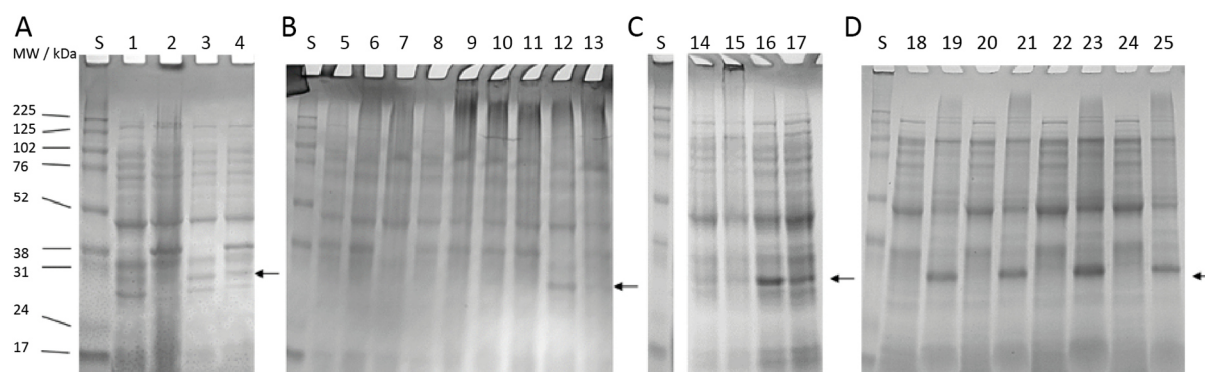


Figure 33: SDS-PAGEs showing optimizations of the expression of full-length GlpG wild type. S: Amersham ECL Rainbow Marker. The arrows indicate the GlpG band matching ~33 kDa. **A.** Expression of GlpG (pET25) in labelled M9 medium using the alternative protocol (see chapter 4.9.2), growing the cells in LB medium and transferring them to labelled M9 medium in 100% D₂O directly. 1: LOBSTR cells before induction, 2: C43 cells before induction, 3: LOBSTR cells after expression, 4: C43 cells after expression. **B.** Expression of GlpG (pET25) in M9 medium using the adaptive protocol. Before induction and after expression, respectively: 5, 10: Lemo21 cells, 6, 11: C41 cells, 7, 12: LOBSTR cells, 8, 13: BL21 cells. **C.** GlpG expression in unlabelled M9 medium in D₂O using LOBSTR cells. 14: Alternative protocol plus BioExpress, 15: alternative protocol plus ISOGRO, 16: adaptive protocol, 17: alternative protocol. **D.** Different incubation times before induction with IPTG in M9 medium according to the alternative protocol. Each an aliquot before and after induction was taken. 18-19: one hour, 20-21: two hours, 22-23: three hours, 24-25: four hours incubation time.

After selecting the cell strain, different GlpG expression protocols and media supplements were tested in combination with M9 medium (Figure 33, panels C and D). The protocol used for expression in deuterated media, which slowly adapts the cells to grow in D₂O, was called

“adaptive protocol” (chapter 4.9.2). A protocol to circumvent time-consuming adaptation steps and to lower the chances of arrested cell growth^[167] was called “alternative protocol”. Here, the cells were grown to a high density in large LB medium cultures, then pelleted, washed, and resuspended in isotopically labelled medium. After transferring the cells to the new medium, the expression was only induced after a certain time. The M9 media were supplemented with D-, ¹³C-, ¹⁵N-labelled BioExpress or ISOGRO, directly providing labelled amino acids which the *E. coli* cells can readily use for translation. Surprisingly, the addition of both supplements did not result in increased GlpG expression yields comparing the relative intensities of the band corresponding to GlpG with a non-identified band at ~50 kDa (Figure 33, panel C). Looking at the adaptive and alternative protocols, the expression culture using the adaptive protocol reached higher GlpG levels for a similar cell density (Figure 33, panel C). On the other hand, roughly three- to four-fold higher cell densities were reached with the alternative protocol. Therefore, and because the expression takes half as long, the alternative protocol was chosen. However, this strategy bears the risk of delivering lower isotopic labelling levels. Moreover, the incubation time before induction with IPTG was optimized (Figure 33, panel D). It could be shown that the yields were slightly higher after three hours of incubation. ¹H-¹⁵N NMR spectra revealed that slightly higher peak intensities were achieved for the samples corresponding to three to four hours incubation time (data not shown). Therefore, GlpG was subsequently expressed in LOBSTR cells using M9 medium without supplements at 18 °C overnight according to the procedures of the alternative protocol with three hours incubation time after resuspending the cells.

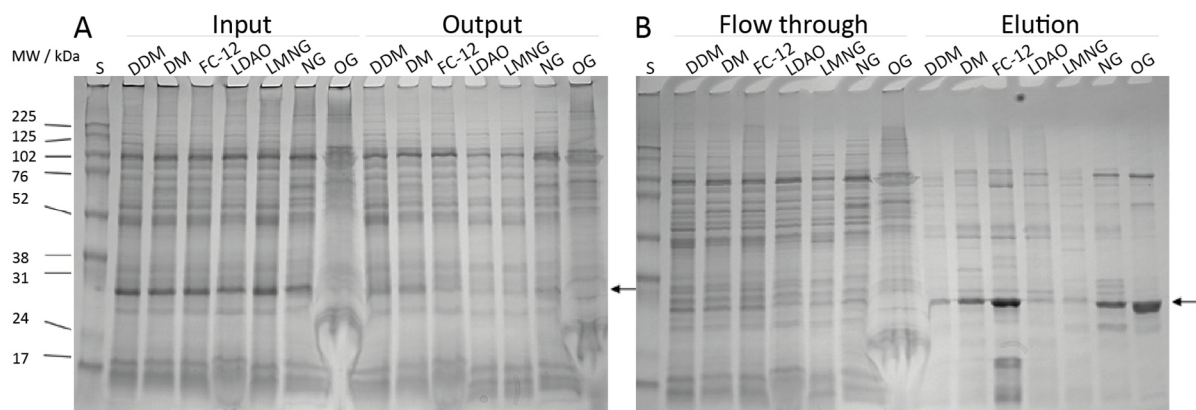


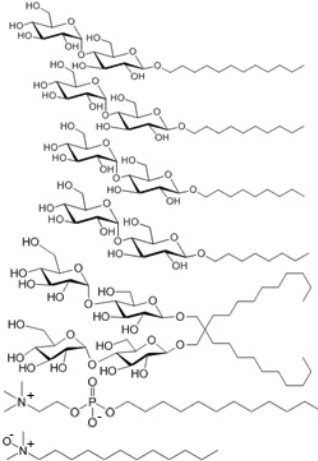
Figure 34: SDS-PAGEs showing the solubilization results using different detergents (Table 6). The arrows indicate the GlpG band of ~33 kDa. S: Amersham ECL Rainbow Marker. **A.** Input and output of the solubilization procedure (before and after ultracentrifugation at 100,000 g) for samples treated with each one detergent. **B.** Flow through and elution fractions of the affinity chromatography procedure using HisPur Cobalt resin for samples treated with each one detergent.

Next, the purification of GlpG was established and optimized, based on a published protocol^[168]. First, the cell membranes were isolated with various centrifugation steps (chapter 4.10.4). Next, the solubilization in detergent micelles was a crucial step to extract the membrane protein from the prepared membranes. For every membrane protein, the detergent needs to be selected individually (chapter 3.1.1). In Figure 34 panel A, the solubilization yields are shown comparing the supernatant before and after solubilization and centrifugation. The detergents DDM, DM, and NG worked best during solubilization as the relative band intensities after solubilization and centrifugation are higher than for samples treated with other detergents. The samples solubilized with OG showed smearing, which can be attributed to the higher viscosity of the buffer with OG. Increased viscosity can hamper the purification efficiency and is likely unsuitable for NMR measurements in solution, if not exchanged for a different detergent during the purification. Subsequently, the purification of GlpG in the different detergent micelles was tested using HisPur Cobalt resin, because the arrangement of micelles around membrane proteins can, for example, influence the accessibility of the fusion tag (Figure 34, panel B). The best results were achieved for detergents DM, FC-12, NG, and OG. It should be noted that after solubilization with FC-12, protein bands at roughly 24 and 17 kDa appeared during the elution from the resin. This can indicate that GlpG was unstable after solubilization in FC-12 micelles, resulting in degradation products. Alternatively, these could be other solubilized proteins, which the other detergents could not extract. Also, the band for GlpG in the elution fraction after solubilization with OG is thicker than for other detergents, e.g. NG. This likely reflects the high viscosity and/or heterogeneity of GlpG in OG micelles. Taken together, DM and NG performed best during solubilization and subsequent purification with HisPur Cobalt resin.

Unfortunately, the ideal detergent for solubilization is not necessarily the best choice for crystallization or solution NMR experiments. NG was reported to be well-suited for the crystallization of GlpG^[144], keeping the native fold and activity of GlpG but also enabling intermolecular interactions necessary for crystals to form. For solution NMR experiments, the smallest possible micelle is often desirable to avoid slow tumbling and short T_2 relaxation times (chapter 1.1.1). The risk for small micelles, however, is that they can disrupt the native protein structure as they usually correspond to harsher detergents. To track the state of GlpG, an activity fluorescence assay was performed, measuring the cleavage of a substrate with a

fluorescence tag^[168]. The substrate used for the assay was soluble in water and its concentration was thus not micelle-dependent, which enabled the comparison of GlpG activity in different detergent micelles.

Table 6: Summary of the characteristics of the used detergents (Anatrace Products) and the results for GlpG solubilization. In addition to ¹H-¹⁵N HSQC spectra, the sample quality was checked visually after three hours incubation, and activity assays^[169] were performed (see Figure 62). For the samples in DM, DDM, and LMNG micelles, the NMR sample concentrations were too low to record any spectra. Thus, for the sample in DM micelles, a four-week-old GlpG sample with a concentration of 303 μM was used. n.d.: not determined.



Detergent	CMC / % (w/v)	MW of micelle / kDa	NMR sample conc. / μM	Dispersion in NMR spectrum	Sample quality after 3h incubation at 37°C	Activity in fluorescence assay
DDM	0.009	4 – 76	27	n.d.	n.d.	Medium
DM	0.087	33.3	14	OK	Very good	Medium
NG	0.2	40.8	40	OK	Precipitate	High
OG	0.53	7.8 - 29	57	OK	Precipitate	Medium
LMNG	0.001	Large	8	n.d.	n.d.	Low
FC-12	0.047	19	44	Low	Good	High
LDAO	0.023	17	112	Low	Precipitate	None

The molecular weight, detergent-specific CMC, proteolytic activity, stability, and the visual appearance of the HSQC spectra of GlpG samples with different detergents were considered to select the optimal detergent for solution NMR experiments (Table 6). All these parameters determine the feasibility of NMR measurements and the quality of the resulting spectra. The smallest micelles of the detergents are formed by LDAO and FC-12 with sizes of 17 and 19 kDa, resulting in molecular weights around 50 kDa of the GlpG-detergent complexes, respectively. The largest micelles are formed by DDM and LMNG with GlpG-detergent complex sizes of ≥100 kDa, which were thus not well-suited. Concerning the CMC of detergents, detergents with high CMC are preferred as they can be exchanged and removed more easily compared to detergents with low CMC values. In turn, very high detergent concentrations can reduce the sensitivity of NMR measurements. Therefore, NG and OG were considered less suitable. GlpG activity was confirmed in all detergent micelles except for LDAO and LMNG (Figure 62), which were therefore excluded. In terms of stability, DM and FC-12 revealed to stabilize GlpG better than NG, OG, and LDAO, because the latter showed signs of precipitation. Especially for 3D NMR experiments, a high sample stability over time is essential. Independent of precipitation, the quality of the ¹H-¹⁵N spectra of GlpG was best for detergents DM, NG, and

OG, while it was worse for FC-12 and LDAO (Table 6). The spectrum of GlpG solubilized in FC-12 showed low spectral dispersion with broad and overlapping peaks (Figure 35, panel A). It can be assumed that the GlpG structure was either distorted or even partially unfolded when solubilized in FC-12 or LDAO, leading to higher sample heterogeneity. Both FC-12 and LDAO are zwitterionic detergents and therefore harsher than non-ionic detergents. On the other hand, the GlpG sample in DM micelles achieved high spectral dispersion and good peak separation (Figure 35, panel A). Overall, the detergents DM and FC-12 satisfied the chosen criteria best and were thus studied in more detail.

Comparing the ^1H - ^{15}N spectra of GlpG in FC-12 and DM micelles (Figure 35, panel A), the peak numbers and the peak dispersion differ drastically. The peak number in DM micelles corresponded to roughly one third of the expected peak number for GlpG amide groups. Interestingly, this peak number and the pattern matched well with assignments known for the N-terminal domain of GlpG.

To achieve the best purification and NMR measurement conditions using both detergents, it was tested whether GlpG could be solubilized in DM and then transferred to FC-12 micelles for solution NMR experiments. Indeed, the detergent exchange over a Size Exclusion Chromatography (SEC) column after Talon affinity chromatography was implemented successfully and did not decrease the purification yields. The resulting spectrum (Figure 35, panel B, blue spectrum) was outstandingly better than the previously acquired spectra in DM or FC-12 achieving both roughly twice the number of peaks than for GlpG in DM micelles (panel A) and good spectral dispersion. Both peaks belonging to the N-terminal domain and the core were detected when comparing it with a spectrum of GlpG core (red spectrum in panel B). Comparing the red spectrum of GlpG core in panel B with the blue spectrum of full-length GlpG in DM micelles in panel A showed distinct peak patterns, indicating that no peaks belonging to the GlpG core were detected in DM micelles. The complex size of GlpG in DM micelles was likely too large to be detected under these slow-tumbling conditions compared to GlpG core, whereas the N-terminal domain seems to have additional flexibility, allowing its detection. On the other hand, the green spectrum in panel A, corresponding to GlpG in FC-12 micelles, resembled the spectrum of GlpG core. However, the low spectral dispersion corresponds more likely to partly denatured heterogeneous GlpG structures caused by the harsh detergent, which showed signs of degradation before (Figure 34, panel B).

The improvement of the ^1H - ^{15}N HSQC for GlpG by combining the solubilization in DM and the exchange to FC-12 was mainly attributed to the faster tumbling time in FC-12 micelles but also smooth solubilization with DM (comparing the blue spectrum in panel B with the spectra in panel A, Figure 35). However, roughly 87 peaks for the backbone plus additional sidechain amide peaks were still not visible. In addition, the activity of GlpG core and full-length GlpG was compared via fluorescence activity assays (Figure 35, panel F). Both GlpG constructs were solubilized in DM and then transferred to FC-12 micelles. It was revealed that both full-length GlpG and GlpG core retained their actives. GlpG core showed higher conversion rates than full-length GlpG at the same concentration. It should be noted that the estimation of the protein concentrations via UV absorption is prone to small inaccuracies due to the detergents.

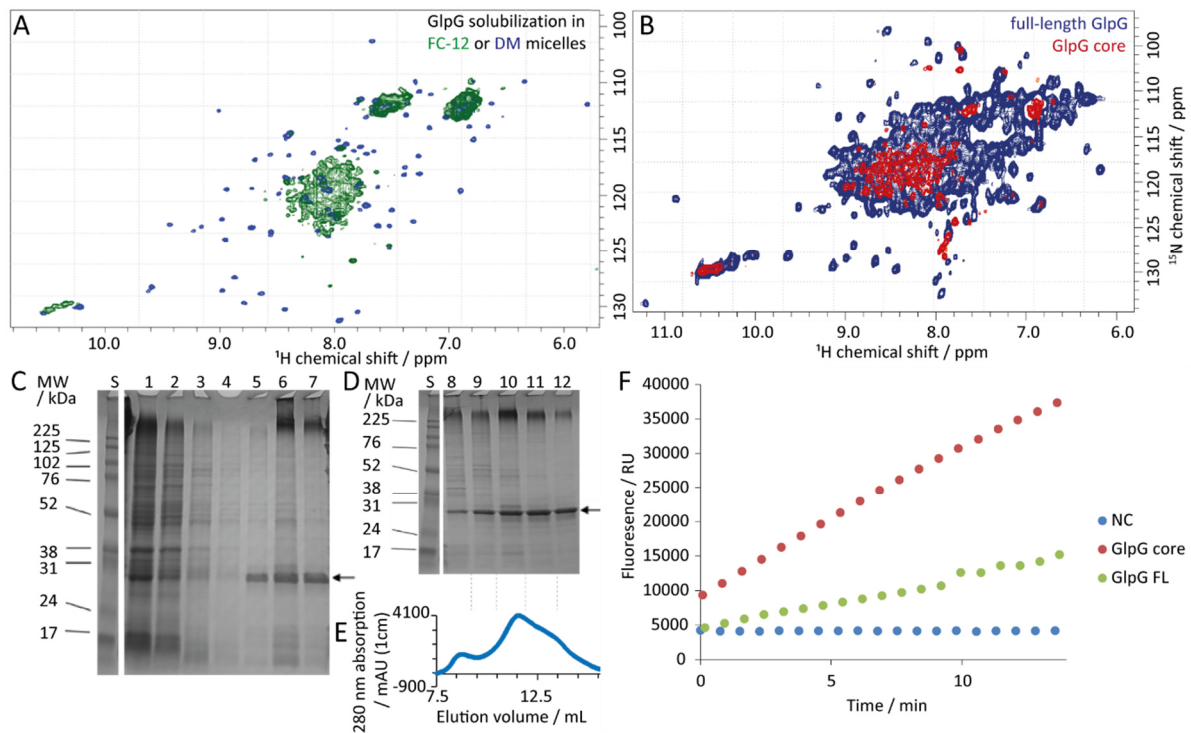


Figure 35: Different GlpG constructs in various detergent micelles. S: Amersham ECL Rainbow Marker. **A.** Overlay of two ^1H - ^{15}N HSQC spectra recorded on a 700 MHz spectrometer at 37 °C. The blue spectrum refers to 303 μM ^{15}N -labelled GlpG solubilized in DM which was stored at 4 °C for one month (recorded for 40 min). The green spectrum corresponds to 44 μM ^{15}N -labelled GlpG solubilized in FC-12 (recorded for 15 hours) in 25 mM TRIS, 150 mM NaCl, and 10% D_2O . **B.** Overlay of two ^1H - ^{15}N HSQC spectra comparing GlpG core (red, recorded on a 700 MHz spectrometer) and full-length GlpG (blue, recorded on a 800 MHz spectrometer). Both samples were ^{13}C -, ^{15}N -labelled and kept in 25 mM TRIS, 150 mM NaCl, 0.1% FC-12, and 10% D_2O at pH 7.8-8.0 at 40 °C during the measurements. The GlpG core sample had a concentration of 175 μM and was produced cleaving full-length GlpG with chymotrypsin (chapter 4.10.4). Full-length GlpG was used at a concentration of 360 μM . **C, D.** SDS-PAGE tracking the purification of D-, ^{13}C -, ^{15}N -labelled full-length GlpG wild type. HisPur Cobalt resin purification: 1: input, 2: flow through, 3-4: wash fractions, 5-7: elution fractions; 8-12: SEC fractions of D-, ^{13}C -, ^{15}N - and ILVA-methyl labelled full-length GlpG S201T in 50mM TRIS buffer, 150 mM NaCl, and 0.11% d_{38} -FC-12 micelles at pH 8.0, corresponding to panel E. The arrows indicate the band for full-length GlpG of ~33 kDa. **F.** Activity assay^[169] measuring the fluorescence increase upon Ksp63 cleavage in buffer (NC), with GlpG core and full-length GlpG in 25 mM phosphate buffer, 150 mM NaCl, and 0.1% FC-12 at pH 7.4.

Comparing ^1H 1D spectra to estimate the efficiency of detergent exchange, some DM molecules were still present after SEC despite the large excess of FC-12 in the SEC buffer, (Figure 36). For the GlpG sample in d_{38} -FC-12 micelles, represented by the green spectrum, and the DM sample, represented by the blue spectrum, similar peak patterns were detected. These spectra differ from the red spectrum of the FC-12 sample and the purple spectrum of the GlpG sample in non-deuterated FC-12 micelles. In the GlpG sample in FC-12 micelles, no proton peaks for DM were visible, although it was prepared the same way as GlpG in d_{38} -FC-12 micelles. Here, the FC-12 protons dominated the signal, as expected, because traces of DM detergent molecules were only visible after SEC when FC-12 was deuterated. Most likely, mixed detergent micelles formed after detergent exchange. Potentially, the retention of DM during SEC could indicate that GlpG prefers to interact with DM instead of FC-12. Detergent molecules with no preferred binding to GlpG would be expected to be easily exchanged by the abundant FC-12 molecules during SEC. In the end, the solubilization under mild conditions with DM, the decreased GlpG-detergent complex size after detergent exchange with FC-12, and potentially also the mixed micelles improved the peak count and dispersion of the ^1H - ^{15}N -HSQC spectra of full-length GlpG (blue spectrum in panel B, Figure 35).

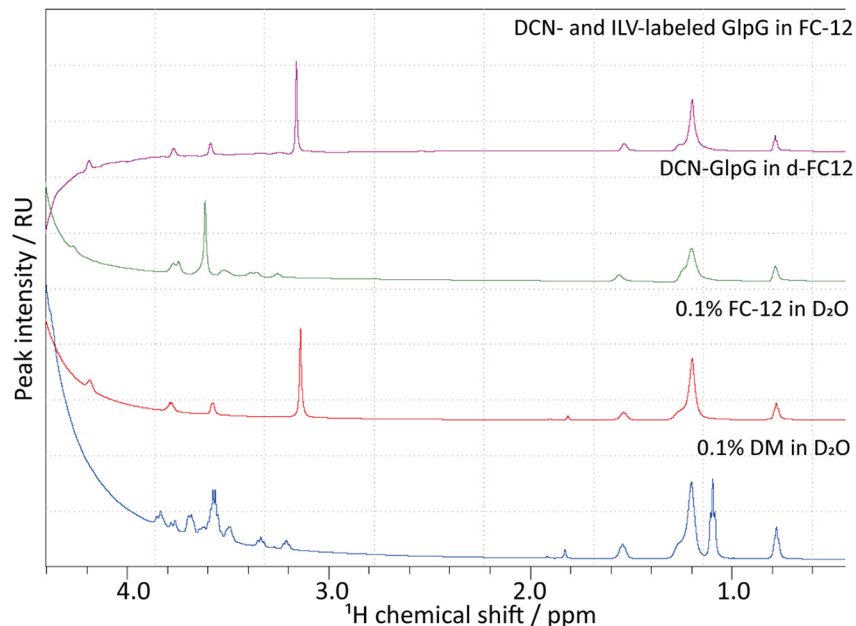


Figure 36: 1D ^1H spectra of 0.1% FC-12 in D_2O at 20 °C (red), 0.1% DM in D_2O at 20 °C (blue), 630 μM D-, ^{13}C -, ^{15}N -labelled full-length GlpG wild type recorded at 40 °C (green), and 112 μM D-, ^{13}C -, ^{15}N -labelled full-length GlpG L123A and I237A mutant with stereospecific H_3^{13}C -ILV-methyl labelling at 40 °C (purple). Both GlpG samples were kept 150 mM NaCl, 10% D_2O , and 0.11% FC-12 at pH 8 after solubilization in DM and detergent exchange. The GlpG wild type and mutant samples were buffered by 50 mM TRIS in d_{38} -FC-12 or phosphate buffer in FC-12 micelles, respectively. The water peak at ~ 4.7 ppm distorted the baseline of the 1D spectra in the region close to the peak.

The purification of GlpG is depicted exemplarily in panels C to E in Figure 35. Apart from the ~33 kDa band corresponding to GlpG, many other bands representing impurities are visible which decrease and disappear during affinity chromatography with HisPur Cobalt resin and SEC. In the flow-through and wash fractions using HisPur Cobalt resin, mostly impurities were detected indicating that the majority of GlpG molecules was captured via the His-tag by the resin. The remaining impurities in the elution fractions were removed by separating the proteins by their hydrodynamic radii during SEC. Very large proteins and mostly aggregates elute at the void volume of ~8 mL corresponding to lane 8 on the SDS-PAGE in panel D. Smaller proteins enter the pores and elute later according to their size. As often encountered for membrane proteins, the elution peaks of the SEC were not well-separated. Here, small amounts of GlpG were detected in most fractions, likely because the column was overloaded, multimers and smaller aggregates formed, the GlpG population was heterogeneous, or GlpG interacted with other proteins. These effects can lead to broad elution peaks. Especially in the first three lanes 8-10, many bands in addition to the expected ~33 kDa GlpG monomer were detected. While the dimer size of GlpG should be ~66 kDa, the most dominant band on the SDS-PAGE has a size of >225 kDa which likely identifies this band as aggregates. The intensity of this band at high molecular weight seems to correlate with the intensity of the GlpG monomer band at ~33 kDa, compared to other protein bands. These aggregates likely correspond to GlpG and formed during SDS-PAGE preparation, because large complexes should have eluted early on the SEC column. The last two lanes 11-12 correspond to the smallest GlpG-detergent complexes and showed the highest purity. The samples of lanes 11-12 were consolidated and resulted in an NMR sample of full-length GlpG wild type of roughly one milligram. In general, the purification yields for full-length GlpG with high purity ranged between 0.7 and 2.8 mg per litre expression culture with an average yield of 1.3 mg/L (n = 5). The yields of GlpG wild type were slightly higher than for the inactive S201T mutant. Comparing the GlpG core construct, which was directly expressed and not obtained by proteolytic cleavage, and the cleaved full-length GlpG, the expression yields were similar.

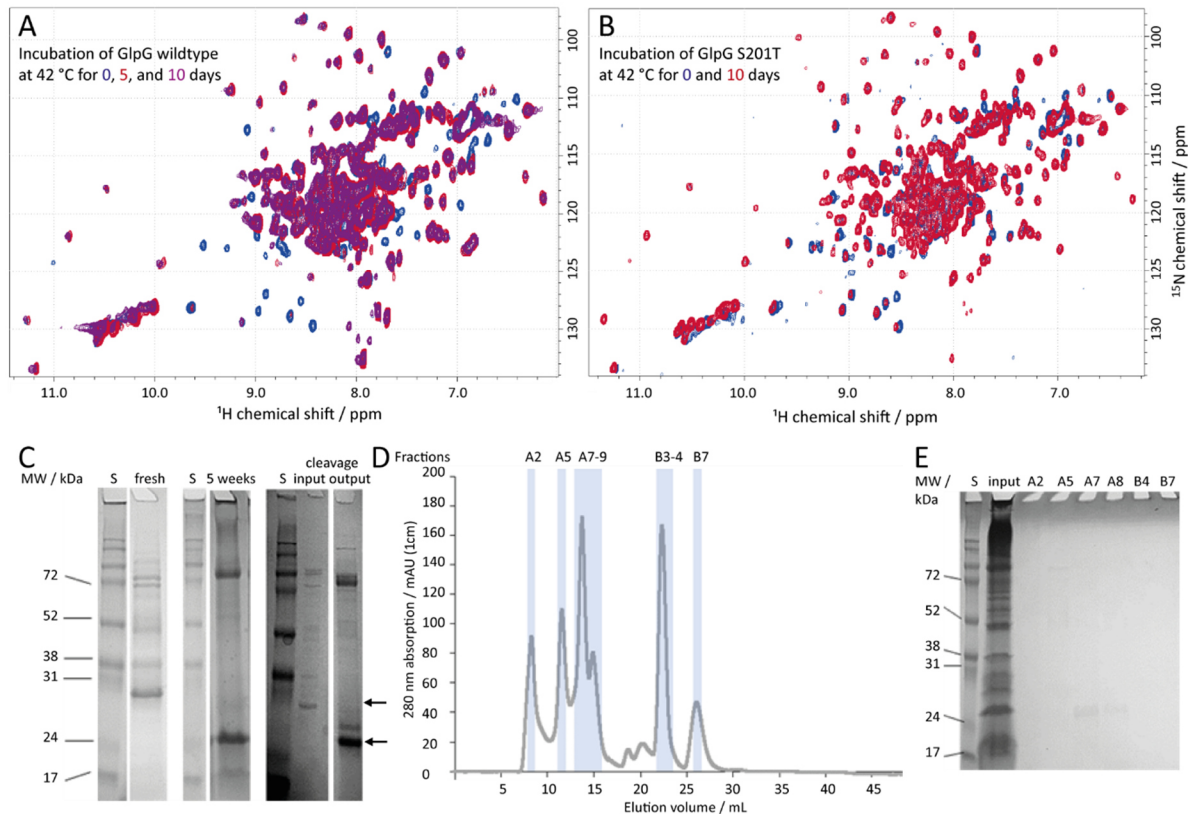


Figure 37: Stability of GlpG over time. S: Amersham ECL Rainbow Marker. **A.** ^1H - ^{15}N HSQC spectra of 667 μM D-, ^{13}C -, ^{15}N -labelled full-length GlpG wild type in 50 mM TRIS, 150 mM NaCl, 10% D_2O , and d_{38} -FC-12 micelles (solubilized in DM) at pH 8 before (blue), after 5 days (red) and after 10 days (purple) of NMR measurement time at 40 °C on an 800 MHz spectrometer. **B.** ^1H - ^{15}N HSQC spectra of 240 μM D-, ^{13}C -, ^{15}N -labelled (and stereospecific H_3^{13}C -ILVA-methyl labelled) full-length GlpG S201T in 50 mM TRIS, 150 mM NaCl, 10% D_2O , and d_{38} -FC-12 micelles (solubilized in DM) at pH 8 before (blue), and after 10 days (red) of NMR measurement time at 40 °C on an 800 MHz spectrometer. The red spectrum was recorded for 5 hours while the blue spectrum was recorded for 20 min only. **C.** SDS-PAGEs of fresh and five weeks old full-length GlpG wild type samples corresponding to panel A, compared to full-length GlpG (input) and cleaved GlpG (output) with chymotrypsin (chapter 4.10.4). The arrows indicate the bands of full-length GlpG (~33 kDa) and GlpG core (~22 kDa). **D, E.** SEC chromatogram and corresponding SDS-PAGE of the five weeks old NMR sample shown in panel C.

The sensitivity of NMR spectra can be further optimized by choosing the highest acquisition temperature possible during NMR measurements to accelerate the molecular tumbling speed. High temperatures can, however, compromise the sample quality. Consequently, the sample stability was tested at high temperatures over time (Figure 37). Panel A shows spectra of GlpG wild type at full-length incubated at 40 °C over the course of eleven days. After five days, the intensities of peaks belonging to the N-terminal domain were depleted suggesting degradation or unfolding. SDS-PAGEs and SEC chromatograms revealed that the size of GlpG in the sample decreased by roughly ~5-10 kDa over time (Figure 37, panels C-E). This size reduction matched the difference between full-length GlpG and GlpG core after chymotrypsin cleavage. Thus, the sizes of GlpG core and GlpG after incubation at 40 °C were very similar inferring that GlpG lost a fragment of similar size as the N-terminal domain. The SEC

chromatogram (panel D) and the bands for the SEC input sample on the SDS-PAGE (panel E) reveal the heterogeneous nature of the NMR sample after incubation at 40 °C. It displays a peak at the void volume of 8 mL (fraction A2), peaks of additional aggregates of larger sizes (fraction A5), two peaks at the size expected for full-length GlpG or core (fractions A7-9), and peaks referring to smaller peptides (fractions B3-4 and B7). The protein in fractions A7-9 could be confirmed to correspond to ~24 kDa, which is similar to the size of GlpG core. Unfortunately, the protein concentrations in the other fractions were too low to be detected on the SDS-PAGE. The smaller fragments in fractions B3-4 and B7 might refer to degradation products of full-length GlpG, which would be peptides from the N- or C-terminal domain. Indeed, most missing peaks in the ^1H - ^{15}N -HSQC spectrum of full-length GlpG wild type after incubation at 40 °C were identified as residues of the N-terminal domain (compare purple and blue spectra in panel A in Figure 37 and panel A in Figure 35, respectively). To rule out any connection to the proteolytic activity of GlpG itself, the inactive mutant S201T^[168] was prepared and incubated at the same conditions. Surprisingly, the relative peak intensities were stable over the course of eleven days, in contrast to the peaks of the GlpG wild type sample. Consequently, it can be suggested that deactivating GlpG protects the N-terminal domain from what appears to be self-cleavage or cleavage of one by another GlpG molecule as the only difference between the two samples is the S201T point mutation. The sample stability of S201T was lower than for GlpG wild type, as the NMR signal declined over time and precipitate formed. A lower stability of the S201T mutant compared to GlpG wild type has been reported previously^[137]. Otherwise, the peak patterns of GlpG wild type and the S201T mutant were very similar, which allowed to neglect chemical shift differences between the two constructs (Figure 37, panels A and B). It was considered most important to optimize the construct for NMR and to accept slightly lower expression yields and minor precipitation of GlpG S201T.

Finally, GlpG deuteration and different pulse sequences were implemented to assess their effect on the quality of the amide spectra (Figure 38). Deuterating both the FC-12 detergent and GlpG, the ^1H - ^1H dipolar coupling network was diluted, improving the T_2 relaxation properties for back-exchanged amide protons in the core region of GlpG. This was reflected by narrower peak widths and better peak separation of the amide peaks of GlpG core residues in the region of 7.0-9.0 ppm for ^1H and 115-122 ppm for ^{15}N (Figure 38, panel A). The ^1H linewidth of F127 improved by 10 Hz from 0.063 to 0.050 ppm in the ^1H dimension, comparing

protonated and deuterated full-length GlpG and FC-12, respectively. Furthermore, the peaks corresponding to residues F146, L148, and G222, chosen as representatives, gained higher signal-to-noise ratios.

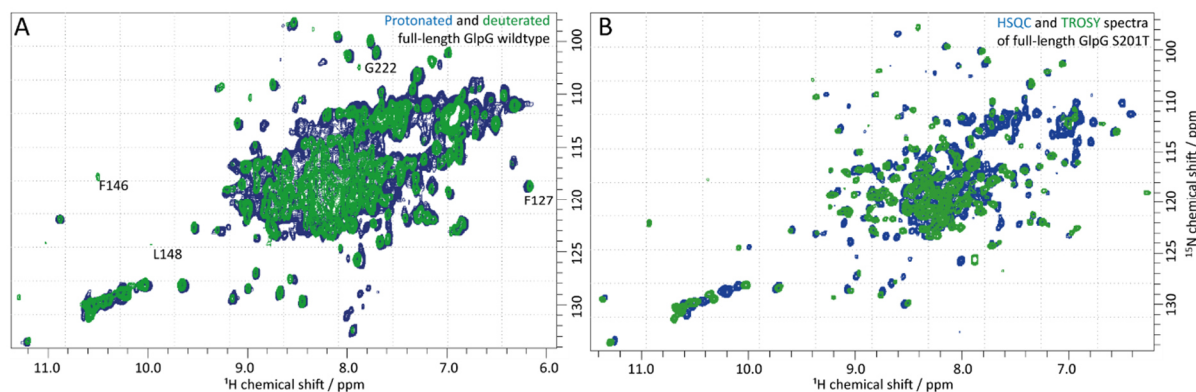


Figure 38: Optimization of ^1H - ^{15}N 2D spectra for full-length GlpG via deuteration and choosing different NMR pulse sequences. **A.** Overlaid spectra of $360\ \mu\text{M}$ ^{13}C -, ^{15}N -labelled GlpG wild type (blue) in 0.11% protonated FC-12 micelles and $667\ \mu\text{M}$ D-, ^{13}C -, ^{15}N -labelled GlpG wild type in 0.11% (d_{38} -)FC-12 (green) using buffer containing 25-50 mM TRIS, 150 mM NaCl, and 10% D_2O at pH 8. The experiments were both recorded on an 800 MHz spectrometer at $40\ ^\circ\text{C}$. Some peaks were labelled with the assigned residues (Figure 64). **B.** Overlaid ^1H - ^{15}N 2D HSQC and TROSY spectra of GlpG S201T. The blue HSQC spectrum refers to $260\ \mu\text{M}$ D-, ^{13}C -, ^{15}N - and stereospecifically ILVA-methyl labelled GlpG in 50 mM TRIS, 150 mM NaCl, 0.11% d_{38} -FC-12, 12 μM DSS, and 10% D_2O at pH 8, recorded at $40\ ^\circ\text{C}$ for 1 h 23 min. The green TROSY spectrum corresponds to $450\ \mu\text{M}$ D-, ^{13}C -, ^{15}N - and stereospecifically ILV-methyl labelled GlpG without His-tag in 50 mM d_5 -TRIS, 150 mM NaCl, 0.11% d_{38} -FC-12, 3.8 mM DSS, and 10% D_2O at pH 8, recorded with half the scans compared to the HSQC at $40\ ^\circ\text{C}$ for 23 min.

In addition to classical ^1H - ^{15}N -HSQC experiments, the TROSY experiment selects components with the lowest T_2 relaxation contributions to increase sensitivity. The TROSY experiment achieved a signal-to-noise ratio similar to the HSQC spectrum. However, especially for the peaks of the GlpG core with ^1H chemical shifts around 7.0-9.0 ppm and ^{15}N chemical shifts around 115-122 ppm, the peaks in the TROSY spectrum were better separated than in the HSQC spectrum. The peak numbers for the backbone amides were otherwise similar.

In summary, the best NMR spectrum was obtained with the deuterated GlpG S201T mutant, which was solubilized in DM and exchanged for d_{38} -FC-12, using a ^1H - ^{15}N TROSY experiment and recording at $40\ ^\circ\text{C}$. These conditions allowed to retain the near-native structure and activity of full-length GlpG but also achieved high-resolution spectra at reasonable acquisition times.

3.2.2 Amide and ILV-methyl sidechain peak assignments for full-length GlpG

Proteins can only be characterized via NMR spectroscopy with atomic resolution if assignments are available which connect the protein residues to the peaks in the NMR spectra. For this purpose, both the chemical shifts of the amide backbone and methyl

sidechains were determined. The amide assignments were based on chemical shifts published for the N-terminal domain in solution by the Goto group^[140b] and for the GlpG core in the solid state by the Lange group^[164]. Unfortunately, *de novo* assignments have not been possible, because only an incomplete data set for backbone assignments could be recorded within reasonable time due to the large size of the GlpG-detergent complex. Hence, the published assignments were transferred for 3D HNCA, HNCOC, and HNCACB spectra. For the methyl sidechains of GlpG core, no chemical shifts have been published yet. *De novo* methyl assignments for GlpG core residues were obtained combining mutagenesis and different methyl labelling schemes. For the N-terminal domain, methyl sidechain assignments have been published^[140b] and could be transferred to the 2D ¹H-¹³C HMQC recorded here.

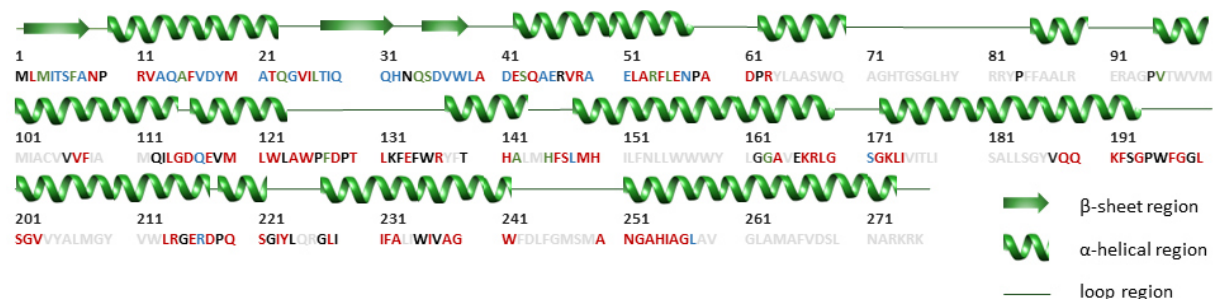


Figure 39: Assessment of the chemical shift assignment transfer of published chemical shifts^[140b, 164] to 2D ¹H-¹⁵N TROSY, 3D TROSY-HNCA, 3D TROSY-HNCO, and 3D TROSY-HNCACB spectra recorded here. The grey and black letters designate residues (Figure 28) for which no matching peaks were published or found in the spectra, respectively. The green, blue, and red letters refer to assignments grouped according to different tolerances. If the ¹H and heteronuclear chemical shifts (¹⁵N, ¹³CO, and/or ¹³CA) deviated by <0.2 ppm and <1 ppm, respectively, the letters were coloured in green. If the ¹H chemical shifts were off by <0.2 ppm but the heteronuclear chemical shifts were off by ≥1 ppm, the residues were coloured in blue. The red residues finally correspond to less reliable assignments with the ¹H and heteronuclear chemical shifts deviating by >0.2 ppm and >5 ppm, respectively. The displayed sequence corresponds to full-length GlpG wild type aligned with secondary structure elements from AlphaFold (P09391^[19]) predictions. For more information, see Table 31.

In Figure 39, the success of matching peaks from 2D and 3D backbone NMR experiments with published chemical shifts^[2,3] is displayed as a function of GlpG sequence. Based on the expected number of peaks, only ~75% of the peaks were visible in the 2D ¹H-¹⁵N TROSY experiment. In the HNCO and HNCA spectra, ~75% and ~59% of the expected peaks (excluding inter-residual peaks) were detected, respectively. The residues with undetected peaks were likely either located in a very rigid part of the protein and therefore suffered from fast *T*₂ relaxation, overlapped strongly with other peaks, or moved on a timescale which induced severe peak broadening. The assignment transfer was also limited by the number of available chemical shift assignments^[140b, 164]. 69% of residues 1-86 of the N-terminus had been assigned as the construct used by the Goto group only covers residues 1-61. The available assignments

for the GlpG core amount to ~62% of the core sequence. It is worth noting that low assignment coverage of the protein sequence comes with a higher chance for false assignments. To transfer the published chemical shift assignments to the obtained spectra, peaks corresponding to the published assignments were depicted in the spectra and referenced. Next, the individual peaks were matched considering different tolerances for deviations and the amino acid-specific reference chemical shifts for C_{α} and C_{β} . Overall, the assignment success was moderate, although all data sets were recorded on an optimized GlpG sample and solely TROSY-based 3D experiments were utilized. Here, 30% of the core residues (aa 87-276) have been assigned but only 4% were assigned reliably (residues coloured in blue or green in Figure 39). The strong deviations for the proton chemical shifts are likely related to differences between solution and solid-state NMR spectra of the data set obtained here and the published chemical shifts. Moreover, the published chemical shifts were based on GlpG core in liposomes^[164]. Concerning the N-terminal domain (aa 1-86), 69% of the residues have been assigned with 45% of reliable assignments (residues coloured in blue or green in Figure 39). Both the data sets acquired here and the published chemical shifts for the N-terminal domain were based on solution NMR experiments. In addition, the N-terminal domain is not part of the micelles or liposomes. In total, roughly 42% of the residues of full-length GlpG were generally assigned, with high reliability for 17% of the assignments. It remains questionable if sequential backbone experiments for full-length GlpG in solution can be performed to improve these assignments. Either longer measurement times and higher stability over time or higher sensitivity due to a smaller construct size or higher temperatures could achieve this goal. For this purpose, other GlpG mutants might be worth investigating. Longer measurement time could resolve low-intensity peaks but there are limited options available for resolving crowded peak regions in the ^1H dimension.

Especially for large proteins in solution, methyl sidechain groups serve well as reporters on protein structures and dynamics. Due to the additional flexibility of the sidechain, methyl groups can be assumed to outperform backbone amides in terms of sensitivity (chapter 1.1.6). The ^1H - ^{13}C SOFAST-HMQC experiment was chosen to detect methyl groups of GlpG due to shaped selective ^1H -pulses and Ernst-angle excitation^[104c]. Both improve the sensitivity per unit time by shortening the recycle delay of the experiments^[104c]. Here, pro(*R*)-precursor of 2-acetolactate and α -ketobutyrate were used to stereospecifically label the $C_{\delta 1}$ and $C_{\gamma 1}$ methyl

groups of leucine and valine residues, respectively, as well as the $C_{\delta 1}$ methyl groups of isoleucine residues^[45d]. Based on this labelling scheme, 71 methyl peaks were expected, corresponding to 35 leucine, 20 valine, and 16 isoleucine peaks. Two of these leucine residues were part of the fusion tag. The resulting ^1H - ^{13}C HMQC spectrum showed well-dispersed methyl peaks with roughly 16 isoleucine and 55 leucine/valine peaks, as expected. Most peaks above 1.2 ppm in the proton dimension correspond to detergents and buffers, except for one methyl peak with ^1H and ^{13}C chemical shifts of ~ 1.3 ppm and ~ 19 ppm. For 3D and relaxation experiments, deuterated chemicals were used for labelling. All other methyl groups have ^1H chemical shifts of -0.2 to 1.2 ppm. It was concluded that the ILV-methyl labelling protocol (chapter 4.9.3) was successfully established with reproducible results. The precursor for leucine and valine was added 2 hours after transferring the cells to isotopically labelled media and the precursor for isoleucine was added 40 min later as the uptake of 2-acetolactate (precursor for leucine and valine) drops in the presence of α -ketobutyrate (precursor for isoleucine)^[45d].

To group the methyl peaks by residue type and domain, different NMR samples have been prepared (Figure 40). First, the isoleucine peaks can be distinguished from the leucine and valine peaks via their ^{13}C chemical shifts. Isoleucine peaks have ^{13}C chemical shifts between 7-14 ppm while valine and leucine peaks are found at >14 ppm. To the GlpG core, four isoleucine peaks could be assigned. Ten leucine and valine peaks were assigned to the N-terminal domain, which harbours seven leucine and three valine residues. Two additional valine peaks were expected for the N-terminal domain, which likely were falsely assigned to the GlpG core. False assignments can originate from missing peaks for a residue or, conversely, several peaks for different conformations of one residue. The chemical shifts and the ratio of populations of different sidechain conformations and dynamics can differ between constructs as well. Comparing an ILV- with an IV-methyl labelled sample, 35 leucine and 20 valine peaks could be identified for full-length GlpG (Figure 40, panel B). All in all, the isoleucine residues match the expectations while some discrepancies between the expected and picked number of peaks were found for leucine and valine residues.

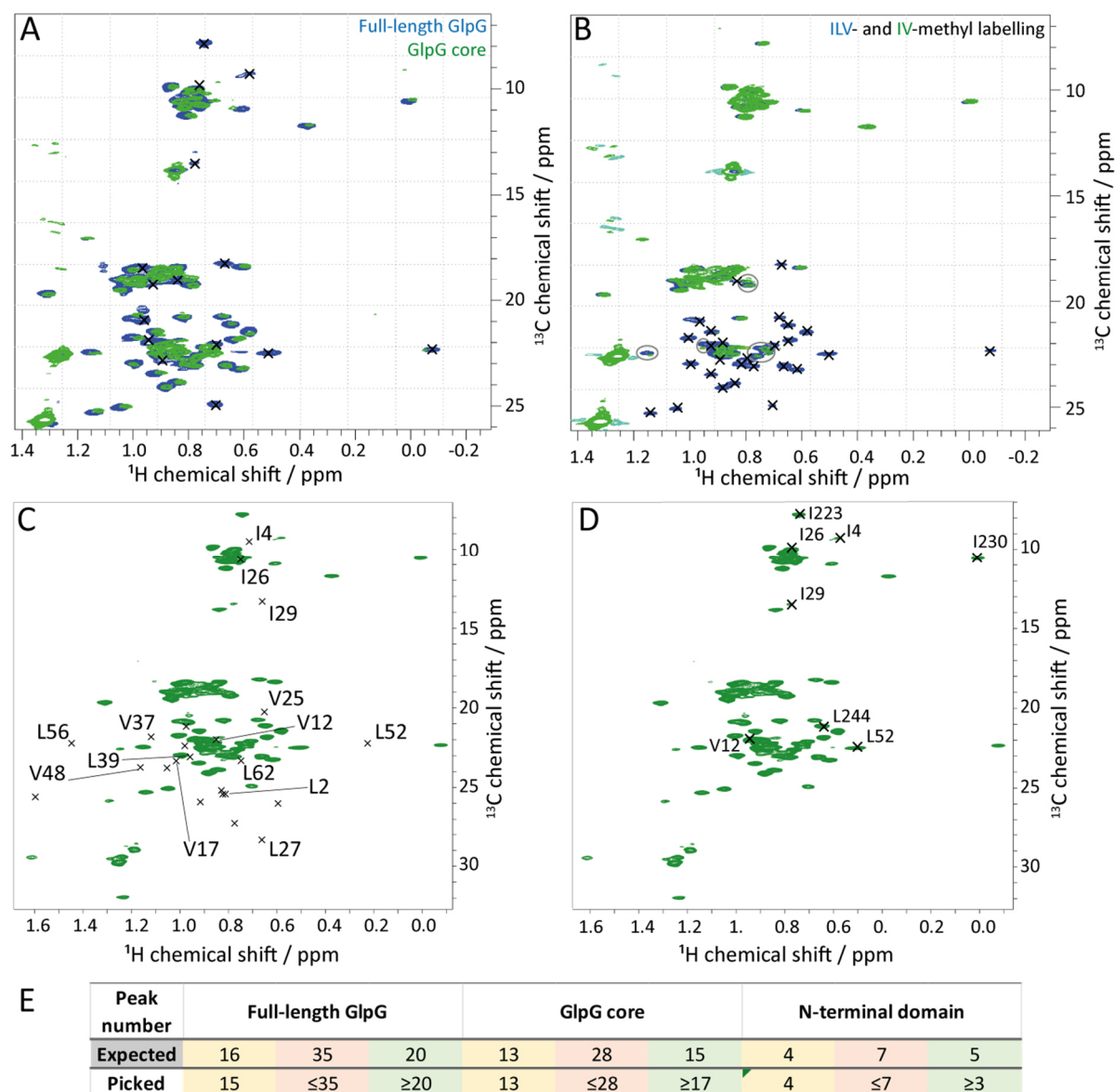


Figure 40: ^1H - ^{13}C SOFAST-HMQC spectra characterizing the ILV-methyl sidechains of GlpG recorded on an 800 MHz spectrometer at 40 °C. **A.** Overlaid spectra of D-, ^{13}C -, ^{15}N - and H_3^{13}C -ILV-methyl labelled 450 μM full-length GlpG S201T (blue) and 18 μM GlpG GlpG core (green). The sample corresponding to the blue spectrum was kept in 50 mM d_5 -TRIS, 150 mM NaCl, 0.11% d_{38} -FC-12, and 10% D_2O at pH 8. The green spectrum refers to a sample in 25 mM phosphate buffer, 150 mM NaCl, 0.2% FC-12, and 10% D_2O at pH 7.4. Markers indicate peaks assigned to the N-terminal domain. **B.** Overlaid spectra of 450 μM D-, ^{13}C -, ^{15}N - and ILV-methyl labelled (blue) or 59 μM IV-methyl labelled full-length GlpG S201T (green). The samples were kept in 25-50 mM TRIS, 150 mM NaCl, 0.11% FC-12, and 10% D_2O at pH 8. The sample belonging to the blue spectrum contained d_5 -TRIS and d_{38} -FC-12. Picked peaks were identified as leucine residues. Peaks with ambiguous leucine assignments were circled in grey. Simulated peaks according to published assignments^[140b] (**C**) and assigned peaks due to mutations (**D**). The spectrum is the same as the blue spectrum in panel A. The published assignments of simulated pro(*R*)-methyl groups are shown in panel C. **E.** Numbers of expected and picked peaks for isoleucine (yellow), leucine (orange), and valine (green), sorted for different GlpG constructs and domains.

Ambiguous leucine assignments correspond to peaks with the lowest intensities for the IV-methyl labelled GlpG sample (Figure 40, panel B), which are circled in grey. Most likely, the suppression of leucine labelling by adding unlabelled leucine was not entirely successful and

a small population of leucine residues was labelled. In addition, picking peaks in crowded regions with ^{13}C chemical shifts of roughly 18-19 and 22-23 ppm was ambiguous. It is thus possible that leucine and valine peaks were not distinguished correctly in these regions.

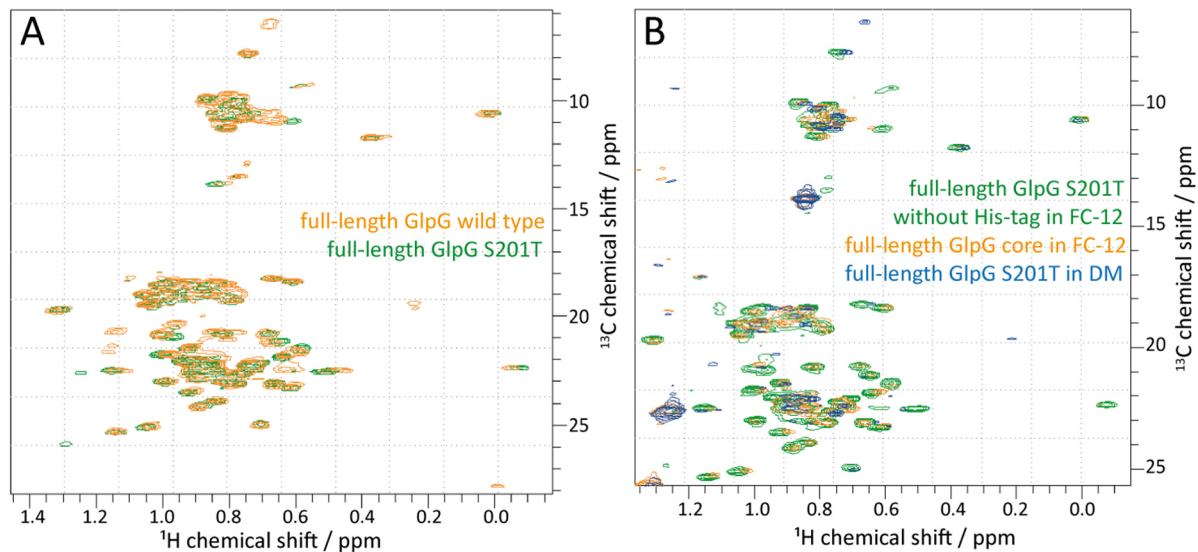


Figure 41: ^1H - ^{13}C SOFAST-HMQC spectra comparing different GlpG constructs and micelles. The spectra were recorded on an 800 MHz spectrometer at 313 K. **A.** 260 μM full-length GlpG wild type (orange) and 450 μM GlpG S201T without His-tag (green) in 50 mM (d_5 -)TRIS buffer, 150 mM NaCl, 2% and 0.11% d_{38} -FC-12, 10% D_2O , and <4 mM DSS at pH 8.0, respectively, with deuterated TRIS buffer for GlpG S201T. Both spectra were recorded for 30 min. **B.** Overlay of methyl spectra of GlpG with and without His-tag in DM or FC-12 micelles. The green spectrum refers to 450 μM full-length GlpG S201T without His-tag in 50 mM d_5 -TRIS buffer, 150 mM NaCl, 0.11% d_{38} -FC-12, 10% D_2O , and 4 mM DSS at pH 8.0. The orange spectrum corresponds to a sample with 18 μM GlpG core wild type in 50 mM TRIS, 150 mM NaCl, 0.2% FC-12, 50 mM DSS, and 10% D_2O at pH 7.4. The dark blue spectrum refers to 76 μM full-length GlpG S201T in 50 mM phosphate buffer, 150 mM NaCl, 0.2% DM, and 10% D_2O at pH 8.0. The spectra were recorded for 27 min, ~ 15 h, and 3 h, respectively.

Moreover, the conditions optimized for ^1H - ^{15}N spectra of GlpG (chapter 3.2.1) were also used to obtain the methyl spectra (Figure 41). Evidently, the spectra of GlpG wild type and S201T are very similar (Figure 41, panel A), only showing a few CSPs and a few additional peaks for GlpG wild type. These changes are rather minor considering that similar differences can also be expected for different expressions and purifications of the same construct under the same conditions. Looking at GlpG samples in DM, 57 peaks were detected, which is very close to the number of detected peaks in FC-12 micelles (Figure 41, panel B). The chemical shifts of the visible peaks also seem very similar to the chemical shifts in FC-12 micelles. Due to the larger molecular weight of the DM micelles, it was speculated that the weaker peaks belong to the GlpG core. However, comparing it with the ^1H - ^{13}C spectrum taken of the GlpG core (yellow spectrum), this hypothesis did not hold true as many GlpG core residues were observed with high intensity in DM micelles. Notably, two peaks with ^1H chemical shifts of roughly 0.2 and 0.65 ppm and ^{13}C chemical shifts of 19.5 and 6.5 ppm were detected for GlpG

in DM micelles, which coincide with some of the additional peaks shown for GlpG wild type in FC-12 micelles (Figure 41, panel A, yellow spectrum). It was not possible to assign the appearance and absence of these peaks to a specific condition or event. It rather seems they represent residues suffering from exchange broadening, low-populated conformations, and/or excited states of some residues. In conclusion, the optimizations selected for amides also hold true for the methyl groups. Again, the similarities between the ^1H - ^{13}C spectra of GlpG wild type and S201T were confirmed, indicating comparable structures for the constructs, as discussed in chapter 3.2.1. Conformational similarities can also be deduced from the corresponding methyl chemical shifts of GlpG in FC-12 and DM micelles. This supports the finding that GlpG can withstand harsher ionic detergents like FC-12 without impairing its fold, as has been suggested by activity assays (Figure 62).

Table 7: GlpG mutants prepared via QuikChange Lightning Multi Site-Directed Kit, replacing single amino acids by alanine (chapters 4.8.4, 4.8.5). For all mutants, full-length GlpG S201T (pET25) was used as template. All mutations were confirmed by sequencing.

Mutant	I	II	III	IV	V	VI	VII	VIII	IX	X
Site 1	L244	V204	L123	V99	V211	L89	L131	L143	L200	L229
Site 2	I177	I223	I237	I151	I255	V203	V96	V165	V267	V119
Site 3	-	-	-	-	-	I109	I235	I180	I230	I175
Sample quality	Low	OK	OK	OK	OK	OK	OK	Low	Low	OK

In total, mutations for 25 residues were achieved, exchanging isoleucine, leucine, and valine to alanine (Table 7). To reduce the number of constructs, double and triple mutants were prepared with each an isoleucine, leucine, and/or valine mutation. The mutated residues in the same construct were selected to be far apart to minimize the changes of the chemical environment for the surrounding residues. Looking at the SEC chromatograms, it is apparent that, apart from a few exceptions, the overall yields and thus the purities differed for the mutants (Figure 42, panel E). As before (Figure 35, panel E), the SEC peak for GlpG is relatively broad with two maxima. Substituting isoleucine, leucine, or valine with alanine induced likely destabilization of the GlpG monomer in FC-12 micelles leading to poorer NMR sample quality. This effect was most drastic for mutants IX and VIII, which revealed low dispersion in the NMR spectra (Figure 42, panel D; Figure 65, panel C). These mutants also show the lowest ratios between the GlpG peak and a peak corresponding to higher molecular weight at ~ 11.5 mL elution volume in the SEC chromatogram, hinting towards aggregation, although the purification otherwise was performed like for GlpG wild type. It can be speculated that alanine

substitution of L143 and L200 led to destabilization as they are spatially close to each other. For mutant I, the isoleucine peaks did not appear in the spectrum when it was prepared for the first time. Repeated expression and purification only resulted in low dispersion spectra similar to mutants VIII and IX, unfortunately. Even when the same expression and purification conditions were chosen, the properties of the first sample could not be reproduced.

It was observed that the ^1H - ^{13}C methyl spectra of the mutants also changed in TRIS buffer at pH 8 or phosphate buffer at pH 7.4 over time (Figure 43). As mentioned before, peak intensities are modulated by various parameters, for example, the time-regime of chemical exchange and the ratios between differently populated conformations. In more detail, it could mean that the populations of different conformations result in several peaks per residue with different intensities. Moreover, multiple conformations observed in a freshly prepared sample could shift to the thermodynamically more stable conformation after several hours or days. Also, sample degradation can cause peaks to disappear. Deviating peak intensities or chemical shifts for different buffer conditions can for instance indicate altered water networks, protonation states, or detergent contacts of individual residues. These effects can be interlinked as well for residues in protein regions which are generally more sensitive to sample preparation conditions. However, if peaks were consistently missing for the mutants screened here, they were assigned to the mutated residues. Generally, the NMR spectra of all mutants shared similar peak patterns (Figure 65).

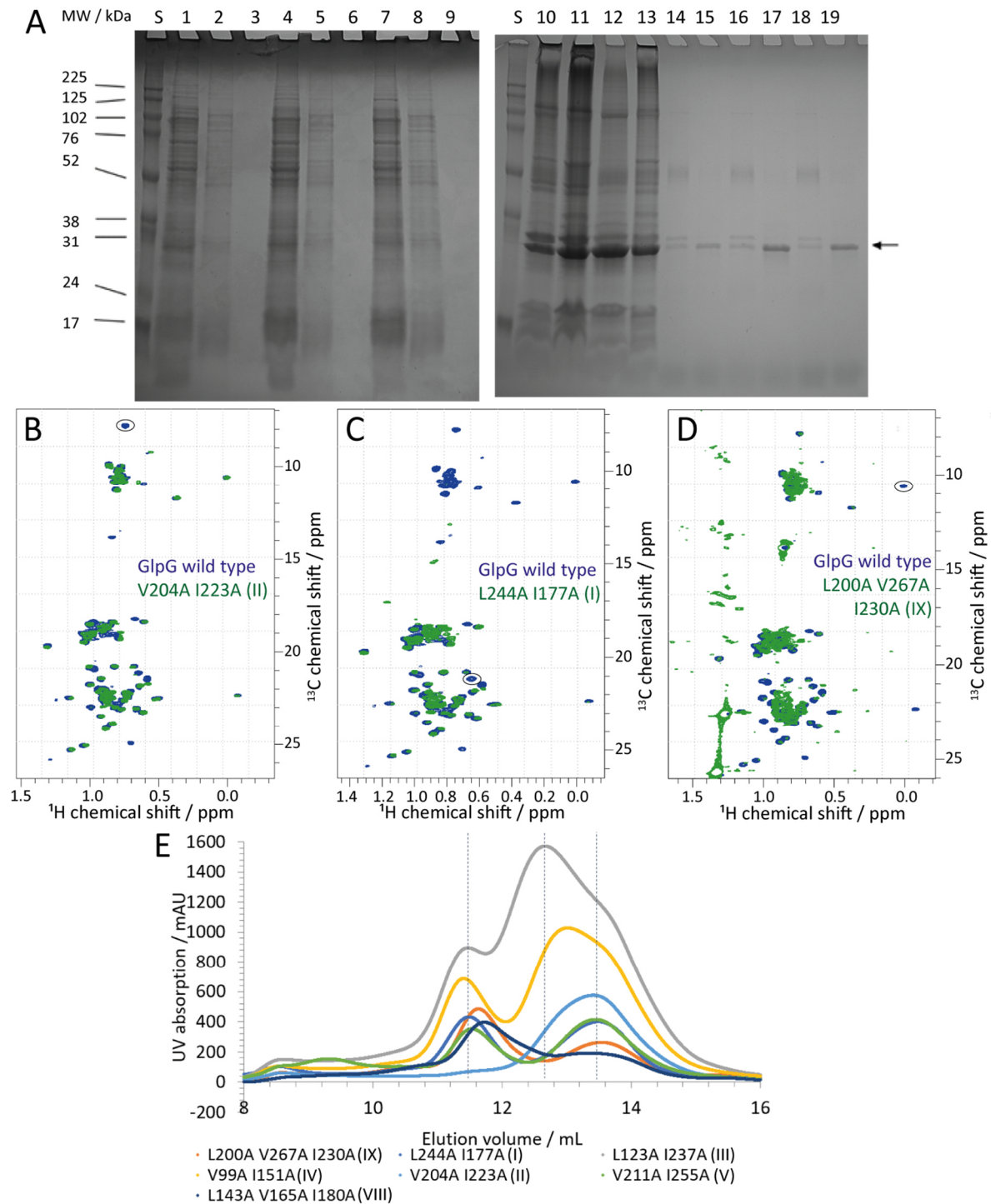


Figure 42: Characterization of D-, ^{13}C -, ^{15}N - and ILV-methyl labelled full-length GlpG mutants via solution NMR and SEC chromatography. **A.** SDS-PAGEs showing the purification of two GlpG mutants and IV-methyl labelled GlpG. S: Amersham ECL Rainbow Marker. Mutant IX: 1: flow through of HisPur Cobalt resin, 2-3: wash fractions 1 and 2 of HisPur Cobalt resin, SEC input: 10, SEC fractions: 14-15; Mutant I: 4: flow through, 5-6: wash fractions 1 and 2, SEC input: 11, SEC fractions: 16-17; IV-methyl labelled full-length GlpG S201T: 7: flow through, 8-9: wash fractions 1 and 2, 12: SEC input, 13: pellet of insoluble proteins removed from SEC input sample, 18-19: SEC fractions. The arrow indicates the expected GlpG monomer size of ~ 33 kDa. **B, C, D.** Overlaid ^1H - ^{13}C SOFAST-HMQC spectra of $450\ \mu\text{M}$ full-length GlpG S201T without His-tag in $50\ \text{mM}$ (d_5 -)TRIS buffer, $150\ \text{mM}$ NaCl, 2% and 0.11% d_{38} -FC-12, 10% D_2O , and $<4\ \text{mM}$ DSS at pH 8.0 (blue, in B, C, D; recorded for 30 min) with $42\ \mu\text{M}$ mutant II (green spectrum in panel B, recorded for 2), $34\ \mu\text{M}$ mutant I (green spectrum in panel C, recorded for 17 h), or $29\ \mu\text{M}$ mutant IX (green spectrum in panel D, recorded for 1 d 22 h) in 25-50 mM TRIS buffer, $150\ \text{mM}$ NaCl, 0.11% d_{38} -FC-12, 10% D_2O , and $1\ \text{mM}$ DSS at pH 8. **E.** SEC chromatograms of the indicated mutants.

In total, three methyl peaks of GlpG core could be assigned to I223, I230, and L244 (Figure 40, panel D). The assignments were achieved comparing the ^1H - ^{13}C SOFAST-HMQC spectra of the mutants to each other and the spectrum of GlpG wild type (Figure 42, panels B-D). A peak was assigned if no or negligibly low intensity was detected close to the expected peak position while other peaks with usually similar peak intensities were clearly visible. This was especially challenging when the protein concentrations and consequently also the signal-to-noise ratios were low. In these cases, usually several candidates for each mutation were found. The spectra of mutant II were consistently missing the peak at ^1H chemical shift of 0.74 and ^{13}C chemical shift of 7.78 ppm in the isoleucine-assigned region despite otherwise well-dispersed methyl peaks and even in different buffers and over time (Figure 42, panel B). This peak was assigned to I223. This was not the case for any other mutant. The methyl sidechain of I223 is sandwiched between TM2 and TM4 and is likely shielded from the detergent molecules. Surprisingly, this peak also disappeared when the N-terminal domain was cut off (Figure 40, panel A). Relying on the unambiguity of assignments via mutation in this case, the missing peak in the core construct is likely caused by chemical exchange on the intermediate timescale. Such changes would make sense as I223 is part of L4 and faces the N-terminal domain in the full-length GlpG structure. In turn, this could imply that the N-terminal domain for full-length GlpG directly or indirectly interacts with GlpG core. Alternatively, the flexible residues attached to TM1 of GlpG core could interact with L4 and to some extent influence its motion, while they are more restricted when the sequence is extended by the N-terminal domain for full-length GlpG. L244 was assigned using the spectrum of mutant I, which only showed a peak at ^1H chemical shift of 0.65 ppm and ^{13}C chemical shift of 21.23 ppm at the intensity level of noise peaks, thus being identified as the only leucine peak missing for this spectrum (Figure 42, panel C). L244 belongs to L5, which is close to the active site of GlpG. The IV-methyl labelled sample confirmed that this peak corresponds to a leucine residue (Figure 40, panel B). I230 could be assigned with the spectra of mutant IX despite the low spectral dispersion, as all isoleucine peaks were detected apart from the peak at ^1H and ^{13}C chemical shifts of 0.01 and 10.55 ppm, respectively (Figure 42, panel D). I230 is located at the outer rim of GlpG at TM5 and is likely in contact with TM4 and detergent molecules.

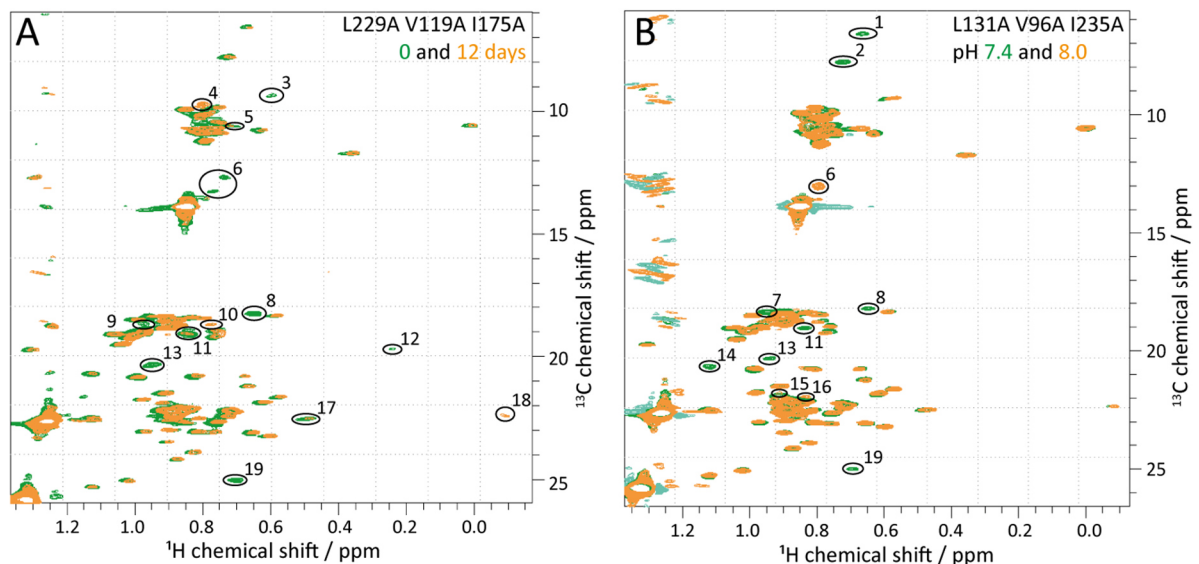


Figure 43: Changes in the ^1H - ^{13}C SOFAST-HMQC spectra of different full-length GlpG mutants over time or with different buffers. Peaks with strong peak intensity changes were labelled and circled in black. **A.** Overlays of spectra recorded for the fresh (green) and the twelve days old (stored at 4°C , yellow) sample of mutant X. The sample consisted of $30\ \mu\text{M}$ D-, ^{13}C -, ^{15}N - and ILV-methyl labelled protein in $50\ \text{mM}$ phosphate buffer, $150\ \text{mM}$ NaCl, 0.2% FC-12, and 10% D_2O at pH 7.4. **B.** Overlays of spectra recorded for GlpG mutant VII in phosphate buffer at pH 7.4 (green) and TRIS buffer at pH 8.0 (yellow). The samples correspond to $78\ \mu\text{M}$ D-, ^{13}C -, ^{15}N - and ILV-methyl labelled GlpG in $150\ \text{mM}$ NaCl, 0.2% FC-12, and 10% D_2O .

Moreover, the methyl spectra of the mutants changed in different buffers and over time for mutants II, III, IV, and VII. Only time-dependent changes were tracked for mutants VI and X. Different buffers were tested for mutants I, VIII, and IX, but no time dependence was tested. In general, some peaks were clearly susceptible to external changes. Overlays of spectra with the most pronounced changes are shown exemplarily comparing the fresh and 12-days-old sample for mutant X (Figure 43, panel A) and comparing different buffers for mutant VII (Figure 43, panel B). Peaks labelled as 6, 8 (sometimes with low intensity), 11, 13, and 19, appear to vary both over time and at different pH values. These might be generally more susceptible to changes in their local chemical environments. Peaks 1, 6, 12, 13, and 14 only appear in a few spectra and might correspond to residues with chemical exchange near the intermediate regime or additional conformations. Especially peak 1 was only visible for mutants VII and X when freshly prepared at pH 7.4. Taken together with information identifying peaks from the N-terminal domain and GlpG core, many changing peaks belong to the N-terminal domain, namely peaks 3, 6, 7, 8, 11, 15, 17, and 19. Only peaks 2 and 18 could be assigned to GlpG core. Peak 2 corresponds to I223, which is part of L5 and therefore considered more flexible overall. This might indicate that the loop flexibility was directly or indirectly impacted by the pH value. As mentioned before, an interaction with the N-terminal

domain cannot be excluded as well. For peaks 4, 5, 9, 10, and 16, either the chemical environment or regime of motion changed, resulting in perturbed chemical shifts or lower peak intensities. Interestingly, the detected changes due to different buffers were reversible. In general, more peaks were visible at pH 7.4 compared to pH 8.0, which might be related to the protonation states of polar sidechains like histidine, water networks, and the phosphate groups of the FC-12 detergents close by. An exception to these changes is mutant II, which did not experience changes for peaks 9, 10, 15, and 16, while CSPs were measured for peaks 5 and 17 (data not shown). Peaks 6, 7, 8, 13, 14, and 19 did not show up at either condition. It can be suggested that mutant II might prefer one conformation and was therefore less sensitive towards pH value changes.

The overall methyl assignment success with point mutations is relatively low with only 12%. On one hand, peak overlap in the crowded regions impeded the assignments, as only 45% of all peaks were well-dispersed. On the other hand, changes of the chemical environments over time and for different buffers with only a pH difference by ~ 0.6 impaired the identification of the mutated residues. Moreover, low yields and spectral dispersion reduced the spectral quality for some mutants. Most of these challenges are associated with general characteristics of α -helical membrane proteins, especially in detergent micelles, due to low chemical shift dispersion in α -helices, decreased stability of the folded 3D structures, and increased sensitivity towards changing external parameters (chapter 3.1.1).

The published methyl sidechain assignments for the N-terminal domain^[142] were compared to the peaks of the ^1H - ^{13}C -spectrum for full-length GlpG S201T to identify peaks belonging to the N-terminal domain (Figure 40, panel C). Five published assignments for methyl sidechains of the N-terminal were transferred on the spectra obtained here, namely I4, V12, I26, I29, and L52^[142]. For isoleucine assignments, the transfers were less ambiguous due to only one methyl sidechain peak per residue, while two assignments needed to be considered for each leucine or valine peak. Although the published assignments differentiate the two geminal methyl groups for valine and leucine, they were likely not assigned stereospecifically. The pro(*R*)-methyl group of V12 was assigned, because both pro(*R*)- and pro(*S*)-methyl group assignments were in proximity to a peak which was identified as valine of the N-terminal domain (Figure 40 panel B), peak circled in grey. For the pro(*R*)-methyl group of L52, on the other hand, only one published assignment was located close to a single leucine peak in this region of the spectrum. It should be noted, however, that the assignment deviated by

0.35 ppm in the ^1H -dimension rendering the assignment transfer more ambiguous. For the peak at ^1H chemical shift of 0.70 ppm and a ^{13}C chemical shift of 24.93 ppm, five assignments match, namely the *pro(R)*-methyl group of residue L2 or the *pro(S)*-methyl groups of residues L27, L52, L56, or L62. Assuming that the stereospecificity for the methyl groups was determined correctly by the Goto group^[140b] and considering that only *pro(R)*-methyl labelled groups were labelled here, the L2 residue matches this peak. As expected, the published ^{13}C chemical shifts seem to align better with the recorded spectra compared to the ^1H chemical shifts. Of course, assigning peaks from published chemical shifts is more error-prone than mutation-derived assignments.

In summary, 17% of the amide backbone and 12% of the ILV-methyl sidechain peaks for full-length GlpG could be assigned reliably. Despite these low numbers, the three methyl assignments derived from mutations are placed in functionally interesting regions for GlpG activity, while, for example, L5 and TM5 are only sparsely covered by amide peak assignments. In addition, methyl groups are more advantageous reporters for characterizing large complexes via solution NMR compared to backbone amide groups.

3.2.3 Chemical shift perturbations for full-length GlpG induced by substrate-like inhibitors

As mentioned in Figure 32, several GlpG core complex structures with substrate-like inhibitors have been solved. Even so, substrate recognition and cleavage are not yet fully understood. In this chapter, the amide and methyl chemical shifts of full-length GlpG in apo and inhibitor-bound states are compared, focussing on three substrate-like inhibitors. The inhibitor STS1775 has a short peptide chain comprised of four amino acids and deactivates GlpG wild type by forming a covalent bond with S201. Similar inhibitors have already been studied in complex with the GlpG core^[163]. Ksp152a has a more extended peptide chain of 30 amino acids, can be cleaved by GlpG wild type, and was derived from HybA. The GlpG-Ksp152a complex was obtained using the deactivated mutant S201T. The third and fourth inhibitors are called Ksp144 and Ksp145 and were cross-linked with the GlpG constructs called pPR16 and pPR14, respectively. pPR16 contains the mutations C104A, S201T, and M247C. For crosslinking, a disulfide bond between the cysteine residue in the P2 position of the inhibitor and the cysteine residue M247C of GlpG is formed. Similarly, pPR14 has the mutations C104A, W196C, and S201T, and was crosslinked to Ksp145 via a cysteine residue in the P4 position.

While no structure of full-length GlpG and STS1775 has been obtained, the X-ray complex structure of GlpG core with a similar inhibitor, called Ac-RVRHA-phenylethyl has been published (PDB: 5MT6^[163]). This inhibitor binds to GlpG with a K_D value of 220 nM^[163] and is extended by one amino acid at the N-terminus compared to STS1775. As it had been shown that removing this amino acid did not significantly affect the ligand potency^[163], the small difference can be neglected. The main recognition motif for binding includes positions 1-5 (chapter 3.1.4).

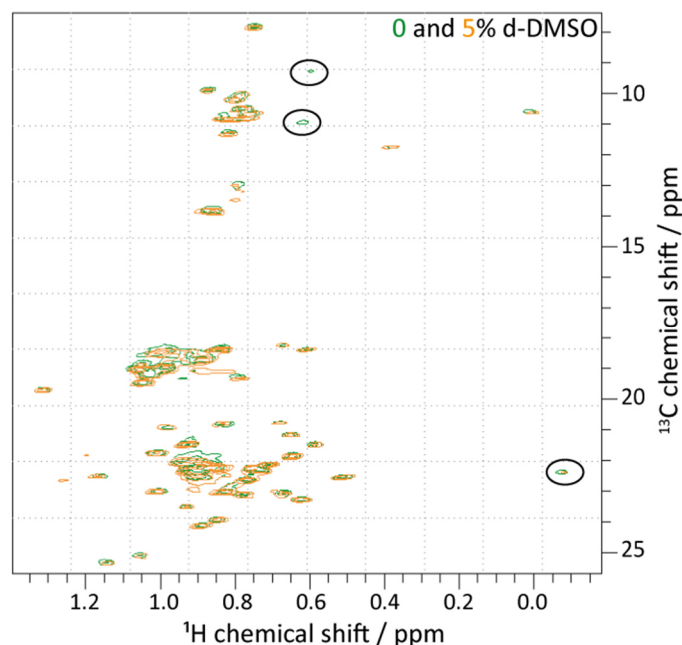


Figure 44: Overlay of ^1H - ^{13}C SOFAST-HMQC spectra of GlpG in presence of 0% (green) and 5% d_6 -DMSO (orange). A 360 μM D-, ^{13}C -, ^{15}N - and ILV-methyl labelled full-length GlpG S201T sample without His-tag (cleaved off by TEV) was used, which was eleven days old and kept in 50 mM d_5 -TRIS buffer, 150 mM NaCl, 0.11% d_{38} -FC-12, 10% D_2O , and 3.8 mM DSS at pH 8.0. The spectra were recorded at 40 $^\circ\text{C}$ on an 800 MHz spectrometer for 26 min.

The inhibitors were added from DMSO stocks. To factor in chemical shift changes of the methyl groups induced by DMSO, a titration was performed to solely estimate the influence of DMSO on the chemical shifts of each residue of GlpG (Figure 44). Comparing the ^1H - ^{13}C NMR spectra of GlpG S201T with 0 or 5 % d_6 -DMSO, it can be stated that no CSPs were measured for most peaks. Solely decreased peak intensities for one isoleucine, I4, and one leucine from the core were visible when adding DMSO. Likely, unspecific interactions with DMSO or already low signal-to-noise ratios led to these observations. ^1H - ^{15}N NMR spectra were acquired as well but the signal-to-noise ratios were insufficient.

Upon binding of STS1775 to full-length GlpG wild type, CSPs were detected for many amide and methyl peaks (Figure 45). Most amide peaks reflect the slow-exchange regime for

STS1775 binding (Figure 45, panel A). To visualize which residues were impacted most, CSP plots were prepared, revealing changes especially throughout the GlpG core region (Figure 45, panels C and E). The largest changes were experienced by residues V96, L114, D128, H145, R168, F192, G194, and G240. These residues have in common that they are all located at the edge of a transmembrane helix or in a loop. Out of these residues, H145, F192-G194, and G240 are in immediate proximity to the active binding site, as expected. Although L114 and D128 are placed on the periplasmic site like the active binding site, they can be estimated to be >10 Å away from the inhibitor. These findings imply that the rearrangements of GlpG upon inhibitor binding are extensive at the periplasmic site. V96 and R168 experienced larger changes in their chemical environments, although they are facing the extracellular site of the transmembrane region. These rearrangements were supposedly indirectly caused by inhibitor binding. Most likely, larger changes in the chemical environment of residues do not only affect one amino acid but several residues in proximity, as already noticed for F192 and G194 at TM3. For V96 and L240, very few neighbouring residues in the sequence were assigned. The proximity to the binding site for L240 at the tip of TM5 and chemical shift changes for R128, which is spatially close to V96, support the reliability of the measured changes to be consequences of ligand binding. Focussing at D128 in L1, residue A124 in L1 underwent chemical shift changes as well. Looking at H145 reveals higher CSPs for H141 and A142 as well. The chemical shift changes of R168 are supported by higher perturbations for several residues close by, especially S171. While the assignments at L5 and TM5 are sparse (Figure 39), residues E216, L229, I255, and G257 were assigned and revealed changes in their chemical environment, too. E216 and L229 are placed at the tips of TM4 and TM5, respectively, at the extracellular site of the transmembrane region. These two residues might be indirectly affected by inhibitor binding. I255 and G257, on the other hand, could be affected both directly as they are placed close to the tip of TM6 as well as indirectly via changes propagated from the tip of TM3 with F192 and G194, showing changes in their chemical environment.

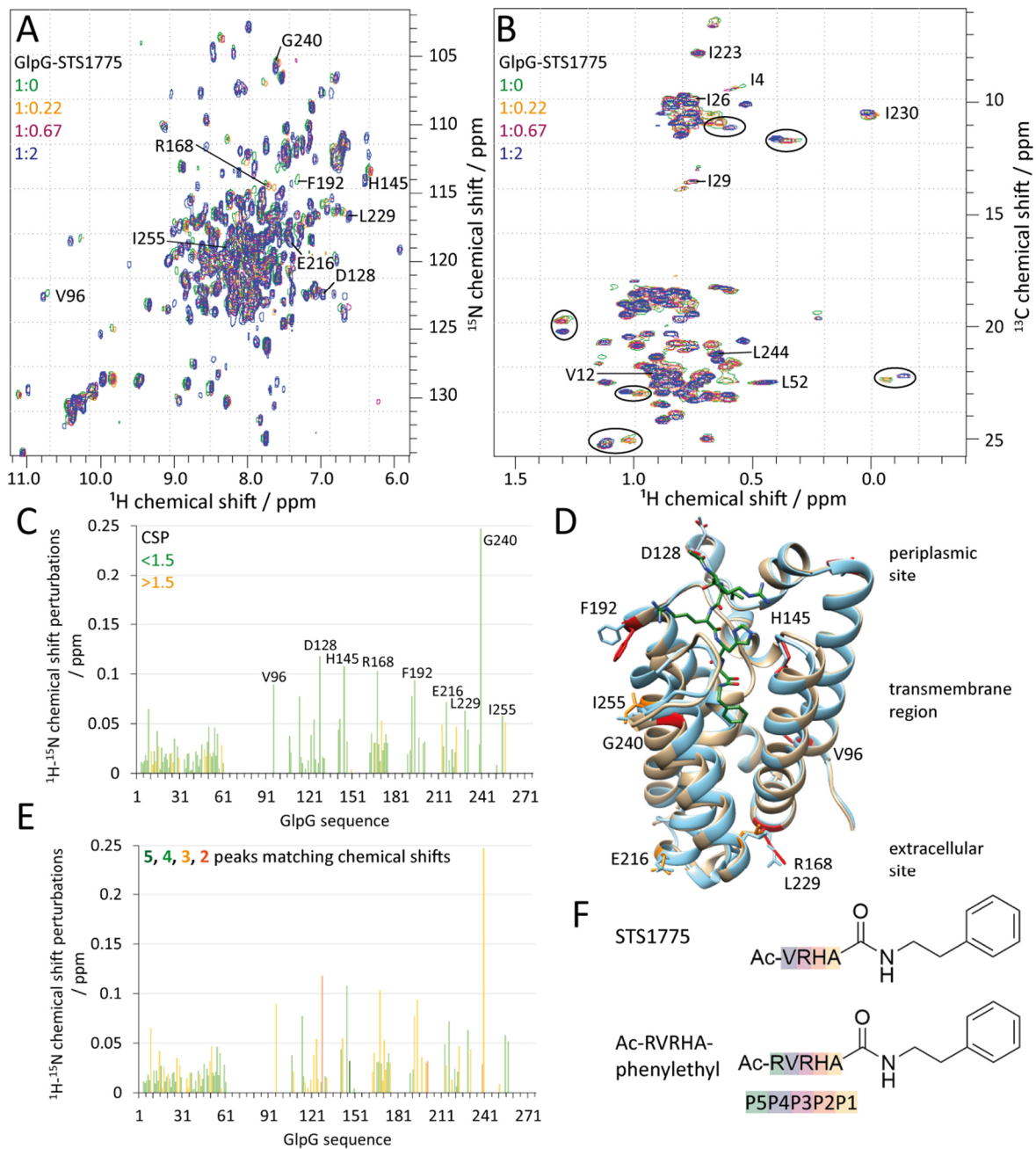


Figure 45: STS1775 ligand titration to full-length GlpG wild type recorded at 40 °C on an 800 MHz spectrometer. The sample contained 240 μM D-, ^{13}C -, ^{15}N - and ILV-methyl labelled GlpG in 50 mM TRIS buffer, 150 mM NaCl, 0.11% d_38 -FC-12, 1 mM DSS, and 10% D_2O at pH 8.0. Molar GlpG-to-STS1775 ratios of 1:0 (green), 1:0.22 (yellow), 1:0.67 (purple), and 1:2 (blue) were chosen with $\leq 4\%$ DMSO. Methyl and amide assignments were indicated for peaks with the largest CSPs. **A, B.** ^1H - ^{15}N TROSY and ^1H - ^{13}C SOFAST-HMQC spectra recorded for 1 h 20 min and 27 min, respectively. **C, E.** CSP plot coloured according to the HN, HNC0, HNC1, and HNCACB assignments (Figure 39). The assignment certainty was judged by the CSP values below (green) or above (yellow) the threshold of 1.5 compared to the published ^1H and ^{15}N chemical shifts^[140b, 164] (**C**) or by the number of nuclei of each residue matching the published chemical shifts (H, N, CA, CO, CB) with 5 (dark green), 4 (light green), 3 (yellow), and 2 (red) matches (**E**). **D.** X-ray structures of apo-GlpG (blue, PDB: 2IC8^[138a]) and the GlpG-Ac-RVRHA-phenyl-ethyl complex, a STS1775-like inhibitor (beige and inhibitor in green, PDB: 5MT6^[163]). The five residues with large CSPs were highlighted in red and three residues with lower CSPs were coloured in orange. **F.** Ligand structures of the ligand STS1775 used here and ligand Ac-RVRHA-phenyl-ethyl (PDB: 5MT6^[163]) for comparison. Positions 1-5 (P1-P5) were labelled via coloured backgrounds.

Most perturbations for the N-terminal domain upon addition of STS1775 were minor with the largest changes being observed for N9. In the predicted full-length GlpG structure, N9 is placed at the tip of the N-terminal domain facing GlpG core and the region linking N-terminal domain and core. It can be speculated that the N-terminal domain might interact directly or indirectly with GlpG core as suspected based on the deviating exchange regime for I223 comparing full-length GlpG and GlpG core. However, small CSPs can also occur via unspecific interactions with the inhibitor or DMSO. As the amide peak assignments were transferred from published chemical shifts^[140b, 164], the assignment uncertainty was considered as well (Figure 45, panels C and E). In panel C, chemical shift changes per residue were weighed against the published chemical shifts. In panel E, the assignment certainty was derived from the number of assigned nuclei per residue matching the published chemical shifts. For the N-terminal domain, many assignments were considered reliable according to both coloured plots. For the core, the assignments in panel C appeared more reliable compared to the plot in panel E as the threshold was chosen more generously.

The CSPs of the methyl peaks indicate both slow- and fast-exchange regimes upon binding of STS1775 (Figure 45, panel B). No CSP plot was prepared as only few peaks have been assigned. Overall, ~63% of the N-terminal methyl groups show CSPs of <0.05, which can be considered negligibly small. I26 and L52 experienced small changes in their chemical environments both being close to the β -sheet network of the N-terminal domain of GlpG. For L52, the chemical exchange occurred in the fast regime. I26 and L52 did not experience CSPs during the titration with DMSO. Hence, their perturbations when adding STS1775 might derive from the altered interactions between the N-terminal domain and the core. The CSPs for I4 also correspond to the fast regime but are more extensive. Already in the apo state, the peak for I4 was broadened. I4 is located at the N-terminal tip of the GlpG sequence and is part of the network of three β -sheets in the N-terminal domain of GlpG. Likely, the dynamics of I4 are not important for the catalytic function of GlpG, being far away from the binding site. Instead, these CSPs likely correspond to unspecific interactions with DMSO, as concluded during DMSO titrations (Figure 44).

The core methyl groups experienced more extensive changes in their chemical environments compared to the N-terminal domain (Figure 45, panel B, peaks with largest changes circled in black). Unfortunately, most of these peaks have not been assigned. One isoleucine

experienced fast exchange. Perturbations in the slow-exchange regime were observed for one isoleucine, one valine, and two leucine residues. Interestingly, one of the circled peaks at a ^{13}C chemical shift of ~ 25 ppm was slightly broadened and then narrowed again during the ligand titration. A second peak close by, however, disappeared at higher ligand concentrations. It can be hypothesized that these peaks merged during the titration with STS1775. Furthermore, these peaks might correspond to two different conformations of the same leucine residue. Two conformations would accordingly be present in the apo state while only one conformation would be favoured in the ligand-bound state. Surprisingly, only small chemical shift changes were detected for I230. I230 is part of TM5, which was suspected to bend upon inhibitor binding^[162], so strong CSPs were expected. No changes were visible for I223 either. The peak for L244 is placed in a spectral region in which many peaks change their position according to the slow-exchange regime, hampering the peak assignment in the presence of the ligand. Interestingly, in the spectral region with ^{13}C chemical shifts of 20-21.8 ppm, more peaks appeared after adding STS1775 to GlpG with two-fold molar excess than were visible in apo-GlpG state. This implies that either more than one conformation per residue is adapted or that previously invisible peaks, for example, due to intermediate exchange, appear due to slowed down (or accelerated) dynamics of a methyl side chain upon inhibitor binding. The chemical shift of the L244 peak itself did not change in the titration series. It remains ambiguous whether a different peak took its place or if no changes occurred for L244. In summary, adding STS1775 induced chemical shift changes at the beginning for the core (V96), L1 (D128, H145), the tip of TM2 (R168), the tip of TM3 (F192 and G194), the tip of TM4 (R217), TM5 (L229, I230, G240), and TM6 (I255) of full-length GlpG. Most of these residues are located next to the active site and/or at the edges of the transmembrane regions, providing them with more freedom to move and being less braced by the detergent molecules compared to sidechains deeply embedded in the micelle.

Furthermore, inhibitors with a more elongated peptide sequence were tested, namely Ksp152a and Ksp144, which have been synthesized by our collaborators from the Strisovsky group^[168]. Amino acids P1-5 of these peptides are essential for the recognition by GlpG^[163]. A proline is placed four residues upstream. The introduction of prolines facilitates substrate cleavage by destabilizing the α -helix^[152]. While Ksp144 can covalently bind to the GlpG cysteine mutant called pPR16, Ksp152a is either cleaved by GlpG wild type or non-covalently

bound to GlpG S201T. To achieve high saturation of GlpG with these inhibitors, different amounts of substrate were required for covalent or transient binding.

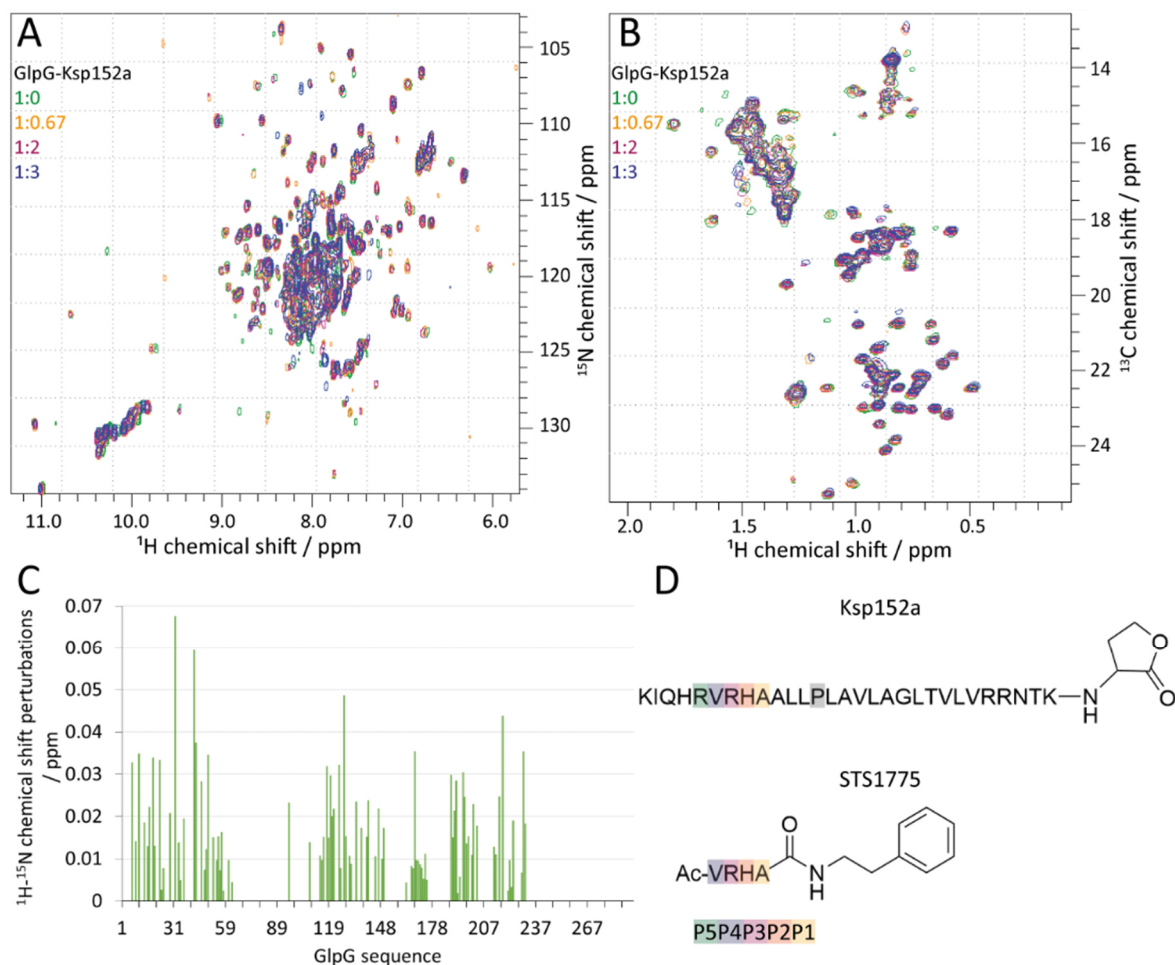


Figure 46: Ksp152a ligand titration to full-length GlpG S201T at different molar ratios recorded at 40 °C on an 800 MHz spectrometer. The sample consisted of 260 μM D-, ^{13}C -, ^{15}N - and ILVA-methyl labelled GlpG in 50 mM TRIS buffer, 150 mM NaCl, 0.11% $\text{d}_{38}\text{-FC-12}$, 10% D_2O , and 12 μM DSS at pH 8.0. The molar GlpG-to-Ksp152a ratios were 1:0 (green), 1:0.67 (yellow), 1:2 (purple), and 1:3 (blue) with $\leq 7\%$ $\text{d}_6\text{-DMSO}$. **A**, **B**. ^1H - ^{15}N TROSY and ^1H - ^{13}C SOFAST-HMQC spectra recorded for 1 h 20 min (amides), 27 min (methyl groups, green), or 54 min (methyl groups, yellow, purple, blue). **C**. CSP plot of backbone amides, corresponding to panel A. **D**. Structures of STS1775 and Ksp152a. Positions 1-5 (P1-P5) and a helix-destabilizing proline were labelled via coloured backgrounds.

The titration of Ksp152a to full-length GlpG S201T was tracked for backbone amides and ILVA-methyl sidechains (Figure 46, panels A and B). It should be noted that the methyl labelling scheme included alanine residues here. While a similar peak pattern for leucine and valine peaks was observed for both ILV- and ILVA-methyl labelling, the isoleucine sidechain peaks did not match the number of peaks from Figure 45, showing that isoleucine labelling likely did not work sufficiently during ILVA-methyl labelling. Only seven out of 29 alanine peaks were well-dispersed, while 32 peaks were expected in total. Due to the low resolution of alanine

peaks, no alanine assignments were further pursued except matching already published assignments for the N-terminal domain (Figure 67)^[140b]. Both amide and methyl spectra revealed only small CSPs, much smaller compared to when STS1775 was added (Figure 45), even at three-fold molar excess of substrate over protein. Due to the larger substrate size and the larger potential interaction surface between protein and ligand in the transmembrane region, larger CSPs were expected in especially the core region of GlpG for the substrate compared to STS1775 binding. However, the largest perturbations upon substrate binding were detected for H32 and E42 in the N-terminal domain. In addition, F127 and D218 from GlpG core experienced perturbations above 0.04 ppm. For the alanine peaks with ¹H chemical shifts around 1.5 ppm and ¹³C chemical shifts of roughly 17 ppm, CSPs in the fast-exchange regime were detected. According to published assignments (Figure 67), these peaks likely correspond to alanine residues of the N-terminal domain. In summary, it can be suggested that these data do not reflect the binding of Ksp152a. The perturbations rather correspond to unspecific interactions with Ksp152a or DMSO and indicate that the active site of GlpG remained in the apo state for the majority of the GlpG molecules in the sample. In contrast to STS1775, the substrate Ksp152 needs to be surrounded by detergents to shield the hydrophobic transmembrane region, while soluble inhibitors like STS1775 are not dependent on detergent micelles. Therefore, the interaction surface of STS1775 is not shielded by detergents, and their effective affinity does not depend on the detergent concentration. For Ksp152a, the likelihood for GlpG and the substrate to meet and interact productively is low, which is why a three-fold excess of Ksp152a over GlpG was aimed at. In addition, STS1775 binds covalently while Ksp152a binds non-covalently to GlpG. Taken together, these factors reflect that the STS1775 inhibitor interacts more easily with GlpG than Ksp152a and forms a more stable complex, resulting in nearly saturated GlpG in the presence of a two-fold excess of STS1775. Meanwhile, a three-fold excess of Ksp152a over GlpG was apparently not sufficient to increase the likelihood for GlpG and Ksp152a meet and to form a stable complex. To crosslink the Cys-mutants pPR14 and pPR16 with the substrates Ksp145 and Ksp144, respectively, the expression and purification of the new GlpG constructs was established first (Figure 47, panels A and B). For pPR16, similar expression and purification yields could be achieved as for the GlpG S201T construct. For pPR14, lower yields were obtained, thus pPR16 was selected for further optimizations. Reasons for more impurities and low yields for pPR14-based complexes can be low expression yields or low stability of the constructs during

purification. Before crosslinking with labelled pPR16, different ratios between Ksp144 and pPR16 were tested with small batches (Figure 47, panel D). As expected, higher excess of substrate elevated the ratio between crosslinked and apo-pPR16 with most pPR16 already crosslinked at two-fold excess of substrate. However, no strategy for separating apo- and crosslinked pPR16 was available, which is why ideally 100% of GlpG in the sample should be occupied. At five- to ten-fold excess of Ksp144, no clear band of apo-pPR16 was observed anymore, suggesting that the remaining apo-pPR16 amounts were negligibly low. Meanwhile, the amounts of monomeric Ksp144 were similar under all three conditions. Only the amount of cross-linked Ksp144 dimers increased with higher substrate excess over pPR16. As reasonable compromise, a five-fold excess of substrate was chosen for preparing crosslinked samples.

To obtain CSPs, the ^1H - ^{13}C SOFAST-HMQC spectra of apo-pPR16 and pPR16 crosslinked with the Ksp144 substrate in different buffers were compared (Figure 47, panel C). Unfortunately, the spectrum of apo-pPR16 already displayed low-dispersed peaks compared to GlpG wild type or the S201T mutant, indicating increased sample heterogeneity. A possible explanation could be lower construct stability. Surprisingly, no signs of aggregation were observed during purification via SDS-PAGEs or during SEC (Figure 47, panels A and B). Moreover, 5 mM DTT had been added to the apo-pPR16 construct buffer to avoid disulfide bond formation between the M247C residues or GlpG monomers, so dimerization was not the reason for the low quality of the NMR spectrum. While peaks for the leucine and valine residues were missing and showed low dispersion, the peaks for isoleucine residues matched the peak pattern of full-length GlpG S201T quite well (Figure 40 panel D). Hence, the isoleucine peaks remained as reporters on CSPs upon cysteine crosslinking. Due to the low protein concentration in the sample, recording ^1H - ^{15}N spectra was not an option.

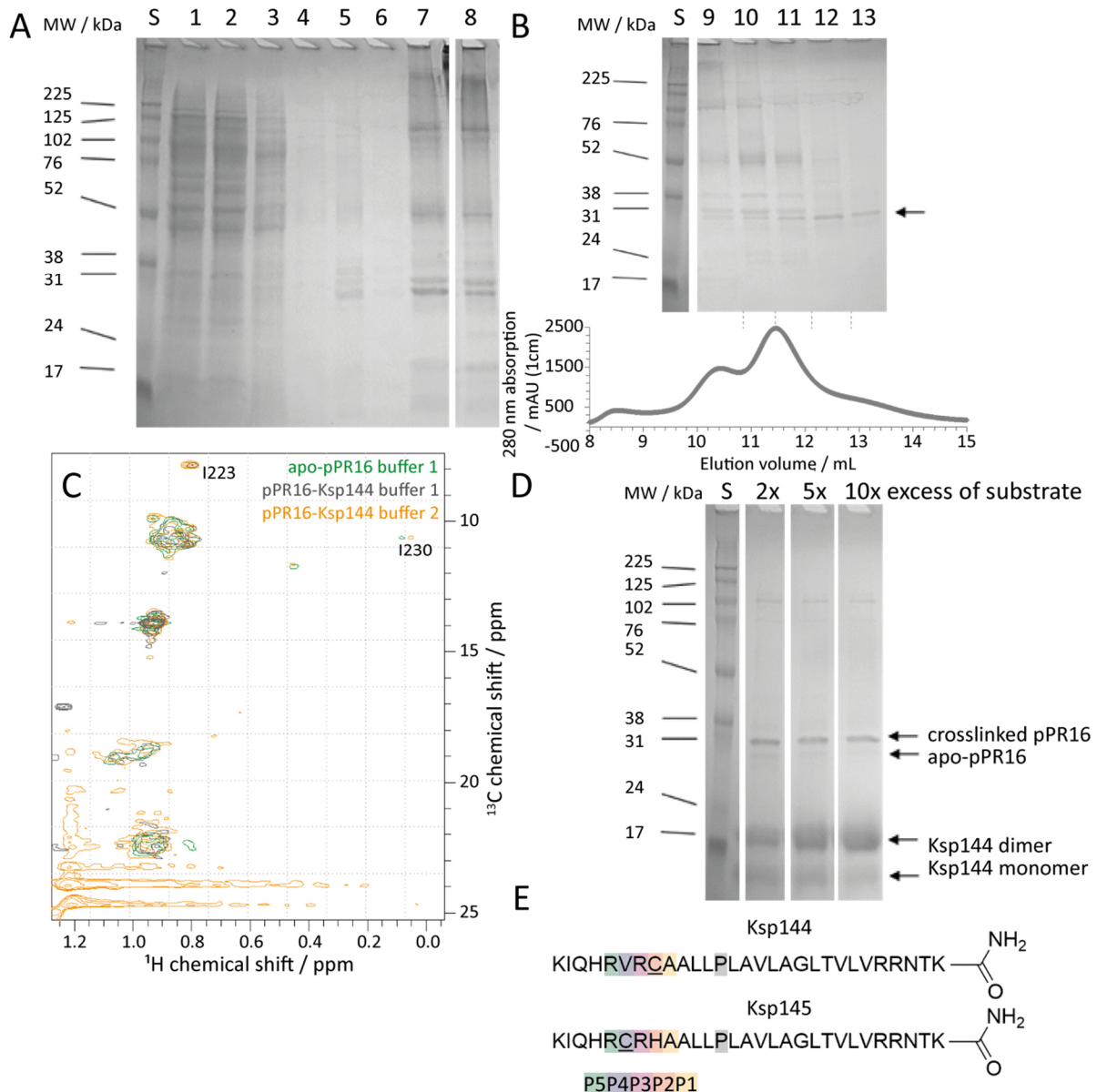


Figure 47: Purification and crosslinking of D-, ¹³C-, ¹⁵N- and stereospecifically ILV-methyl labelled pPR16 with Ksp144. For the SDS-PAGEs, Laemmli buffer without 2-mercaptoethanol was used. S: Amersham ECL Rainbow Marker. **A, B.** SDS-PAGEs and SEC chromatogram of the pPR16 purification. Affinity chromatography with HisPur Cobalt resin: 1: input, 2: flow through, 3, 4: wash fractions 1 and 2, 5, 6: elution fraction 1-2 and 3-4; 7: desalted elution fractions; SEC: 8: input, 9-13: elution fractions corresponding to the SEC chromatogram below. **C.** ¹H-¹³C SOFAST-HMQCs of pPR16 recorded at 40 °C on 700 MHz spectrometer for ~17 h. 50 μM apo-GlpG in 25-50 mM TRIS buffer, 150 mM NaCl, 0.11% d₃₈-FC-12, 5 mM DTT, and 10% D₂O at pH 8 (green), 30 μM GlpG-Ksp144 crosslinked with five-fold excess of Ksp144 in CAPS buffer, 150 mM NaCl, 0.11% FC-12, and 10% D₂O at pH 10 (orange), and 13 μM GlpG-Ksp144 crosslinked with five-fold excess of Ksp144 in the same buffer as apo-pPR16 (grey) were used. **D.** SDS-PAGE of small-scale GlpG crosslinking samples with two-fold, five-fold, and ten-fold excess of Ksp144. **E.** Structures of inhibitors Ksp144 and Ksp145 with cysteine residues introduced for crosslinking in positions P2 and P4, which have been underlined, respectively.

After crosslinking, the complex was examined both in the same buffer as apo-GlpG as well as in the buffer used for crosslinking. In the crosslinking buffer, the spectrum showed that I223 did not experience CSPs. Considering that this peak was sometimes not visible when changing

TRIS and phosphate buffers at pH 8.0 or 7.4 (called peak 1 in Figure 43, panel B), buffer and pH changes of 10.0 and 8.0 were expected to alter the peak intensity for I223. Interestingly, I230 was perturbed by crosslinking and/or buffer exchange. Comparing TRIS and phosphate buffers at pH 8 and 7.4 (Figure 43 panel B), no such changes were detected. To test whether the perturbations originate from choosing a different buffer or from crosslinking, another spectrum was recorded using the same sample buffer as for apo-GlpG. Unfortunately, the well-dispersed isoleucine peaks for the crosslinked pPR16-Ksp144 complex disappeared after changing the buffer and only peaks in the highly crowded region for isoleucine peaks were visible. These changes impeded any peak traceability, likely due to sample loss during the buffer exchange.

Overall, only the titrations with the inhibitor STS1775 could be analysed thoroughly for both amide and methyl groups. Extensive changes in the chemical environments were found, which were not only observed at the active site but extended even towards loops facing the other site of the transmembrane region. While the saturation of GlpG with substrate could be augmented via crosslinking, as shown via SDS-PAGE, only poorly resolved NMR spectra were obtained, which was probably caused by high sample heterogeneity.

3.2.4 Methyl sidechain dynamics of full-length GlpG detected via solution NMR

As indicated by CSPs, the changes upon inhibitor binding were not solely limited to the residues flanking the binding site but were also observed for core residues, which could only be indirectly affected by inhibitor binding. Moreover, different exchange regimes have been detected. These are promising indicators for changes in protein dynamics upon inhibitor binding. GlpG cleaves peptides with a low conversion rate of roughly 0.4/min^[170], implying rather slow conformational changes associated with catalysis. Here, ¹³C-CPMG experiments were implemented to measure slow microsecond and fast millisecond dynamics (in the following referred to as millisecond motion) to determine which parts of the protein are flexible in the apo state and in complex with substrate-like inhibitors on this timescale.

Indeed, millisecond motion was detected for several methyl sidechains of apo-GlpG, looking at the $R_{2,eff}$ rates at different CPMG frequencies. The full-length GlpG S201T sample prepared for the CPMG experiments was methyl labelled as before (chapter 3.2.3, chapter 1.1.6), labelling only the methyl group carbon nuclei as ¹³C, while all other sidechain carbons were

kept as ^{12}C to avoid C-C J -couplings^[28-29]. In addition, the protein was deuterated except for the methyl groups, reducing proton dipolar couplings as much as possible.

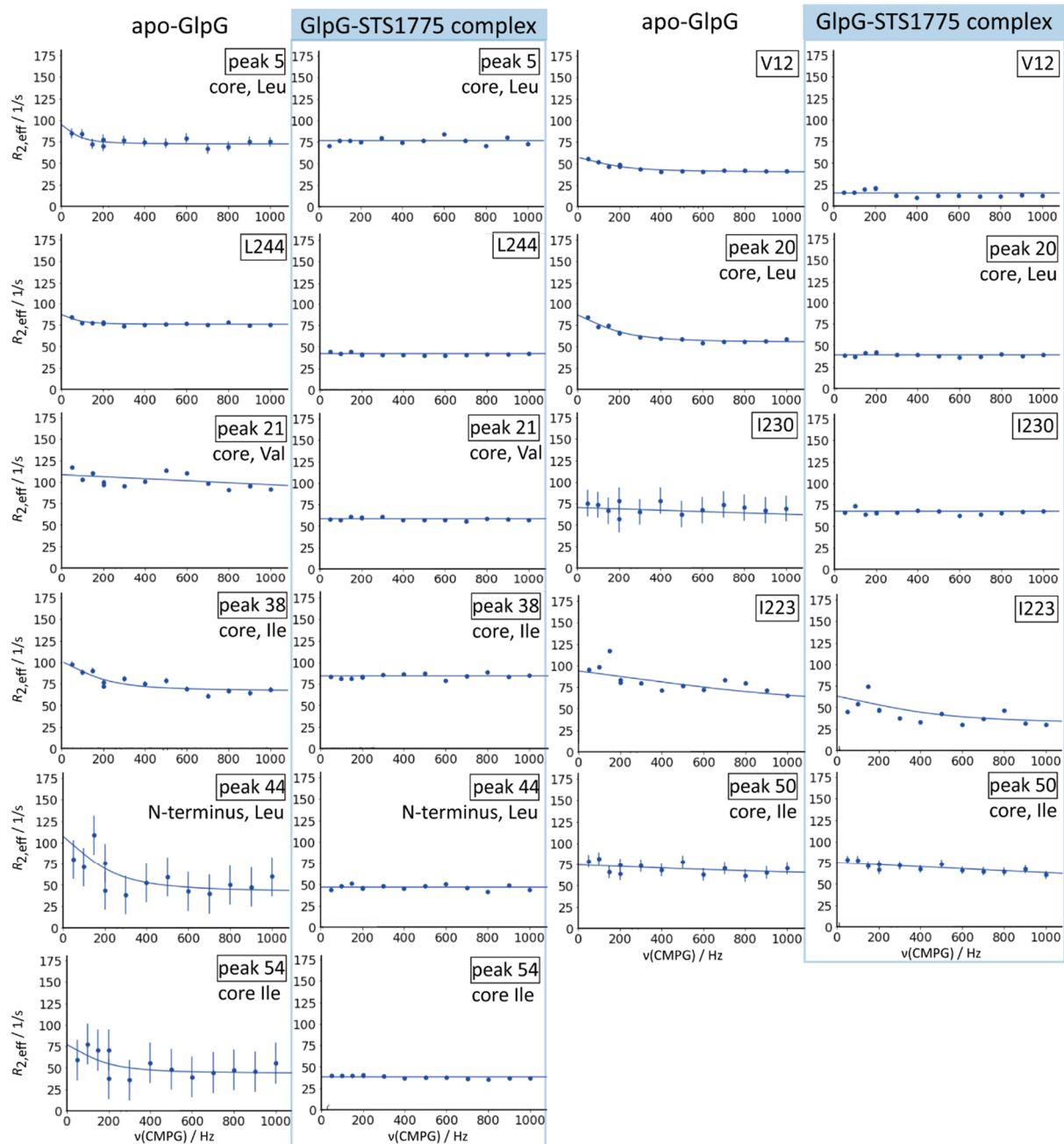


Figure 48: Relaxation dispersion plots of ^{13}C -CPMG experiments acquired for apo-GlpG and the GlpG-ST51775 complex at 313K on an 800 MHz spectrometer. The non-assigned peaks were numbered randomly (Figure 66).

In the apo state, millisecond motion was detected for several methyl groups of GlpG core and the N-terminal domain including V12, I223, and L244 (Figure 48). In addition, two leucine residues, one valine, and one isoleucine residue of the core plus one leucine of the N-terminal domain (peaks 5, 20, 21, 38, and 44, respectively) showed millisecond motion. Instead, no significant millisecond motion was detected for I230. While these motions were present in

the apo state, they were all diminished upon STS1775 binding, indicated by flat lines fitted on the data points of each residue. The only residues for which data fitting still concluded millisecond motion in an inhibitor-bound state were I223 and potentially peak 50, corresponding to an isoleucine residue of GlpG core. These two profiles only show a very weak signature of residual conformational exchange, corroborating that the millisecond timescale exchange present for apo-GlpG is largely abrogated upon inhibitor binding.

In addition, the CPMG experiment also allows to determine the baseline R_2 values for each residue, reporting on pico- to nanosecond timescale motion (in the following referred to as nanosecond motion). These R_2 values are derived from the baseline value reached after the initial decay of the function fitted to the data points at high CPMG frequencies. Evidently, also nanosecond motion decreased, or did not change, upon inhibitor binding. Statistically, the reduction of nanosecond motion proved significant for methyl groups belonging to the GlpG core domain comparing apo and inhibitor-bound states with $p < 0.01$ (Figure 49 panel B). In contrast, no statistically significant change could be shown for methyl groups belonging to the N-terminal domain of GlpG.

The assigned peaks indicate, as local reporters on motion, which parts of GlpG moved on the detected timescales and whether they were influenced by inhibitor binding. The three methyl groups that have been assigned for GlpG core are depicted in panel A in Figure 49. L244 is part of L5 between transmembrane domains TM5 and TM6 and shows a decimation of the baseline R_2 value from 75.61 to 41.95 s^{-1} upon inhibitor binding. The relaxation dispersion curve for L244 also displays slow millisecond motion with a small increase limited to the first point of the plot for apo-GlpG, which was diminished once the inhibitor was bound. L5 was suggested to serve as a gate for inhibitors which can be lifted in the apo state and closed when occupied with a substrate (chapter 3.1.3). Millisecond motion of the overall loop and nanosecond motion of the sidechain would corroborate this theory. A drop in nanosecond dynamics based on the baseline R_2 value was likely induced by a direct or indirect interaction with the inhibitor bound to the active site. According to X-ray studies, all L5 residues except for L244-M247 are rigidified upon binding of an inhibitor with four amino acids^[171]. CSPs upon ligand binding were not detected (chapter 3.2.3).

For I230, neither changes in the milli- nor nanosecond motion regimes were measured. Notably, no millisecond motion was detected in the apo-GlpG state to start with, although

small CSPs were obtained upon inhibitor binding (chapter 3.2.3). X-ray studies indicated that TM5 is first disordered when the inhibitor initially binds and then stabilized again^[171].

I223, which experienced no chemical shift perturbations upon inhibitor binding, changed its methyl sidechain dynamics. Despite a noisy data set, it seems that both milli- and nanosecond-dynamics were altered upon inhibitor binding. The differences on the millisecond timescale remain ambiguous as the fitted curves do not match the noisy data well, showing increased χ^2 values of 1200-1570 (Table 25, Table 26). I223 is part of L4 between transmembrane domains TM4 and TM5. Interestingly, the inhibitor binding pocket is far away from this residue. This might be the reason why both nano- and millisecond motion seem to be decreased but not abrogated upon inhibitor binding. Interestingly, only rotations around the C $_{\alpha}$ -C $_{\beta}$ -bond have been reported via X-ray studies in bicelles^[171], which do not match the observations here.

Moreover, CPMG plots with differences for both nano- and millisecond motion between apo and STS1775-bound state can be reported for peaks 20, 21, and 38, corresponding to leucine, valine, and isoleucine methyl groups of the core. For peaks 20 and 21, the nano- and millisecond motion was depleted upon inhibitor binding. Instead, the CPMG data on peak 38 indicate that the millisecond motion was abrogated but the nanosecond motion did not change. Peaks 5 and 54 represent leucine and isoleucine methyl groups of GlpG core, respectively, for which the nanosecond motion was not impaired by inhibitor binding. Instead, the slow millisecond motion for peaks 5 and 54 was depleted. It should be considered that large error bars are associated with the CPMG plots for peak 54 in apo-GlpG state.

The N-terminal domain was assumed to be flexible, but no influence of inhibitor binding was expected. By contrast, changes of both nano- and millisecond motion were detected for V12, and changes in millisecond motion were also determined for peak 44, a leucine methyl sidechain from the N-terminal domain. V12 is part of an α -helix at the N-terminal domain and experienced a decrease in both slow millisecond and nanosecond motion upon inhibitor binding. The plots for peak 44 reflect millisecond motion in the apo state, which was depleted after inhibitor binding. As mentioned before, such changes, especially on the millisecond timescale, have been thought to be unlikely for residues of the N-terminal domain as they are placed far away from the binding site. These residues might experience unspecific interactions. Alternatively, residues of the N-terminal domain could be indirectly affected, assuming interactions with the GlpG core. Moreover, some assignments for the N-terminal

domain could be wrong so these changes in motion would rather represent methyl sidechains of GlpG core. This also seems reasonable as both peaks overlap with other peaks. It can be concluded that only CPMG dispersion profiles of well-separated peaks should be considered reliable, even if the peak maxima seem to be easily distinguishable (as for peaks 17 and 44, Figure 66).

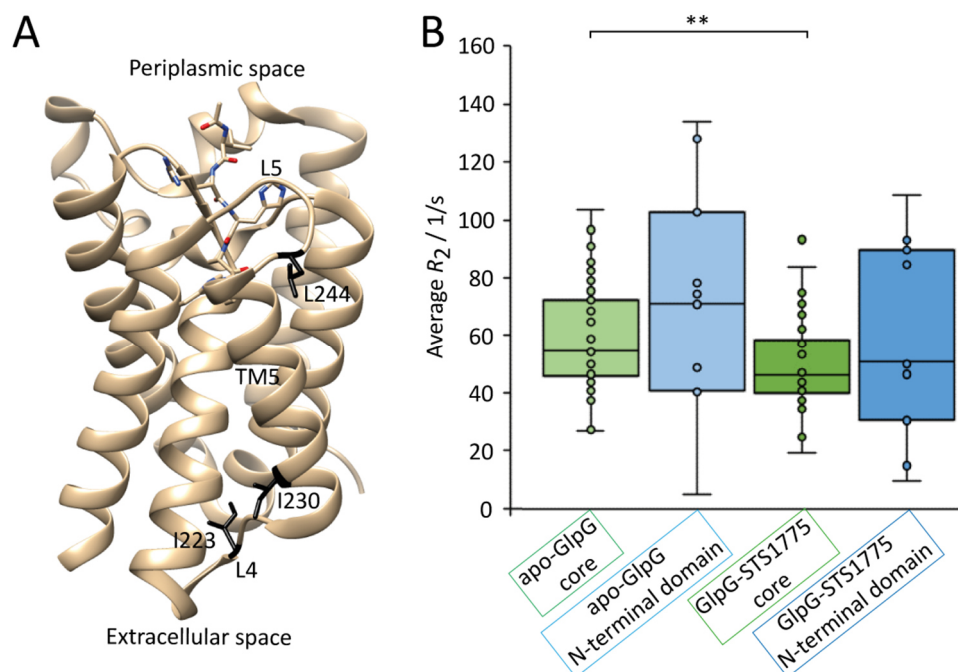


Figure 49: Structure and dynamics features of apo-GlpG and the GlpG-STs1775 complex. **A.** X-ray structure of GlpG core in complex with inhibitor Ac-VRHA-cmk (PDB: 5mt7^[163]). The assigned residues I223, I230, and L244 are highlighted in black, including the names of the secondary structural features. **B.** Boxplots of baseline R_2 rates corresponding to the CPMG data of apo-GlpG (light green and blue) and the GlpG-STs1775 complex (saturated green and blue) dividing the residues into GlpG core (green) and the N-terminal domain (blue) (chapter 3.2.2). A Welch two-sample t-test was performed with $R^{[172]}$ determining a significant difference (** for $\alpha < 0.01$) between the nanosecond dynamics of GlpG core in the apo and the inhibitor-bound state (Table 25, Table 26, and Table 27).

Looking at the exchange regimes determined for the CPMG fits, fast millisecond motion in the regime of 1.7 ± 6 ms was detected for I223 of apo-GlpG. I223 is part of L4, which is placed on the opposite site of the transmembrane region compared to the active site. When inhibitor was bound, this millisecond motion of I223 was apparently slowed down by roughly ten-fold. This could correspond to local loop flexibility which might be involved indirectly in ligand binding or interactions with other molecules in the extracellular space.

Overall, it can be concluded that milli- and nanosecond motion of methyl sidechains of full-length GlpG were diminished when the inhibitor STs1775 was bound. Most of these motions were localized at the GlpG core, for which a significant difference of nanosecond dynamics in

the apo and inhibitor-bound state was determined. It can also be deduced that peaks in overlapped regions and the transferred assignments for the N-terminal domain bear a degree of uncertainty which should be considered during interpretation.

3.2.5 Crystallizing GlpG core for solid-state NMR studies

Alternatively, large molecules or complexes can be studied via solid-state NMR, as these experiments do not rely on a fast tumbling like in solution (chapter 1.1.2). Membrane proteins for solid-state samples are often concentrated into pellets by centrifugation or crystallized. During crystallization, the thermodynamically most favourable protein conformations are populated, which usually results in high sample homogeneity. Therefore, NMR peak line widths can be assumed to be narrower for crystalline samples compared to protein pellets. Crystals of GlpG core in NG micelles have been studied via X-ray crystallography (Table 8) but not yet via solid-state NMR. Crystallization is often a bottleneck during sample preparation for solid-state NMR, as several milligrams of crystals are needed. On the other hand, GlpG core has already been studied in liposomes^[164]. Taken together, it was aimed for solid-state NMR studies with GlpG core crystals.

The expression and purification of GlpG core was established using either the usual full-length GlpG construct or a GlpG core construct^[163, 171] (see chapter 4.10.4). Expressing full-length GlpG, the construct was cleaved with chymotrypsin to remove the His-tag and to obtain GlpG core before exchanging the detergent from DM to NG via SEC. For the GlpG core construct, the chymotrypsin cleavage step was substituted by a thrombin cleavage step to remove solely the His-tag. Otherwise, the purification protocols were identical. Chymotrypsin is a promiscuous protease which cleaves off both the N- and the C-termini except for sequences shielded by the micelle. According to a peptide cutter online tool offered by the Swiss Institute of Bioinformatics^[173], chymotrypsin cleaves off the N-terminal sequence after residues 80 or 87 and the C-terminus after residues 285, 286, or 266 with similar probabilities each. In congruence with these estimations, the chymotrypsin cleavage yields after 19 hours (Figure 50, panel A) and before 55 hours (Figure 50, panel C) were rather heterogeneous. Interestingly, several cleavage products around 25 kDa size appeared after adding chymotrypsin within less than one minute. The upper bands of these ~25 kDa cleavage products fades over time and only the band corresponding to the lowest molecular weight is visible after 55-72 hours of cleavage. Visually, more protein precipitated the longer the

incubation at room temperature lasted. According to literature, an incubation time of 36 hours was sufficient^[138a], which was not the case here despite using the same incubation parameters. Compromising heterogeneity with time efficiency and low precipitation, the chymotrypsin cleavage was always performed for at least 40-50 hours.

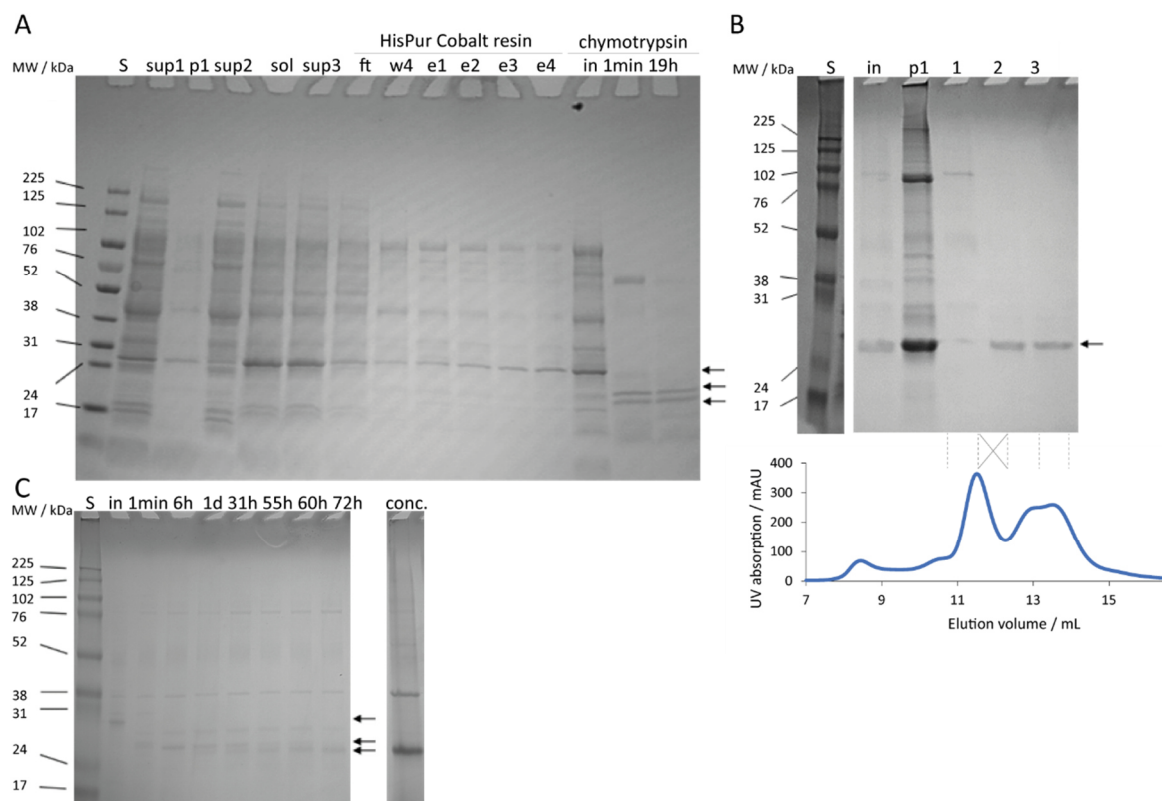


Figure 50: Exemplary purification of D-, ¹³C-, ¹⁵N- and stereospecifically ILV-methyl labelled GlpG core. S: Amersham ECL Rainbow Marker. Arrows indicate full-length GlpG and the cleavage products after treatment with chymotrypsin **A**. SDS-PAGE of samples tracking the purification progress. Sup1 and p1: supernatant and pellet after centrifugation at 10,000 g for 20 min; sup2: supernatant after centrifugation at 130,000 g for two hours; sol and sup3: solubilization supernatants before and after centrifugation at 100,000 g for 30 min. Affinity chromatography purification with HisPur Cobalt resin: ft: flow-through, w4: wash fraction 4, e1-4: elution fractions 1-4. Progress of chymotrypsin cleavage to obtain GlpG core with input (in) and samples taken after 1 min and 19 hours. **B**. Exemplary SEC chromatogram and corresponding fractions 1-3 on an SDS-PAGE for GlpG core. In: input; p1: pellet after centrifugation at 10,000 g for 10 min prior to SEC. **C**. SDS-PAGE with chymotrypsin cleavage fractions over time plus the input (in) and the final concentrated cleavage product (conc.).

Moreover, a GlpG core construct was expressed and purified in a similar fashion (Figure 50, panel B). The His-tag of in total sixteen amino acids with 1.8 kDa size was cleaved off with thrombin but could not be tracked on the SDS-PAGE due to the very small difference in size. The corresponding SEC chromatogram shows four peaks with the void volume at ~8 mL elution volume, some impurities of ~100 kDa molecular weight, and two peaks corresponding to GlpG core. On the SDS-PAGE, the protein sizes in fractions 2 and 3 look similar but the hydrodynamic radii according to the SEC chromatogram differ. These bands could correspond

to non-cleaved and cleaved GlpG core or different conformations. Similarly broad elution peaks had been detected for full-length GlpG (Figure 35) and GlpG mutants (Figure 42), too. Out of precaution, such SEC fractions were crystallized separately, as far as possible, as both sequence lengths and protein conformations can impact the crystallization.

Table 8: Overview on published hanging-drop crystallization conditions for GlpG core from *E. coli*.

PDB	GlpG construct sequence	Crystallization conditions	Buffer components
4NJP ^[170]	87-276	20 °C, pH 4.5	0.1 M sodium acetate, 3 M NaCl, 10% glycerol
2O7L ^[144]	93-272	20 °C, 5 mg/mL protein, pH 7.0	0.1 M Bis-TRIS propane, 3 M NaCl
3B45 ^[148a]	91-270	20 °C, 5 mg/mL protein, pH 7.0	0.1 M Bis-TRIS, 1.5 M NH ₄ Cl
2IC8 ^[138a]	91-272	20 °C, 5 mg/mL protein, pH 7.0	0.1 M Bis-TRIS propane, 3 M NaCl
2XOW ^[138c]	92-270	25 °C	2.5-3 M NH ₄ Cl
2NRF ^[139]	91-272	22 °C, pH 7.4	0.1 M tricine, 50-100 mM Li ₂ SO ₄ , 6% PEG 3,000 (w/w)

Following both procedures, low yields with roughly 0.25-0.5 mg protein per litre of unlabelled expression culture were achieved. When the GlpG core construct or GlpG treated with chymotrypsin were concentrated, small amounts of protein often precipitated. Despite that, the required concentration of >5 mg/mL^[138a] could be reached, and the protein was subjected to crystallization screens. First, crystallization was performed in the hanging-drop setup in 24-well plates at 20 or 22 °C, derived from the published crystallization conditions (Table 8, Figure 52). Two buffers with 0.1 M Bis-TRIS at pH 6.0-8.2 and 0.1 M Bis-TRIS propane at pH 6.0-9.0 were chosen. Each two precipitants were tested with 1-3 M NaCl or 0.5-3 M (NH₄)₂SO₄. Unfortunately, no protein crystals were obtained. During crystallization, most protein precipitated, and the few crystal-like structures did not absorb UV light, revealing that they did not consist of proteins. Consequently, the screening conditions were widened, resorting to classical membrane protein crystallization kits like MemGold1 and 2. These screens were performed in 96-well sitting-drop plates at 1:2, 1:1, and 2:1 protein-precipitant drop ratios at 20 °C. Again, most protein precipitated, and no crystals grew under these conditions.

Overall, the crystallization of GlpG core in NG micelles was not successful. Membrane proteins are known to be difficult to crystallize due to various reasons. For example, membrane protein concentrations are likely to be overestimated due to the presence of detergents which absorb UV light at a similar wavelength. Moreover, samples are often heterogeneous. Although several purification steps have been optimized to yield more homogeneous samples, a certain degree of heterogeneity remained and might have impacted the crystallization.

3.3 Conclusion and outlook

Overall, solution NMR samples of full-length GlpG have been prepared successfully. To the best of our knowledge, amide spectra displaying up to ~75% of the expected backbone peaks and methyl spectra of full-length GlpG have not been reported before with such quality using solution NMR spectroscopy. Newly assigned methyl peaks and transferred amide assignments were utilized to study binding of several substrate-like inhibitors via chemical shift perturbations. Moreover, changes in the fast millisecond dynamics regime of GlpG were detected for methyl sidechains upon binding of the inhibitor STS1775. GlpG crystals for solid-state NMR experiments have not been obtained.

As mentioned in the introduction, membrane proteins are preferably studied in lipid bilayers rather than detergent micelles. Unfortunately, the solution NMR experiments performed here did not allow a large molecular weight. Therefore, FC-12 micelles were chosen resulting in a ~52 kDa GlpG-detergent complex. Arguably, this experimental setup limits the resemblance to the native conditions of GlpG. As an effective compromise, a mild detergent, DM, was used for solubilization and later exchanged for FC-12. SDS-PAGE gels and NMR spectra did not indicate degradation or denaturation of GlpG S201T implying that the chosen conditions did not harm the integrity of the membrane protein. Fluorescence assays confirmed that the proteolytic activity of GlpG was conserved during expression and purification despite the exclusive usage of detergent micelles. Furthermore, GlpG structures crystallized in presence of lipids and detergents looked similar^[151].

In addition, mostly the inactive S201T mutant of GlpG was utilized for NMR measurements. Being inactive, the dynamics of this GlpG construct might deviate from the dynamics of GlpG wild type. In addition, the stability of GlpG S201T is lower than for GlpG wild type^[137]. On the other hand, only small chemical shift perturbations were observed comparing GlpG S201T and wild type. Moreover, both GlpG S201T and wild type induce membrane thinning, so at least the overall folding should be similar^[149a]. Eventually, proteins cannot be studied unless they are isolated and stabilized to allow measurements. The aim should always be to imitate the native conditions as much as possible and to track the state of the protein.

To study GlpG via solution NMR, several unambiguous assignments were obtained for GlpG. The amide assignments were derived from published chemical shifts in solution and the solid state. Consequently, these assignments are less reliable than *de novo* assignments based on,

for example, a backbone walk data set. These experiments were deemed too inefficient due to the large molecular size of GlpG, which resulted in missing peaks in the HNCA spectra. Compared to this experiment, HNcoCA or HNcaCO experiments for backbone assignments would have suffered even more from low sensitivity. As an alternative, methyl assignments were achieved by in total 25 point mutations. Three assignments for GlpG core residues were obtained with these mutants. It is not clear why most mutations did not result in a signal loss of at least one peak in the spectrum. On one hand, the mutation template could have been observed instead of the mutants. This hypothesis, however, is disproved by the sequencing results, obtained for each mutated residue. On the other hand, it could be that the number of peaks detected in the spectrum correspond to possibly several peaks per residue due to different conformations. As consequence, the number of observed peaks in the spectrum would not represent all methyl groups of the GlpG core, although it seems to match the expected peak number.

Out of all titrations with inhibitors including four (STS1775) or 29 (Ksp152a, Ksp144, Ksp145) amino acids, the only significant chemical shift perturbations of the GlpG peaks were obtained with STS1775. For this ligand, large changes in the chemical environments of residues close to the active site and/or at the edges of the transmembrane regions like V96, L229, and G240 were detected. For the more substrate-like inhibitors with extended peptide chains, only low protein occupancies have been achieved for Ksp152a. Achieving higher occupancies via cross-linking with mutant pPR16 and peptide Ksp144, a low protein concentration and likely increased sample heterogeneity did not allow the detection of well-separated peaks. As a consequence, the sample preparation would need to be optimized further, and pPR14 in combination with Ksp145 could be tested again as well.

In the end, the changes on the nanosecond timescale upon inhibitor binding derived from ^{13}C -CPMG experiments revealed global changes in the flexibility of many GlpG methyl sidechains. Using L244, I230, and I223 as locally defined reporters, changes in nanosecond motion were detected in L4 and L5. In TM5, no such changes were detected for I230. Also, fast millisecond dynamics of some GlpG residues like L244 changed once STS1775 was bound, indicating that not only fast sidechain but also slower motion of loops and α -helices were affected. Unfortunately, most of these residues could not be assigned. Consequently, advancing the methyl assignments is an important future goal, either via extended mutation screens or NOESY-based experiments. It would also be interesting to study GlpG mutants with enhanced

activity like TM5 mutants^[10-11], which could be reflected by higher or lower flexibility of the helix. Moreover, the chemical shifts of the excited states obtained via ¹³C-CPMG experiments should be determined by acquiring another data set at a spectrometer with a different magnetic field B_0 . Additionally, ¹³C-CEST experiments could identify dynamics on the slow millisecond and second timescale, which could be hidden under flat ¹³C-CPMG profiles.

As mentioned in chapters 3.1.3 and 3.1.4, the substrate binding process has been proposed to consist of multiple steps of recognizing, unfolding, and binding the substrates. Assuming that inhibitors with a long peptide sequence undergo similar steps, the aim would be to separate the different intermediate states, as it has already been done for an inhibitor with four amino acids^[171]. It could be that changes in the dynamics take place at the same time, are caused by each other, or happen independently upon binding of the interaction partner. Ways to test these hypotheses are, for example, comparing the dynamics of several mutants, which represent different functionally relevant conformations.

In summary, different purification conditions and experimental tools were screened and optimized to further advance the knowledge on GlpG dynamics via solution NMR. Furthermore, additional GlpG peak assignments were obtained. Moreover, chemical shift perturbations and ¹³C-CPMG data indicated that inhibitor binding not only influences residues close to the active site but is even transmitted to residues at the opposite side of the transmembrane region.

4. Materials and methods

4.1 Instrumentation

Table 9: Instruments.

Instruments	Specifications	Manufacturers
Accu-jet pro	Pipetting aid	Brand GmbH & Co KG
ÄKTA pure	Chromatography system	Cytiva
Analytical Balance LA 214	0.1 mg – 220 g	VWR International
Ascend 800 NMR spectrometer	With Cryoprobe, attached to AVANCE NEO console	Bruker
Ascend 700 NMR spectrometer	Attached to AVANCE NEO console	Bruker
Avestin EmulsiFlex C5	High Pressure Homogenization	ATA Scientific Pty Ltd.
Avanti J-15R	Table centrifuge	Beckman Coulter
BioDrop Duo	UV/VIS spectrophotometer	SERVA Electrophoresis GmbH
BP analytical balance 221S	0.0001 g – 220 g	Sartorius AG
Burner, teclu		JUCHHEIM Laborgeräte GmbH
Diaphragm pump	Vacuum pump MZ 2 NT	VACUUBRAND GMBH + CO KG
Dounce Tissue Grinders, Wheaton	Loose and tight pestles, 40 mL	DWK Life Sciences
Dragonfly crystal	Pipetting robot	SPT Labtech Ltd.
Drigalski spatulas		Thermo Fisher Scientific
EpTIPS Biopur racks	For tips of 10 µL, 200 µL and 1000 µL volume	Thermo Fisher Scientific
Eppendorf Research plus	0.1 - 2.5 µL, 2 - 20 µL, 20 - 200 µL, 100 - 1000 µL	Eppendorf AG
Erlenmeyer flasks	Narrow neck, baffled, 2L	SciLabware Ltd
FiveEasy Standard pH Meter		Mettler Toledo
Fixed-Angle Aluminum Rotor	JA-25.50	Beckman Coulter
Fluidizer	Cell lysis	Made available by Rauh group
Freeze Dryer	Stainless Steel Alpha 1-2 LDplus	Martin Christ Gefriertrocknungsanlagen GmbH
Freezers	-20 °C	Liebherr-Hausgeräte GmbH
Fridges	4 °C	Liebherr-Hausgeräte GmbH
Gyro-rocker	SSL3	Cole-Parmer
Illuminated table Prolite basic 2		KAISER
Inoculating loop holder and loops	Brass and steel wire	Heinz Herenz Medizinalbedarf GmbH
J-LITE Fixed-Angle Aluminum JLA-8.1000 Rotor		Beckman Coulter
JANSi UVEX microscope	Brightfield and UV	SWISSCI
LaboGene ScanLaf Mars 1200	Class 2 bio-safety cabinet	LMS Consult GmbH & Co. KG
Liquid Dewar	LD25	Worthington Industries
Nananalysis 60 MHz benchtop NMR spectrometer	60PRO	Nanalysis Corp.
Natural Refrigerant –86 °C Ultra-Low Temperature Freezers	MDF-DU702VH-PA	PHC Corporation
New Brunswick Innova 44R	Stapled incubation shakers	Eppendorf AG
New Brunswick Innova 43	Incubation shakers	Eppendorf AG
Magnetic hotplate stirrers, VMS Advanced series	VMS – C7, VMS – C4	VWR International
Magnetic stirrer bars	Cylindrical	VWR International
Mastercycler X50a	PCR thermal cycler	Eppendorf
Microwave	MW 7873	Severin Elektrogeräte GmbH
Mini-PROTEAN Tetra Cell		Bio-Rad Laboratories

Mosquito crystal	Pipetting robot	SPT Labtech Ltd.
Optima XPN-80	Ultracentrifuge used at up to 130,000 g	Beckman Coulter
Precision balance LSP-1502	0.01 g – 1500 g	VWR International
Precision scale M-Prove AY412	0.01 g – 410 g	Sartorius AG
ROCK IMAGER 1000		FORMULATRIX
Sonorex Digiplus Ultrasounds bath	DL 1028 H	BANDELIN electronic GmbH & Co. KG
Spoon spatulas		Thermo Fisher Scientific
Swinging-Bucket Rotor	SW 32 Ti	Beckman Coulter
Tecan Infinite M1000	Fluorescence reader	Tecan Trading AG, provided by Rauh group
Thermometer, easy read	-20 - 150 °C ± 1 °C	VWR International
Tube rotator		VWR International
UVP UvSolo touch	Gel documentation	Analytik Jena AG
UVP UvSolo touch accessories	UVP Visi-White Converter plate Emission filter 510 - 560 nm	Analytik Jena AG
Vortex 2		IKA-Werke GmbH & CO. KG

4.2 Biological and chemical substances

Table 10: Protein ligands and substrates.

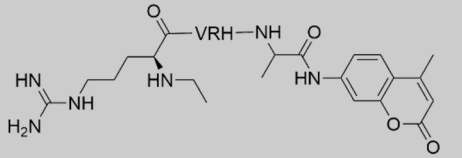
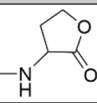
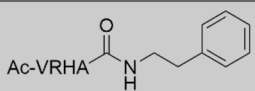
Ligand abbreviations/PDB ligand and IUPAC names (interaction partner)	MW / g/mol	Structures	Manufacturers or providers
M25 <i>N</i> -{2-[4-(aminosulfonyl)phenyl]ethyl}acetamide (binds to hCAII)	242		Merck KGaA
SBR <i>(R)</i> - <i>N</i> -(3-indol-1-yl-2-methyl-propyl)-4-sulfamoyl-benzamide (binds to hCAII)	372		Synthesized by Klebe group
Ksp63 Ac-RVRHA-4MC ^[169] (binds to GlpG)	837		Synthesized by Strisovsky group
Ksp152a (binds to GlpG)	3327	KIQHRVRHAALLPLAVLAGLTVLVRRTK-NH- 	Synthesized by Strisovsky group
STS1775 Ac-VRHA-CONH-phenylethyl (binds to GlpG)	655	Ac-VRHA- 	Synthesized by Strisovsky group

Table 11: Isotopically labelled chemicals.

Chemicals	Specifications	Manufacturers
L-Alanine	3- ¹³ C, 99%; 2-[D], 96% (solution NMR)	Cambridge Isotope Laboratories
¹⁵ N-Ammonium chloride	98%	Cambridge Isotope Laboratories
¹⁵ N-BioExpress	98%	Cambridge Isotope Laboratories
¹³ C-, ¹⁵ N-BioExpress	98%	Cambridge Isotope Laboratories
D-, ¹³ C-, ¹⁵ N-BioExpress		Cambridge Isotope Laboratories
D ₂ O	99.8%	Merck KGaA
d ₆ -Dimethyl sulfoxide	99.5%	Merck KGaA
d ₃₈ -Dodecylphosphocholine	98%	Cambridge Isotope Laboratories
Ethyl-2-hydroxy-2-methyl-3-oxobutanoate	4- ¹³ C, 99%; 4- ¹³ C, 2-methyl-[D] ₃ , 98%, pro-(R) precursor (solution NMR)	Cambridge Isotope Laboratories
D-, ¹³ C-labelled D-Glucose	≥97%	Cambridge Isotope Laboratories
¹³ C-ISOGRO powder	≥98%	Merck KGaA
¹³ C- ¹⁵ N-ISOGRO powder	≥98%	Merck KGaA
D-, ¹³ C- ¹⁵ N-ISOGRO powder	≥98%	Merck KGaA
α-ketobutyrate, sodium salt	3- ¹³ C, 99%; 3,3,4,4-[D] ₄ , 98% (solid-state NMR)	Cambridge Isotope Laboratories
α-ketobutyrate, sodium salt	4- ¹³ C, 99%; 4- ¹³ C, 3,3-[D] ₂ , 98% (solution NMR)	Cambridge Isotope Laboratories
α-ketoisovalerate, sodium salt	dimethyl- ¹³ C ₂ , 98%; 3-methyl-[D] ₂ , 4,4-[D] ₂ , 98% (solid-state NMR)	Cambridge Isotope Laboratories
d ₄ -methanol	99%	Merck KGaA
d ₅ -Tris(hydroxymethyl)amino-methane (TRIS)	98%	Merck KGaA

Table 12: Detergents and lipids, see also Table 5.

Detergents and lipids	Specifications	Critical Micelle Concentrations / % (v/v)	MW of micelles / kDa	Manufacturers
<i>n</i> -Dodecylphosphocholine (FC-12)	Sol-grade, F308S	0.047	~18.98	Anatrace
Full lipid extract from <i>E. coli</i>				Anatrace
<i>n</i> -Decyl-β-D-maltopyranoside (DM)	Sol-grade, D322S	0.087	33.30	Anatrace
<i>n</i> -Dodecyl-β-D-maltopyranoside (DDM)	Sol-grade, D310S	0.0087	39.83 – 76.08	Anatrace
<i>n</i> -Dodecyl- <i>N,N</i> -Dimethylamine- <i>N</i> -Oxide (LDAO)	Sol-Grade, D360S	0.023	~17.43	Anatrace
2,2-Didecylpropane-1,3-bis-β-maltopyranoside (LMNG)	NG310	0.001	Large	Anatrace
<i>n</i> -Nonyl-β-D-maltopyranoside (NG)	Anagrade, N330	0.28	~25.77	Anatrace
<i>n</i> -Octyl-β-D-maltopyranoside (OG)	Sol-grade, O310S	0.89	~21.36	Anatrace

All detergent stock solutions were prepared as 20% (w/v) solutions.

Table 13: Chemicals used for buffers and general applications.

Buffer components	Specifications	Manufacturers
Acetic acid		Carl Roth GmbH + Co. KG
Aceton		VWR International
Agar-agar		Carl Roth GmbH + Co. KG
Amersham ECL Rainbow Marker	Full range protein standard	Merck KGaA
Ammonium chloride		Carl Roth GmbH + Co. KG
Ampicillin sodium salt		Carl Roth GmbH + Co. KG
Basal Medium Eagle vitamin concentrate	100x	MP Biomedicals
Boric acid	≥99.8%	Carl Roth GmbH + Co. KG
Bradford Reagent	5x concentrate	SERVA Electrophoresis GmbH
Bromphenol blue sodium salt	for SDS-PAGEs	Carl Roth GmbH + Co. KG
Calcium chloride		Carl Roth GmbH + Co. KG
Carbenicillin disodium salt		Merck KGaA
Chloramphenicol		Carl Roth GmbH + Co. KG
Clontech Labs 3P TALON Metal Affinity Resin		Thermo Fisher Scientific
Cobalt(II) chloride hexahydrate	≥99 %	Carl Roth GmbH + Co. KG
cOmplete ULTRA Tablets	Mini, EDTA-free, EASYpack	Roche Deutschland Holding GmbH
Coomassie Brilliant Blue R250		Carl Roth GmbH + Co. KG
Copper(II) chloride dihydrate		Carl Roth GmbH + Co. KG
ddH ₂ O	MilliQ water	Produced in house
Dimethylsulfoxide (DMSO)		Carl Roth GmbH + Co. KG
1,4-Dithiothreitol	≥99%	MP Biomedicals
Ethanol	≥99.8%	VWR International
Ethanol absolute for HPLC	≥99.9%	Th. Geyer GmbH & Co. KG
6-Ethoxy-2-benzothiazolesulfonamide	>97%	Merck KGaA
Ethylendiamintetraacetic acid (EDTA)	≥99%	Carl Roth GmbH + Co. KG
GelGreen Nucleic Acid Gel Stain	10,000x	Biotium
Gel Filtration Calibration Kit	Low molecular weight range of 6.5 to 75 kDa	Cytiva
D-Glucose		Carl Roth GmbH + Co. KG
L-Glutathione	reduced	VWR International
Glycerol	≥99.5%, anhydrous	Carl Roth GmbH + Co. KG
Glycine	≥99%	Carl Roth GmbH + Co. KG
HisPur Cobalt Resin	His-tag protein purification (equivalent to Talon beads)	Thermo Fisher Scientific
Hydrogen chloride		Carl Roth GmbH + Co. KG
N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES)	PUFFERAN ≥99.5%	Carl Roth GmbH + Co. KG
Imidazole		Carl Roth GmbH + Co. KG
Iron(III) chloride hexahydrate		Carl Roth GmbH + Co. KG
Isopropanol	≥99.8%	Carl Roth GmbH + Co. KG
Isopropyl β-D-1-thiogalactopyranoside	≥99%	Carl Roth GmbH + Co. KG
Kanamycin		Carl Roth GmbH + Co. KG
L-Leucine		Merck KGaA
Magnesium sulphate	≥99%	Carl Roth GmbH + Co. KG
Manganese(II) chloride tetrahydrate	≥99%	Carl Roth GmbH + Co. KG
2-Mercaptoethanol	≥98%	Alfa Aesar (Thermo Fisher Scientific Inc.)
Ni-NTA agarose beads	His-tag protein purification	Thermo Fisher Scientific

Pepton from Casein		Carl Roth GmbH + Co. KG
Phenylmethylsulfonyl fluoride (PMSF)	≥99%	Carl Roth GmbH + Co. KG
Pierce Glutathione Agarose	GST-tag protein purification	Thermo Fisher Scientific
Potassium chloride	≥99.5%	Carl Roth GmbH + Co. KG
Potassium dihydrogen phosphate	≥99%	Carl Roth GmbH + Co. KG
di-Potassium hydrogen phosphate	≥99%	Carl Roth GmbH + Co. KG
L(+)-Rhamnose monohydrate	99%	MP Biomedicals
Sample buffer, Laemmli	2x concentrate	Merck KGaA
SOC medium	1x	New England Biolabs
Sodium chloride		Carl Roth GmbH + Co. KG
Sodium dihydrogen phosphate	≥96%	Thermo Fisher Scientific
di-Sodium hydrogen phosphate	≥99%	Carl Roth GmbH + Co. KG
Sodium hydroxide	≥99%	Carl Roth GmbH + Co. KG
Thiamin hydrochloride	≥98.5%	Carl Roth GmbH + Co. KG
3-(trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS)		Merck KGaA
Tris(hydroxymethyl)aminomethane (TRIS)		Carl Roth GmbH + Co. KG
TRIS-HCl		Carl Roth GmbH + Co. KG
Trypton from Casein		Carl Roth GmbH + Co. KG
peqGOLD DNA Ladder Mix	10,000-100 bp	VWR International
Yeast extract		Carl Roth GmbH + Co. KG
Zinc(II) chloride	≥97%	Carl Roth GmbH + Co. KG

Instruments and chemicals for crystallization were kindly provided by the crystallization department of the MPI for molecular physiology in Dortmund, Germany, run by Dr. Raphael Gasper-Schoenenbruecher.

4.3 Consumables

Table 14: Consumable materials.

Consumable materials	Specifications	Manufacturers
Amicon Ultra-15 Centrifugal Filter Units	MW limit 10 kDa and 50 kDa	Merck KGaA
Amicon Ultra-4 Centrifugal Filter Units	MW limit 10 kDa and 50 kDa	Merck KGaA
Amicon Ultra-2 Centrifugal Filter Units	MW limit 10 kDa and 50 kDa	Merck KGaA
Amicon Ultra-0.5 Centrifugal Filter Units	MW limit 10 kDa and 50 kDa	Merck KGaA
Carbonic Anhydrase Activity Assay Kit		PromoCell GmbH
Cellulose Acetate Membrane Filters	Type 11107	Sartorius AG
Cryogenic vials Cryo.s Star foot	Outer thread, 2 mL	Carl Roth GmbH + Co. KG
Disposable PD 10 Desalting Columns	Cross-linked dextra, 85-260 µm size of wet particles	Merck KGaA
EpTIPS Racks Biopur pipette tips	10 µL, 200 µL, and 1000 µL volume	Thermo Fisher Scientific
EpTIPS Biopur racks	For tips of 10 µL, 200 µL, and 1000 µL volume	Thermo Fisher Scientific
Falcon conical centrifugal tubes	50 mL, 15 mL	Thermo Fisher Scientific

Greiner 96-well plate	V-shaped, transparent	Greiner, donated by Rauh group
Greiner 96-well plate	Flat bottom black imaging plate	Greiner, donated by Rauh group
Open-Top Thickwall Polycarbonate Tube	31 mL for ultracentrifuge	Beckman Coulter
Polypropylene Bottles	1 L, for JLA-8.1000 rotor	Beckman Coulter
Open-Top Thinwall Ultra-Clear Tube	31 mL volume	Beckman Coulter
MemGold MD1-39		Biozol Diagnostics Vertrieb GmbH
MemGold2 MD1-63		Qiagen
Mini-PROTEAN TGX Precast Protein Gels	12 % gel, 15 wells	Bio-Rad Laboratories
PD-10 desalting column		Cytiva
Pierce 660 nm protein assay	ionic detergent compatibility reagent	Thermo Fisher Scientific
Plasmid Miniprep Kit I (S-Line)	peqGOLD	VWR International
QuikChange (Multi) Site-Directed Mutagenesis Kit		Agilent Technologies
Seed Bead Kit	HR2-320	Hampton Research
SnakeSkin Dialysis Clips	Polyamide derivative, 10 mm x 16 mm	Thermo Fisher Scientific
Spectra/Por Dialysis membrane	Pre-wetted tubing, MWCO 10 kDa	Thermo Fisher Scientific
Vivaspin 15 Centrifugal Concentrator	15 mL starting volume, 10,000 kDa MWCO, Polyethersulfone membrane	Vivaproducts
Vivaspin 4 Centrifugal Concentrator	4 mL starting volume, 10,000 kDa MWCO, Polyethersulfone membrane	Vivaproducts

4.4 DNA and proteins

All plasmids shown below were sequenced at least once (chapter 4.8.3) and showed full alignment with the expected DNA sequences.

Table 15: Plasmid names, characteristics, and origins.

Plasmid names	Sizes / bp	Antibiotic resistances	Purification tags	Cleavage sites	Origins
pGEX-4T1-hCAII-GST	5743	Ampicillin	GST	Thrombin	Klebe group
pET25_GlpG_wt	6561	Ampicillin	6xHis	TEV protease	Strisovsky group
pET15b_GlpG_wt_core_Urban	6276	Ampicillin	6xHis	Thrombin	Strisovsky group
pET25_GlpG_S201T	6561	Ampicillin	6xHis	TEV protease	Strisovsky group
pET25_GlpG_S201T_L244A_I177A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_V204A_I223A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_L123A_I237A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_V99A_I151A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_V211A_I255A	6561	Ampicillin	6xHis	TEV protease	Part of this work

pET25_GlpG_S201T_L89A_V203A_I109A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_L200A_V267A_I230A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_L131A_V96A_I235A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_L143A_V165A_I180A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pET25_GlpG_S201T_L229A_V119A_I175A	6561	Ampicillin	6xHis	TEV protease	Part of this work
pPR14_pET25+M_GlpG_C104A+W196C+S201T	6296	Ampicillin	6xHis	TEV protease	Strisovsky group
pPR16_pET25+M_GlpG_C104A+S201T+M247C	6296	Ampicillin	6xHis	TEV protease	Strisovsky group
pWaldo-GlpG-GFPe-SH1446 plasmid	7488	Kanamycin	GFP, 8xHis	TEV protease	Strisovsky group

Table 16: DNA sequences of PCR primers and genes. The genes for hCAII and GlpG on the plasmids and the three nucleotides of mutated amino acid of the mutagenesis primers are highlighted in bold letters. The nucleotides deviating from the wild type are underlined. All displayed sequences start with the 5'- and finish with the 3'-end.

Plasmid or primer names	Plasmid and primer DNA sequences
Excerpt of the pGEX-4T1-hCAII-GST plasmid	ATGGCCATCATA CGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGT GCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTCGAGAATTGCA TATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAA TGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTC ATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCTGGATGCGTTCCTCAA AATTAGTTTGTTTAAAAAACGATTGAAGCTATCCACAAATTGATAAGTACTTGAAATCCAG CAAGTATATAGCATGGCCTTTCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCC AAAATCGGATCTGGTTCGCGTGGATCCCCGGAATTCATGTCCCATCACTGGGGGTACGGCA AACACAACGGACCTGAGCACTGGCATAAAGGACTTCCCCATTGCCAAGGGAGAGCGCCAGTC CCCTGTTGACATCGACTCATAACAGCCAAGTATGACCCTTCCCTGAAGCCCCTGTCTGTTTCC TATGATCAAGCAACTTCCCTGAGGATCCTCAACAATGGTCATGCTTCAACGTGGAGTTTGTAT GACTCTCAGGACAAAGCAGTGCTCAAGGGAGGACCCCTGGATGGCACTTACAGATTGATTC AGTTTCACTTTCCTGTTTCACTTGGTTCCTGATGGACAAGGTTGAGAGCATACTGTGGATAAAAAG AAATATGCTGCAGAACTTCACTTGGTTCCTGGAACACCAAATATGGGGATTTGGGAAAGC TGTGCAGCAACCTGATGGACTGGCCGTTCTAGGTATTTTTTGAAGGTTGGCAGCGCTAAAC CGGGCCTTCAGAAAGTTGTTGATGTGCTGGATTCCATTAACAAGGGCAAGAGTGCTGA CTTCACTAACTTCGATCCTCGTGGCCTCCTCCTGAATCCTTGGATTACTGGACCTACCCAGGC TCACTGACCACCCTCCTCTTCTGGAATGTGTGACCTGGATTGTGCTCAAGGAACCCATCAGC GTCAGCAGCGAGCAGGTGTTGAAATTCGTAACCTTAACTTCAATGGGGAGGGTGAACCCG AAGAACTGATGGTGGACAACCTGGCGCCAGCTCAGCCACTGAAGAACAGGCAAATCAAAGC TTCTTCAAATAA
pGEX-3 primer for hCAII sequencing	GGAGCTGCATGTGTCAGAGG
pGEX-5 primer for hCAII sequencing	CTGGCAAGCCACGTTTGG
Excerpt of the pET25_GlpG_wt plasmid	ATGTTGATGATTACCTCTTTTGCTAACCCCGCGTGGCGCAGGCGTTTGTGATTACATGGCG ACGAGGGTGTATCCTCAGATTCAACAACATAACCAAAGCGATGTCTGGCTGGCGGATGA GTCCAGGCCGAGCGGTACGGGCGGAGCTGGCGCGTTTTCTCGAAAACCCGGCAGATCCG CGTTATCTGGCGGCGAGCTGGCAGGCAGGCCATACCGGCAGTGGCCTGCATTATCGCGTTA TCCTTTCTTGGCCCTTGGTGAACGCGCAGGTCCGGTAACCTGGGTGATGATGATCGCCTG CGTGGTGGTGTATTGCCATGCAAATCTCGGCGATCAGGAAGTGTGTTATGGCTGGCCT GGCATTTCGATCCAACACTGAAATTTGAGTTCTGGCGTACTTCACCCACGCGTTAATGCACT

	TCTCGCTGATGCATATCCTCTTTAACCTGCTCTGGTGGTGGTATCTCGGCGGTGCGGTGGAAA AACGCCTCGGTAGCGGTAAGCTAATTGTCATTACGCTTATCAGCGCCCTGTTAAGCGGCTAT GTGCGCAAAAATTCAGCGGGCCGTGGTTTGGCGGGCTTTCTGGCGTGGTGTATGCGCTGAT GGGCTACGTCTGGCTACGTGGCGAACCGGATCCGCAAAGTGGCATTACCTGCAACGTGGG TTAATTATCTTTGCGCTGATCTGGATTGTCGCCGGATGGTTTGATTGTTGGGATGTCGATG GCGAACGGAGCACACATCGCCGGTTAGCCGTGGGTTAGCGATGGCTTTTGTGATTGCGT CAATGCGCGAAAAACGAAAAGCTAGCCTCGAGAGAGAAAACTTGTATTCCAGGGCAGCAGCC ACCACCACCACCACCACTGA
T7 primer for GlpG sequencing	TAATACGACTCACTATAGGG
T7term primer for GlpG sequencing	CTAGTTATTGCTCAGCGGT
Reverse primers for L244A and I177A	GTCGCCGGATGGTTTGGTATGCGTGGGATGTCGATGGC TAGCGGTAAGCTAATTGCTACGCTTATCAGCGCCCTG
Reverse primers for V204A and I223A	GCTTACTGGCGTGGCGTATGCGCTGATGG CGATCCGCAAAGTGCCGCTTACCTGCAACGTGGG
Reverse primers for L123A and I237A	GATCAGGAAGTGATGTTATGGCGGCCTGGCCATTCG TCTTTGCGCTGATCTGGGCTGTCGCCGGATGGTTTG
Reverse primers for V99A and I151A	GTCCGGTAACCTGGCGGATGATGATCGCCTG ATGCACTTCTCGCTGATGCATGCCCTCTTTAACCTGCTCTG
Reverse primers for V211A and I255A	CTGATGGGCTACGCCCTGGCTACGTGGC CGAACGGAGCACACGCCGCCGGGTTAGCCG
Reverse primers for L89A, V203A, and I109A	CCTTCTTTGCCGCCGCGCGTGAACGCGCAGG GGGCTTACTGGCGCGGTGTATGCGCTG GATGATCGCCTGCGTGGTGGTGTGTTGCTGCCATGCAAATT
Reverse primers for L200A, V267A, and I230A	CGTGGTTTGGCGGGGCTACTGGCGTGGTGT TTTAGCGATGGCTTTTGGTATTGCTCAATGCGC TTTACCTGCAACGTGGGTTAGCTATCTTTGCGCTGATCTGG
Reverse primers for L131A, V96A, and I235A	CTGGCCATTCGATCCAACAGCGAAATTTGAGTTCTGGCGT CGCGCAGGTCCGGCAACCTGGGTGATG GTTAATTATCTTTGCGCTGGCCTGGATTGTCGCCGGATGG
Reverse primers for L143A, V165A, and I180A	CGTTACTTACCCACGCGCAATGCACTTCTCGCTGAT TCGGCGGTGCGCGGAAAAACGCCTC GCGGTAAGCTAATTGTCATTACGCTTGGCAGCGCCCTGTAA
Reverse primers for L229A, V119A, and I175A	GCATTTACCTGCAACGTGGGGCAATTATCTTTGCGCTGATCT TCGGCGATCAGGAAGCGATGTTATGGCTGGC CGCCTCGGTAGCGGTAAGCTAGCTGTCATTACGCTTATCA

Table 17: Molecular weights and extinction coefficients of the expressed and purified proteins.

Proteins	MW / kDa	Extinction coefficients for 0.1% (=mg/mL)
hCAII with affinity tag	55.91	1.673
hCAII, affinity tag cleaved off	29.76	1.694
Full-length GlpG with affinity tag	33.71	2.607
Full-length GlpG, affinity tag cleaved off	32.66	2.691
GlpG core with affinity tag	23.43	2.964
GlpG core after treatment with chymotrypsin	22.80	3.046
GlpG core, affinity tag cleaved off	21.63	3.211

Table 18: Purchased proteins for DNA or protein cleavage and for assays.

Proteins	Specifications	Manufacturers
Albumin fraction V	≥98%, lyophilized	Carl Roth GmbH + Co. KG
Benzonase Nuclease	>99%, solution	Merck KGaA
α-Chymotrypsin	Lyophilized, from bovine pancreas, ≥40 units/mg powder	Merck KGaA
DNase I	From bovine pancreas, grade II	Roche Diagnostics GmbH
TEV protease	10,000 units/mL	New England Biolabs
TEV protease with His-tag	MBP and His-tag	Kindly provided by the Protein Chemistry Facility at MPI Dortmund
Thrombin from bovine plasma	Lyophilized powder	Merck KGaA
Trypsin	-	Kindly provided by Rauh group

4.5 Cell lines

Table 19: Detailed information on the used competent cell lines.

Cell lines	Specifications	Original cell line	Genotypes	Manufacturers
5-alpha cells	High efficiency	<i>E. coli</i>	<i>E. coli fhuA2 Δ(argF⁻ lacZ)U169 phoA glnV44 φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	New England Biolabs
BL21-CodonPlus(DE3)-RIL cells	Chloramphenicol resistance	<i>E. coli</i>	<i>E. coli B F⁻ ompT hsdS(r_B⁻ m_B⁻) dcm⁺ Tet^R gal endA Hte [argU ileY leuW Cam^r]</i>	Agilent Technologies
BL21(DE3) cells	Tetracycline resistance	<i>E. coli</i>	<i>B F⁻ dcm ompT hsdS(r_B⁻ m_B⁻) gal l(DE3)</i>	Agilent Technologies
OverExpress C41(DE3) cells	-	<i>E. coli</i>	<i>F⁻ ompT hsdSB (r_B⁻ m_B⁻) gal dcm (DE3)</i>	Kindly donated by Musacchio group
C43 cells	-	<i>E. coli</i>	<i>F⁻ ompT hsdSB (r_B⁻ m_B⁻) gal dcm (DE3)</i>	Kindly donated by Musacchio group
Rosetta 2 cells	Chloramphenicol resistance	<i>E. coli</i>	<i>F⁻ ompT hsdSB(r_B⁻ m_B⁻) gal dcm (DE3) pRARE2 (Cam^R)</i>	Merck KGaA
Lemo21 (DE3) cells	Chloramphenicol resistance	<i>E. coli</i>	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo(Cam^R) λ DE3 = λ sBamHlo ΔEcoRI-B int:::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo= pACYC184-PrhaBAD⁻lysY</i>	New England Biolabs
LOBSTR cells	Chloramphenicol resistance	<i>E. coli</i>	Same as for BL21(DE3) plus modified arnA and slyD copies with reduced affinities to Ni and Co resins	Kerafast
XL-10 Gold cells	Kanamycin resistance	<i>E. coli</i>	<i>Tetr Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacI^QΔM15 Tn10 (Tet^R) Tn5 (Kan^R) Amy]</i>	Agilent Technologies

4.6 Buffer and media composition and preparation

In this chapter, the composition and preparation of all media and buffers for protein expression, purification, crystallisation etc. are displayed.

Table 20: Composition of the used buffers and media. All were prepared in ddH₂O or D₂O. All media components were sterilized via filtration or autoclaving.

Buffers/media	Components	pH
Coomassie Blue stain	25% isopropanol, 10% acetic acid, 0.5 g/L Coomassie Brilliant Blue R250	-
Cryo buffer	75 mM (d ₅ -)TRIS buffer, 1.5-1.45 mM (NH ₄) ₂ SO ₄ , 10% glycerol	8.0-7.8
Detergent-free Laemmli buffer	278 mM TRIS-HCl, 100 mM DTT, 44.4 % glycerin, 0.02 % bromphenol blue	6.8
GlpG buffer A	25 mM phosphate buffer, 300 mM NaCl	7.4
GlpG buffer B	25 mM phosphate buffer, 150 mM NaCl, 1.25% DM	7.4
GlpG buffer C	25 mM phosphate buffer, 150 mM NaCl, 0.2% DM	7.4
GlpG buffer W1	25 mM phosphate buffer, 500 mM NaCl, 10 mM imidazole, 1 mM CaCl ₂ , 1 mM MgCl ₂ , 0.2% DM	7.4
GlpG buffer W2	25 mM phosphate buffer, 150 mM NaCl, 10 mM imidazole, 0.2% DM	7.4
GlpG buffer E	25 mM phosphate buffer, 150 mM NaCl, 0.2% DM, 250 mM imidazole	7.4
GlpG SEC buffer	25 mM phosphate buffer, 150 mM NaCl, 0.2% of FC-12 or 0.11% d ₃₈ -FC-12	7.4-8.0
GlpG reaction buffer	25 mM TRIS, 150 mM NaCl, 0.2% DM (or other detergents)	8.0
hCAII buffer E	50 mM TRIS, 10 mM glutathione, 1 mM DTT	8.0
hCAII SEC buffer	50 mM ammonium bicarbonate	7.8
PBS buffer	10.0 mM Na ₂ HPO ₄ , 1.8 mM KH ₂ PO ₄ , 139.3 mM NaCl, 2.7 mM KCl	7.5
10x Running buffer (SDS-PAGE)	250 mM TRIS, 1.92 M glycine, 1% SDS	-
Laemmli buffer	278 mM TRIS-HCl, 4.4% LDS, 44.4% glycerin, 0.02%, bromphenol blue	6.8
LB medium	5 g/L yeast extract, 10 mg/L tryptone, 0.5 g/L NaCl	7.0
M9 medium	1x salt mix, 1x trace elements, 2 g/L glucose, 0.5 g/L NH ₄ Cl, 1 mM MgSO ₄ , 0.3 mM CaCl ₂ , 1 µg/mL thiamine, 1 µg/mL biotin, 100 µg/mL ampicillin/carbenicillin	7.5
M9* medium	1x salt mix, 2 g/L glucose, 0.5 g/L NH ₄ Cl, 1 mM MgSO ₄ , 0.3 mM CaCl ₂ , 1 µg/mL thiamine, 1 µg/mL biotin, 60 mM ZnCl ₂ , antibiotics (100 µg/mL ampicillin/carbenicillin or 100 µg/mL kanamycin or 30 µg/mL chloramphenicol)	7.5
Salt mix, 10x	15 g/L Na ₂ HPO ₄ , 7.5 g/L KH ₂ PO ₄ , 1.25 g/L NaCl	7.5
LB agar plates	37 g/L LB agar with antibiotics (100 µg/mL ampicillin/carbenicillin or 100 µg/mL kanamycin or 30 µg/mL chloramphenicol)	
TB medium	12 g/L tryptone, 24 g/L yeast extract, 0.5% glycerol, 2.3 g/L KH ₂ PO ₄ , 12.5 g/L K ₂ HPO ₄	7.5
Trace elements, 100x	5 g/L EDTA, 0.83 g/L FeCl ₃ · 6H ₂ O, 84 mg/L ZnCl ₂ , 13 mg/L CuCl ₂ · 2H ₂ O, 10 mg/L CoCl ₂ · 6H ₂ O, 10 mg/L H ₃ BO ₃ , 1.6 mg/L MnCl ₂ · 6H ₂ O	7.5

All solutions were either autoclaved at ≥121 °C and 1.4 bar for ≥30 min or sterilized by filtration using 0.22 µm filters. Antibiotics were only added to liquids which had cooled down to ≤60 °C. All solutions were stored at -20 °C, unless autoclaved. Agar plates were stored at 4 °C and used for up to 1.5 months. When preparing M9 media, the MgSO₄ solution was

always added before the CaCl₂ and ZnCl₂ solutions. In between these steps, the solution was mixed thoroughly to minimize salt precipitation.

4.7 Software

Table 21: Software.

Software	Applications
Aria ^[101]	Solid-state NMR structure calculations
CCPN2 ^[174]	Analyzing NMR spectra
CCPN3 ^[175]	Analyzing NMR spectra
Coot ^[99]	X-ray structure model building
MethylFLYA ^[8]	Methyl assignments
NESSY ^[176]	Analyzing NMR Relaxation Dispersion data
Phaser ^[177]	X-ray structure data processing
Phenix ^[97]	X-ray structure calculations
R: A Language and Environment for Statistical Computing ^[172]	Statistical analysis
ROCK MAKER (FORMULATRIX)	Protein crystallization
Topspin 3.5 and 4.1.3 (Bruker)	Acquiring, processing, and analyzing NMR spectra
UCSF Chimera ^[178]	Visualizing protein structures
UNICORN 7.3 (Cytiva)	Operation of ÄKTA pure chromatography system
XDS ^[179]	X-ray structure data processing

4.8 Molecular biology and cell biology methods

4.8.1 Estimation of DNA concentrations

To estimate DNA concentrations, the BioDrop Duo instrument was used. The sample port for micro-volume measurements was washed regularly with ddH₂O and a 50% isopropanol solution. First, a reference measurement was taken using 2 μL of ddH₂O. Then, a 2 μL drop of DNA solution was applied and the absorption at 260 nm wavelength was measured.

4.8.2 Plasmid DNA amplification and isolation

The plasmid DNA was prepared according to the high copy-number plasmid isolation protocols described in the Plasmid Miniprep Kit I (S-Line) by Peqlab (VWR). The DNA was eluted adding 5-10 μL ddH₂O and centrifuging the assemblies at 5,000 g for 1 min. To maximize the yield, the elution step was repeated. The final DNA concentrations were measured using the BioDrop Duo (chapter 4.8.1). The plasmid DNA was stored at -20 °C.

4.8.3 DNA sequencing

DNA sequencing was performed by Eurofins Scientific SE via the SupremeRun using standard primers. For hCAII pGEX-3 and -5 primers and for GlpG T7 and T7term primers were utilized

(Table 16). Sequencing was done for every newly received plasmid (for example, from collaborators and colleagues) or after mutagenesis.

4.8.4 Single Site-directed mutagenesis

The following protocol corresponds to the instructions for the QuikChange Lightning Single Site-directed mutagenesis kit offered by Agilent Technologies for plasmids smaller than 5 kbp. QuikChange Lightning reaction buffer with 1x final concentration, 100 ng of plasmid DNA template, 125 ng of each the forward and backward primers for the desired mutation site, 1 μ L of deoxynucleotide triphosphates (dNTP) mix and 1.5 μ L of QuikSolution reagent were mixed for the Polymerase Chain Reaction (PCR). The sample volumes were adjusted to 50 μ L using ddH₂O. Each 1 μ L QuikChange Lightning enzyme was added right before starting the reaction. The PCR was conducted as follows: After 2 min incubation time at 95 °C, a three-step cycle followed, which was repeated 18 times. During each cycle, the samples were heated to 95 °C to separate the DNA double strands for 20 s, then cooled down to 60 °C to allow the primers to anneal for 10 s, and finally the elongation reaction followed at 68 °C for 30 s per kb of plasmid length (here: 192 s). After the cycles were completed, another elongation period followed at 68 °C for 5 min. The samples were subsequently stored at 4 °C. The PCR products were treated with 2 μ L of the restriction enzyme Dpn I at 37 °C to digest the plasmid templates for at least 10 min. Subsequently, the treated plasmid DNA was transformed into XL-10 Gold cells (chapter 4.8.7). After the transformation, the cells were incubated on agar plates with 100 μ g/mL kanamycin at 37 °C overnight. The cells ingested the plasmids, repaired the plasmid nicks left after the PCR, and multiplied the copies of plasmids. On the next day, single colonies of the plates were picked and grown separately in LB medium containing 100 μ g/mL carbenicillin at 37 °C while shaking at 180 rpm overnight. On the next day, the plasmid DNA was extracted using the peqGold Plasmid MiniPrep Kit I (chapter 4.8.2) and sent for sequencing (chapter 4.8.3).

4.8.5 Multi site-directed mutagenesis

The following protocol corresponds to the instructions for the QuikChange Lightning multi site-directed mutagenesis kit offered by Agilent Technologies. The PCR samples with <5 kbp templates were prepared with the following ingredients: QuikChange Lightning reaction buffer with 1x final concentration, 50 ng of plasmid DNA template, 100 ng of either the

forward or backward primers for, here, up to three mutation sites, 0.5 μL of dNTP mix and 0.75 μL of QuikSolution reagent. The final sample volume was 25 μL upon addition of ddH₂O and 1 μL QuikChange Lightning Multi enzyme blend. The PCR was started immediately afterwards. The first step of the reaction consisted of a 2 min incubation time at 95 °C followed by the actual PCR cycle which was executed 30 times. The cycle consisted of an incubation at 95 °C to separate the DNA double strands for 20 s, a primer annealing step at 55 °C for 30 s, and finally an elongation step at 65 °C for 30 s per kb of plasmid length (here: 192 s). Subsequently, an additional elongation period followed at 65 °C for 5 min. Then, the samples were stored at 4 °C. The plasmid templates were selectively digested with 1 μL of the restriction enzyme Dpn I at 37 °C for >10 min. Subsequently, the treated plasmid DNA was transformed into XL-10 Gold cells (chapter 4.8.7). After the transformation, the cell suspensions were spread on agar plates containing 100 mg/mL carbenicillin and incubated at 37 °C overnight. Here, the cells replicated the mutated plasmids. On the next day, several single colonies were picked from each plate and grown separately in LB medium with 100 mg/mL carbenicillin at 37 °C while shaking at 180 rpm overnight. The plasmid DNA was extracted from the overnight cultures with the peqGold Plasmid MiniPrep Kit I (chapter 4.8.2) and sent for sequencing (chapter 4.8.3).

4.8.6 Preparation of chemically competent cells

Chemically competent cells were purchased or prepared in house^[180].

To obtain chemically competent cells, they were grown in a small LB culture at 37 °C while shaking at 200 rpm overnight. On the next day, 100 mL of liquid LB medium were inoculated with cells from the overnight culture and grown to an optical density at 600 nm (OD_{600}) of 0.5 to 0.8 at 37 °C while shaking at 200 rpm. The cell suspension was subsequently centrifuged at 7000 g and 4 °C for 10 min. The cell pellet was washed with 20 mL of cold, sterilized 50 mM CaCl₂ solution, and centrifuged again. Next, the cell pellet was resuspended in 20 mL of cold 50 mM CaCl₂. An incubation step followed on ice for 30 min. The cell suspension was centrifuged under the same conditions as before and the pellet was washed with 5 mL cold 50 mM CaCl₂ solution. Another incubation step followed on ice for 15 min. Next, a cold, sterilized glycerol solution was added to the cell suspension to yield a 15% (v/v) glycerol suspension. Aliquots of 200 μL were transferred to reaction tubes and promptly frozen in liquid nitrogen. The competent cells were stored at -80 °C.

4.8.7 Transformation of competent 5-alpha and XL-10 cells

The competent cells were thawed on ice for 10 min. Each cell aliquot was treated with ~200 ng plasmid DNA and carefully mixed. The cell suspensions were incubated on ice for 30 min. Afterwards, the cells were incubated at 42 °C for 30 s and placed on ice for 5 min. After adding 1 mL of sterile SOC or LB medium without any antibiotics, the cells were incubated at 37 °C while shaking at 250 rpm for 60 min. A small aliquot was spread over agar plates with the appropriate antibiotics (Table 19) and kept at 37 °C overnight.

4.8.8 Transformation of competent BL21-Condon Plus(DE3)-RIL cells

The cells were thawed on ice for 20-30 min. Around 200 µg of plasmid DNA was added to each cell aliquot containing the gene of interest. The suspension was mixed and placed on ice for 30 min. Each tube with cells was incubated at 42 °C for 20 s and then kept on ice for 2 min. Around 1 mL of prewarmed, sterilized SOC or LB medium without antibiotics was added. The cells were grown at 37 °C while shaking at 200 rpm for 60 min. Afterwards, a small aliquot was spread over LB agar plates containing the required antibiotics, here 100 µg/mL ampicillin or carbenicillin and 30 µg/mL chloramphenicol. The plates were incubated at 37 °C overnight.

4.8.9 Transformation of competent Rosetta 2 cells

The cells were thawed on ice for 20-30 min. Then, around 200 µg DNA was added to each cell aliquot and mixed. The mixtures were incubated on ice for 5 min, then at 42 °C for 30 s, and again on ice for 2 min. Next, 1 mL prewarmed and sterilized LB or SOC medium (without antibiotics) was added. The cells grew at 37 °C while shaking at 200 rpm for 60 min. A small aliquot was spread on agar plates containing 100 µg/mL ampicillin or carbenicillin and 30 µg/mL chloramphenicol and incubated at 37 °C overnight.

4.8.10 Transformation of competent Lemo21(DE3) cells

The cells were thawed on ice for 10 min. Around 200 µg of plasmid DNA was added each and mixed with the cell suspension. The cells were incubated on ice for 30 min, then at 42 °C for 10 s, and again on ice for 5 min. Around 1 mL of prewarmed, sterilized SOC or LB medium was added to each tube and transferred to an incubation chamber which was operating at 37 °C while shaking at 200 rpm for 60 min. Afterwards, a small aliquot was spread over LB agar plates with 100 µg/mL kanamycin and 30 µg/mL chloramphenicol and incubated at 37 °C overnight.

4.8.11 Transformation of competent LOBSTR(DE3) cells

The cells were thawed on ice for 10 min. Around 200 µg of plasmid DNA was added and mixed with the cell suspension. The cells were incubated on ice for 30 min and then heat shocked at 42 °C for 20 s. The cells were incubated on ice for another 5 min. Around 1 mL of prewarmed, sterilized SOC or LB medium was added to each tube and transferred to an incubation chamber at 37 °C while shaking at 200 rpm for 60 min. Afterwards, a small aliquot was spread over LB agar plates with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol and incubated at 37 °C overnight.

4.9 Protein expression

Studying proteins requires considerable amounts of highly pure protein. The expression in bacterial cells dedicated to the straight-forward and large-scale production of proteins facilitates the process. When these proteins are not natively produced in the cells, it is called heterologous expression. For hCAII, the protein was derived from human cells and was subsequently heterologously expressed in *E. coli*. GlpG, on the other hand, originates from and was expressed in *E. coli*. To overexpress the protein, the introduced plasmid containing the corresponding gene of interest includes a promoter which allows the expression to be induced in the presence of IPTG. This way, the starting time and extent of the protein expression can be regulated.

4.9.1 Expression of hCAII or GlpG in LB and M9 or M9* medium

The BL21-Condon Plus (DE3)-RIL or Rosetta 2 cells freshly transformed with the pGEX-4T1_HCAII-GST plasmid and LOBSTR or Lemo21 cells freshly transformed with a pET25_GlpG or the pWaldo-GlpG-GFPe-SH1446 plasmid, were grown in LB medium with 100 µg/mL ampicillin or carbenicillin at 180 rpm and 37 °C overnight. A large LB, M9 (for GlpG), or M9* (for hCAII) culture with the same antibiotic was inoculated at $OD_{660\text{ nm}} = 0.05-0.1$ with the cell suspension grown overnight. In case the M9 or M9* medium contained isotopically labelled components, the cells were pelleted by centrifugation at 4000 g for 20 min, washed with M9 or M9* medium, centrifuged again, and afterwards added to the large culture. This cell culture was incubated at 37 °C and 180-220 rpm until the cells reached a density of $OD_{660\text{ nm}} = 0.4-0.6$. The protein expression was induced by adding 1 mM IPTG and 1 mM ZnCl₂ (hCAII) or 0.4 mM IPTG (GlpG). Subsequently, the culture was incubated at 30 °C (hCAII) or 18 °C (GlpG) while shaking at 180 rpm overnight. The cells were harvested by centrifugation

at 6,240 g and 4 °C for 25 min. The cell pellet was flash-frozen in liquid nitrogen for further storage at -80 °C.

4.9.2 Adapting and growing cells in D₂O

Fully deuterated protein was yielded by growing the freshly transformed cells in D₂O before starting the expression. As D₂O-based media stress the cells and may even arrest cell replication or lead to cell death, the cells were slowly adapted to the new conditions using various strategies. The first adaptation strategy, mainly used for the expression of hCAII, is schematically shown in Figure 51.

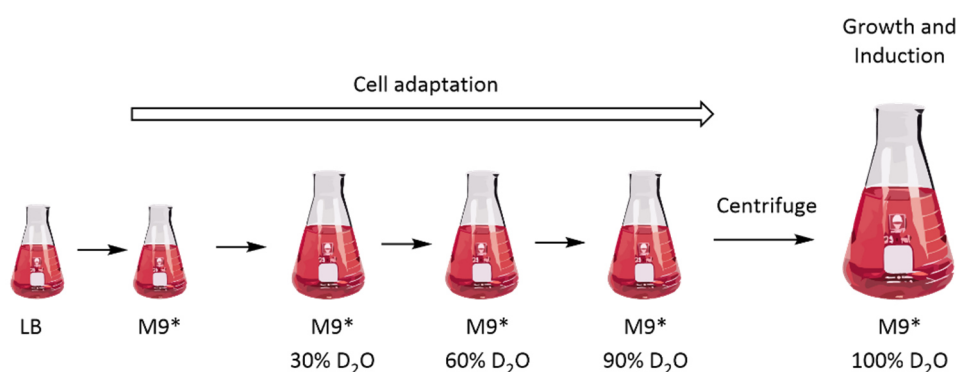


Figure 51: Adaptation of *E. coli* cells to grow in M9* medium based on 100% D₂O.

An overnight culture was prepared in LB medium at 37 °C and 180 rpm and used to inoculate M9 (for GlpG) or M9* medium (for hCAII) at $OD_{660\text{ nm}} = 0.05-0.1$. The same cell density was established upon inoculation during the following steps. When the cell density reached $OD_{660\text{ nm}} = 0.6$, the cells were transferred to 100mL of M9 or M9* medium with 30% D₂O either taking an aliquot or centrifuging the cells at 4000 g for 20 min and resuspending the cells in the new medium. The culture was incubated at 37 °C while shaking at 200 rpm until reaching $OD_{660\text{ nm}} = 0.4-0.5$. The same OD value needed to be reached during all subsequent amplification steps until the cells were transferred to a new medium. Next, 100 mL M9 or M9* medium with 60% D₂O were prepared and inoculated. At sufficient density, the cells were used to inoculate a 300 mL solution of M9 or M9* medium with 90% D₂O. The cells were grown overnight to proceed with the final medium on the next day. The the cells were centrifuged and resuspended as before to inoculate the final medium containing 100% D₂O. Subsequently, the expression proceeded as described in chapter 4.9.1.

The second strategy, called alternative protocol and mainly used for GlpG, started with the inoculation of an LB culture at $OD_{660\text{ nm}} = 0.05-0.1$ based on an overnight culture. The culture

was agitated at 37 °C and 220 rpm until $OD_{660\text{ nm}} = 0.7$ was reached. Next, the cells were centrifuged at 5000 g for 25 min. The pellet was washed with 1x M9 salts and centrifuged again. The resulting pellet was suspended in isotopically labelled M9 medium based on 100% D_2O with four times less volume compared to the previous LB culture. The culture was shaken at 220 rpm and 37 °C for in total 2.5-3 h. Subsequently, the expression was induced with 1 mM IPTG. Isotopically labelled ISOGRO or BioExpress supplements were added one hour before inducing the cells with IPTG. Afterwards, the incubation followed at 180 rpm and 18 °C overnight. On the next day, the cells were harvested via centrifugation at 5,000 g at 4 °C for 20 min. The supernatant was discarded, the cell pellet frozen with liquid nitrogen and stored at -80 °C.

4.9.3 ILV-methyl labelling

To specifically label methyl side chains of proteins, precursors were added to the D -, ^{13}C -, ^{15}N -labelled cell cultures at a specific timepoint before induction. Different precursors were used to prepare solution and solid-state NMR samples^[45d].

For hCAII, ILV-methyl labelling was performed non-stereospecifically yielding two, for leucine and valine, and one 1H -, ^{13}C -labelled methyl group for isoleucine. Here, α -ketobutyrate (4- ^{13}C , 3,3,4,4- $[D]_4$) and α -ketoisovalerate (dimethyl- ^{13}C , 3-methyl- $[D]_2$, 4,4- $[D]_2$) were used to label isoleucine and leucine/valine residues, respectively. The leucine/valine-precursor was added one hour before induction, while the isoleucine-precursor was added 20 min before induction with 1 mM IPTG. The leucine/valine-precursor required a concentration of 60 mg/L and the isoleucine-precursor concentration was 50 mg/L in the final culture.

For GlpG, stereospecific 1H -, ^{13}C -ILV-methyl labelling was performed with each one methyl group per leucine, valine, and isoleucine residue. For isoleucine labelling α -ketobutyrate (4- ^{13}C , 3,3- $[D]_2$) and for leucine/valine labelling ethyl-2-hydroxy-2-methyl-3-oxobutanoate (4- ^{13}C , 2-methyl- $[D]_3$, pro- (R)) was used. Alternatively, the methyl labelling kits TLAM-Id1LV^{proR} and DLAM-LV^{proR} were utilized for leucine/valine labelling, resulting in the same labelling scheme. The leucine/valine-precursor was added with 300 mg/L concentration one and a half hours after transferring the cells into the final M9 medium. After incubating the cells for 40 min in the shaker, the isoleucine-precursor was added at 50 mg/L concentration. Another 20 min later, the protein production was induced with 1 mM IPTG and it was continued with the expression as usual.

Table 22: Stereospecific methyl labelling kits.

Chemicals/kits	Ethyl-2-hydroxy-2-methyl-3-oxobutanoate (4- ¹³ C], 2-methyl-[D] ₃ , pro-(R)) and α-ketobutyrate (4- ¹³ C], 3,3-[D] ₂)	DLAM-LV ^{proR} and α-ketobutyrate (4- ¹³ C], 3,3-[D] ₂)	TLAM-Id1LV ^{proR}
Amounts / mg/L	300 mg/L and 50 mg/L	1 kit / L and 50 mg/L	1 kit / L
Manufacturers	NMR-BIO	NMR-BIO and Cambridge Isotopes	Cambridge Isotopes

Ethyl-2-hydroxy-2-methyl-3-oxobutanoate was hydrolyzed via addition of NaOH before usage according to the instructions included by the provider upon purchase (CDLM_10508_instructions_for_use). Each 0.25 g was diluted with 1 mL 0.06 M K₂HPO₄, mixed, and carefully transferred to a reaction tube. Proton NMR 1D spectra were measured before and during the reaction to trace the progress. While stirring the solution with a stirring bar, drops of 10 M NaOH were added in a stepwise fashion to hydrolyze the ester of the compound. To avoid undesired side reactions, pH 13 was never exceeded. When the pH value decreased below 9.5, more NaOH was added. Upon pH stabilization, full conversion was reached. After confirming via NMR that no educt was left, the pH of the solution was adjusted to a value of 7.4 using a 6 M HCl solution.

4.10 Biochemistry methods

4.10.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The protein samples were diluted in Laemmli buffer and cooked at 95 °C for 30 min. The electrophoresis gel system was assembled and filled with running buffer for SDS-PAGEs. As protein standard, 3 μL of Amersham ECL Rainbow Marker was applied in the first well of the Mini-Protean TGX gel. The cooked samples were centrifuged shortly at 10,000 g for 10 s and loaded into the remaining wells. The system was switched on and electrophoresis was performed at 200 V for 35 min. Afterwards, the gel was retrieved, washed thoroughly in ddH₂O, and covered with Coomassie Blue staining solution. The gel was heated in the microwave for 5 min and left shaking at 30 rpm and room temperature for 30 min. The gel was destained by washing the gel with ddH₂O. To speed up the destaining process, the gel was boiled in ddH₂O in the microwave for 5 min and left shaking at 30 rpm for more than ten minutes.

4.10.2 Cell lysis via high pressure homogenization

The cell pellets were thawed on ice, diluted at least 5x with ice-cold buffer (cooled PBS buffer with 1 mM DTT at pH 7.5 for hCAII and cooled GlpG buffer A for GlpG). Before cell lysis, 10 μ g DNase per mL cell suspension as well as protease inhibitors (1 mM PMSF and/or cOmplete ULTRA Tablets) were added. Subsequently, the suspensions were stirred for 30-60 min at 4 °C to further homogenize the suspension. Meanwhile, the Emulsiflex instrument was flushed twice with 50 mL ddH₂O and equilibrated with at least 100 mL buffer. An ice bucket was filled with ice and water to cool the tubing of the Emulsiflex. The samples were each filtered through a 1 mm pore sieve before lysis with the Emulsiflex. Then, the cell suspension was pumped through the Emulsiflex at a pressure of 10,000-15,000 psi. The lysate was collected and passed through the system for another two or three times. Air bubbles were avoided to enter the system when possible. A lower viscosity and turbidity of the suspension indicated that the lysis was successful. The Emulsiflex system was washed with buffer and subsequently ddH₂O. In the end, the system was equilibrated with 100% isopropanol for storage.

4.10.3 hCAII purification

The cell lysate was centrifuged at 6,240 g at 4 °C for 30 min. The resulting supernatant was loaded on a GST-tag binding affinity chromatography column. The chromatography was done using gravity flow and glutathione agarose beads with a binding capacity of 10 mg/mL. For a 1 L expression, 2 mL of glutathione agarose beads were used. The resin was prepared by washing it five times with two column volumes of ddH₂O and equilibrating it five times with two column volumes of PBS supplemented with 1 mM DTT at pH 7.5. The supernatant remaining after lysis and centrifugation was added to the beads and incubated while rotating slowly at 4 °C for one hour. Next, the column was washed again until the 280 nm absorption of the fractions reached a baseline. Only then, it was proceeded with the elution. The protein was eluted three times with one column volume of hCAII buffer E. The progress of the elution was tracked by measuring the 280 nm absorption of the fractions as before. The elution fractions with the highest UV absorbance at 280 nm were pooled and incubated with a few milligrams of lyophilized thrombin at room temperature overnight. The beads were washed with ddH₂O and recycled for the next use. To control the progress of the proteolysis, the sample contents were visualized via SDS-PAGEs. The protein samples were concentrated to >5 mL as soon as sufficient cleavage was obtained, filtered through 0.22 μ m pores, and

purified by SEC. Here, a HiLoad Superdex 16/600 75 µg column was used with a bed volume of 120 mL at a flow rate of 1 mL/min. The column was washed with at least two column volumes of degassed and filtered ddH₂O first and equilibrated with two column volumes of degassed and filtered hCAII SEC buffer. After preparing both sample and column, the sample was injected manually using a 5 mL loop. The fractions were collected, frozen with liquid nitrogen, and lyophilized with a Freeze Dryer. The sample was lyophilized until no visible liquid was left, which took roughly 2.5 h/mL. In case solvents with lower volatility than water were present, e.g. DMSO, the lyophilization was extended by at least half a day.

4.10.4 Full-length GlpG and GlpG core purification

The cell lysate was centrifuged at 10,000 g and 4 °C for 20 min. Subsequently, the supernatant was centrifuged again to isolate the cell membranes at 130,000 g and 4 °C for two hours. The resulting pellet was either frozen with liquid nitrogen and stored at -80 °C or directly resuspended in GlpG buffer A including protease inhibitors, for example half a cOmplete ULTRA tablet per litre of expression culture. A dounce homogenizer was used to homogenize the membrane buffer mix. Then, the Pierce 660 nm Protein Assay was utilized to measure the protein concentration in presence of cell membranes. The samples were diluted to yield a <3.5 mg/mL protein solution in GlpG buffer B. Next, the membrane proteins were solubilized in a beaker while stirring slowly at 4 °C for 1-3 hours. The remaining cell membrane patches were spun down at 100,000 g for 30 min. The solubilized GlpG in the resulting supernatant was subsequently separated from proteins without a His-tag using HisPur Cobalt beads for affinity chromatography. The resin was prepared by washing it twice with two column volumes of ddH₂O and twice with one column volume of GlpG buffer C. On average, 1 mL pure resin with a binding capacity of 5-15 mg/mL was used for a sample resulting from 1 L expression culture. The solubilized membranes were mixed with the resin and incubated in a beaker while stirring at 4 °C for one hour. Afterwards, the suspension was transferred to a small manual column and washed with two column volumes of GlpG buffer W1 and W2 each until the fractions reached a baseline at 280 nm UV absorption. Next, the protein was eluted with three times one column volume of GlpG buffer E. The elution fractions with significant protein content according to their absorption at 280 nm were pooled and desalted using PD-10 desalting columns, which were washed and equilibrated before use according to the protocol provided by the manufacturer.

Depending on the desired GlpG construct, chymotrypsin (pWaldo-GlpG-GFPe-SH1446 plasmid or pET25_GlpG_wt for GlpG core), TEV (pET25_GlpG_wt or pET25_GlpG_S201T for GlpG FL) or thrombin cleavage (pET15b_GlpG_wt_core_Urban) was performed. For cleavage with chymotrypsin, a 1:50 (w/w) chymotrypsin solution in water with 1 mM HCl and 2 mM CaCl₂ was prepared. To the protein solution, enough chymotrypsin solution was added to yield a 1:50 (w/w) ratio of target protein to chymotrypsin. After adding the chymotrypsin mixture, the sample was incubated at room temperature for 72 h. The reaction was stopped by adding 1 mM PMSF. TEV cleavage, on the other hand, proceeded by adding TEV in a 1:1 molar ratio to the target protein. The mixture was incubated at room temperature overnight. Thrombin cleavage was performed by adding a few mg of lyophilized thrombin and incubating the solution at room temperature overnight. The success of each proteolysis step was tracked by SDS-PAGE, if the size difference between educt and product was large enough.

Lastly, the resulting GlpG samples were concentrated using centrifugal concentrators with a molecular weight cut-off (MWCO) of 50.000, which have been rinsed with water before. In case the samples showed signs of precipitation, the sample was not concentrated further, if possible. Before the SEC, the protein sample was prepared by centrifugation at 10,000 g for 10 min. The sample was applied manually onto the column with a 500 µL loop. A designated Superdex 200 Increase 10/300 GL column with ~25 mL column volume was used for the SEC experiment. Before the run, the column was washed with at least two column volumes of degassed and filtered ddH₂O and equilibrated with two column volumes of degassed and filtered GlpG SEC buffer (with deuterated detergents, if required). The run was performed at 0.5 mL/min.

4.10.5 Estimation of protein concentrations

Whenever possible, each two of the following tools were used to confirm the measured protein concentration. Especially when working with detergents and lipids in addition to the protein, protein concentration measurements tend to be less accurate.

4.10.5.1 Estimation via BioDrop Duo

Protein concentrations were measured with the BioDrop Duo instrument based on the protein-specific extinction coefficients. The sample port for micro-volume measurements was washed with ddH₂O before use. First, 2 µL ddH₂O or buffer were used as reference. Next, a 2 µL drop of protein solution was applied and the absorption at 280 nm wavelength was

measured. To receive accurate measurements, the procedure was repeated. Afterwards, the sample port was washed again with ddH₂O and 50% isopropanol in ddH₂O.

4.10.5.2 Estimation via Bradford assay

The assay was performed using cuvettes and the BioDrop instrument at 595 nm absorption wavelength as described in the product manual by SERVA Electrophoresis GmbH. A standard curve was created, measuring and plotting five BSA samples of concentrations of 0.025 - 1 mg/mL. The sample with the unknown protein concentration was diluted 1:30 and 1:50 (v/v) in buffer. The background signal was accounted for using the dilution buffer only. The Bradford reagent was added to the BSA samples and the protein samples at 1:5 ratio. All solutions were mixed and incubated at room temperature for 5 min. Subsequently, the absorption was measured for each solution and the unknown protein concentrations were determined by extrapolating the data of the standard curve.

4.10.5.3 Estimation via Pierce 660 nm assay

This assay is especially well-suited for samples containing detergents and reducing agents. It was performed using cuvettes and the BioDrop instrument at 660 nm absorption wavelength as described by the manufacturer. A standard curve was prepared with five BSA samples of concentrations between 0.025 - 2 mg/mL. The samples with unknown protein concentration were diluted by 1:30 and 1:50 (v/v) in dilution buffer. The background signal was accounted for using the dilution buffer only. The Pierce reagent was added to all samples at 1:15 ratio. All solutions were mixed, protected from light, and incubated at room temperature for 5 min. Like for the Bradford assay, the absorption was measured, and the unknown protein concentrations were determined by extrapolation.

4.10.5.4 Estimation via Äkta purification system

The protein concentration was estimated via the Äkta purification software integrating the elution peaks of each protein fraction of interest. Using the protein-specific extinction coefficients, the protein concentration and total protein amount was calculated by the software UNICORN.

4.11 Activity assays

To ensure protein integrity and functionality after expression and purification, activity assays were used to characterize hCAII and GlpG samples.

4.11.1 Promo Cell activity assay for hCAII

The instructions for the Carbonic Anhydrase Activity Assay Kit attached by the manufacturer were followed to perform the assay. First, a nitrophenol standard curve was prepared with 0, 8, 16, 24, 32, and 40 nmol/well nitrophenol diluted in PBS. The solutions were mixed well and absorbance at 405 nm wavelength was measured. Reference samples (buffer only), positive controls (provided in the kit), negative controls (hCAII plus inhibitors), and protein samples diluted at 1:20 or 1:30 ratio were prepared in PBS buffer with final volumes of 95 μ L each. The samples were incubated at room temperature for 15 min. Then, 5 μ L carbonic anhydrase substrate was added each and mixed with the solutions. Again, the absorbance at 405 nm was measured in a kinetic mode for one hour at room temperature. The nitrophenol standard curve was used to estimate the specific carbonic anhydrase activity (a_{CA}), based on the calculated amount of released nitrophenol in nmol (B), the dilution factor (D), the reaction time in minutes (Δt), and the sample volume in microliters (V).

$$a_{CA} = B * \frac{D*1000}{\Delta t*V} \quad (11)$$

4.11.2 Activity assay for GlpG^[169]

First, a 20 mM stock of the substrate Ksp63 was diluted in GlpG reaction buffer (25 mM TRIS, 150 mM NaCl, 0.2% DM (or other detergents) at pH 8.0) to yield a final concentration of 200 μ M. Upon substrate cleavage, a previously quenched fluorophore is released which allows to track the cleavage reaction via an increase in fluorescence. The DMSO concentration was adjusted to 10%. Samples of GlpG with each 25 μ L of 800 nM concentration were prepared in GlpG reaction buffer. As negative control, 25 μ L GlpG reaction buffer was used. The solutions, the black imaging plate, and the Tecan instrument chamber were incubated at 37 $^{\circ}$ C before the measurement for at least 5 min. Then, each 25 μ L protein and substrate solution were mixed on the imaging plate and the measurements were started.

The fluorescence measurements for the assay were performed with a Tecan Infinite M1000 fluorescence reader with kinetic intervals of 30 s for 40-60 min at 37 $^{\circ}$ C. An excitation wavelength of 355 nm and emission wavelength of 450 nm at 20 nm bandwidth was applied.

Multiple reads per well were conducted with a gain of 70%. After 24 s of orbital shaking and 150 ms settle time, the samples were excited with 30 flashes at 100 Hz frequency in a 2x2 square format with 750 μm border. The fluorescence was averaged over multiple reads per well and plotted against time for each sample.

4.12 Crystallization and X-ray diffraction of protein crystals

Crystallization was used to solidify proteins to obtain their 3D structure in the crystal via X-ray diffraction or to study the protein structure and dynamics via solid-state NMR spectroscopy. Different strategies and methods were applied to crystallize the apo protein or protein complexes with ligands or other interaction partners.

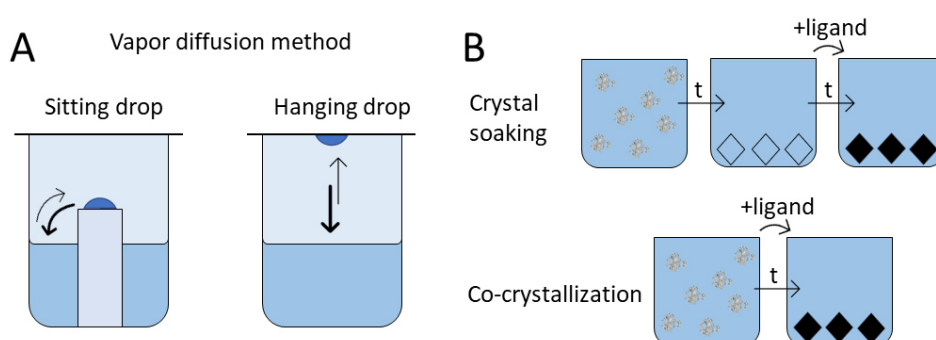


Figure 52: Different systems used for crystallization. **A.** Crystallization via vapor diffusion either in a sitting drop or hanging drop format. **B.** Strategies to obtain crystals of protein-ligand complexes via soaking or via co-crystallization over the time t .

4.12.1 Preparation of hCAII-SBR crystals

To crystallize the hCAII-SBR complex, different strategies were explored. Namely, co-crystallization and crystal soaking were performed using different isotopically labelled components in the sitting-drop apparatus for vapor diffusion. All protocols included protein solutions of 10-20 mg/mL hCAII in 50 mM (d_5 -)TRIS buffer at pH 8.0 which were mixed with precipitant solution in a 1:1 ratio. The precipitant solution contained 100 mM (d_5 -)TRIS buffer and 3.0-2.7 M $(\text{NH}_4)_2\text{SO}_4$ at pH 8.0-7.8. Each crystallization plate was incubated at 4 $^\circ\text{C}$ for at least one week.

Co-crystallization was performed by adding the ligand SBR to the hCAII solution before crystallization (Figure 52, panel B). A one-fold molar excess of the ligand SBR was incubated with hCAII on ice for at least one hour prior to crystallization. In case the crystals were soaked, the ligand SBR was added after crystallization (Figure 52, panel B). Here, the desired amount of ligand was added to the precipitant solution including the crystals and incubated for at

least 2 days at 4 °C. For both approaches, 10 or 100 mM stocks of SBR in (d_6 -)DMSO were used. The final DMSO concentration did not exceed 10%.

For 24-well plates, 5 μ L protein and precipitant solution were mixed in a 1:1 ratio while for 96-well plates, ≥ 0.1 μ L drops each were consolidated. The Dragonfly pipetting robot assisted in distributing the precipitant solutions of ~ 400 μ L volume for 24-well plates and ~ 40 μ L volume for 96-well plates. The protein solution was added manually for the 24-well plates or with the Mosquito pipetting robot for the 96-well plates. Protein and precipitant solutions were thoroughly homogenized. The crystallization plates were sealed with plastic foil to ensure an airtight system for each condition and incubated at 4 °C. The crystallization was manually monitored, taking images with Brightfield and UV microscopes for the 24-well plates. For 96-well plates, images were acquired with white and polarized light according to the Fibonacci Nightly imaging schedule by the ROCK IMAGER which also stored the plates.

4.12.2 X-ray diffraction and data processing for hCAII-SBR crystals

The crystals were picked with 0.2 mm loops and treated with cryo buffer, which consisted of precipitant buffer and 10% glycerol. The drops containing the crystals were flash-frozen in liquid nitrogen. The samples were sent to the Swiss Light Source synchrotron (SLS) in Villigen, Switzerland. The X-ray diffraction data at the SLS were collected on a PXII X10SA instrument with an EIGER2 XE 16M detector with 12.4 keV excitation voltage at 100 K with a wavelength of 1 Å. The data were processed using XDS^[179] and Phaser^[177]. The structures were subsequently refined and constructed using Phenix^[97] and Coot^[99].

4.12.3 Crystallization of GlpG core

Several crystallization conditions for GlpG core have been published. These conditions were revised and reproduced to yield GlpG core crystals. As no screening of small-scale hanging drop crystallization conditions was possible, the screening was performed with sitting drops. First, solutions of 6 mg/mL GlpG core in 25 mM TRIS buffer, 150 mM NaCl, and 0.5-1% NG at pH 8.0 were subjected to sitting-drop crystallization trials. Here, full-length GlpG wild type was either cleaved by chymotrypsin to yield the GlpG core or the GlpG core construct was expressed directly. Then, 96-well TTPIQ plates were prepared mixing each ≥ 0.1 μ L of protein solution with precipitant solutions of the Mem Gold 1 or 2 kits at 1:1, 1:2, and 2:1 ratios. The precipitant solution was distributed manually with each 40 μ L per well. The protein solution, on the other hand, was added and mixed with the precipitant solution using the Mosquito

pipetting robot. The plates were covered with plastic foil to prevent them from drying out and incubated at 20 °C. The crystallization success was monitored using RockMaker which enabled to visualize the crystallization solutions with both white and polarized light.

When crystals or crystal-like structures with UV absorbance were observed for one of the screened conditions, these conditions were reproduced in both sitting and hanging drop setup in 24-well plates mixing each 2 μL of protein solution at the previously used ratio with the same precipitant as before. Here, 400 μL precipitant solution in total were applied per well with the Dragonfly pipetting robot while the protein solution was added manually and mixed with precipitant. The plates were incubated at 20 or 22 °C. Crystals were identified with white and UV light using a JANSi UVEXm microscope.

4.13 NMR spectroscopy

All solution NMR experiments were recorded with an Ascend 800 MHz spectrometer connected to a Cryoprobe, if not indicated differently. All solid-state NMR experiments were performed on an Ascend 700 NMR spectrometer with an AVANCE NEO console. The peak widths were determined accounting for the line broadening during processing.

4.13.1 Solution NMR sample preparation

Solution NMR samples were prepared in buffers with low salt concentration and a buffering agent which ideally did not contain any NMR active nuclei. This is easily achieved for ^{13}C and ^{15}N nuclei as their NMR active isotopes have low natural abundance. For ^1H spectra, deuterated solvents and buffering agents were added, if needed. The field was locked with 10% of D_2O via the deuterium channel. Usually, DSS was added to the sample for referencing. The pH value of the sample was adapted depending on the stability of the protein of interest. The protein concentration itself was increased as much as possible depending on the expression and purification yields and the sensitivity of the measurement. Samples with 0.5 mL volume were filled into 5 mm NMR tubes. For 5 mm Shigemi NMR tubes, 250 μL volume was needed. For 3 mm NMR tubes, >130 μL volume was required.

4.13.2 Solid-state NMR sample preparation

When protein crystals successfully grew, the crystals were harvested from the wells of the crystallization plate with a pipette tip. Stirring the drop containing the crystals with the tip enabled to detach the crystals from the well surface and to break the larger crystals into

smaller pieces which could be more easily transferred with a pipette and, more importantly, packed into solid-state NMR rotors. The crystal suspension from wells with similar conditions was combined and used to fill the 1.3 mm or 0.7 mm solid-state NMR rotors. This procedure was conducted using the MAS rotor tools and the respective protocol provided by Bruker with the help of Prof. Dr. Rasmus Linser. In several, consecutive iterations, the crystals were filled into the rotor at increasing centrifugation speed. Each time, the liquid on top was removed and another few drops of precipitant solution with crystals were added on top. Most importantly, the winged cap was added last, and the rotor was cleaned and thoroughly inspected before use.

5. Appendix

5.1 Precursors, ligands, additional titrations, and water contacts for hCAII

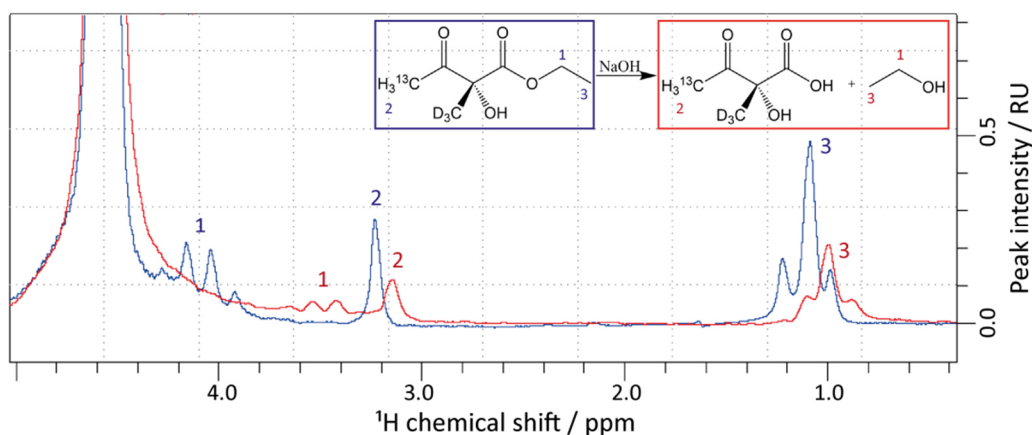


Figure 53: 1D ^1H -spectra of the ethyl-2-hydroxy-2-methyl-3-oxobutanoate precursor for H_3^{13}C -LV-methyl labelling before (blue) and after (red) hydrolysis upon stepwise addition of in total an equimolar amount of NaOH. After hydrolysis, the methylene quadruplet is shifted from ~ 4.1 to ~ 3.5 ppm (1) while the methyl groups (2) and (3) are each shifted upfield by ~ 0.1 ppm. The signal intensity was decreased after hydrolysis due to sample dilution.

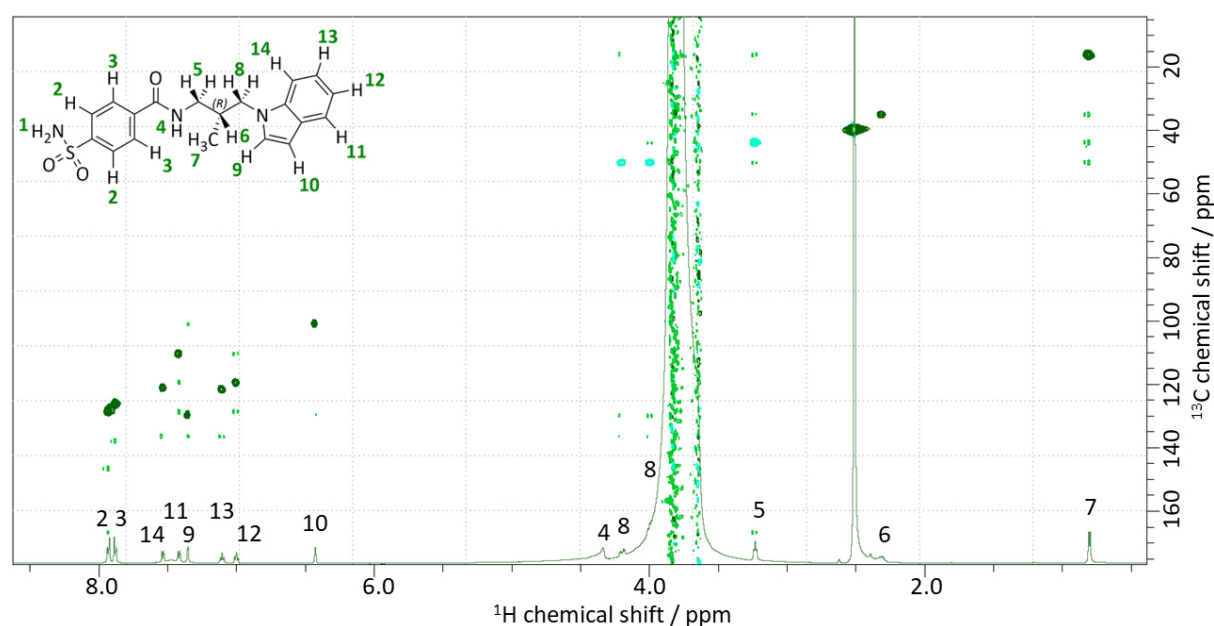


Figure 54: 2D ^1H - ^{13}C HMBC spectrum (light green), 1D ^1H -spectrum (medium green), and 2D ^1H - ^{13}C -HSQC spectrum (dark green for positive and cyan for negative peaks, corresponding to methylene groups) of 9.7 mM SBR in d_6 -DMSO, recorded in solution on a 700 MHz spectrometer at 25 $^\circ\text{C}$. The peak at ~ 2.5 ppm corresponds to a signal from non-deuterated DMSO. The ligand structure is depicted as well with the spin numbers used for peak assignment.

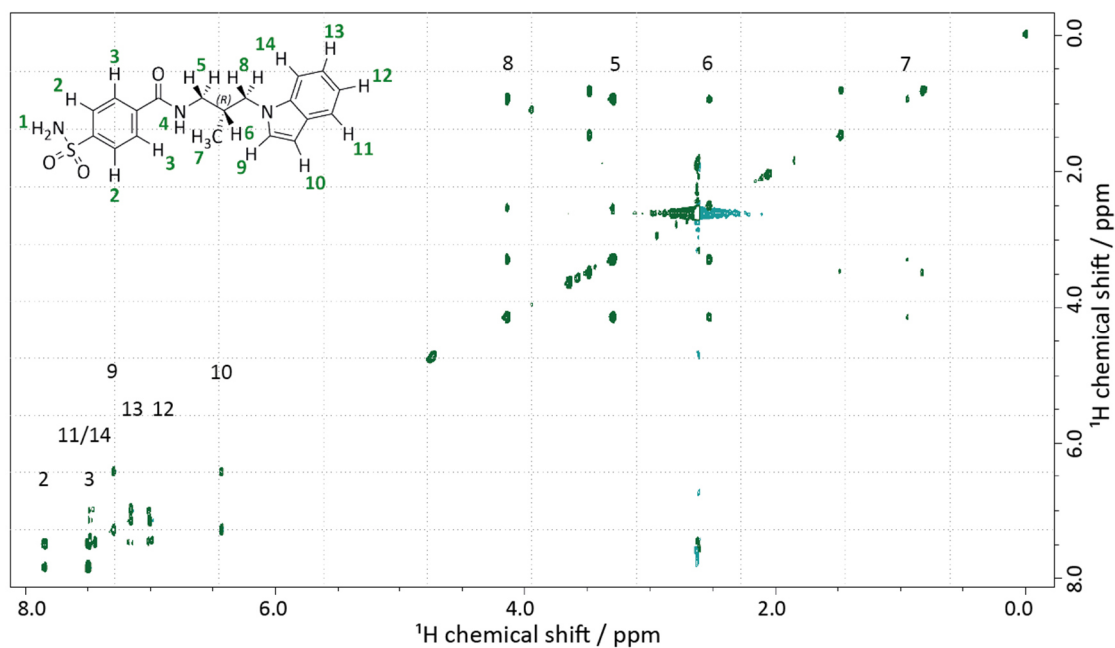


Figure 55: 2D ^1H - ^1H TOCSY spectrum of 200 μM SBR in D_2O recorded in solution on a 700 MHz spectrometer at 25 $^\circ\text{C}$. The ligand structure is depicted as well with the spin numbers used for peak assignment. The DMSO solvent peak is visible at ~ 2.6 ppm.

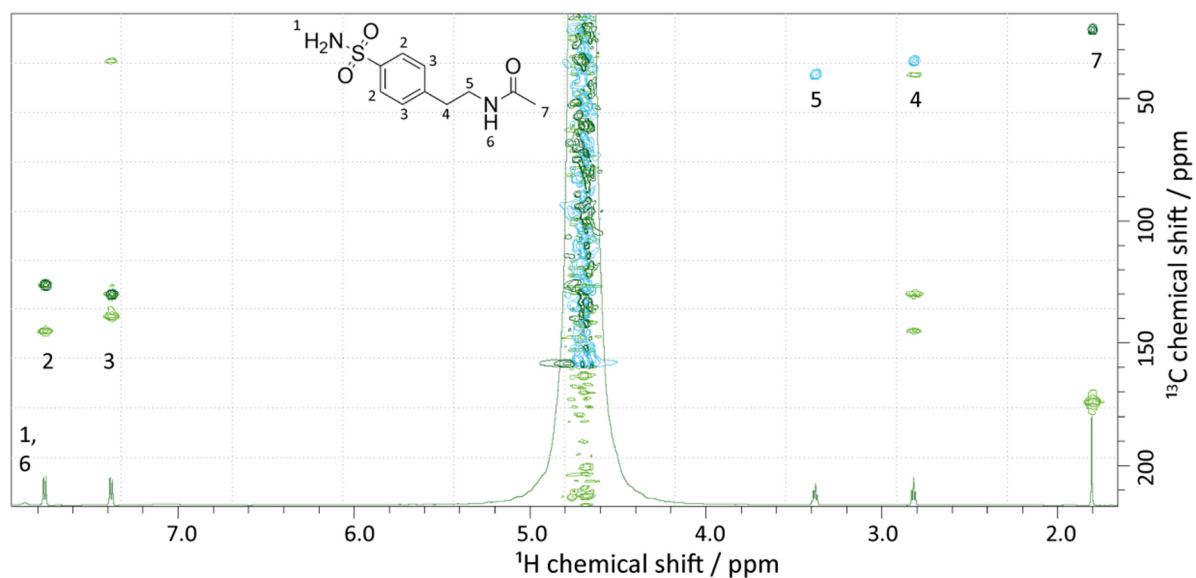


Figure 56: 2D ^1H - ^{13}C HMBC spectrum (light green), 1D ^1H -spectrum (medium green), and 2D ^1H - ^{13}C -HSQC spectrum (dark green for positive and light blue for negative peaks, corresponding to methylene groups) of 5 mM M25 in 50% D_2O , recorded in solution on a 700 MHz spectrometer at 25 $^\circ\text{C}$. The ligand structure is depicted as well with the spin numbers used for peak assignment.

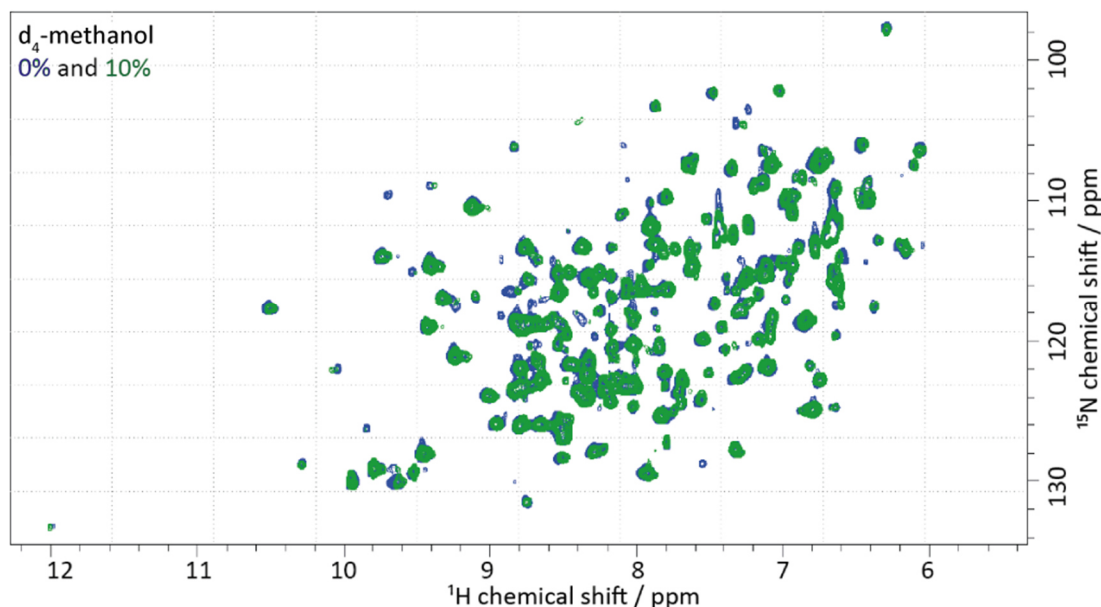


Figure 57: Overlay of two 2D ^1H - ^{15}N HSQC solution spectra comparing hCAII-SBR without (blue) and with 10% d_4 -methanol (green). The sample consisted of $480\ \mu\text{M}$ ^{13}C -, ^{15}N -labelled hCAII with SBR in a 1:2 molar ratio in 80 mM phosphate and 20 mM NaCl at pH 8.0, recorded at $7\ ^\circ\text{C}$ on an 800 MHz spectrometer. The hCAII-SBR sample was 10 months old and stored at $4\ ^\circ\text{C}$.

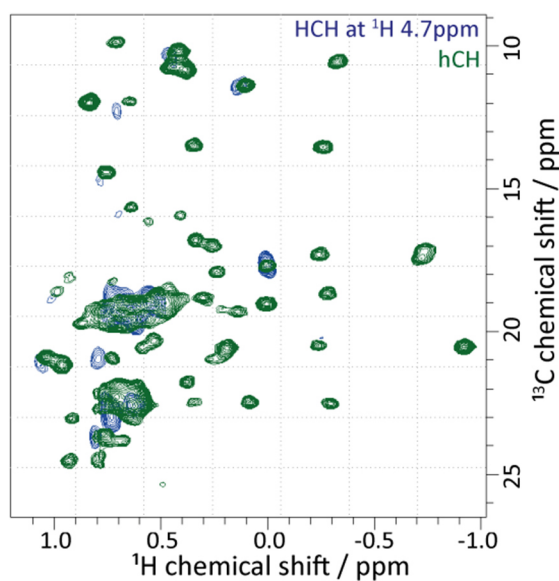


Figure 58: Overlay of solid-state NMR spectra of a 2D ^1H - ^{13}C methyl spectrum of hCAII and water contacts derived from a plane of the 3D HCH RFDR spectrum. The green spectrum corresponds to an hCH HMQC spectrum of a crystallized hCAII sample in 2.9 M $(\text{NH}_4)_2\text{SO}_4$ and 50 mM d_5 -TRIS buffer at pH 7.9 measured in a 1.3 mm rotor at 55 kHz recorded at $23\ ^\circ\text{C}$ on a 700 MHz spectrometer. The blue spectrum refers to a plane at the water ^1H chemical shift of 4.7 ppm in the direct proton dimension of a 3D HCH RFDR spectrum, measured on a crystalline hCAII sample, which has been soaked with ligand SBR. The sample was kept in 2.9 M $(\text{NH}_4)_2\text{SO}_4$ and 50 mM d_5 -Tris buffer at pH 7.9 in a 0.7 mm rotor. The spectrum was recorded at 100 kHz and $\sim 23\ ^\circ\text{C}$ on a 700 MHz spectrometer for roughly three days.

5.2 SEC column calibration

The SEC column used to purify GlpG was calibrated (see Figure 59). Plotting the elution volume against the logarithm of the size of known proteins, the molecular size of the GlpG-FC-12 complex (~12.5-13.5 mL elution volume) was estimated to be around 60-37 kDa which matches the theoretical size of 52 kDa.

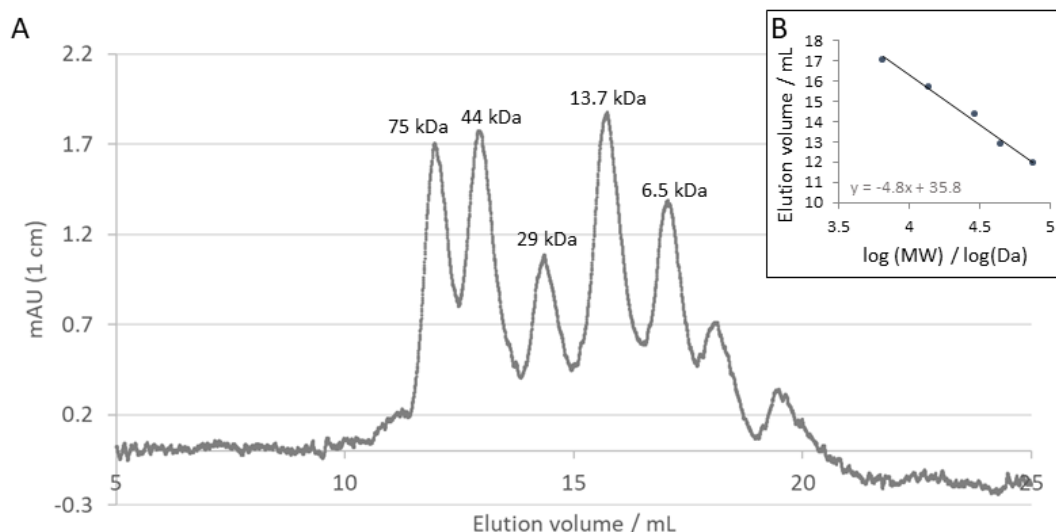


Figure 59: Calibration of the Superdex 200 Increase 30 / 100 column using the Gel Filtration Calibration Kit. Here, the peaks correspond to conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), ribonuclease (13.7 kDa), and aprotinin (6.5 kDa), here sorted by increasing elution volume. The last two peaks are noisy and were not considered to correspond to one of the marker proteins.

5.3 X-ray crystallography data

Table 23: Information on X-ray data (PDB: 8R1I) collected and processed for the co-crystallized and soaked hCAII-SBR complex. All datasets were obtained with waves of 1 Å at an excitation voltage of 12.4 keV at 100 K with an EIGER2 XE 16M detector at a PXII X10SA at the Swiss Light Source. Single crystals with P 1 21 1 space groups were used.

Data collection parameters	Co-crystallized hCAII-SBR complex	hCAII crystals soaked with SBR
Unit cell dimensions	42.352 41.595 72.243 90 104.343 90	42.339 41.615 72.286 90 104.364 90
Resolution / Å	1.459 – 32.098	1.481 - 41.015
Number of reflections	42546	39468
Completeness / %	99.77	96.77
R _{work} , R _{free}	0.1680, 0.1917	0.1779, 0.1996
RMS (bonds, angles)	0.009, 1.00	0.0046, 0.75
Ramachandran (favored, outliers) / %	95.70, 0.39	95.69, 0.00
Clashscore	1.67	4.32

5.4 Expression, activity, temperature sensitivity, and assignments of GlpG

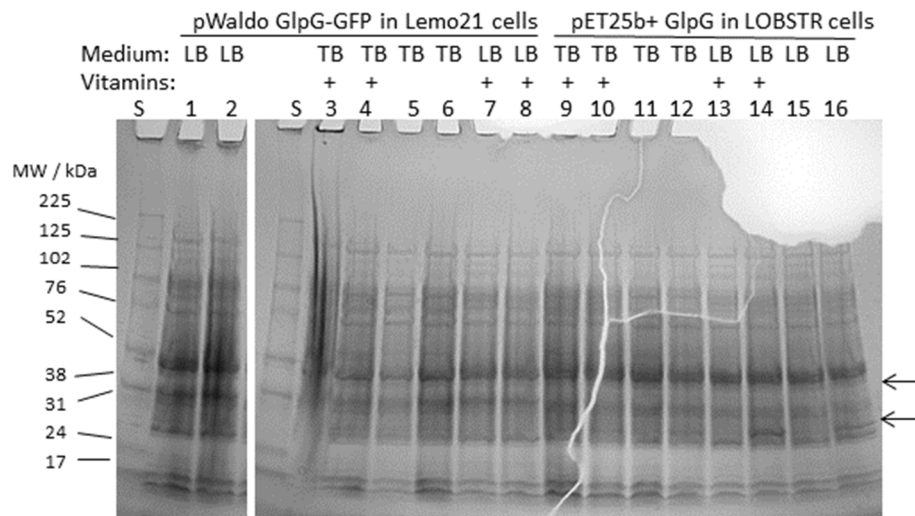


Figure 60: SDS-PAGE comparing the different unlabelled expression conditions for the 61 kDa GlpG-GFP expressed via the pWaldo plasmid in Lemo21 cells and the 33 kDa full-length GlpG wild type expressed via the pET25 plasmid in LOBSTER cells. The expression was tested in LB and TB media with and without 1x Basal Medium Eagle vitamin concentrate. For each condition, the first lane corresponds to an aliquot before and second lane to an aliquot taken after the induction with 1 mM IPTG. Unfortunately, the GlpG-GFP band is not clearly visible.

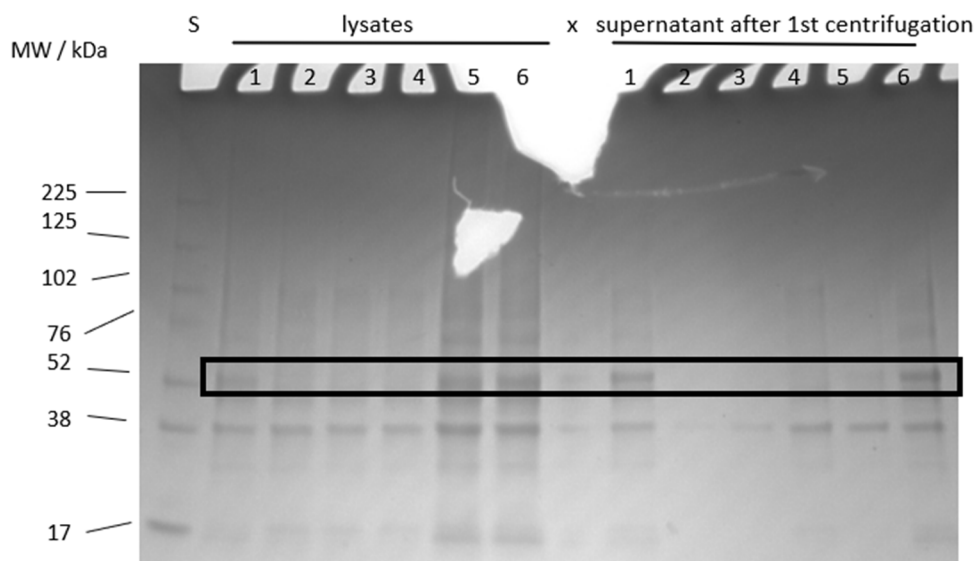


Figure 61: SDS-PAGEs showing the lysate and supernatant after centrifugation of different expression conditions of GlpG-GFP (pWaldo plasmid) in Lemo21 cells. The expression can be regulated by adding 1 mM rhamnose upon induction with 0.4 mM IPTG at 18 °C (1, 2, 5, 6) or 25 °C (3, 4) overnight. 1, 5, 6: no rhamnose, 2, 4: 1 mM rhamnose, 3: the cells were incubated on ice for 10 min before induction and adding 1 mM rhamnose.

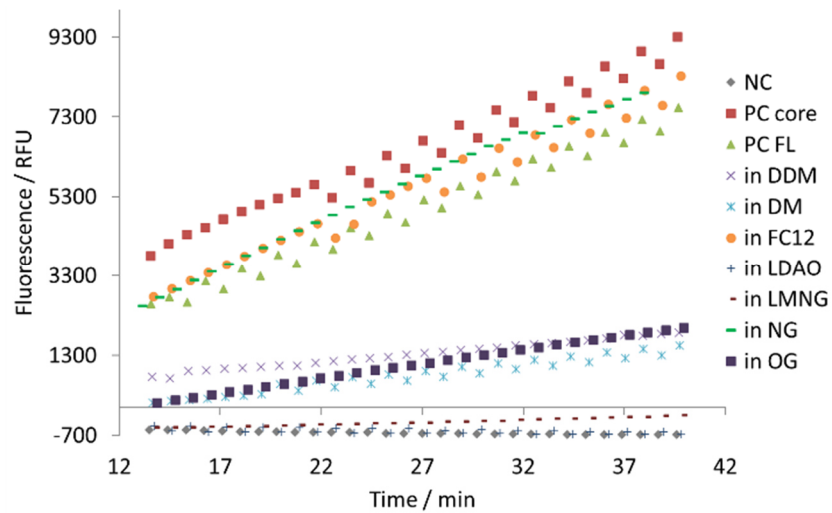


Figure 62: Activity assay^[169] of GlpG in different detergent micelles (chapter 4.11.2). The negative control contained buffer and DM micelles only. The positive controls consisted of full-length GlpG wild type (PC FL) or GlpG core wild type (PC core) in DM micelles. Due to an initial temperature difference of the plate and the Tecan fluorescence imager, the first 12 min are not shown. As all samples were affected similarly, a qualitative assessment is possible.

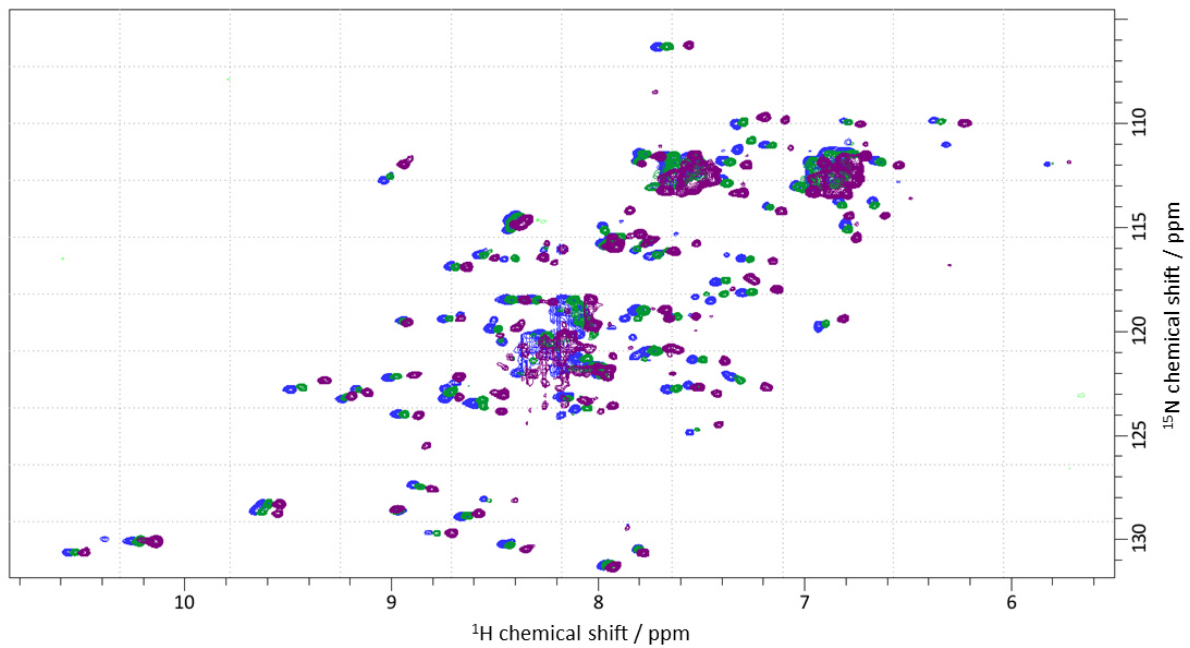


Figure 63: 2D ^1H - ^{15}N HSQC spectra of $300\ \mu\text{M}$ ^{15}N -labelled full-length GlpG wild type in 25 mM TRIS buffer, 150 mM NaCl, and 0.2 % DM at pH 8 recorded at 25 °C for 40 min (purple), at 37 °C for 40 min (green) and at 42 °C for 12 hours (blue), roughly one month old sample, which was stored at 4 °C.

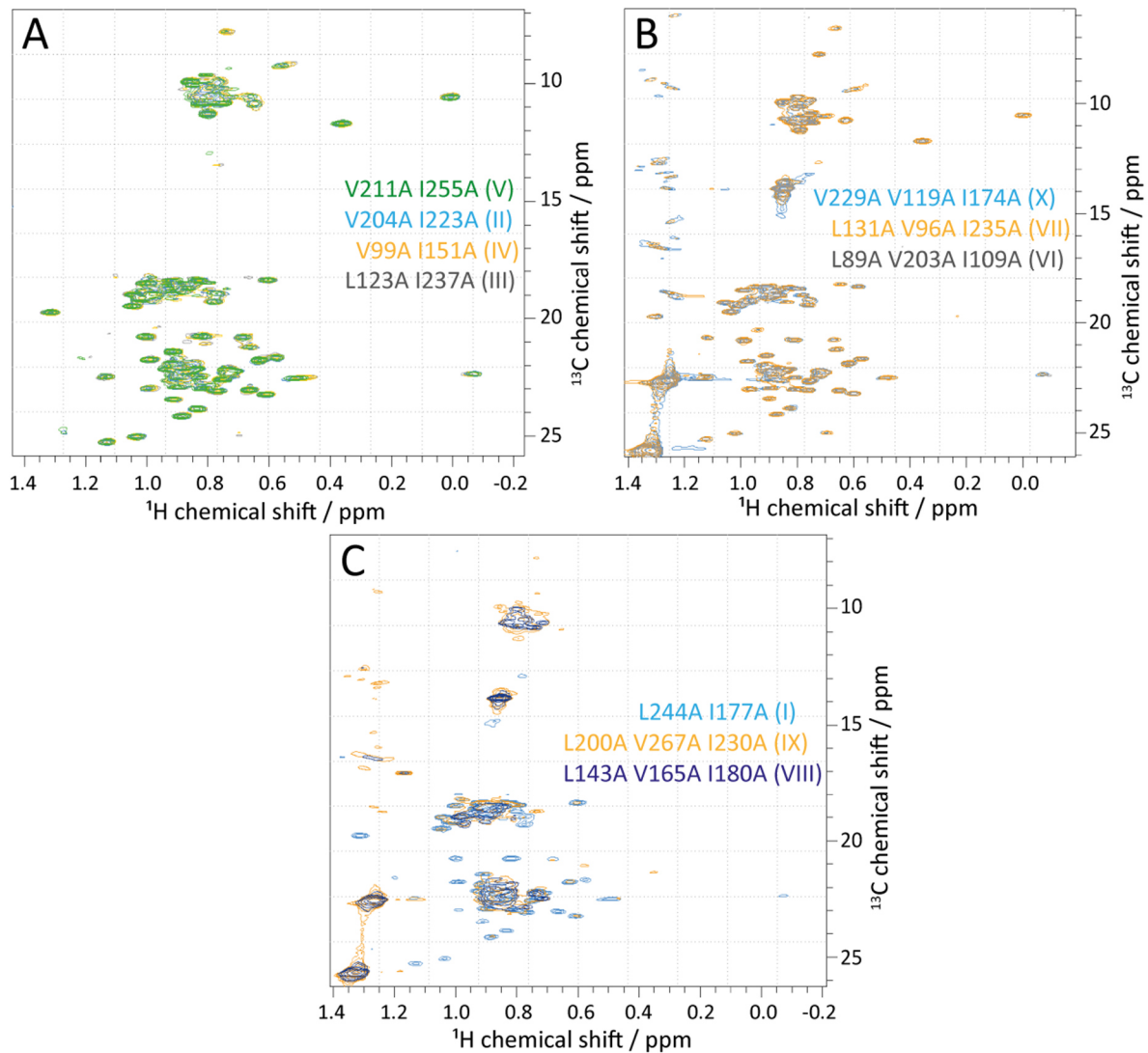


Figure 65: 2D ^1H - ^{13}C SOFAST-HMQC spectra of all ^2H -, ^{13}C -, ^{15}N - and H_3^{13}C -ILV-methyl labelled full-length GlpG ILV-mutants. **A.** ^1H - ^{13}C SOFAST-HMQC spectra of 24 μM mutant V (green, recorded for 17 h), 42 μM mutant II (light blue, recorded for 4), 120 μM mutant IV (yellow, recorded for 3.5 h), and 120 μM mutant III (grey, recorded for 3 h) in 50 mM TRIS buffer, 150 mM NaCl, 0.11% $\text{d}_{38}\text{-FC-12}$, 10% D_2O , and 1 mM DSS at pH 8. **B.** ^1H - ^{13}C SOFAST-HMQC spectra of 38 μM mutant VI (grey, recorded for 17 h), 78 μM mutant VII (yellow, recorded for 6 h), 30 μM mutant X (light blue, recorded for 7 h) in 50 mM phosphate buffer, 150 mM NaCl, 0.2% FC-12, and 10% D_2O at pH 7.4. Here, no SEC run was performed. **C.** ^1H - ^{13}C SOFAST-HMQC spectra of 34 μM mutant I (light blue, recorded for 17 h), 29 μM mutant IX (yellow, recorded for 1 d 22 h), and mutant VIII (dark blue, recorded for 17 h) in 25-50 mM TRIS buffer, 150 mM NaCl, 0.11% protonated (deuterated for mutant I) FC-12, and 10% D_2O (plus 1 mM DSS for mutant I) at pH 8.

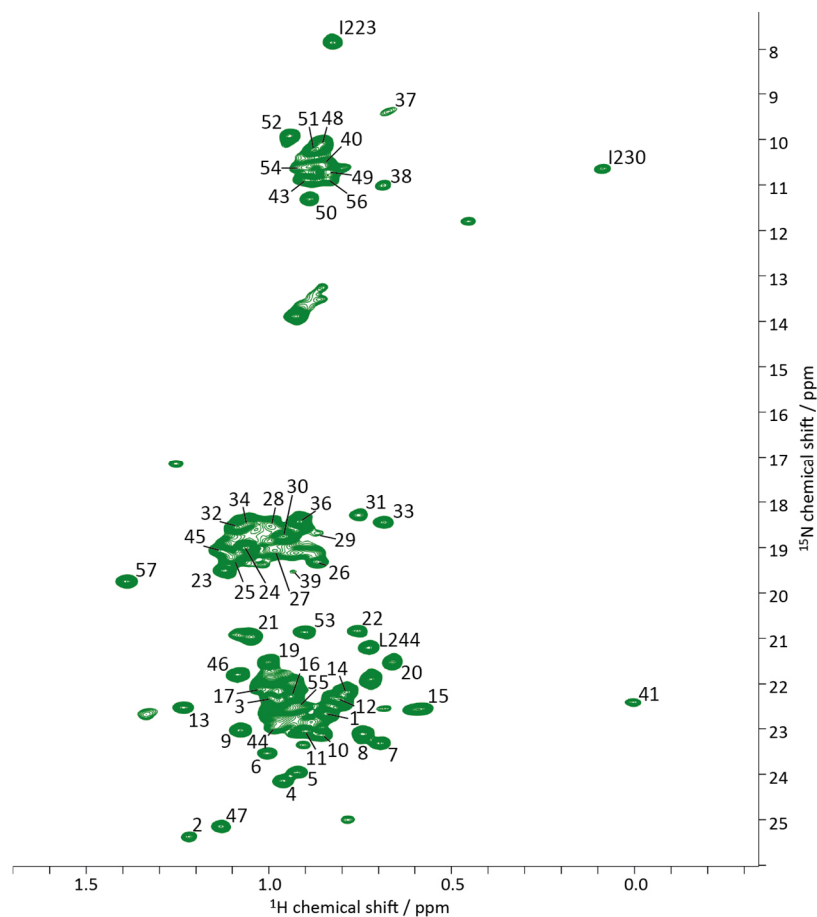


Figure 66: 2D ^1H - ^{13}C HMQC spectrum of $450 \mu\text{M}$ ^2H -, ^{13}C -, ^{15}N - and H_3^{13}C -ILV-methyl labelled full-length GlpG in 50 mM d_5 -TRIS, 150 mM NaCl, 0.11% d_{38} -FC-12, and 10% D_2O at pH 8 with peaks assigned according to methyl assignments (I223, I230, L244) or according to the peak number also referred to in Figure 83-86 and Table 25-26. The spectrum was recorded at 40°C on an 800 MHz spectrometer.

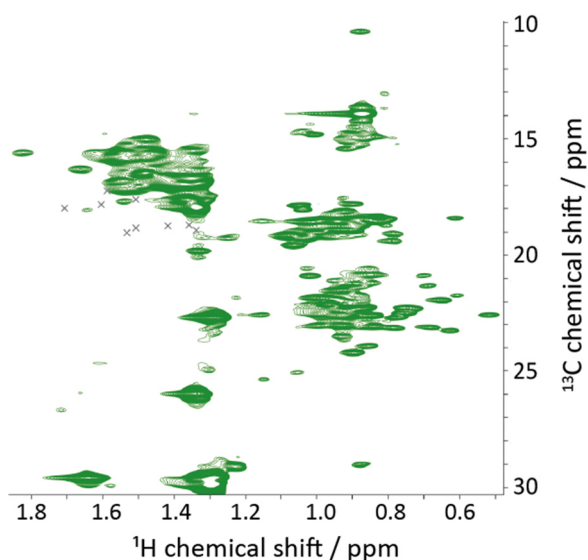


Figure 67: 2D ^1H - ^{13}C HMQC spectrum of $260 \mu\text{M}$ ^2H -, ^{13}C -, ^{15}N - and H_3^{13}C -ILVA-methyl labelled full-length GlpG S201T in 50 mM Tris, 150 mM NaCl, 0.11% d -FC12, 10% D_2O , and $12 \mu\text{M}$ DSS at pH 8.0. The spectrum was recorded at 40°C on an 800 MHz spectrometer. The grey peaks were simulated from published assignments for the N-terminal domain^[142].

5.5 Raw data of relaxation dispersion plots for SBR ligand dynamics

All plots were obtained using Monte Carlo error estimation, provided by Suresh Vasa.

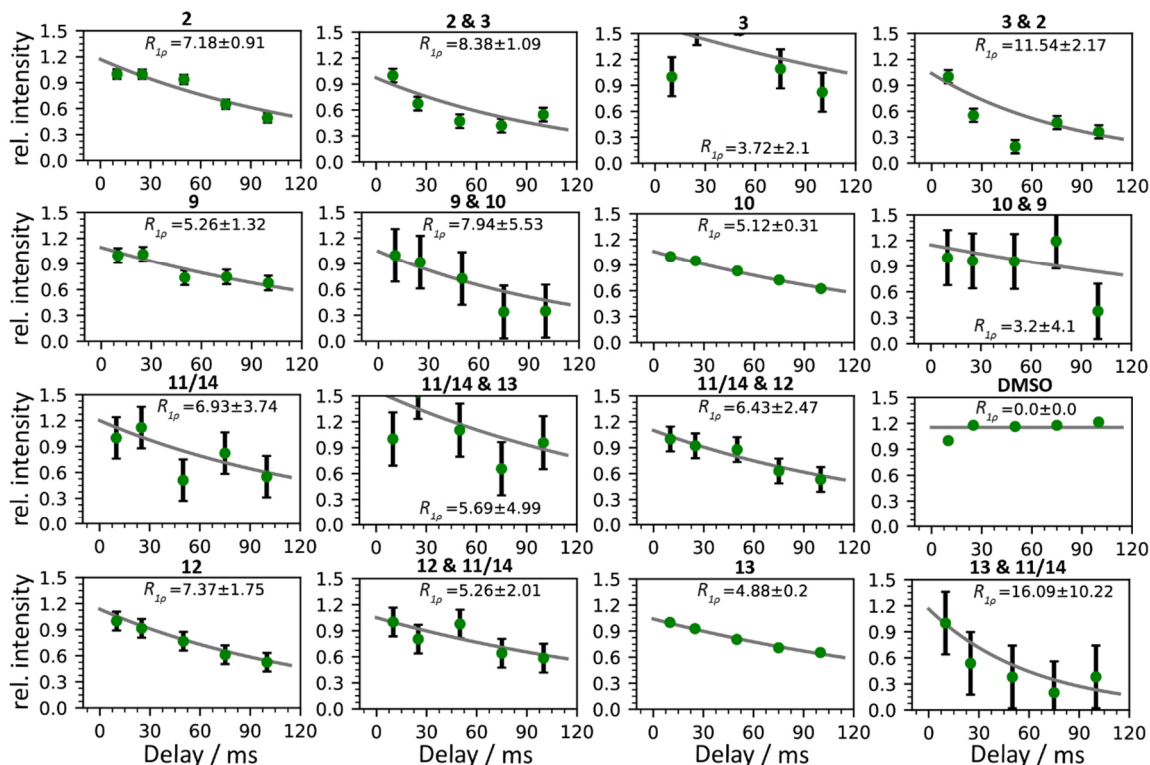


Figure 68: Plots estimating $R_{1\rho}$ at 1 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

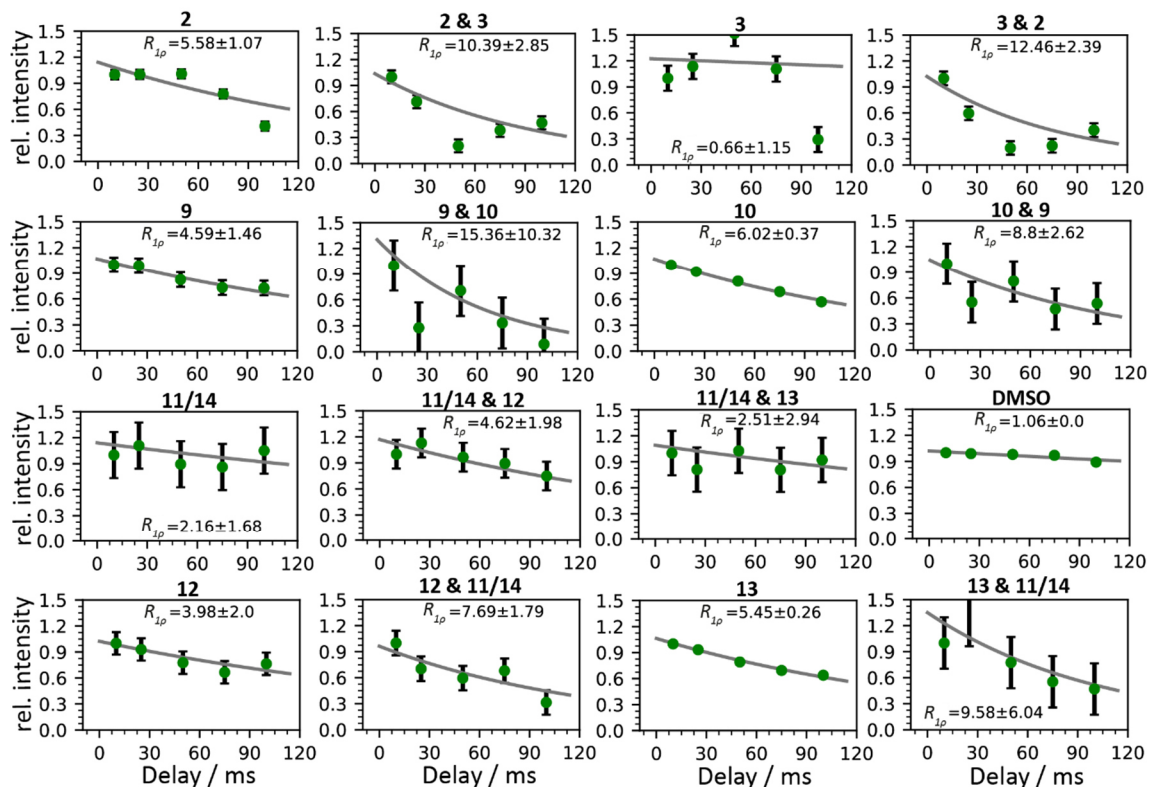


Figure 69: Plots estimating $R_{1\rho}$ at 1.5 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

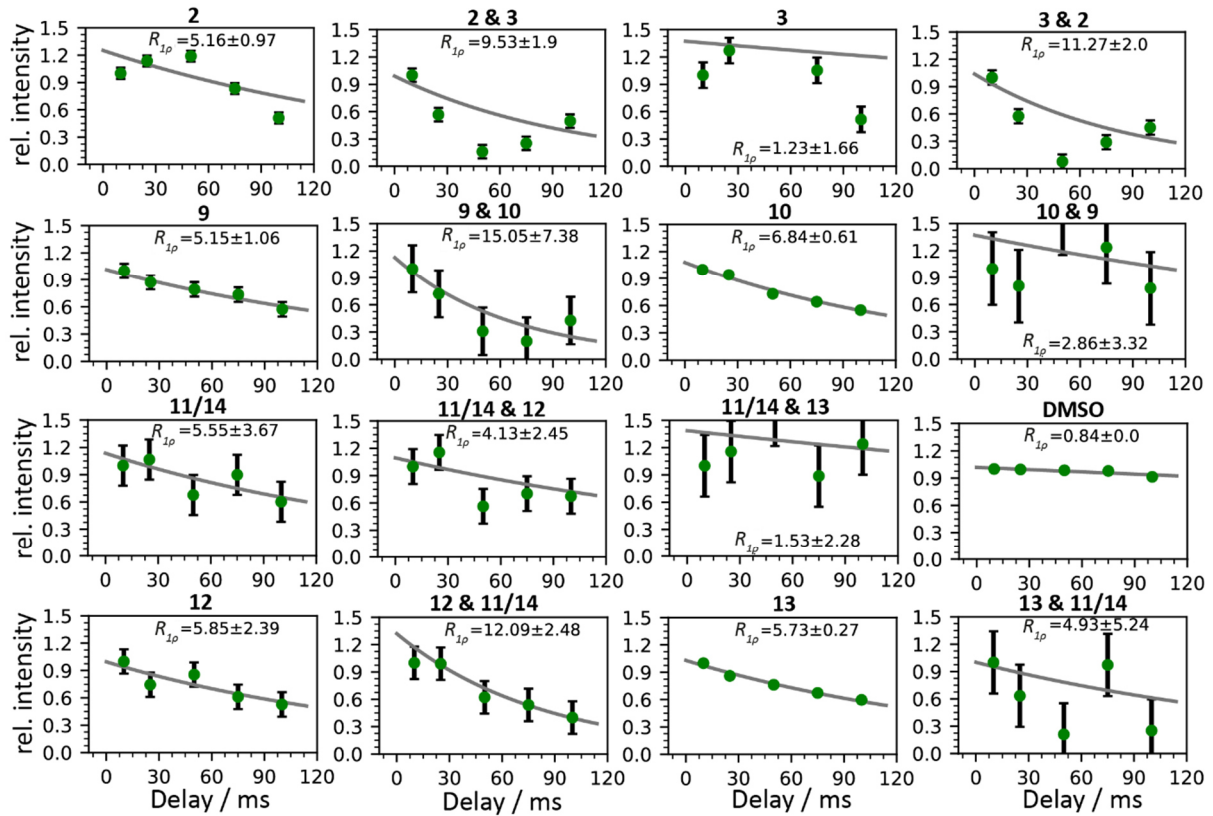


Figure 70: Plots estimating $R_{1\rho}$ at 2 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

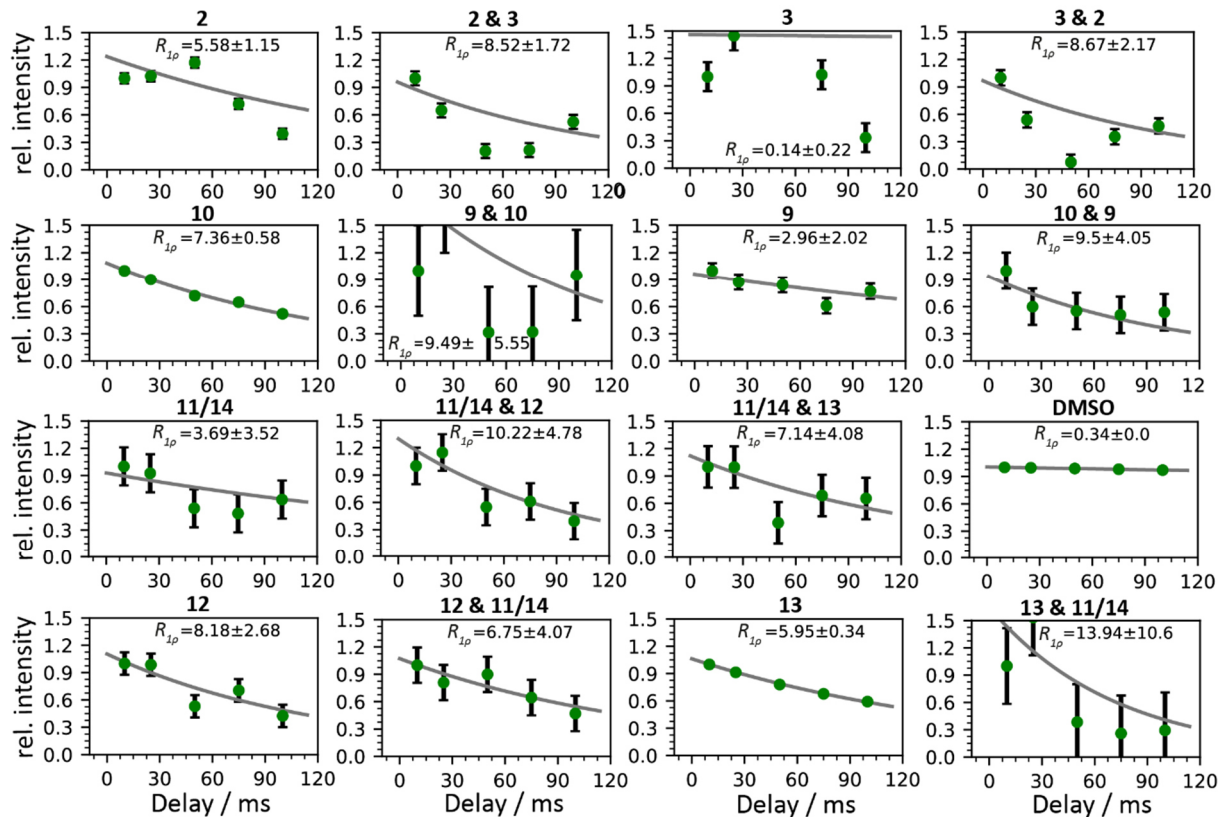


Figure 71: Plots estimating $R_{1\rho}$ at 2.5 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

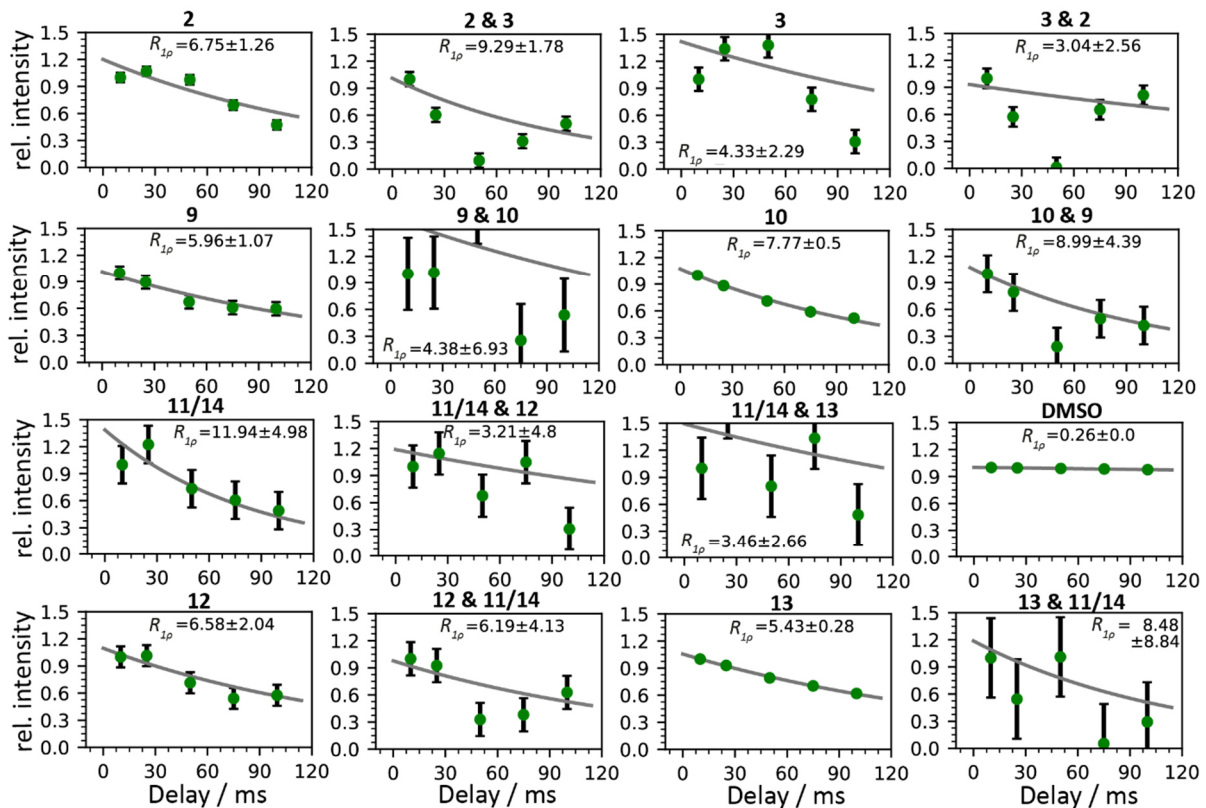


Figure 72: Plots estimating $R_{1\rho}$ at 3 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

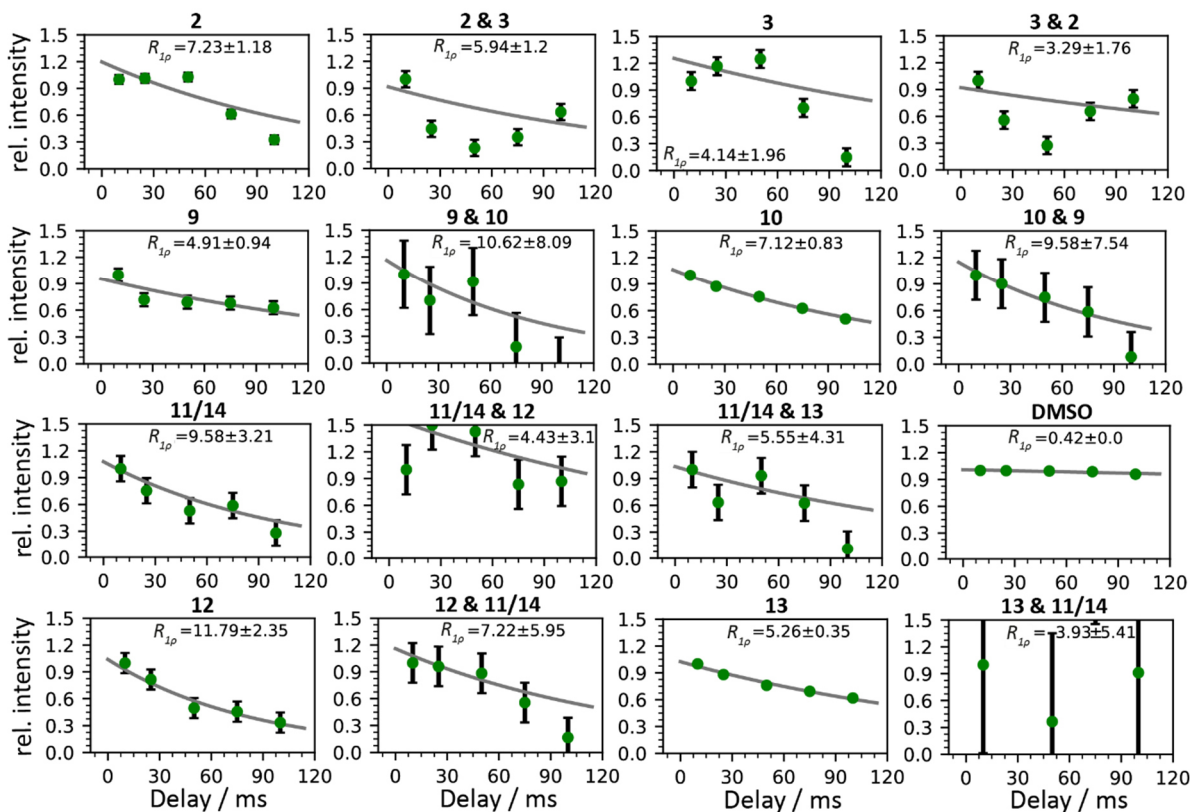


Figure 73: Plots estimating $R_{1\rho}$ at 5 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

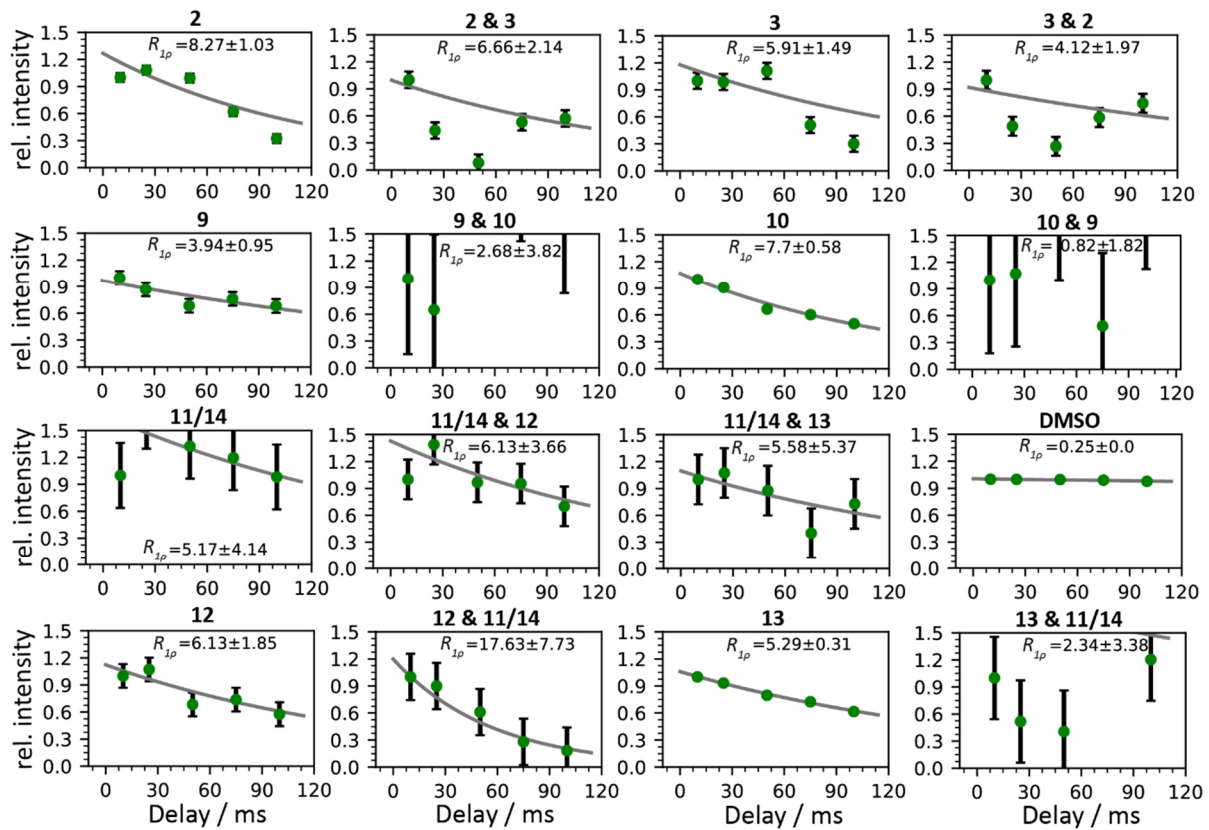


Figure 74: Plots estimating $R_{1\rho}$ at 7 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

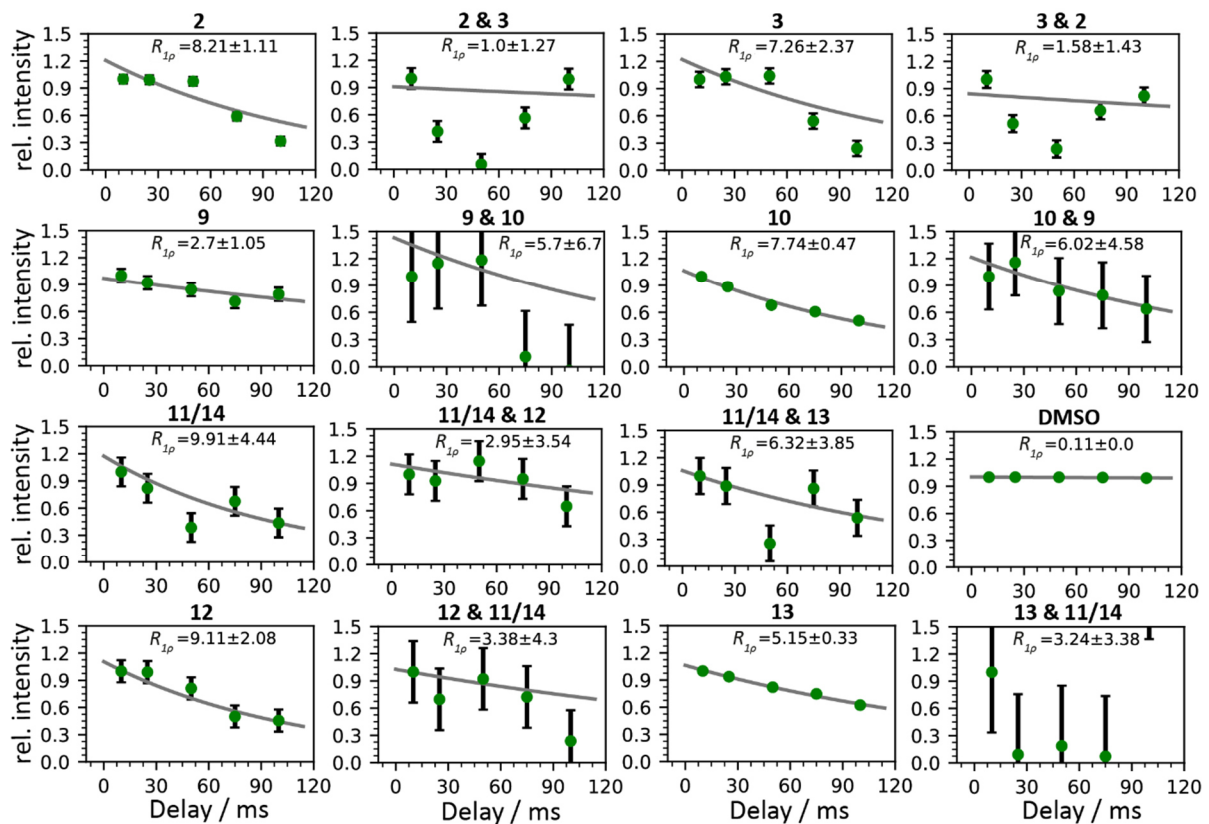


Figure 75: Plots estimating $R_{1\rho}$ at 10 kHz for the different ^1H -BMRD conditions of the ligand SBR while bound to hCAII in solution.

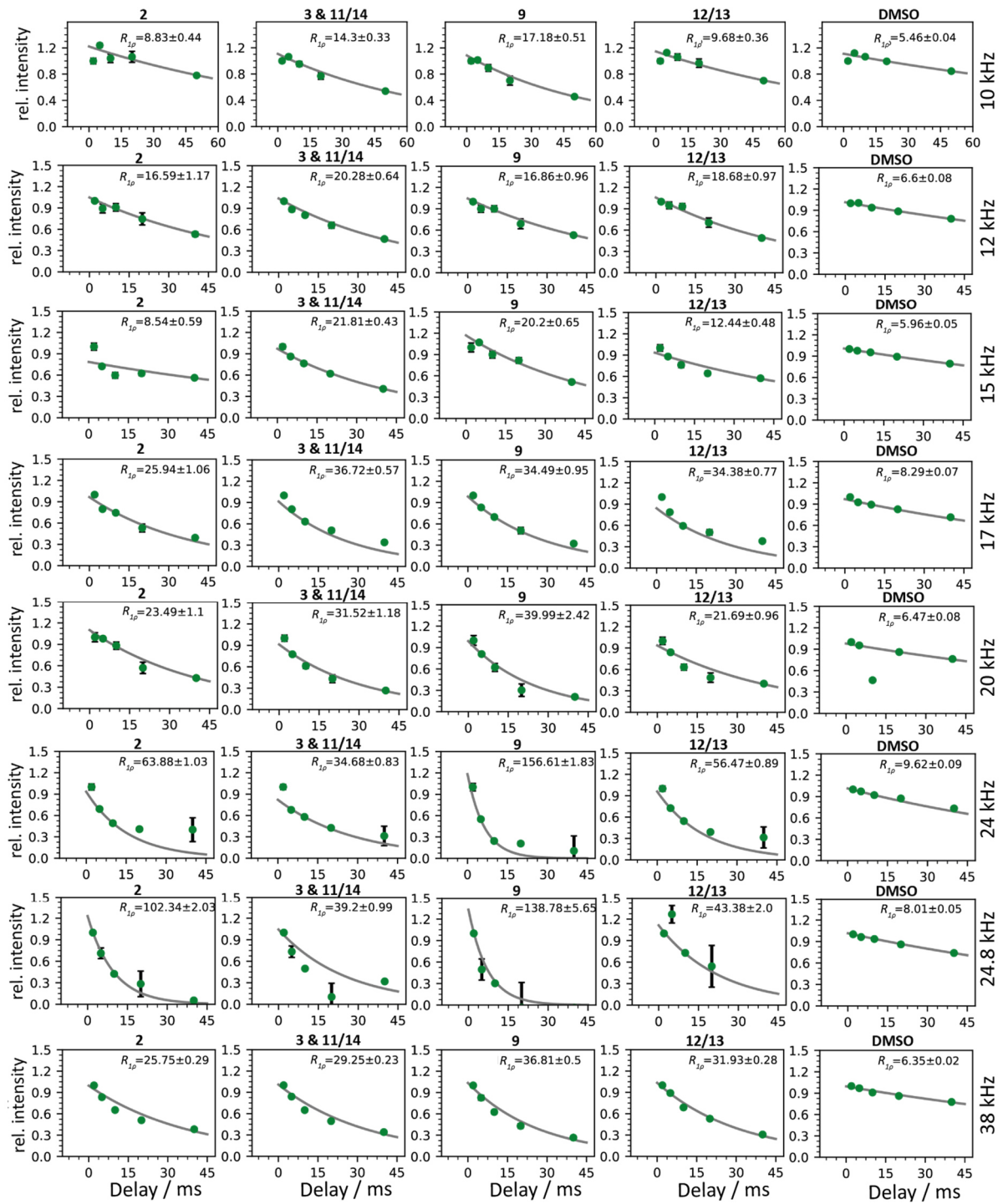


Figure 76: Plots estimating $R_{1\rho}$ for the different ^1H -NERRD conditions of the ligand SBR while bound to hCAII in the solid state. The data were acquired at the spin lock field strengths written on the right-hand side of the rows.

5.6 CPMG plots for hCAII and GlpG in apo and inhibitor-bound state

All plots were obtained with NESSY^[176]. Model 1 assumes no exchange, while models 2 and 3 represent fast and slow exchange between two states, respectively.

Table 24: Parameters of the CPMG profiles for apo-hCAII shown in Figure 77-82 fitted with model 1.

Residue	Baseline R_2 / 1/s	Baseline R_2 error / 1/s	χ^2 value	Residue	Baseline R_2 / 1/s	Baseline R_2 error / 1/s	χ^2 value
5	20.39	0.10	10.75	130	17.82	0.02	4.26
6	22.18	0.01	172.40	131	18.89	0.03	42.96
7	24.13	0.06	18.60	132	19.29	0.06	17.54
8	18.53	0.02	23.72	133	20.51	0.07	10.50
12	20.29	0.10	12.48	134	20.43	0.05	10.28
15	21.99	0.03	11.05	135	19.38	0.03	7.87
16	20.48	0.04	13.81	136	18.94	0.05	13.61
17	22.45	0.06	17.60	138	18.20	0.03	11.11
18	20.29	0.07	9.63	141	20.24	0.05	18.05
19	21.20	0.02	20.35	142	21.43	0.02	44.11
20	19.10	0.01	36.87	143	23.57	0.18	15.74
22	19.99	0.01	26.37	144	22.21	0.02	60.26
23	21.38	0.03	22.99	145	22.85	0.06	15.89
24	21.20	0.02	20.35	146	23.83	0.34	14.88
25	20.73	0.05	16.04	147	21.40	0.06	12.92
26	20.99	0.10	11.26	148	20.88	0.02	63.55
27	21.61	0.03	9.90	149	21.97	0.04	6.10
28	19.74	0.03	5.01	150	20.22	0.07	6.86
29	22.61	0.06	30.23	153	22.46	0.11	7.88
31	22.92	0.09	17.05	156	18.00	0.03	4.41
32	21.85	0.02	15.92	157	26.11	0.10	15.71
33	22.46	0.09	27.66	159	20.24	0.05	8.86
34	19.37	0.05	30.70	160	20.77	0.03	18.31
35	21.31	0.01	52.70	162	19.44	0.02	28.49
37	19.19	0.04	9.40	163	22.30	0.04	13.85
38	17.81	0.04	11.43	164	22.47	0.03	19.98
39	19.43	0.01	15.83	165	21.45	0.10	8.73
40	41.92	0.60	240.32	166	19.05	0.09	8.60
41	21.29	0.06	18.50	167	19.52	0.01	22.87
43	18.71	0.01	20.52	168	22.67	0.03	214.02
44	19.53	0.05	4.56	169	17.82	0.02	4.26
47	21.09	0.05	8.52	170	19.43	0.02	33.10
48	19.00	0.00	69.55	172	20.31	0.02	49.25
49	22.49	0.04	29.25	174	24.28	0.10	14.13
50	18.47	0.01	24.50	175	21.37	0.07	25.33
51	21.89	0.03	12.63	176	21.46	0.03	15.27
53	28.23	0.06	65.31	177	20.97	0.06	15.27

54	17.82	0.02	4.26	178	19.37	0.06	4.99
55	28.22	0.12	24.07	179	34.27	0.12	109.11
56	21.06	0.03	24.91	181	21.85	0.02	15.92
57	23.16	0.20	13.18	182	23.14	0.03	88.34
58	21.20	0.02	20.35	183	19.52	0.01	22.87
59	22.19	0.01	91.80	184	18.94	0.04	9.05
61	21.97	0.04	6.10	186	22.47	0.04	19.98
62	21.39	0.03	40.17	187	15.38	0.03	7.72
63	20.93	0.07	11.06	189	20.48	0.03	13.81
64	18.20	0.03	11.11	190	23.06	0.03	16.86
65	18.47	0.02	24.50	191	22.41	0.05	32.43
66	21.20	0.09	13.66	192	20.97	0.07	15.27
67	22.19	0.01	91.80	193	21.91	0.01	61.34
68	22.44	0.10	14.46	195	21.16	0.05	23.24
70	20.79	0.10	8.61	196	20.01	0.04	11.07
73	21.99	0.03	11.05	197	19.10	0.01	36.87
74	19.01	0.03	9.83	198	22.22	0.06	20.08
77	19.43	0.01	15.83	199	21.97	0.04	22.64
78	21.28	0.04	8.85	202	23.54	0.05	36.88
79	20.24	0.04	18.05	203	22.01	0.07	22.94
81	20.27	0.04	49.00	204	22.31	0.04	14.11
82	21.68	0.05	12.12	205	20.31	0.08	8.26
84	19.51	0.02	18.20	207	21.78	0.12	8.14
86	27.13	0.04	111.66	208	21.90	0.09	23.32
87	22.37	0.04	24.11	209	22.21	0.07	13.00
88	21.53	0.06	8.87	210	22.73	0.11	34.62
89	20.67	0.03	10.68	211	22.35	0.06	14.06
90	23.89	0.11	26.37	212	18.83	0.03	16.40
91	21.21	0.05	17.47	215	22.20	0.03	36.54
92	22.44	0.09	7.88	217	20.68	0.08	10.59
93	22.16	0.03	21.99	219	20.99	0.08	11.26
94	22.33	0.08	18.14	221	18.94	0.05	13.61
95	21.13	0.03	55.78	222	20.77	0.03	18.31
96	21.99	0.02	11.05	223	19.99	0.01	26.37
97	20.26	0.05	16.29	224	20.49	0.04	28.86
98	27.41	0.17	22.09	225	20.08	0.06	5.48
99	22.35	0.10	18.25	226	21.99	0.07	18.19
100	19.34	0.02	44.70	227	19.98	0.05	29.03
101	20.01	0.04	11.07	228	20.64	0.06	4.42
102	28.67	0.12	32.13	229	21.13	0.07	25.78
104	20.41	0.02	61.72	230	23.16	0.19	13.18
105	20.43	0.03	36.18	231	22.66	0.12	10.90
107	21.68	0.05	69.59	232	19.74	0.11	11.60
108	20.43	0.05	10.28	235	16.88	0.02	15.24
109	20.77	0.09	8.37	237	19.43	0.01	15.83
110	21.30	0.05	61.57	238	20.61	0.14	15.75

111	20.01	0.14	15.59	239	22.30	0.03	158.32
112	20.41	0.02	108.52	241	20.58	0.04	20.13
113	20.24	0.02	60.91	242	21.20	0.02	20.35
114	21.50	0.09	13.97	243	19.00	0.00	69.55
115	22.34	0.05	21.94	244	18.46	0.01	33.54
116	20.10	0.03	14.70	245	23.49	0.17	13.20
117	21.76	0.09	12.61	247	19.99	0.01	26.37
118	24.31	0.15	33.10	248	22.44	0.09	14.46
121	22.54	0.02	87.12	250	20.67	0.04	13.11
122	19.69	0.04	35.48	253	15.76	0.03	8.14
123	21.09	0.04	9.57	254	26.58	0.03	116.26
124	20.88	0.02	63.55	255	20.33	0.06	18.00
125	19.10	0.01	36.87	257	21.91	0.03	31.56
126	19.80	0.05	9.31	258	21.42	0.03	42.40
127	19.74	0.03	5.01	259	19.99	0.01	56.94
128	21.53	0.03	21.93	260	18.07	0.01	23.76
129	18.00	0.03	4.41				

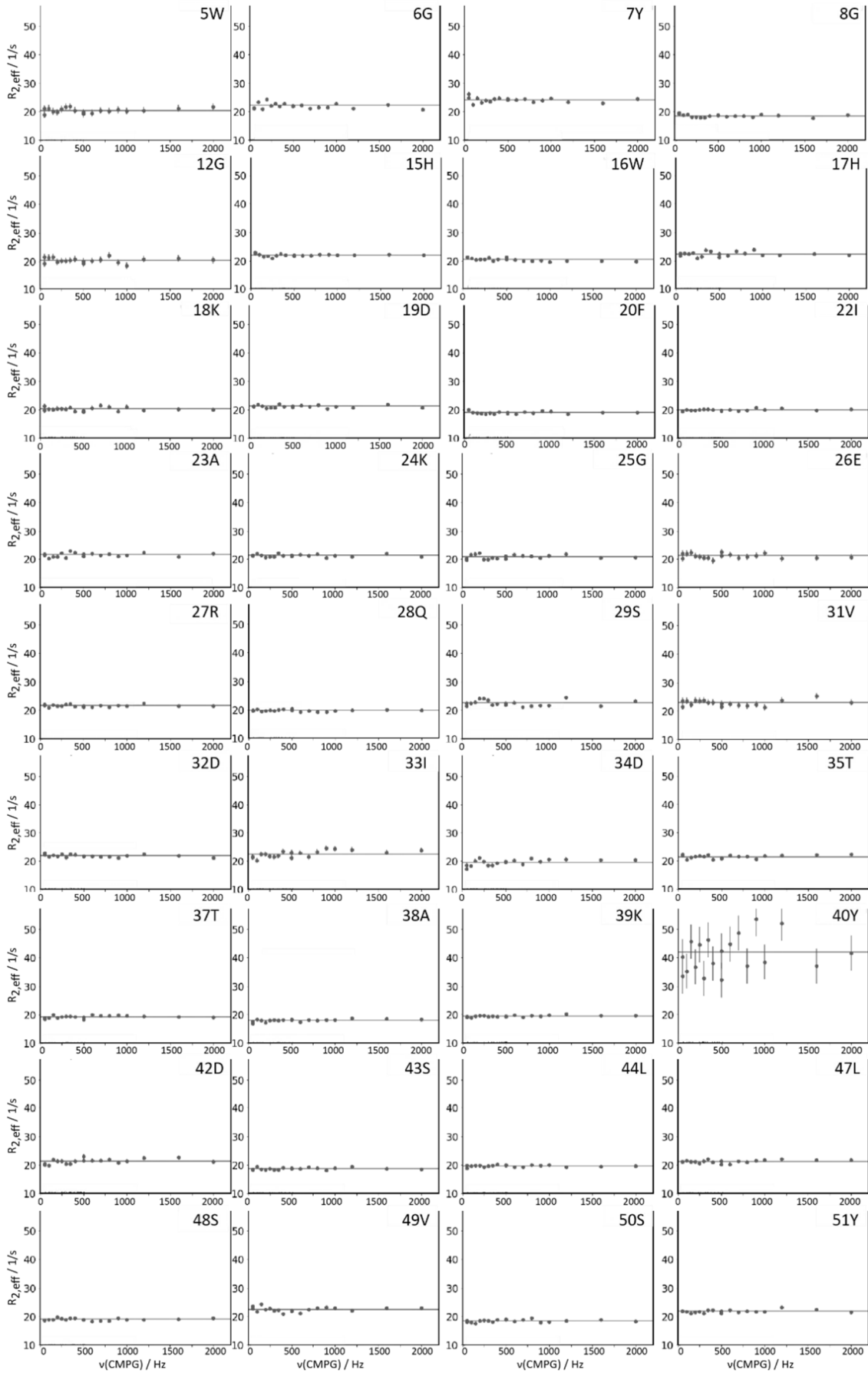


Figure 77: ^{15}N -CPMG relaxation dispersion plots for the assigned residues 5-51 of apo-hCAII.

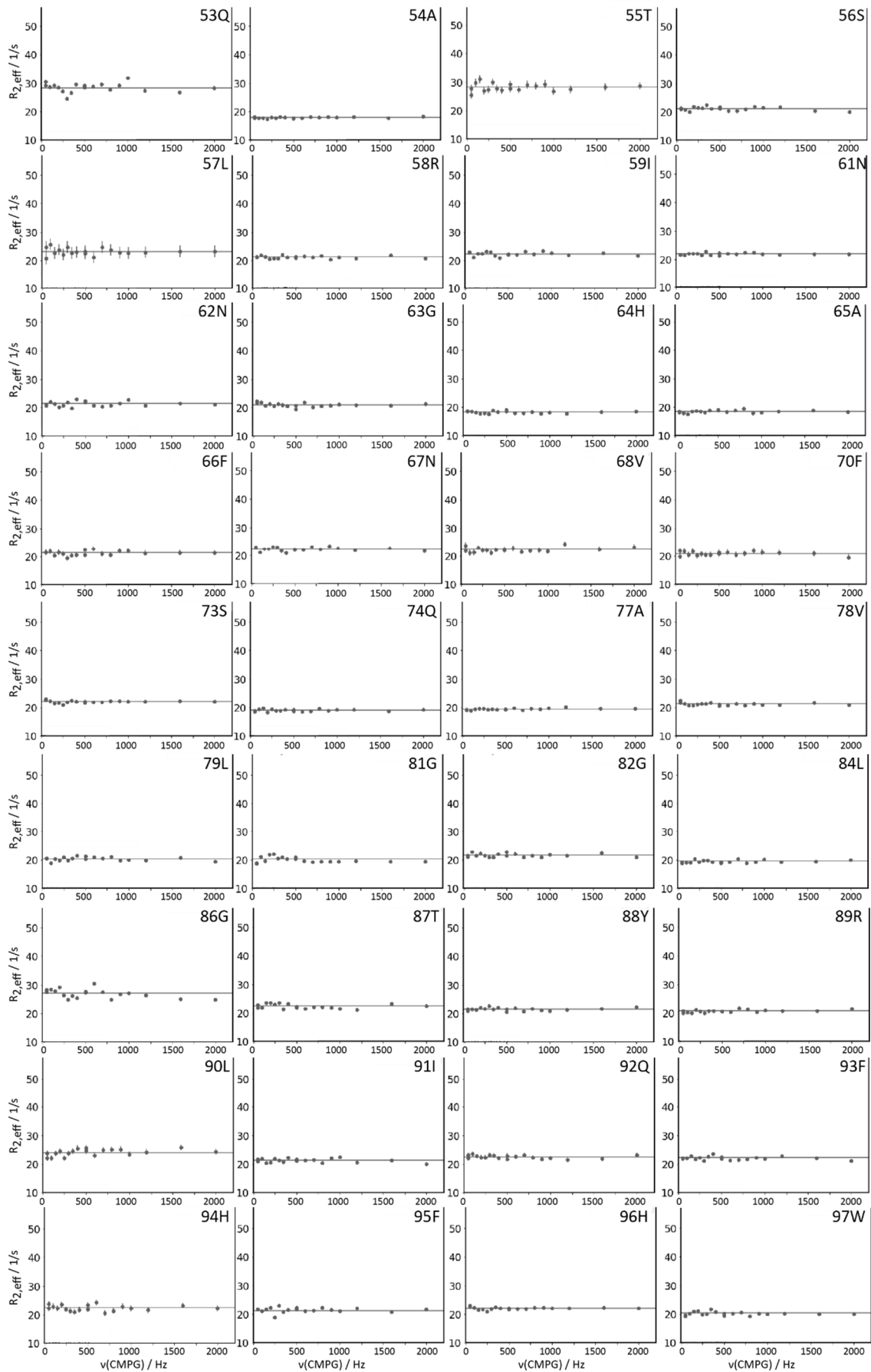


Figure 78: ^{15}N -CPMG relaxation dispersion plots for the assigned residues 53-97 of apo-hCAII.

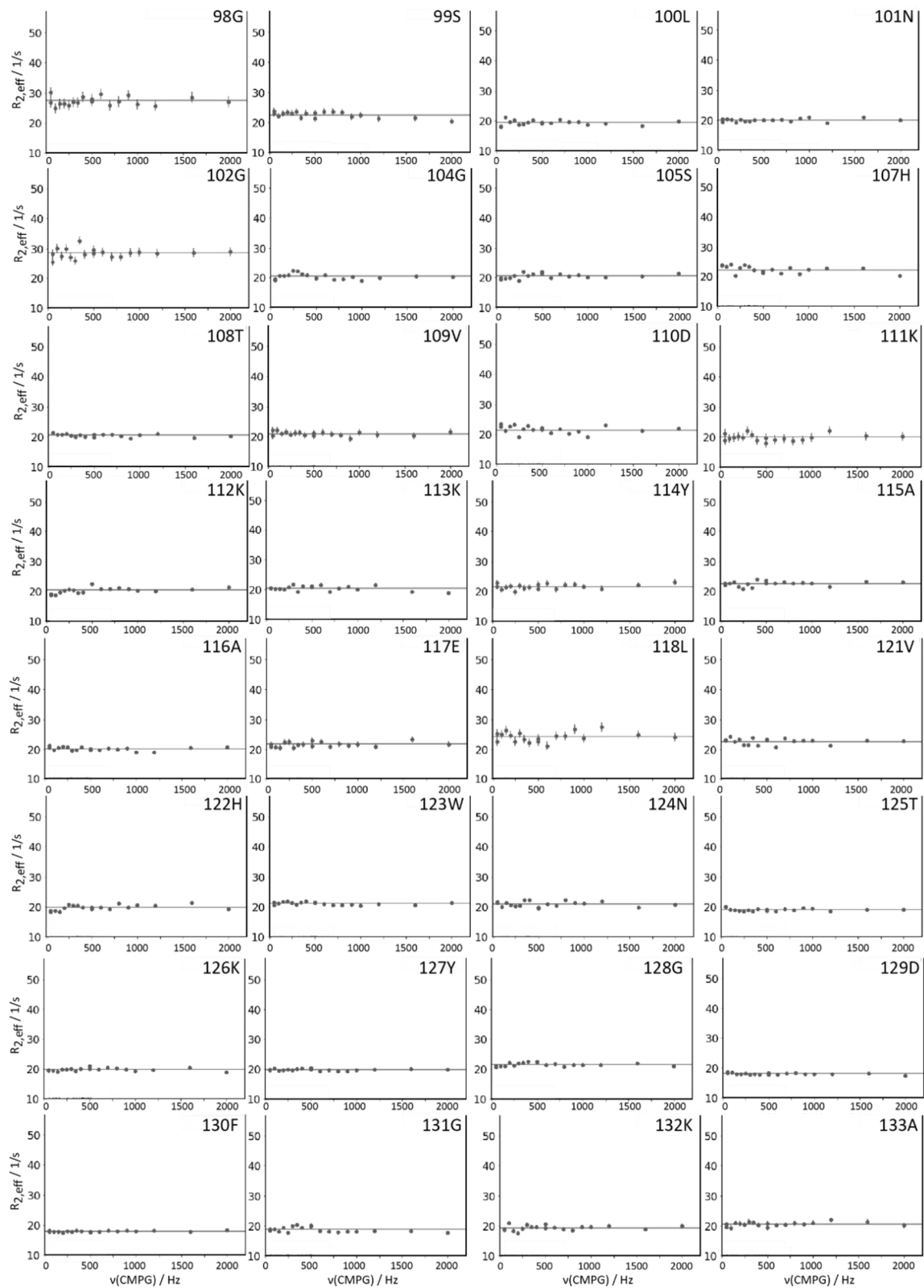


Figure 79: ^{15}N -CPMG relaxation dispersion plots for the assigned residues 98-133 of apo-hCAII.

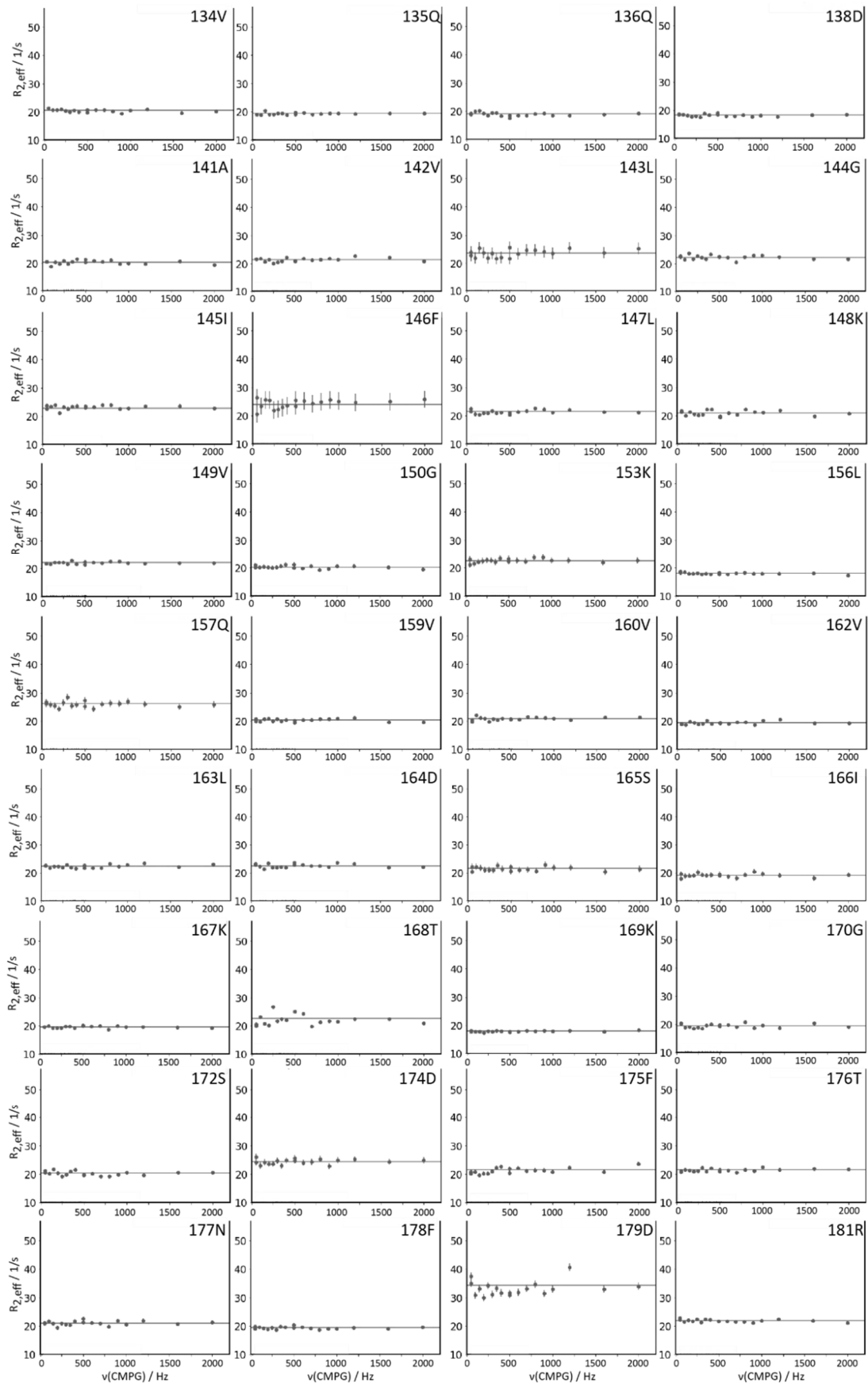


Figure 80: ^{15}N -CPMG relaxation dispersion plots for the assigned residues 134-181 of apo-hCAII.

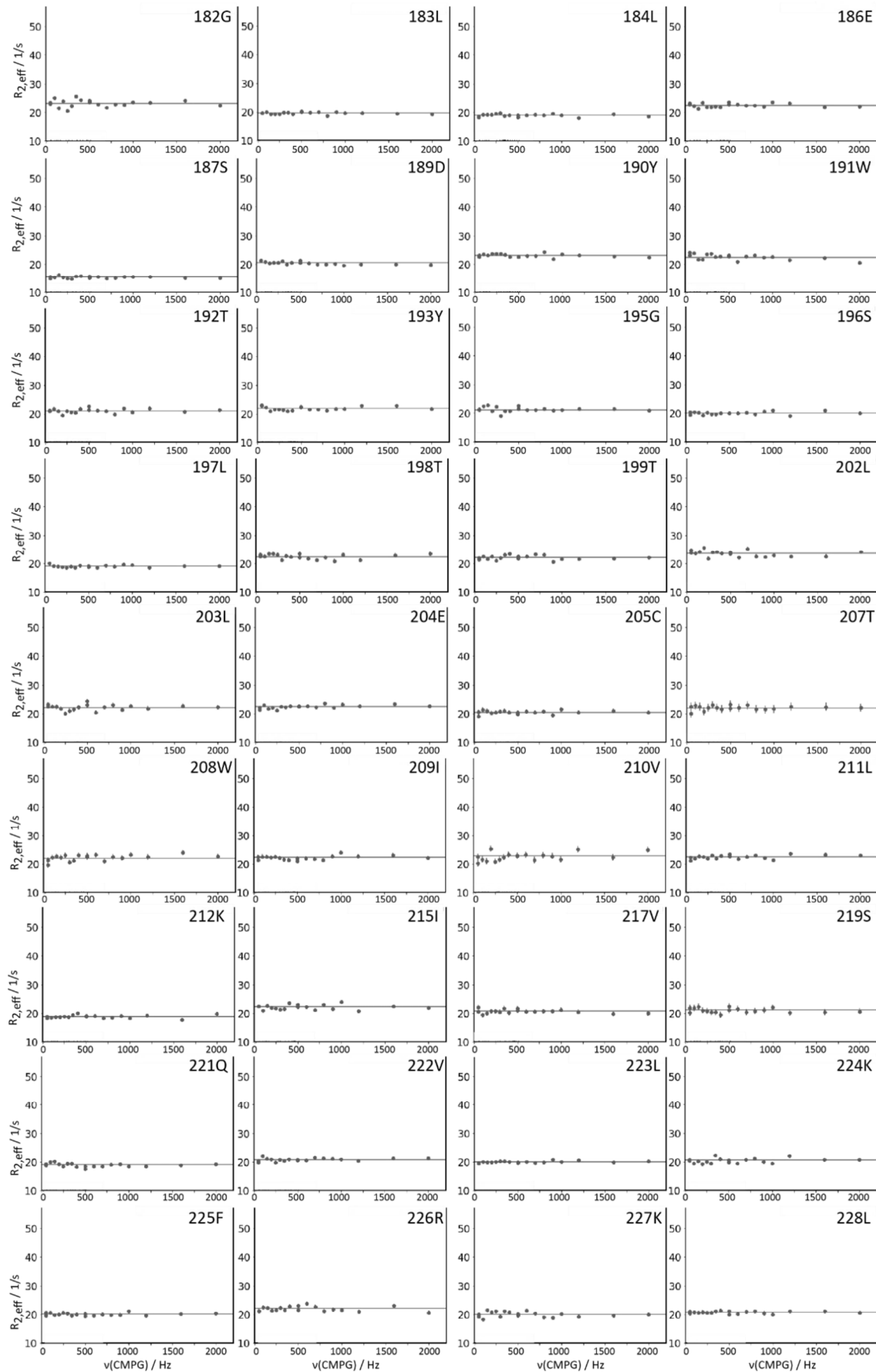


Figure 81: ^{15}N -CPMG relaxation dispersion plots for the assigned residues 182-228 of apo-hCAII.

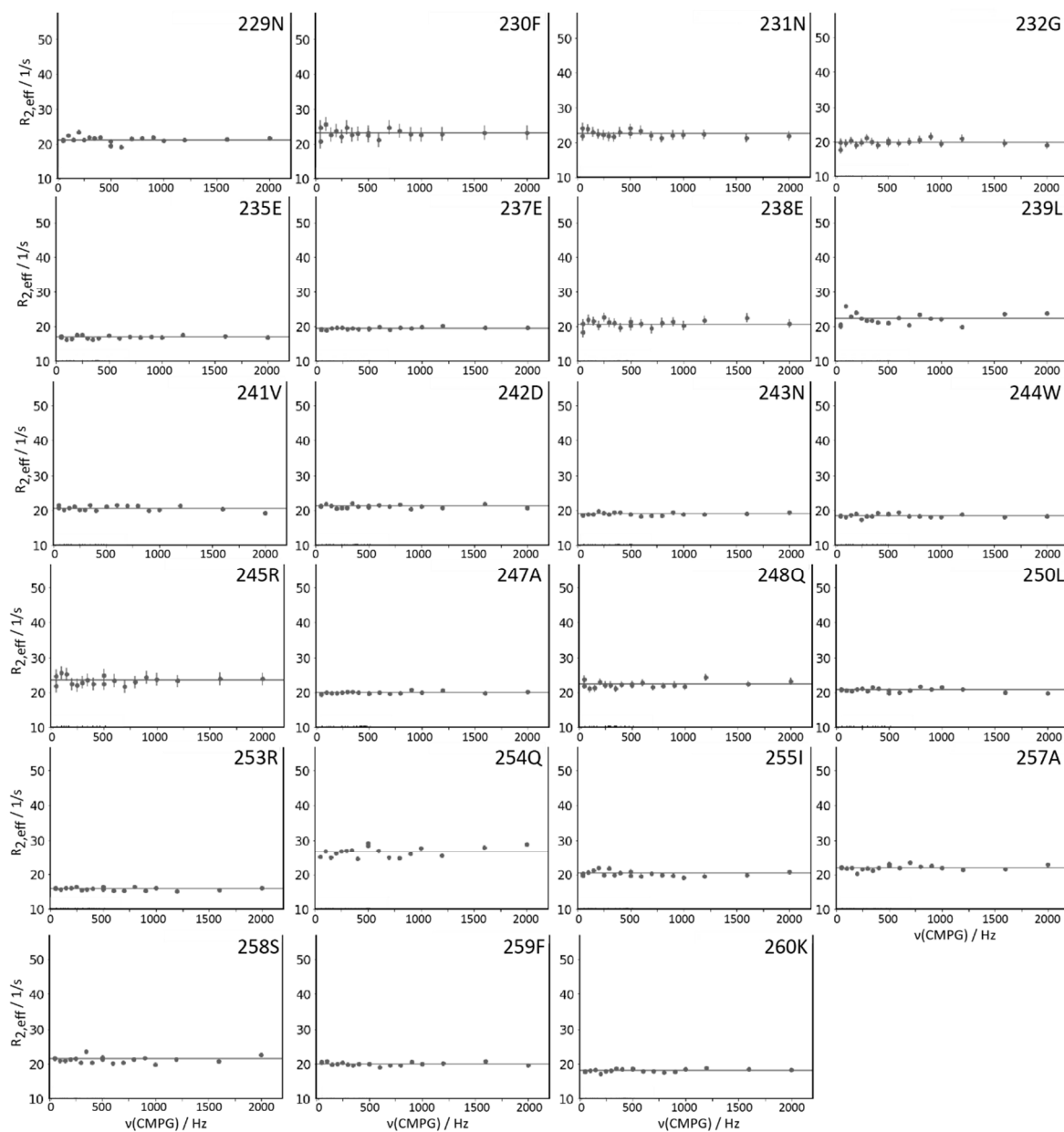


Figure 82: ^{15}N -CPMG relaxation dispersion plots for the assigned residues 229-260 of apo-hCAII.

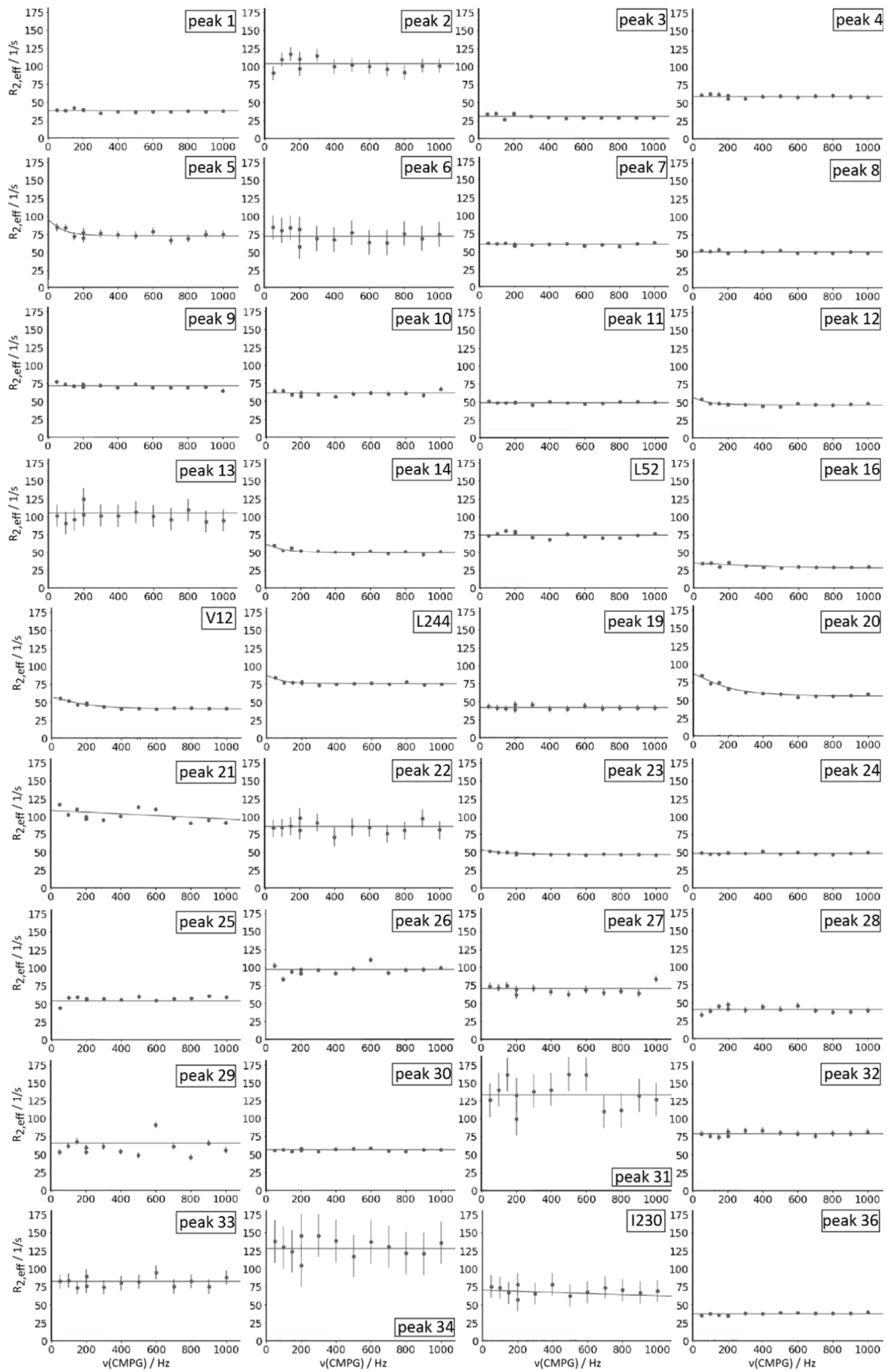


Figure 83: ^{13}C -CPMG profiles of apo-GlpG, residues 1-36.

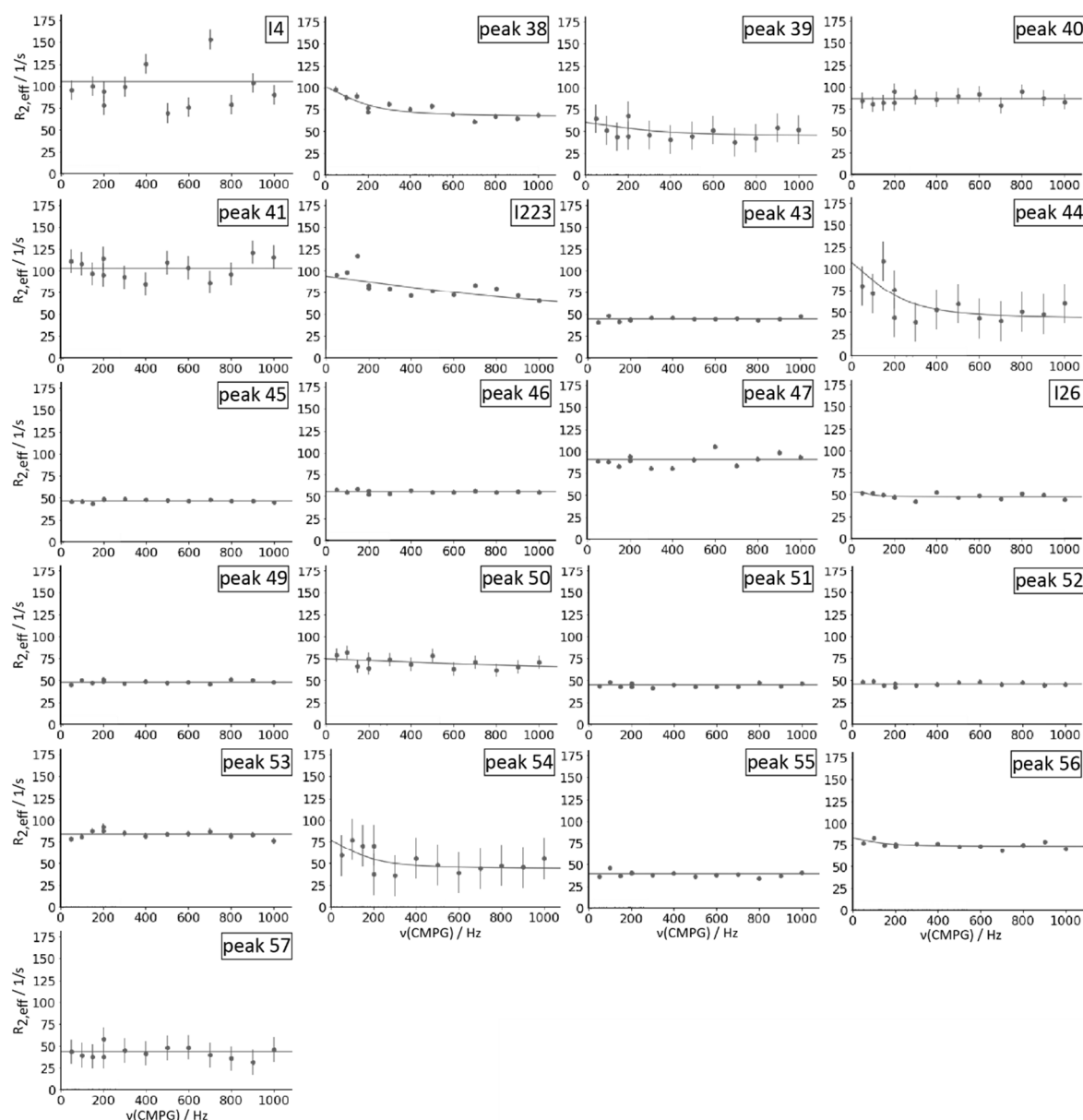


Figure 84: ^{13}C -CPMG profiles of apo-GlpG, residues 37-57.

Table 25: Parameters of the CPMG profiles for apo-GlpG shown in Figure 83 and Figure 84.

Residue	Model	Baseline R_2 and error / 1/s	k_{ex} / 1/s	χ^2 value
1	1	38.12 ± 0.04	-	145.47
2	1	103.64 ± 1.12	-	176.30
3	3	26.83 ± 0.20	15.43 ± 61.04	105.73
4	1	58.86 ± 0.37	-	34.41
5	3	72.42 ± 1.04	47.21 ± 154.59	63.57
6	1	72.51 ± 1.95	-	105.65
7	1	59.86 ± 0.18	-	46.49
8	3	50.00 ± 0.01	1668.71 ± 53.23	1387.29
9	3	64.78 ± 5.64	3000 ± 363.00	35.71
10	1	61.33 ± 0.41	-	69.57
11	1	48.73 ± 0.03	-	105.99

12	3	46.12 ± 0.10	87.93 ± 74.62	34.15
13	3	78.14 ± 30.36	95.30 ± 170.78	122.10
14	3	48.76 ± 0.15	968.24 ± 161.39	69.60
15/L52	1	74.56 ± 0.16	-	240.73
16	3	27.12 ± 0.04	19.18 ± 2.18	280.86
17/V12	3	40.24 ± 0.49	1155.26 ± 312.22	11.23
18/L244	2	75.61 ± 0.15	500 ± 69.84	106.57
19	1	42.32 ± 0.56	-	13.22
20	3	54.74 ± 0.06	1183.73 ± 26.12	535.30
21	3	70.40 ± 4.31	121.31 ± 361.41	501.38
22	1	85.46 ± 1.51	-	114.50
23	2	46.45 ± 0.33	785.61 ± 287.61	2.90
24	1	48.44 ± 0.00	-	486.38
25	1	54.17 ± 0.09	-	735.37
26	1	96.69 ± 0.44	-	262.72
27	1	70.93 ± 0.59	-	190.57
28	1	40.71 ± 0.51	-	48.01
29	1	65.11 ± 0.49	-	878.48
30	1	56.21 ± 0.18	-	36.27
31	1	133.95 ± 2.87	-	408.27
32	1	79.04 ± 0.46	-	48.70
33	1	82.49 ± 1.12	-	108.74
34	1	128.08 ± 3.32	-	122.18
35/I230	1	68.95 ± 1.68	-	62.94
36	1	37.31 ± 0.07	-	45.82
37/I4	3	5.00 ± 19.34	487.11 ± 277.24	561.34
38	3	66.87 ± 0.87	1182.43 ± 287.13	187.80
39	1	50.00 ± 0.00	-	59.07
40	1	86.74 ± 1.08	-	73.62
41	1	102.63 ± 1.57	-	237.87
42/I223	3	50.00 ± 0.60	610.28 ± 170.00	1205.51
43	1	44.38 ± 0.9	-	77.64
44	3	40.85 ± 6.12	160.48 ± 600.04	115.08
45	1	46.47 ± 0.02	-	140.46
46	1	55.69 ± 0.29	-	24.80
47	1	90.96 ± 0.37	-	441.85
48/I26	3	47.32 ± 0.08	17.94 ± 168.36	184.77
49	1	48.38 ± 0.16	-	36.56
50	3	68.50 ± 1.76	44.76 ± 137.97	94.89
51	1	44.74 ± 0.27	-	24.55
52	1	45.64 ± 0.35	-	16.33
53	1	84.29 ± 0.38	-	132.83
54	3	44.08 ± 6.80	83.39 ± 405.01	53.16
55	1	39.27 ± 0.05	-	277.41
56	3	73.24 ± 0.36	24.91 ± 102.78	111.81
57	1	43.70 ± 1.61	-	39.95

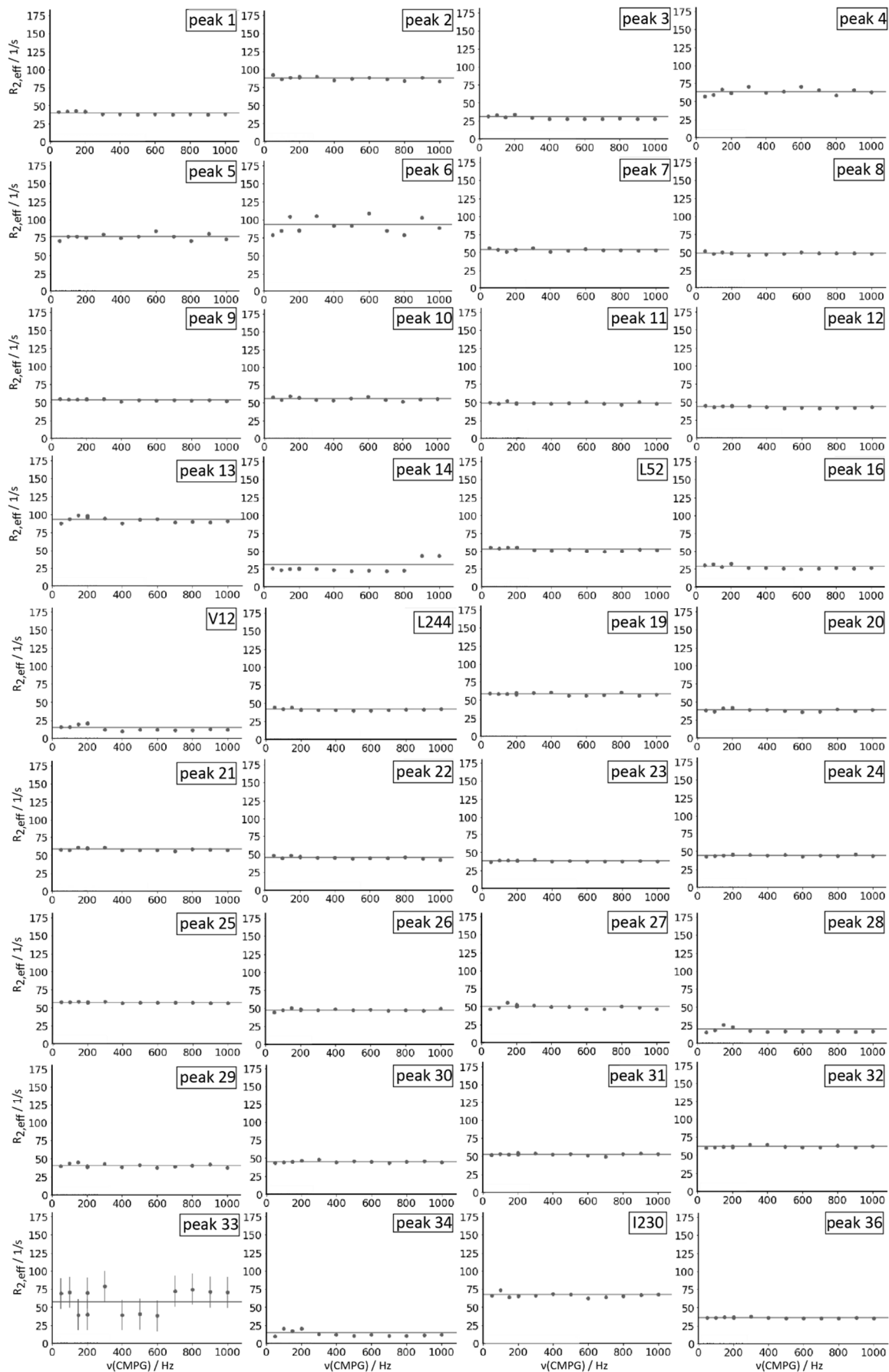


Figure 85: ^{13}C -CPMG profiles of GlpG in complex with inhibitor STS1775, residues 1-36.

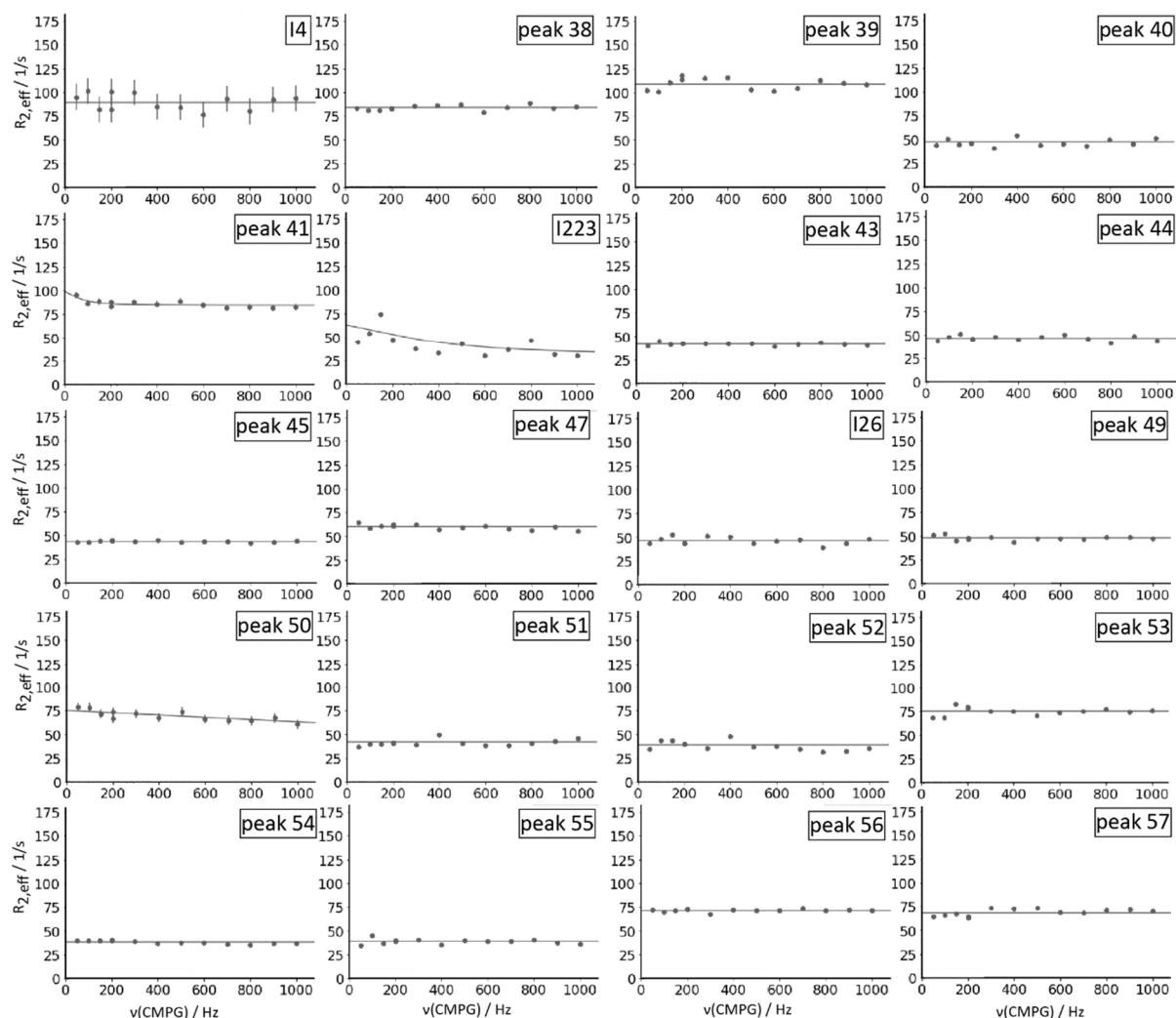


Figure 86: ^{13}C -CPMG profiles of GlpG in complex with inhibitor STS1775, residues 37-57. No relaxation dispersion curve was observed for peak 46 as negative $R_{2,\text{eff}}$ values were obtained.

Table 26: Parameters of the CPMG profiles for the GlpG-STS1775 complex shown in Figure 85 and Figure 86.

Residue	Model	Baseline R_2 and error/ 1/s	k_{ex} / 1/s	χ^2
1	3	37.50 ± 0.08	16.42 ± 48.01	14.65
2	3	62.41 ± 10.32	3000 ± 179.64	76.75
3	3	26.53 ± 0.05	19.13 ± 2.28	124.10
4	1	64.03 ± 0.07	-	736.17
5	1	76.59 ± 0.01	-	6745.23
6	1	93.26 ± 0.10	-	2829.80
7	1	53.75 ± 0.05	-	158.92
8	2	48.08 ± 0.02	500 ± 19.49	123.97
9	1	53.39 ± 0.07	-	51.43
10	1	55.72 ± 0.02	-	868.09
11	1	49.04 ± 0.01	-	208.85
12	1	43.24 ± 0.04	-	62.52
13	1	93.01 ± 0.21	-	195.63
14	1	30.47 ± 0.04	-	2488.38
15/L52	3	50.89 ± 0.09	15.29 ± 30.49	61.50

16	3	24.63 ± 0.04	19.21 ± 1.96	137.91
17/V12	3	9.48 ± 0.10	29.49 ± 34.14	179.39
18/L244	1	41.95 ± 0.10	-	35.53
19	1	58.53 ± 0.14	-	60.02
20	1	38.92 ± 0.06	-	87.39
21	1	58.33 ± 0.09	-	81.34
22	3	43.72 ± 0.20	10.00 ± 490.95	25.20
23	1	38.23 ± 0.06	-	14.52
24	1	44.33 ± 0.05	-	32.31
25	1	57.25 ± 0.08	-	16.94
26	1	47.37 ± 0.12	-	30.13
27	1	50.00 ± 0.00	-	64.18
28	1	19.30 ± 0.00	-	8567.28
29	1	40.87 ± 0.12	-	65.32
30	1	45.42 ± 0.04	-	63.19
31	1	52.28 ± 0.16	-	37.23
32	1	62.21 ± 0.13	-	45.26
33	1	57.24 ± 2.39	-	321.07
34	1	14.80 ± 0.01	-	3013.08
35/I230	1	67.32 ± 0.10	-	230.94
36	1	36.17 ± 0.02	-	21.59
37/I4	1	89.68 ± 1.58	-	128.06
38	1	83.87 ± 0.09	-	208.50
39	1	108.73 ± 0.32	-	304.97
40	1	47.02 ± 0.00	-	6990.30
41	3	84.64 ± 0.58	331.94 ± 297.65	42.76
42/I223	3	34.49 ± 0.13	84.34 ± 11.47	1566.49
43	1	41.93 ± 0.00	-	739.45
44	1	46.40 ± 0.04	-	237.35
45	1	43.83 ± 0.05	-	26.93
46	-	-	-	-
47	2	58.17 ± 0.27	1813.46 ± 378.46	77.31
48/I26	1	46.25 ± 0.02	-	1162.08
49	2	46.25 ± 0.14	520.67 ± 75.83	57.83
50	2	45.02 ± 3.72	10000 ± 1135.63	55.12
51	1	42.42 ± 0.06	-	323.77
52	1	38.92 ± 0.00	-	7397.12
53	1	74.83 ± 0.18	-	276.11
54	1	29.94 ± 0.04	-	8736.51
55	1	38.05 ± 0.05	-	79.74
56	1	39.13 ± 0.11	-	102.62
57	1	42.42 ± 0.06	-	323.77
58	1	71.04 ± 0.00	-	40629.05
59	1	68.75 ± 0.13	-	302.87

A Welch two-sample t-test, well-suited for unequal variances, was performed using the software R^[172] to assess whether the difference between the means of the baseline R_2 values was significant. Comparing the apo state and the GlpG-STS1775 complex, the hypothesis was that the baseline R_2 values decline when the inhibitor binds to GlpG. It was also proposed that the dynamics of GlpG core differ from the dynamics of the N-terminal domain.

Table 27: Welch two-sample t-test was performed for the baseline R_2 values of the CPMG experiments for apo state and STS1775-bound GlpG at full length with the software R^[172]. The differences were deemed significant for $\alpha < 0.01$. The raw data is shown in Table 25 and Table 26, while the corresponding boxplot is displayed in Figure 49, panel B. df: degrees of freedom.

	Apo state versus complex		N-terminus versus GlpG core	
	N-terminus	GlpG core	Apo state	Complex
t	0.84099	2.8151	0.62826	0.68649
df	19.387	73.305	11.992	11.073
p-value	0.4106	0.006258	0.5416	0.5065
	N-terminus, apo state	GlpG core, apo state	N-terminus, complex	GlpG core, complex
mean	70.28580 (n=11)	62.49581 (n=45)	57.30796 (n=11)	50.33955 (n=45)

5.7 Chemical shifts for hCAII and its ligand SBR as well as GlpG

Table 28: ^1H chemical shifts of the ligand SBR obtained via solution and solid-state NMR.

Solution NMR		Solid-state NMR	
SBR nuclei	Chemical shift / ppm	SBR nuclei	Chemical shift / ppm
2	7.84	2	8.08
3	7.48	3 and 11/14	7.72
5	3.30	-	-
6	2.53	-	-
7	0.94	-	-
8	4.14	-	-
9	7.29	-	-
10	6.43	10	6.84
11/14	7.44	-	-
12	7.01	12/13	7.14
13	7.16	-	-

Table 29: ^{13}C chemical shift differences between published solution backbone-to-sidechain assignments^[66a, 70] and the 3D XXH solution NMR experiment-based methyl assignments for apo-hCAII. The difference was taken to compare the deviations. Chemical shift deviations <0.5 ppm were coloured in green, while larger deviations were coloured in orange.

hCAII residues	^{13}C chemical shifts / ppm		Absolute values of $\Delta\delta(^{13}\text{C})$
	Backbone-based assignments ^[66a, 70]	Methyl NOESY assignments	
22	14.20	11.26	2.94
31	18.04	17.94	0.10
31	12.76	13.25	0.50
33	11.19	9.82	1.37
35	19.40	-	-
37	19.28	-	-
44	18.61	22.13	3.52
47	21.05	-	-
47	20.69	-	-
49	16.07	16.42	0.35
49	16.49	16.70	0.21
55	19.48	-	-
57	22.77	-	-
57	19.19	22.43	3.23
59	12.30	-	-
60	21.49	-	-
60	21.49	-	-
68	17.69	-	-
68	17.81	-	-
78	18.29	-	-
78	14.10	-	-
79	21.52	-	-

79	20.39	-	-
84	19.12	-	-
84	17.31	-	-
87	20.00	-	-
90	22.78	-	-
91	12.04	10.56	1.68
100	21.99	21.91	0.09
100	20.29	20.47	0.18
108	19.59	-	-
109	18.39	18.61	0.22
109	18.39	-	-
118	26.30	-	-
118	18.16	-	-
120	24.09	-	-
120	23.19	-	-
121	18.13	-	-
121	18.24	-	-
125	19.59	-	-
134	17.65	17.34	0.31
134	16.79	17.25	0.46
140	21.05	-	-
140	21.10	-	-
142	20.09	-	-
142	18.54	-	-
143	21.09	-	-
143	21.69	-	-
145	11.99	-	-
147	24.19	-	-
147	21.71	20.90	0.81
149	17.89	17.99	0.10
149	17.89	17.77	0.12
156	22.29	20.78	1.52
156	20.79	21.57	0.78
159	19.19	-	-
159	19.19	-	-
160	18.89	-	-
160	18.89	-	-
162	17.89	-	-
162	17.89	18.31	0.41
163	16.99	17.61	0.62
163	16.99	16.85	0.14
166	15.18	-	-
183	22.19	20.78	1.42
183	19.19	(or L156)	-
188	22.19	-	-

188	19.59	-	-
203	21.13	-	-
203	17.98	-	-
206	19.99	20.06	0.07
206	17.09	17.17	0.07
209	12.25	-	-
210	18.49	-	-
210	16.12	-	-
211	19.79	-	-
211	20.09	-	-
215	10.55	9.52	1.03
217	18.19	20.90	2.71
217	16.81	15.88	0.93
222	18.62	17.17	1.45
222	16.73	18.46	1.73
223	21.19	18.46	2.74
223	21.19	-	-
228	23.79	-	-
228	22.19	-	-
239	21.79	-	-
239	18.89	-	-
241	20.29	-	-
241	15.06	-	-
250	21.89	21.89	0.00
250	19.79	19.78	0.02
255	10.32	10.99	0.67

Table 30: Amide chemical shifts of hCAII in solution and the solid state. sc: sidechain.

Residues		Solution NMR		Solid-state NMR	
Numbers	Type	¹ H chemical shifts / ppm	¹⁵ N chemical shifts / ppm	¹ H chemical shifts / ppm	¹⁵ N chemical shifts / ppm
5	W	7.61	120.80	7.06	120.42
5	W _{sc}	10.24	128.28	9.76	127.67
6	G	8.05	109.27	7.54	109.72
7	Y	8.66	119.85	8.28	119.31
8	G	8.28	110.60	7.84	110.29
9	K	-	-	8.48	119.62
12	G	7.44	108.44	6.98	108.98
14	E	-	-	8.39	121.07
15	H	8.72	115.51	8.55	114.38
16	W	7.39	121.99	7.14	121.46
16	W _{sc}	9.87	130.38	9.33	129.20
17	H	8.86	114.41	8.59	114.95
18	K	6.93	119.82	6.81	119.20
19	D	7.21	114.60	6.71	114.12
20	F	8.01	116.68	7.51	116.58
22	I	7.98	120.74	7.69	120.79

23	A	8.70	123.04	8.36	122.84
24	K	7.19	114.40	6.80	114.05
25	G	8.31	108.76	7.87	108.22
26	E	9.31	117.58	8.82	117.12
27	R	8.84	120.76	8.34	120.93
28	Q	7.68	115.38	7.34	115.76
29	S	8.06	118.37	7.66	117.50
31	V	6.31	106.95	5.82	106.43
32	D	-	-	7.94	118.22
33	I	8.72	128.26	8.27	127.74
34	D	7.42	128.33	7.10	127.59
35	T	10.29	122.20	9.70	122.07
36	H	-	-	7.90	119.63
37	T	7.63	108.17	7.08	108.51
38	A	7.47	127.32	6.59	124.12
39	K	8.18	122.61	7.90	121.87
40	Y	8.68	126.73	8.55	128.14
41	D	7.72	128.75	7.15	128.29
43	S	8.51	115.09	7.92	114.06
44	L	7.05	123.03	6.56	122.29
45	K	7.51	122.88	7.31	123.41
47	L	8.94	125.51	8.49	124.86
48	S	8.48	119.70	8.00	118.30
49	V	8.35	127.15	8.14	125.35
50	S	8.39	122.81	7.98	121.91
51	Y	9.04	123.59	8.63	123.83
52	D	8.87	121.36	8.38	121.35
53	Q	8.03	114.30	8.26	114.96
53	Q _{sc}	6.81	112.28	-	-
53	Q _{sc}	7.56	112.05	-	-
54	A	7.43	120.18	6.79	118.89
55	T	9.39	120.49	8.81	120.15
56	S	9.93	126.23	9.30	126.26
57	L	8.92	117.67	8.37	116.77
58	R	7.15	114.70	6.68	116.80
59	I	9.04	122.14	8.79	122.59
60	L	8.43	123.56	8.58	123.49
61	N	8.40	120.83	7.99	122.03
61	N _{sc}	6.98	107.57	-	-
61	N _{sc}	6.82	107.58	-	-
62	N	8.04	123.21	7.57	121.92
63	G	9.64	108.82	9.24	108.22
64	H	8.33	116.68	7.94	116.43
65	A	8.59	123.25	8.13	123.07
66	F	7.17	109.91	6.76	111.20
67	N	8.87	120.79	8.62	120.84
67	N _{sc}	7.44	105.36	-	-
67	N _{sc}	5.87	105.39	-	-
68	V	8.64	121.78	8.43	121.46
69	E	8.50	124.03	8.17	122.71
70	F	8.71	118.62	8.07	117.79
71	D	8.50	115.17	7.99	117.93
72	D	8.99	131.48	-	-
73	S	8.76	115.36	8.28	115.07

74	Q	7.69	117.65	7.36	117.09
74	Q _{sc}	6.83	111.78	-	-
74	Q _{sc}	7.56	111.79	-	-
75	D	-	-	7.70	120.35
76	K	7.97	123.95	7.45	123.60
77	A	8.22	123.43	8.18	123.15
78	V	8.41	115.94	8.36	115.15
79	L	9.04	123.29	8.72	124.10
80	K	8.39	120.75	8.08	121.07
81	G	8.98	106.43	8.55	105.94
82	G	7.09	108.56	6.57	108.42
84	L	7.81	120.11	7.34	119.67
85	D	8.90	124.44	8.65	123.79
86	G	7.83	109.26	7.25	108.12
87	T	8.42	117.02	8.07	116.58
88	Y	8.51	126.13	8.08	125.42
89	R	8.61	122.76	7.95	121.98
90	L	-	-	8.14	123.16
91	I	8.91	122.91	8.36	123.27
92	Q	7.34	115.29	6.69	114.84
93	F	8.62	113.48	8.16	113.83
94	H	-	-	7.92	111.72
95	F	8.93	116.91	8.79	116.68
96	H	8.80	115.24	8.26	115.21
97	W	9.56	119.16	9.28	119.03
97	W _{sc}	10.22	130.54	9.77	130.25
98	G	-	-	7.42	107.40
99	S	8.46	111.95	7.79	110.45
100	L	7.39	119.02	6.72	118.62
101	D	8.94	120.28	8.64	122.41
102	G	7.66	102.72	6.91	103.82
103	Q	7.86	114.85	7.46	113.52
103	Q _{sc}	7.40	116.04	-	-
103	Q _{sc}	8.85	116.01	-	-
104	G	8.24	106.33	7.84	106.08
105	S	7.34	107.62	6.92	107.30
106	E	8.25	118.96	8.08	119.64
107	H	10.72	117.71	10.28	117.91
108	T	7.38	109.19	6.95	109.03
109	V	7.66	118.47	7.26	118.22
110	D	9.69	132.07	9.23	131.67
111	K	9.84	109.68	9.42	109.48
112	K	8.26	124.35	7.73	124.94
113	K	7.91	121.80	7.63	122.65
114	Y	8.16	121.10	7.75	120.25
115	A	7.62	120.92	6.97	120.91
116	A	8.17	112.12	7.77	112.00
117	E	9.45	121.05	9.05	121.02
118	L	9.75	130.33	9.58	128.60
119	H	8.96	125.39	8.29	122.82
120	L	9.00	123.60	8.55	125.68
121	V	9.20	126.52	8.67	127.93
122	H	8.53	124.25	8.24	123.45
123	W	9.10	119.19	9.02	118.64

123	W _{sc}	12.20	133.33	11.68	132.70
124	N	9.03	119.69	8.32	119.03
124	N _{sc}	7.06	110.02	-	-
124	N _{sc}	6.60	110.10	-	-
125	T	8.18	117.16	7.92	117.46
126	K	7.96	122.83	7.58	122.78
127	Y	7.62	115.37	7.08	114.44
128	G	7.83	107.30	8.64	103.94
129	D	7.52	116.51	6.96	115.38
130	F	7.53	119.89	7.00	120.08
131	G	8.07	103.66	7.81	104.24
132	K	7.70	119.48	7.37	119.54
133	A	8.05	124.32	7.56	122.55
134	V	7.40	108.07	6.93	106.84
135	Q	6.64	113.97	6.22	114.12
135	Q _{sc}	7.32	111.74	-	-
136	Q	7.58	118.22	6.98	118.30
136	Q _{sc}	6.60	110.22	-	-
136	Q _{sc}	7.06	110.26	-	-
138	D	7.91	114.91	7.09	115.19
139	G	7.96	107.59	7.14	106.18
140	L	8.86	115.83	8.43	116.43
141	A	8.88	122.92	7.96	125.55
142	V	7.24	124.16	6.59	124.16
143	L	8.49	128.16	-	-
144	G	9.92	114.12	9.55	113.04
145	I	9.17	124.24	8.67	126.40
146	F	10.09	128.81	9.90	128.17
147	L	8.25	120.29	7.96	119.42
148	K	8.98	119.62	8.61	119.40
149	V	8.43	121.40	8.07	122.04
150	G	9.11	117.91	8.90	118.26
151	S	8.20	123.02	8.29	126.67
152	A	8.35	121.24	7.96	119.32
153	K	8.71	123.84	8.19	122.32
155	G	8.69	104.69	8.30	104.35
156	L	7.54	116.73	6.89	115.96
157	Q	7.83	122.18	7.42	122.35
157	Q _{sc}	6.76	115.97	-	-
157	Q _{sc}	8.15	115.91	-	-
158	K	8.83	115.83	8.21	116.04
159	V	7.05	113.18	6.76	114.05
160	V	7.35	114.49	6.83	113.87
161	D	8.38	117.60	7.89	117.18
162	V	7.16	113.85	6.64	113.49
163	L	7.09	121.51	6.59	120.35
164	D	8.18	115.90	7.70	115.56
165	S	8.01	113.40	7.62	113.29
166	I	7.48	116.33	6.95	115.74
167	K	7.31	117.24	6.98	117.37
168	T	6.62	98.77	6.12	97.95
169	K	7.48	120.52	6.96	120.38
170	G	8.86	116.52	8.73	116.17
171	K	7.77	120.18	7.42	119.84

172	S	8.30	113.52	7.97	114.04
173	A	9.01	123.20	8.57	121.42
174	D	8.61	121.88	8.40	121.19
175	F	7.99	125.22	7.48	125.55
176	T	8.08	114.82	7.55	113.29
177	N	9.67	115.55	9.68	113.38
177	N _{sc}	7.55	112.26	-	-
177	N _{sc}	6.85	112.49	-	-
178	F	8.25	118.82	7.95	118.63
179	D	7.95	128.04	7.79	128.60
181	R	8.34	118.31	8.04	117.82
182	G	7.33	102.65	6.75	102.92
183	L	7.28	117.28	6.75	116.29
184	L	6.66	113.08	5.91	111.38
186	E	8.17	116.30	7.89	119.17
187	S	7.64	111.29	7.13	108.47
188	L	8.61	125.42	8.55	123.92
189	D	7.38	121.72	7.37	122.37
190	Y	8.65	119.65	8.11	117.74
191	W	9.55	117.12	8.89	116.48
191	W _{sc}	10.48	128.79	9.93	128.23
192	T	9.70	115.10	9.27	114.68
193	Y	8.09	125.55	7.61	124.81
195	G	9.41	110.60	8.91	110.44
196	S	8.72	120.54	8.46	121.09
197	L	8.09	118.17	7.57	117.70
198	T	-	-	8.22	110.47
199	T	6.99	108.21	6.59	108.31
202	L	9.21	120.40	8.58	119.82
203	L	5.98	110.83	5.29	110.36
204	E	8.75	124.61	8.39	124.62
205	C	7.65	114.61	7.13	112.95
206	V	7.28	118.97	6.72	118.99
207	T	8.46	125.03	7.90	124.43
208	W	8.26	129.37	7.68	128.00
208	W _{sc}	10.11	130.51	9.56	129.53
209	I	8.79	127.10	8.08	124.52
210	V	9.72	128.42	9.71	127.63
211	L	8.97	126.76	8.45	125.24
212	K	7.39	122.24	6.97	121.14
213	E	8.87	121.57	8.60	123.72
215	I	8.91	113.20	8.45	114.69
216	S	8.07	116.62	7.85	119.60
217	V	8.54	115.70	8.18	117.65
218	S	8.46	116.37	7.94	114.33
219	S	9.28	117.31	8.88	116.66
220	E	-	-	7.91	117.95
221	E	7.54	117.79	7.06	117.22
221	Q _{sc}	7.44	109.83	-	-
221	Q _{sc}	6.86	109.77	-	-
222	V	7.37	114.56	6.77	114.15
223	L	7.95	120.64	7.78	120.93
224	K	6.93	115.00	6.43	114.83
225	F	7.14	118.58	6.60	118.98

226	R	6.70	109.04	6.25	109.14
227	K	6.87	113.95	6.20	113.52
228	L	7.13	119.35	6.56	118.98
229	N	8.89	118.37	8.40	118.31
229	N _{sc}	6.46	108.03	-	-
229	N _{sc}	7.63	107.87	-	-
230	F	8.85	117.28	8.30	116.30
231	N	6.94	116.24	7.70	109.85
231	N _{sc}	8.17	110.61	-	-
232	G	8.69	104.69	8.47	104.27
233	E	8.34	121.09	7.93	121.75
234	G	-	-	8.40	112.49
235	E	7.35	119.94	6.85	119.06
237	E	8.21	123.20	7.93	122.67
238	E	9.01	129.76	8.55	127.94
239	L	8.65	126.01	8.34	125.79
240	M	8.58	122.00	8.24	122.86
241	V	6.74	114.99	6.21	114.38
242	D	7.18	114.63	6.20	113.57
243	N	8.33	119.06	8.01	119.15
244	W	6.49	114.02	5.93	113.73
244	W _{sc}	10.83	126.85	-	-
245	R	10.48	126.81	10.42	126.26
247	A	8.02	120.54	6.72	118.81
248	Q	8.71	121.84	8.20	121.50
250	L	8.67	125.75	8.41	125.47
251	K	8.87	114.53	8.49	114.46
252	N	6.95	113.47	6.25	118.41
252	N _{sc}	9.07	119.14	-	-
253	R	7.33	118.45	6.93	118.16
254	Q	8.53	118.16	8.11	117.49
254	Q _{sc}	7.62	112.77	-	-
254	Q _{sc}	6.89	112.71	-	-
255	I	8.99	131.30	8.54	130.25
256	K	8.78	126.38	8.38	125.80
257	A	8.45	123.74	8.25	123.03
258	S	8.94	116.22	8.36	115.29
259	F	6.70	117.97	6.22	117.80
260	K	7.70	124.19	7.47	123.49

Table 31: Amide chemical shifts of full-length GlpG wild type in solution transferred from published assignments^[142, 164].

Residues		Chemical shifts of the nuclei indicated below / ppm				
Numbers	Type	¹ H	¹⁵ N	¹³ C _α	¹³ C'	¹³ C _β
2	L	9.30	122.10	-	176.94	40.48
3	M	8.80	127.77	-	-	25.09
4	I	9.05	126.25	-	172.97	31.89
5	T	7.32	109.51	-	176.77	-
6	S	8.04	114.38	-	175.01	58.91
7	F	9.11	120.54	57.95	-	37.06
8	A	9.33	122.81	-	173.67	12.80

9	N	8.45	114.54	53.22	-	33.73
11	R	8.27	117.69	55.41	173.91	23.77
12	V	7.45	121.36	51.95	172.09	25.73
13	A	7.10	118.73	-	-	12.68
14	Q	7.95	114.98	-	-	-
15	A	7.94	120.49	53.28	173.44	49.20
16	F	8.35	119.44	51.99	-	32.67
17	V	8.40	119.69	-	175.16	25.27
18	D	9.39	121.18	-	172.91	33.73
19	Y	8.34	120.87	57.44	174.90	32.03
20	M	8.56	119.04	55.06	173.85	22.26
21	A	8.43	122.24	-	174.41	11.23
22	T	7.66	111.27	62.01	-	66.43
23	Q	7.38	117.43	50.96	176.24	22.84
24	G	7.69	106.05	42.36	177.54	-
25	V	7.82	120.04	48.59	-	27.66
26	I	9.08	128.02	65.00	172.79	30.87
27	L	8.50	129.36	50.85	177.75	38.79
29	I	8.88	120.95	59.16	171.80	34.59
30	Q	9.95	126.04	-	174.52	24.43
32	H	8.63	125.42	-	177.93	23.03
33	N	7.70	112.85	55.85	173.44	-
34	Q	8.00	115.48	53.05	171.39	-
35	S	8.73	115.38	-	-	57.59
36	D	8.80	126.38	-	-	35.86
37	V	9.06	121.18	-	171.68	27.30
38	W	9.68	127.74	53.45	171.85	26.25
39	L	9.61	121.68	-	170.85	40.94
40	A	9.03	127.46	-	175.65	13.01
41	D	7.88	116.11	48.34	172.96	-
42	E	8.43	122.49	55.62	175.94	-
43	S	8.29	115.63	-	173.96	-
44	Q	7.87	117.37	54.40	173.50	22.20
45	A	7.79	121.59	-	176.54	12.71
46	E	8.81	116.42	56.36	176.77	22.68
47	R	8.11	120.98	61.43	176.15	-
48	V	8.55	117.41	56.45	175.07	-
49	R	8.67	119.05	54.48	172.73	-
50	A	8.27	122.73	-	176.43	11.84
51	Q	8.40	118.18	55.11	-	-
52	L	8.98	120.88	52.34	175.07	-
53	A	7.55	117.14	55.28	177.94	11.58
54	R	7.34	116.44	57.55	176.26	-
55	F	8.79	120.54	56.95	174.50	-
56	L	8.20	115.50	57.96	175.94	-
57	E	7.63	117.45	54.35	177.74	-

58	N	7.90	113.47	-	-	32.83
60	A	8.04	121.44	44.61	175.59	-
61	D	8.10	117.64	50.87	176.36	34.89
63	R	8.15	118.72	63.77	171.36	27.96
96	V	10.91	121.83	64.43	-	25.97
107	V	8.32	118.74	63.71	175.07	26.58
108	F	8.98	119.68	55.17	176.89	36.28
113	I	7.91	119.06	51.56	175.68	-
114	L	8.97	116.68	56.06	175.59	55.77
115	G	7.27	108.27	41.50	176.72	41.45
116	D	7.96	117.92	55.10	-	35.30
117	Q	8.59	115.87	55.44	-	-
118	E	8.32	118.85	57.46	176.77	26.53
120	M	8.54	119.51	55.02	173.96	54.98
121	L	7.87	116.47	53.93	173.42	-
122	W	7.43	117.96	54.87	172.57	54.91
124	A	8.86	118.84	50.08	-	55.66
125	W	8.27	119.65	52.28	174.50	-
127	A	6.14	118.64	51.77	174.17	-
128	D	7.16	121.77	51.77	175.89	41.05
130	T	8.73	109.09	61.25	173.14	66.36
131	L	7.93	120.13	51.90	175.08	38.69
134	E	7.50	119.48	51.98	-	19.23
135	F	7.59	116.26	53.45	172.91	-
137	R	7.25	121.48	48.95	176.20	25.41
140	T	9.50	106.56	-	-	64.61
141	H	7.21	121.00	55.08	173.55	25.50
142	A	6.85	122.57	49.82	169.29	-
145	H	6.49	112.75	53.56	172.81	30.28
147	S	8.13	111.77	54.37	171.63	62.81
148	L	9.93	124.19	54.64	174.13	36.54
150	H	7.89	117.73	55.08	-	29.18
163	G	6.95	106.04	43.91	-	-
164	A	6.93	121.68	52.14	175.30	12.63
167	K	8.45	116.79	54.42	175.20	27.08
168	R	7.85	113.91	53.85	-	-
169	L	8.87	116.70	52.78	174.56	-
170	G	7.64	109.80	42.16	-	-
171	S	8.78	116.42	42.16	-	-
172	G	9.24	109.27	44.38	-	-
173	K	8.15	120.95	53.96	175.07	23.83
174	L	6.95	115.80	55.58	177.59	38.27
175	I	8.78	121.21	63.26	-	31.33
189	Q	8.05	119.60	55.85	174.81	55.62
190	Q	7.72	115.21	55.97	-	25.09
191	K	7.34	120.67	54.65	-	26.07

192	F	7.47	113.46	56.92	-	-
193	S	8.42	110.34	42.55	171.21	54.56
194	G	8.24	111.22	41.92	172.48	-
196	W	8.07	121.66	49.70	172.87	22.31
197	F	7.26	121.58	52.38	-	52.18
198	G	8.63	103.43	44.76	-	35.30
200	L	8.72	119.28	53.90	-	36.69
201	S	8.44	114.50	53.23	-	61.27
203	V	8.21	121.96	-	-	27.07
213	L	8.25	117.58	54.51	174.95	-
214	R	8.75	118.67	57.24	-	26.27
216	E	7.34	116.25	54.57	175.20	25.06
217	R	8.22	114.87	-	178.58	-
218	D	8.33	117.96	55.17	179.79	38.49
220	Q	7.96	116.14	55.34	-	35.84
221	S	7.96	114.49	61.49	174.27	-
223	I	7.84	119.73	55.87	-	-
224	Y	7.71	119.71	48.98	-	51.31
225	L	9.21	124.16	-	174.91	52.23
229	L	6.86	115.93	53.46	172.79	-
230	I	8.07	121.81	55.27	172.97	31.93
231	I	7.09	115.16	63.18	-	32.31
233	A	9.45	126.46	53.31	-	11.97
239	A	8.76	118.50	55.52	173.4	-
240	G	8.07	104.51	44.06	-	-
241	W	8.52	124.96	-	171.15	23.90
250	A	10.82	129.76	-	175.38	12.91
251	N	8.66	121.08	52.46	-	31.44
253	A	7.18	121.66	-	-	13.98
254	H	7.30	117.73	-	-	25.23
255	I	8.41	118.31	63.85	172.97	-
256	A	8.11	120.78	-	-	13.01
257	G	7.74	104.92	44.85	170.80	-
258	L	8.72	122.94	-	171.80	-

Table 32: Methyl chemical shifts of full-length GlpG wild type in solution derived from mutations (GlpG core) or transferred from published assignments (N-terminal domain)^[142].

Residue number	Residue type	Assignment certainty	¹ H _G or ¹ H _D chemical shifts / ppm	¹³ C _G or ¹³ C _D chemical shifts / ppm
223	I	Higher	0.73	7.78
230	I	Higher	0.00	10.53
244	L	Higher	0.66	21.20
2	L	Low	0.69	24.99
4	I	Low	0.58	9.28
12	V	Low	0.94	22.04
26	I	Low	0.77	9.99

29	I	Low	0.75	13.47
52	L	Low	0.45	22.47
89 or 200	L	Low	0.90	23.44
89 or 229	L	Low	0.90	21.67
89 or 203	L/V	Low	0.82	21.84
119	V	Low	0.82	19.02
200	L	Low	0.97	22.96
229	L	Low	-0.07	22.33
131	L	Low	0.78	22.85

5.8 FLYA assignments

The data were kindly provided by Dr. Alexander Klein.

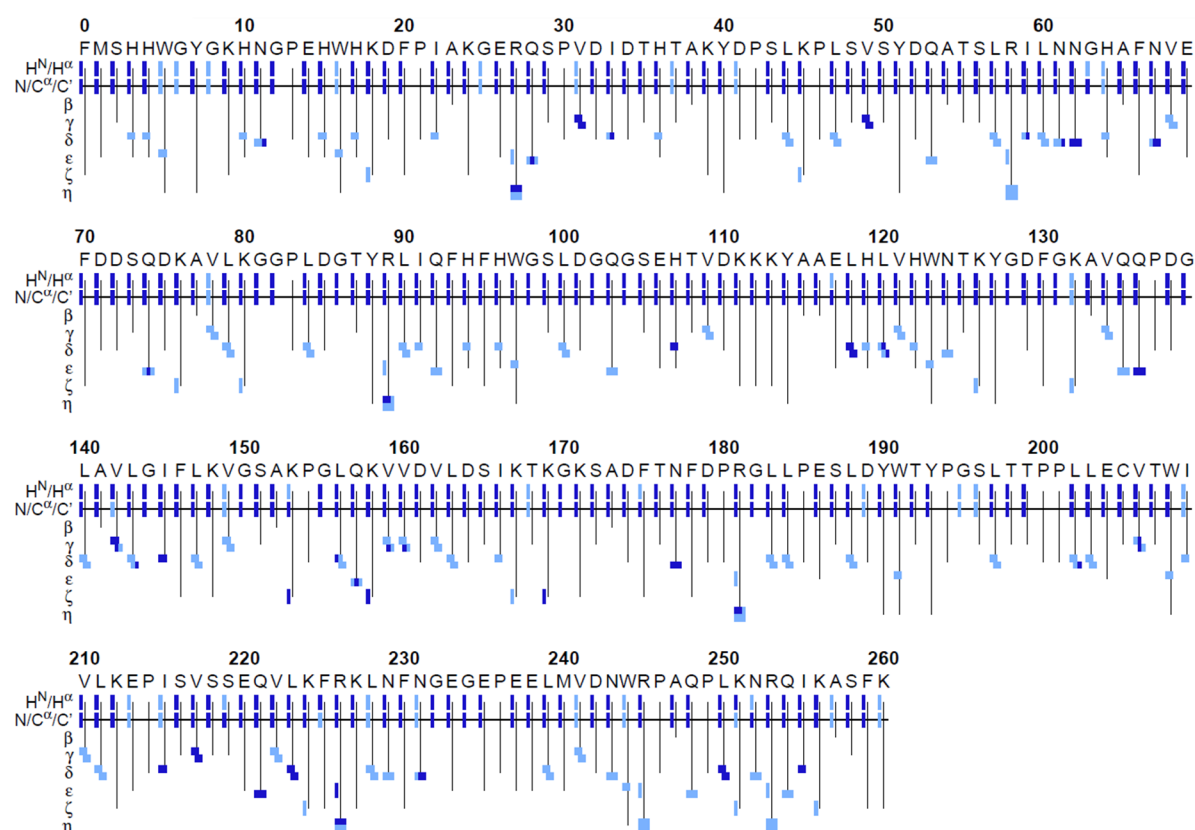


Figure 87: FLYA assignment results for hCAII methyl sidechains and amide groups based on 3D time-shared XXH experiments in solution. Here, for example, C^α corresponds to C_A. Dark blue represents less ambiguous assignments than light blue.

Table 33: FLYA assignments for hCAII methyl sidechains corresponding to Figure 87. For leucine and valine, two methyl assignments are available.

Unambiguous assignments			Ambiguous assignments		
Resi- dues	¹ H chemical shifts / ppm	¹³ C chemical shifts / ppm	Resi- dues	¹ H chemical shifts / ppm	¹³ C chemical shifts / ppm
31	-0.20	13.21	22	0.84	11.21
31	0.51	17.93	33	0.48	19.74
33	-	9.87	44	0.62	22.02

49	-0.25	16.42	44	0.71	19.70
49	0.33	16.60	47	0.63	21.79
59	-	9.94	47	-0.19	-
118	-0.27	18.06	57	0.46	18.46
118	0.53	25.07	57	0.64	21.37
120	0.37	19.20	59	-0.30	18.90
142	0.67	15.17	60	0.70	23.17
142	0.75	18.30	60	0.65	16.20
145	0.68	11.70	68	0.60	20.26
156	0.98	-	68	0.26	17.69
159	-	18.60	78	0.64	17.86
160	-	18.44	78	0.80	-
202	0.66	-	79	0.23	20.47
206	-	18.43	79	-0.20	21.71
206	-	18.63	84	0.78	19.03
215	0.78	9.47	84	0.69	23.94
217	0.45	15.87	90	0.72	18.81
217	0.63	18.35	90	0.86	20.37
223	0.76	21.97	91	0.50	11.31
223	0.93	24.02	100	0.60	19.61
250	0.57	19.71	100	0.76	21.97
250	0.75	21.95	109	0.36	16.22
255	0.44	10.98	109	0.60	18.42
			120	0.74	22.20
			121	0.77	14.08
			121	1.04	20.37
			134	0.57	15.93
			134	0.30	17.36
			140	0.73	18.57
			140	0.70	21.80
			142	0.58	19.21
			143	0.60	23.46
			147	0.97	21.04
			147	0.68	24.09
			149	0.41	16.29
			149	0.62	18.39
			156	0.79	20.86
			159	-0.67	24.04
			159	0.81	17.19
			160	0.81	18.29
			160	0.68	-
			162	0.26	18.57
			162	0.25	20.27
			163	0.001	17.09
			163	-0.20	20.09
			166	0.14	10.79

183	0.97	20.84
183	0.08	22.13
184	0.60	19.21
184	0.37	22.17
188	0.60	20.56
188	0.84	21.76
197	0.75	18.53
197	0.66	21.79
202	0.44	18.39
202	-	23.27
203	0.73	18.54
203	0.83	21.46
206	0.65	-
206	0.13	-
209	0.49	10.70
210	0.31	17.32
210	0.65	18.37
211	0.06	18.78
211	-0.88	20.15
222	0.82	17.34
222	0.31	18.40
228	-0.68	17.01
228	0.72	19.92
239	0.74	18.73
239	0.71	21.92
241	-0.66	17.17
241	1.03	18.55

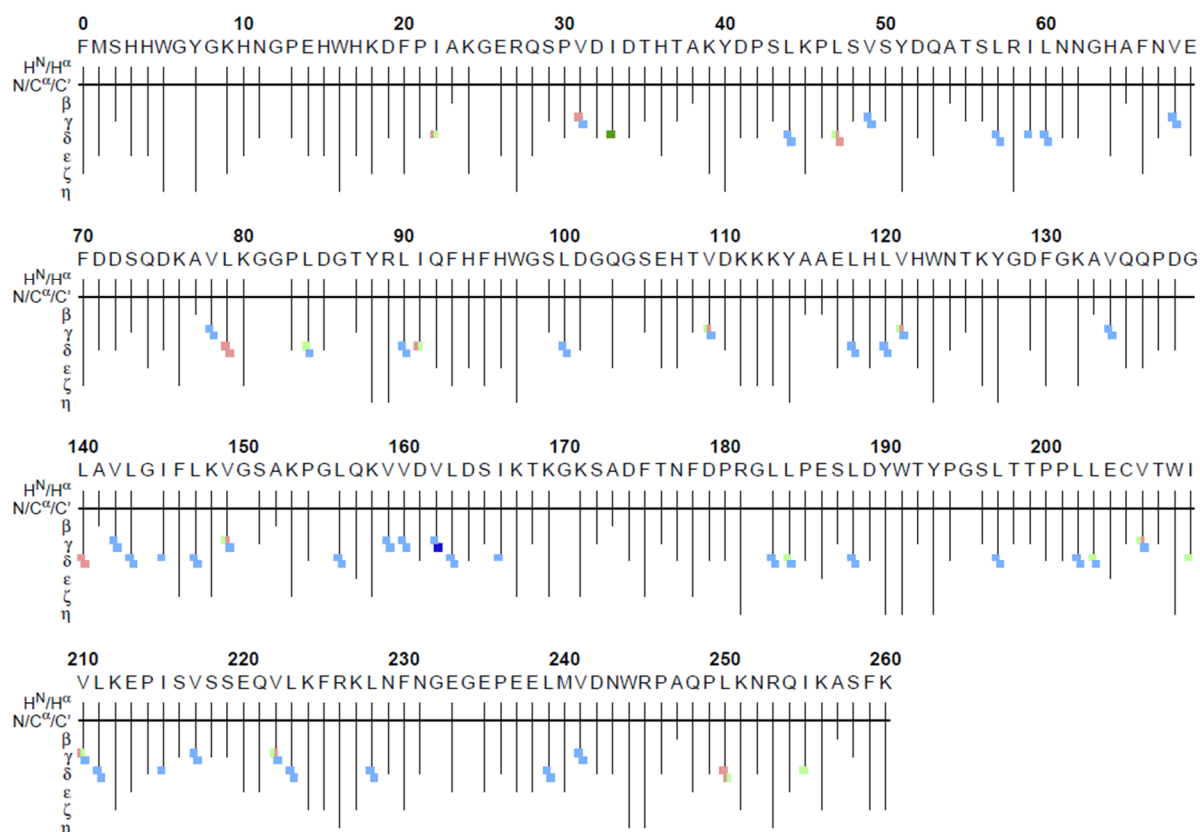


Figure 88: FLYA assignments for hCAII methyl sidechain groups based on 3D time-shared XXH experiments in the solid state. Here, for example, C^α corresponds to C_A. A color code indicates unambiguous (dark blue) and ambiguous assignments (red) with different shades inbetween.

Table 34: FLYA assignments for hCAII methyl sidechains corresponding to Figure 88. For leucine and valine, bearing two methyl groups, two methyl assignments are available.

Unambiguous assignments			Ambiguous assignments		
Resi- dues	¹ H chemical shifts /ppm	¹³ C chemical shifts /ppm	Resi- dues	¹ H chemical shifts /ppm	¹³ C chemical shifts /ppm
33	0.03	13.05	22	0.32	14.71
162	0.35	22.23	31	0.09	21.81
			31	-0.08	19.43
			44	0.30	22.32
			44	0.24	25.40
			47	0.30	25.30
			47	0.27	25.20
			49	-0.09	19.88
			49	-0.67	20.06
			57	0.29	22.22
			57	0.24	25.63
			59	0.37	14.76
			60	0.31	22.31
			60	0.28	25.43
			68	-0.08	21.74

68	0.31	22.02
78	0.25	18.62
78	0.12	21.65
79	0.27	25.25
79	0.30	25.11
84	0.29	25.13
84	-0.02	23.40
90	0.28	23.40
90	0.31	26.73
91	0.33	13.20
100	0.28	22.73
100	0.30	26.97
109	0.19	21.81
109	-0.11	20.63
118	0.33	23.46
118	0.30	25.24
120	0.31	25.30
120	0.27	25.24
121	0.30	21.95
121	0.09	18.75
134	0.31	20.89
134	0.34	22.19
140	0.24	22.08
140	0.28	25.36
142	0.25	18.55
142	0.12	21.85
143	0.27	25.20
143	0.30	25.30
145	-0.29	13.98
147	0.29	23.52
147	0.32	26.80
149	0.61	23.73
149	0.34	17.21
156	0.30	22.13
156	0.27	25.39
159	0.32	21.79
159	0.34	22.15
160	0.32	21.01
160	0.32	21.95
162	0.31	20.93
163	0.30	22.34
163	0.25	25.48
166	0.33	13.23
183	0.30	25.30
183	0.28	25.25
184	0.31	27.05

184	0.28	25.29
188	0.30	23.46
188	0.26	25.26
197	0.29	22.29
197	0.25	25.27
202	0.33	23.79
202	0.30	27.11
203	0.29	22.05
203	0.24	25.31
206	-0.12	20.73
206	0.29	21.76
209	-0.30	14.03
210	0.10	21.92
210	-0.09	19.54
211	0.27	23.34
211	0.31	26.69
215	0.02	13.53
217	-0.05	16.39
217	0.08	21.92
222	0.34	23.77
222	0.62	17.21
223	0.27	22.17
223	0.31	26.79
228	0.30	22.29
228	0.28	25.30
239	0.29	23.55
239	0.31	26.76
241	0.35	17.24
241	0.62	23.78
250	0.25	22.07
250	0.29	25.38
255	0.03	13.30

5.9 Solution NMR pulse sequences

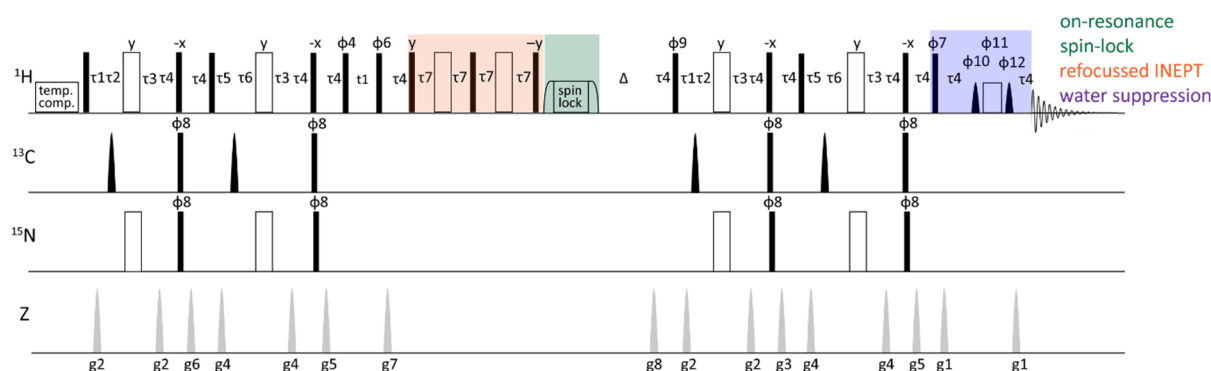


Figure 89: $R_{1\rho}$ relaxation dispersion solution pulse sequence, adapted from “refineptgpphwgxf” from the Bruker library^[35a]. The Z channel displays all implemented gradient pulses. Apart from ^{13}C -, ^{15}N -double filters, the ^1H -BMRD experiment comprises temperature compensation, a refocused INEPT (yellow), an on-resonance spin lock (green) with a constant time delay Δ , and a water suppression module (purple). The delays were defined as $\tau_1=476\ \mu\text{s}$, $\tau_2=204\ \mu\text{s}$, $\tau_3=1.977\ \text{ms}$, $\tau_4=200\ \mu\text{s}$, $\tau_5=375\ \mu\text{s}$, $\tau_6=305\ \mu\text{s}$, and $\tau_7=24.995\ \text{ms}$. The indirect chemical shift evolution period is depicted as t_1 . Spin lock fields were applied at 1.0, 1.5, 2.0, 2.5, 3.0, 5.0, 7.0, and 10.0 kHz. Spin lock times were chosen with 10, 25, 50, 75, and 100 ms. Phase cycling was performed for $\phi_4=0\ 0\ 0\ 2\ 2\ 2\ 2$, $\phi_6=0\ 0\ 0\ 0\ 0\ 0\ 0\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$, $\phi_7=0\ 2$, $\phi_8=0\ 1\ 2\ 3$, $\phi_9=0\ 0\ 2\ 2$, $\phi_{10}=2\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1$, $\phi_{11}=0\ 0\ 0\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 3\ 3\ 3\ 3$, and $\phi_{12}=2\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1$.

Pulse sequence corresponding to Figure 89:

```
;refineptgpphwgxf
;avance-version (12/01/11)
;2D homonuclear Hartman-Hahn transfer
;using refocused INEPT sequence for mixing
;with C-13 and N-15 filter in F1 and F2
;
;H[C12,N14] (t1) -> H[C12,N14] (t2)
;
;phase sensitive
;water suppression using watergate sequence
;
;derived from dipsi2gpphwgxf, v 1.9 2012/01/31
;(A.L. Breeze, Prog. NMR Spectrosc. 36, 323-372 (2000))
;(C. Zwahlen, P. Legault, S.J.F. Vincent, J. Greenblatt, R. Konrat &
;L.E. Kay, J. Am. Chem. Soc. 119 6711-6721 (1997))
;(K. Ogura, H. Terasawa & F. Inagaki, J. Biomol. NMR 8, 492-498 (1996))
;(J. Iwahara, J.M. Wojciak & R.T. Clubb, J. Biomol. NMR 19, 231-241 (2001))
;
```

; List for spinlock pulses

```
define list<power> SLPower=<${VLIST}>
define list<delay> SLdelay=<${VDLIST}>
```

```
"p2=p1*2"
"p22=p21*2"
"d11=30m"
"d12=20u"
"d26=1s/(cnst4*4)"
"d28=1s/(cnst6*4)"
"d29=1s/(cnst7*4)"
"d25=1s/(cnst8*4)"
```

```

;"p18 = 120"

"p18=p9-p17"

"in0=inf1"

"d0=in0/2-p1*4/3.1416"

"DELTA1=d28-p19-d16-p15/2"
"DELTA2=d26-d28-p15/2"
"DELTA3=d26-p19-d16"
"DELTA4=d29-p19-d16-p15/2"
"DELTA5=d26-d29-p15/2"
"DELTA6=d25-p1/2"

1 ze
2 d11
3 d12
  Sldelay*0.01
  10u

  "p17=Sldelay"
  "p18=p9-p17"

;===== Beginning of temperature compensation =====
;20u pl17:f1
  20u SLPower:f1
  10u
  (p18 ph21):f1
  10u
  20u
;===== End of temperature compensation =====

d1 ;cw:f1 ph29
4u ;do:f1

d12 pl1:f1
50u UNBLKGRAD

(p1 ph1)
p19:gp2
d16
DELTA1
(p15:sp18 ph1):f2
DELTA2
(center (p2 ph2) (p22 ph1):f3 )
DELTA3
p19:gp2
d16 pl2:f2
(p1 ph3)

(p3 ph8):f2
(p21 ph8):f3
p16:gp6
d16

```

```

(p1 ph1)
p19:gp4
d16
DELTA4
(p15:sp18 ph1):f2
DELTA5
(center (p2 ph2) (p22 ph1):f3 )
DELTA3
p19:gp4
d16 pl2:f2
(p1 ph3)

```

```

(p3 ph8):f2
(p21 ph8):f3
p16:gp5
d16

```

```

(p1 ph4)
d0

```

```

(p1 ph6)

```

```

4u
p19:gp7
d16

```

```

;===== Beginning of refocused INEPT =====

```

```

(p1 ph16)
DELTA6
(p2 ph17)
DELTA6
(p1 ph18)
DELTA6
(p2 ph19)
DELTA6
(p1 ph20)

```

```

;===== End of refocused INEPT =====

```

```

;===== Beginning of on-resonance spin-lock =====

```

```

10u fq=cnst17:f1
;20u pl17:f1
20u SLPower:f1
10u
(p10:sp15(currentpower) ph21):f1
(p17 ph21):f1
(p10:sp16(currentpower) ph21):f1
10u
20u
10u fq=cnst11:f1

```

```

;===== End of on-resonance spin-lock =====

```

```

;===== Beginning of constant time relaxation =====

```

```

;pl18 should have zero power pl18 = 120db!!!
5u pl18:f1
(p18 ph21):f1
5u pl1:f1

```



```
;===== End of constant time relaxation =====
```

```
4u
p19:gp8
d16 pl1:f1
```

```
(p1 ph9)
p19:gp2
d16
DELTA1
(p15:sp18 ph1):f2
DELTA2
(center (p2 ph2) (p22 ph1):f3 )
DELTA3
p19:gp2
d16 pl2:f2
(p1 ph3)
```

```
(p3 ph8):f2
(p21 ph8):f3
p16:gp3
d16
```

```
(p1 ph1)
p19:gp4
d16
DELTA4
(p15:sp18 ph1):f2
DELTA5
(center (p2 ph2) (p22 ph1):f3 )
DELTA3
p19:gp4
d16 pl2:f2
(p1 ph3)
```

```
(p3 ph8):f2
(p21 ph8):f3
p16:gp5
d16
```

```
;===== Beginning of WATERGATE =====
```

```
(p1 ph7)
4u
p16:gp1
d16 pl0:f1
(p11:sp1 ph10:r):f1
4u
4u pl1:f1
(p2 ph11)
4u pl0:f1
(p11:sp1 ph12:r):f1
4u
p16:gp1
d16
4u BLKGRAD
```

```
;===== End of WATERGATE =====
```

```

go=2 ph31
d11 mc #0 to 2
F1(SLPower.inc,8,SLdelay.inc,5)
;F1(SLPower.inc,1,SLdelay.inc,1)
F1PH(calph(ph4, +90), caldel(d0, +in0))
exit

```

```

ph1=0
ph2=1
ph3=2
ph4=0 0 0 0 2 2 2 2
ph6=0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph7=0 2
ph8=0 1 2 3
ph9=0 0 2 2
ph10=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph11=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph12=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph16=1
ph17=0
ph18=0
ph19=0
ph20=3
ph21=0
;ph23=3
;ph25=1
ph29=0
ph31=0 2 2 0 0 2 2 0 2 0 2 0 2 2 0 0 2

```

```

;p10 : 0W
;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default)
;p13 : f3 channel - power level for pulse (default)
;SLPower: f1 channel - power level for Proton spinlock
;p132: f1 channel - power level for low power presaturation
;sp1: f1 channel - shaped pulse 90 degree
;sp18: f2 channel - shaped pulse 180 degree (adiabatic)
;spnam15: shape for spin-lock adiabatic ramp on
;spnam16: shape for spin-lock adiabatic ramp off
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel - 90 degree high power pulse
;p6 : f1 channel - 90 degree low power pulse
;p9 : max relaxation delay (us)
;p10: 1H adiabatic flip pulse [4000us]
;p11: f1 channel - 90 degree shaped pulse
;p15: f2 channel - 180 degree shaped pulse for inversion (adiabatic)
;p16: homospoil/gradient pulse [1 msec]
;p19: gradient pulse 2 [500 usec]
;p21: f3 channel - 90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;d0 : incremented delay (2D)
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O [30 msec]

```

```

;d12: delay for power switching [20 usec]
;d16: delay for homospoil/gradient recovery
;d26: 1/(4*J(NH))
;d28: 1/(4*J(CH)min)
;d29: 1/(4*J(CH)max)
;cnst4: = J(NH)
;cnst6: = 1J(CH)min
;cnst7: = 1J(CH)max
;cnst8: = 3J(HH)
;inf1: 1/SW = 2 * DW
;in0: 1/(1 * SW) = 2 * DW
;nd0: 1
;ns: 8 * n
;ds: 32
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEQ

;use gradient ratio: gp 1 : gp 2 : gp 3 : gp 4 : gp 5 : gp 6 : gp 7
;                80 : 3 : 7 : 2 : 5 : 30 : 11

;for z-only gradients:
;gpz1: 30%
;gpz2: 3%
;gpz3: 20%
;gpz4: 2%
;gpz5: 30%
;gpz6: 50%
;gpz7: -6%
;gpz8: -8%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100
;gpnam5: SMSQ10.100
;gpnam6: SMSQ10.100
;gpnam7: SMSQ10.100
;gpnam8: SMSQ10.100

;for p15:sp18 use either "Crp60,0.5,20.1" or adiabatic pulse
; with matched sweep
;for matched sweep use
; low to high field sweep,
; carrier shifted to center of spectrum (75 ppm) via spoffs18
; d28 = d29 = 1/(4*J(CH) center of spectrum [1.696 ms])

;set pl32 to 0W when presaturation is not required
; use pl1 + 75 to 80dB to reduce radiation damping

;Processing
;PHC0(F1): 90
;PHC1(F1): -180
;FCOR(F1): 1
;$ld: dipsi2gpphwgxf,v 1.9 2012/01/31 17:49:23 ber Exp $

```

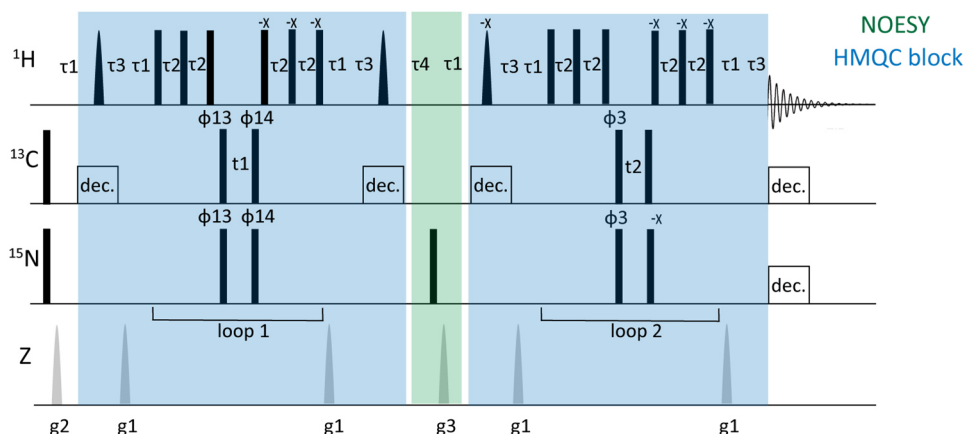


Figure 90: Time-shared pulse sequence for solution including two HMQC modules and one NOESY module, as published by Amero, C., *et al.*^[105], evolving ^{13}C - and ^{15}N -dimensions simultaneously. Initial ^{13}C - and ^{15}N -magnetization was removed before the first HMQC. Loops 1 and 2 are repeated with a centered 180° pulse in the F3 dimension with a γ -phase. The indirect chemical shift evolution periods are depicted as t1 and t2. The delays τ were defined as follows: $\tau_1=200\ \mu\text{s}$, $\tau_2=140\ \mu\text{s}$, $\tau_3=1.87\ \text{ms}$, and $\tau_4=49.2\ \text{ms}$. Phase cycling was performed with $\phi_3=0\ 0\ 2\ 2$, $\phi_{13}=0\ 2$, and $\phi_{14}=0\ 0\ 0\ 2\ 2\ 2\ 2$.

5.10 Solid-state NMR pulse sequences

Pulse sequence corresponding to Figure 21:

```
;avance-version (19/02/08)
;2D homonuclear correlation via dipolar coupling
;dipolar coupling may be due to noe or chemical exchange
;phase sensitive
;selecting C-12 or N-14 bound protons in F1 and F2
;water suppression using watergate sequence
;
;adapted from solution NMR pulse sequence noesygpphwgxf
;(use parameterset NOESYGPPHWGXF)
;M. Ikura & A. Bax, J. Am. Chem. Soc. 114, 2433-2440 (1992)
;M. Piotto, V. Saudek & V. Sklenar, J. Biomol. NMR 2, 661 - 666 (1992)
;V. Sklenar, M. Piotto, R. Leppik & V. Saudek, J. Magn. Reson.,
;Series A 102, 241 -245 (1993)
;
;CLASS=HighRes
;DIM=2D
;TYPE=
;SUBTYPE=
;COMMENT=
```

```
prosol relations=<triple>
```

```
#include <Avance.incl>
#include <Delay.incl>
```

```
"p2=p1*2"
"p22=p21*2"
"d2=1s/(cnst2*2)"
"d11=30m"
"d12=20u"
"d21=1s/(cnst4*2)"
"d24=1s/(cnst2*2)"
```

```

"in0=inf1"
"d0=3u"

"p10 = 4m"

"DELTA=d21-d2-larger(p2,p22)-p3"
"DELTA1=d2*2-d21-p21+larger(p2,p22)+p3-d0"
"DELTA2=d24-p12-d12-8u"
"DELTA3=d21-d24-larger(p2,p22)-p3-p12-d12-8u"
"DELTA4=d24*2-d21-p21+larger(p2,p22)+p3-4u"

"TAU=d8/2--p3/2-p21/2"
"TAU1=d8/2-p21/2-4u"

"cnst63 = plw11"
"plw0 = 0 "

;=====
; Protection for parameters
;=====
1m

    if "p11>101ms" goto Problem
    if "p9>105ms" goto Problem
    if "cnst63>2.8" goto Problem

    goto PassParams
Problem, 1m
    print "Protection: Parameters not accepted, ending."
    goto HaltAcqu
PassParams, 1m

1 d11 ze
2 d11
3 d11

;===== Beginning of temperature compensation =====
"p18=p9-p11"
20u pl11:f1
10u
(p18 ph21):f1
10u
20u
;===== End of temperature compensation =====

d1
4u pl1:f1
(p1 ph10)
d2
(p3 ph6):f2
(center (p2 ph2) (p22 ph7):f3 )
DELTA
(p21 ph6):f3
DELTA1
(p3 ph7):f2
d0

```

```

(p1 ph3)

#ifdef NOESY

TAU
(p3 ph1):f2
(p21 ph1):f3
4u
TAU1

#endif /*NOESY*/

;===== Beginning of on-resonance spin-lock =====
10u fq=cnst17:f1
(p1 pl1 ph22):f1
2u pl11:f1
(p11 pl11 ph21):f1
2u pl1:f1
(p1 pl1 ph22):f1
10u fq=cnst11:f1
;===== End of on-resonance spin-lock =====

;===== Beginning of constant time relaxation =====
5u pl0:f1
(p18 ph21):f1
5u pl1:f1
;===== End of constant time relaxation =====

(p1 ph4):f1
4u
DELTA2 pl0:f1
(p12:sp11 ph11:r):f1
4u
d12 pl1:f1
(p3 ph6):f2
(center (p2 ph1) (p22 ph7):f3 )
4u
d12 pl0:f1
(p12:sp12 ph12:r):f1
4u
DELTA3
(p21 ph9):f3
DELTA4
(p3 ph8):f2
4u ;BLKGRAD
go=2 ph31
d11 mc #0 to 2
F1PH(caliph(ph3, -90), caldel(d0, +in0))

HaltAcqu, 1m
exit

ph1=0
ph2=0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3
ph3=0

```

```

ph4=0 0 1 1 2 2 3 3
ph5=2
ph6=0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph7=0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
ph8=0
ph9=0 0 0 0 2 2 2 2
ph10 = 0
ph11 = 3
ph12 = 2 2 3 3 0 0 1 1
ph22=0
ph21 = 1
ph31=0 2 3 1 2 0 1 3

;p10 : 0W
;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default)
;p13 : f3 channel - power level for pulse (default)
;p111: f1 channel - SPINLOCK pulse power [<10khz]
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel - 90 degree high power pulse
;p9 : f1 channel - max SL relaxation delay (us)
;p11 : f1 channel - SPINLOCK time [< 100 ms]
;p12: f1 channel - water selective 90 degree shaped pulse [1 msec]
;p21: f3 channel - 90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d2 : 1/(2J(CH))
;d8 : NOESY/RFDR mixing time
;d9 : TOCSY/ROESY mixing time [80msec/100msec]
;d11: delay for disk I/O [30 msec]
;d12: delay for power switching [20 usec]
;d16: delay for homospoil/gradient recovery
;d21: 1/(2J(NH))
;d24: 1/(2J'(CH))
;inf1: 1/SW = 2 * DW
;in0: 1/(1 * SW) = 2 * DW
;nd0: 1
;ns: 16 * n
;ds: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEQ

;use gradient ratio: gp 1 : gp 2 : gp 3
; 50 : 30 : 30

;for z-only gradients:
;gpz1: 50%
;gpz2: 30%
;gpz3: 30%
;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;$Id: noesygpphwgxf,v 1.12 2019/02/11 14:04:14 ber Exp $

```

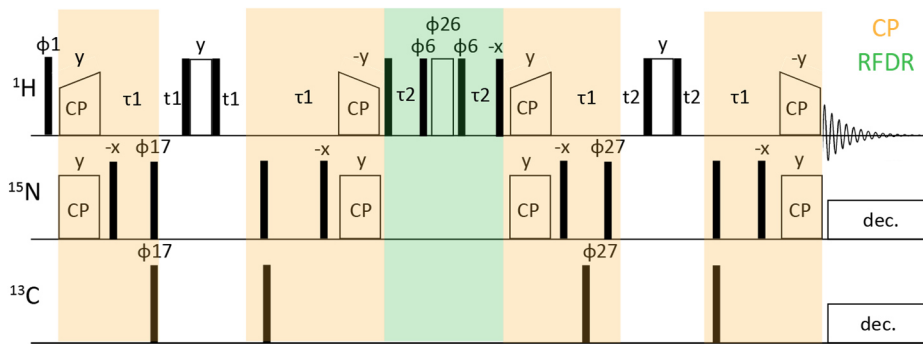


Figure 91: Time-shared pulse sequence for solid state involving CP-transfers (yellow) between ^1H and ^{15}N and HMQC-transfers between ^1H and ^{13}C nuclei, composite 180° pulses for ^1H -decoupling, and one RFDR block (green). The indirect chemical shift evolution periods are depicted as $t1$ and $t2$. The delays τ were defined as follows: $\tau1=3.7$ ms and $\tau2=10.4$ μs . Phase cycling was applied with $\phi1=0\ 0\ 0\ 0\ 2\ 2\ 2\ 2$, $\phi6=1\ 2\ 1\ 2\ 2\ 2\ 3\ 2\ 3\ 3\ 0\ 3\ 0\ 0\ 1\ 0\ 1$, $\phi17=0\ 2$, $\phi26=0\ 1\ 0\ 1\ 1\ 2\ 1\ 2\ 2\ 3\ 2\ 3\ 3\ 0\ 3\ 0$, and $\phi27=0\ 0\ 2\ 2$. The pulse sequence was derived from the solution NMR pulse sequence shown in Figure 90.

Pulse sequence corresponding to Figure 91:

```
;HXH_2D.suvaH
;H to X (t1) to H (t2) transfer;

;=====
;Variables introduction
;=====
;p1 H p90
;p11 H p90 power
;p2 X p90 duration
;p12 X p90 power
;p3 Y p90 duration
;p13 Y p90 power
;l31 =1 (ramp on H) = 2 (ramp on X)
;p15 X 1st ramp-CP power
;p16 1H CP power
;p15 1st CP contact time
;p16 2nd CP contact time
;cnst10 1H carrier for 1st CP
;d2 delay for refocusing
;d0 t1 initial value
;p122 WALTZ decoupling on X (during acquisition)
;p123 WALTZ decoupling on Y (during acquisition)
;p16 water suppression pulse power
;p22 water suppression pulse length
;d8 NOESY mixing time
;cnst5 SPINNING FREQUENCY (in Hz)
;p111 RFDR (low) power level
;p11 RFDR 180 pulse on Proton
;l11 RFDR loop count (=multiple of 16)
;d21 RFDR mixing time

;=====
; Include file for Protection
;=====
#include <Avance.incl>
#include <Delay.incl>
```



```

;=====
; Set variables
;=====
"cnst63 = plw12"
"cnst62 = plw22"
"cnst61 = plw23"
"cnst60 = plw16"
"p4 = 2.0*p1"
"p5 = 2.0*p2"
"p6 = 2.0*p3"
"d11=30m"
"d0=0.2u"
"d10=0.2u"
"d20=0.2u"
"in0 = inf1/2"
"in20 = inf2/2"
"spw2 =spw1"
"spw4 =spw3"
"plw0 =0.0"
"l2=0"
"acqt0 = 0" ; TOPSPIN 3.2 - BASE OPT- first point calculation fid
"DELTA1=(1s/(2*cnst4))"
"plw23 = plw3/((pcpd3/p3)*(pcpd3/p3))"
"plw22 = plw2/((pcpd2/p2)*(pcpd2/p2))"

# ifdef RFDR
"d9 = (1s/(2*cnst5)) - p11/2"
"d21 = l11*(1s/cnst5)"
"DELTA2 =d9-p1/2"
# endif /*RFDR*/

# ifdef RFDRcal
"d9 = (1s/(2*cnst5)) - p11/2"
"d21 = l11*(1s/cnst5)"
"DELTA2 =d9-p1/2"
# endif /*RFDRcal*/

# ifdef CRFDR
"p12 = p11/2"
"d9 = (1s/(2*cnst5)) - p11"
"d21 = l11*(1s/cnst5)"
"DELTA2 =d9-p1/2"
# endif /*CRFDR*/

# ifdef CRFDRcal
"p12 = p11/2"
"d9 = (1s/(2*cnst5)) - p11"
"d21 = l11*(1s/cnst5)"
"DELTA2 =d9-p1/2"
# endif /*CRFDRcal*/

;=====
; Protection for parameters
;=====

```

```

1m
; if "p1 > 100u" goto Problem
; if "p2 > 100u" goto Problem
  if "p3 > 100u" goto Problem
  if "p15 > 9001u" goto Problem
  if "p16 > 9001u" goto Problem
  if "aq > 90m" goto Problem
  if "d1 < 0.3" goto Problem
  if "cnst63 > 4" goto Problem
  if "cnst62 > 7.0" goto Problem
  if "cnst61 > 1.72" goto Problem
  if "cnst60 > 0.7" goto Problem

# ifdef RFDR
  if "d9 < 0" goto Problem
  if "d21 > 300ms" goto Problem
# endif /*RFDR*/

goto PassParams
Problem, 1m
  print "Protection: Parameters not accepted, ending."
  goto HaltAcqu
PassParams, 1m

;---- Relaxation & reset parameters ----
1 ze
  d11
2 d11 do:f3
  40u do:f2
  40u do:f1 ;decouplers off

3 d1
  5u rpp6
  5u rpp7
  5u rpp26
  20u reset:f1 reset:f2 reset:f3
  20u fq=0:f1 fq=0:f2 fq=0:f3 ; 1H carrier freq.
  5u pl1:f1 pl5:f2 pl3:f3

;---- 90 on H ----
  (p1 pl1 ph1):f1

;---- H-X/Y CP-HMQC ----
  1u pl0:f1
  (p15 pl5 ph2):f2 (p15:sp1 ph10):f1
  1u
  (p2 pl2 ph21):f2
  DELTA1
  (center (p2 pl2 ph17):f2 (p3 pl3 ph17):f3)

;---- X (t1) evolution & HY decoupling ----
  2u fq=cnst20:f2 fq=cnst30:f3
  d0
  (p1 pl1 ph22):f1
  (p4 pl1 ph20):f1
  (p1 pl1 ph22):f1
  d0

```

```

2u fq=0:f2 fq=0:f3

;----- X/Y-H CP-HMQC -----
if "I2 %2 == 0"
{
(center (p2 pl2 ph18):f2 (p3 pl3 ph18):f3)
}
if "I2 %2 == 1"
{
(center (p2 pl2 ph19):f2 (p3 pl3 ph18):f3)
}

DELTA1
(p2 pl2 ph21):f2
1u pl5:f2 pl0:f1
(p15 pl5 ph12):f2 (p15:sp2 ph11):f1

;----- NOE/RFDR/ -----
(p1 pl1 ph24):f1

# ifdef NOESY
d8
# endif /*NOESY*/

# ifdef RFDR
DELTA2 fq=cnst10:f1
4 d9
(p11 pl11 ph26^):f1
d9
lo to 4 times l11
DELTA2 fq=0:f1
# endif /*RFDR*/

# ifdef RFDRcal
DELTA2 fq=cnst10:f1
5 d9
(p11 pl11 ph29):f1
d9
lo to 5 times l11
DELTA2 fq=0:f1
# endif /*RFDRcal*/

# ifdef CRFDR
DELTA2 fq=cnst10:f1
6 d9
(p12 pl11 ph6^):f1
(p11 pl11 ph26^):f1
(p12 pl11 ph7^):f1
d9
lo to 6 times l11
DELTA2 fq=0:f1
# endif /*CRFDR*/

```

```

# ifdef CRFDRcal
  DELTA2 fq=cnst10:f1
7 d9
  (p12 pl11 ph6):f1
  (p11 pl11 ph26):f1
  (p12 pl11 ph7):f1
  d9
  lo to 7 times l11
  DELTA2 fq=0:f1
# endif /*CRFDRcal*/

  (p1 pl1 ph25):f1

;----- H-X/Y CP-HMQC -----
  1u pl0:f1
  (p15 pl5 ph2):f2 (p15:sp1 ph10):f1
  1u
  (p2 pl2 ph21):f2
  DELTA1
  (center (p2 pl2 ph27):f2 (p3 pl3 ph27):f3)

;----- X (t1) evolution & HY decoupling -----
  1u pl12:f1 pl2:f2
  d20
  (p1 pl1 ph22):f1
  (p4 pl1 ph20):f1
  (p1 pl1 ph22):f1
  d20
  1u pl16:f1

;----- X/Y-H CP-HMQC -----
  (center (p2 pl2 ph28):f2 (p3 pl3 ph28):f3)
  DELTA1
  (p2 pl2 ph21):f2
  1u pl5:f2 pl0:f1
  (p15 pl5 ph12):f2 (p15:sp2 ph11):f1

;----- Decoupling, acquisition and write data -----
  2u fq=0:f3 fq=0:f1
  1u pl22:f2 pl23:f3
  1u cpd2:f2 cpd3:f3
  go=2 ph31
  d11 do:f2 do:f3 mc #0 to 2
; F1l(iu2,2)
; F1PH(ip17, id0)
; F2PH(ip27, id20)
  F1PH(calph(ph17,+90.0), caldel(d0,+in0))
  F2PH(calph(ph27,+90.0), caldel(d20,+in20))

suspend

HaltAcqu, 1m
exit

;----- Phase cycling -----
ph1= 0 0 0 0 2 2 2 2 ;1H excitation

```

```

ph2= 1
ph10= 1
ph17= 0 2
ph19= 2 2 ;15N along Z after evolution
ph18= 0 0 ;2 2 ;15N after H2O dephase for 2nd CP
ph11= 3 ; 1H in 2nd CP
ph12= 1 ; 15N in 2nd CP
ph20= 1
ph21= 2 ;15N refocus
ph22= 0
ph23= 1
ph24 = 0
ph25 = 2
ph27= 0 0 2 2
ph28= 0 0 ;2 2
ph26 = 0 1 0 1 1 2 1 2 2 3 2 3 3 0 3 0
ph6 = 1 2 1 2 2 3 2 3 3 0 3 0 0 1 0 1
ph7 = 1 2 1 2 2 3 2 3 3 0 3 0 0 1 0 1
ph29 = 0
ph31= 3 1 1 3 1 3 3 1

;=====
;END Pulse program
;=====
;IMPORTANT NOTE
;On AV (avance) machines, it is better to use "spf0" files instead of "sp0".
;The "sp0" files requires lot of time for power settings compared to "spf0".

```

References

- [1] aS. Paschkowsky, M. Hamzé, F. Oestereich, L. M. Munter, *The Journal of biological chemistry* **2016**, *291*, 21903–21912; bT. L. Cheng, Y. T. Wu, H. Y. Lin, F. C. Hsu, S. K. Liu, B. I. Chang, W. S. Chen, C. H. Lai, G. Y. Shi, H. L. Wu, *J Invest Dermatol* **2011**, *131*, 2486–2494; cP. J. Noy, R. K. Swain, K. Khan, P. Lodhia, R. Bicknell, *Faseb j* **2016**, *30*, 2311–2323; dW. Song, W. Liu, H. Zhao, S. Li, X. Guan, J. Ying, Y. Zhang, F. Miao, M. Zhang, X. Ren, X. Li, F. Wu, Y. Zhao, Y. Tian, W. Wu, J. Fu, J. Liang, W. Wu, C. Liu, J. Yu, S. Zong, S. Miao, X. Zhang, L. Wang, *Nat Commun* **2015**, *6*, 8022.
- [2] aJ. Cavanagh, N. J. Skelton, W. J. Fairbrother, M. Rance, A. Palmer, III, *Protein NMR Spectroscopy*, 2nd ed., Elsevier Science, **2007**; bJ. Keeler, *Understanding NMR Spectroscopy*, 2nd ed., John Wiley & Sons, **2010**.
- [3] B. Reif, S. E. Ashbrook, L. Emsley, M. Hong, *Nat Rev Methods Primers* **2021**, *1*, 2.
- [4] H. Kovacs, D. Moskau, M. Spraul, *Prog Nucl Magn Reson Spectrosc* **2005**, *46*, 131–155.
- [5] M. Sattler, J. Schleucher, C. Griesinger, *Prog Nucl Magn Reson Spectrosc* **1999**, *34*, 93–158.
- [6] D. Pantoja-Uceda, J. Santoro, *J Biomol NMR* **2013**, *57*, 57–63.
- [7] E. Schmidt, P. Güntert, *J Am Chem Soc* **2012**, *134*, 12817–12829.
- [8] I. Pritišanac, J. M. Würz, T. R. Alderson, P. Güntert, *Nat Commun* **2019**, *10*, 4922.
- [9] A. Pines, M. G. Gibby, J. S. Waugh, *J Chem Phys* **2003**, *56*, 1776–1777.
- [10] M. J. Bayro, M. Huber, R. Ramachandran, T. C. Davenport, B. H. Meier, M. Ernst, R. G. Griffin, *J Chem Phys* **2009**, *130*, 114506.
- [11] C. J. Giunta, V. V. Mainz, *Discovery of Nuclear Magnetic Resonance: Rabi, Purcell, and Bloch, Vol. 1349*, American Chemical Society, **2020**.
- [12] aL. Müller, A. Kumar, R. R. Ernst, *J Chem Phys* **2008**, *63*, 5490–5491; bW. P. Aue, E. Bartholdi, R. R. Ernst, *J Chem Phys* **2008**, *64*, 2229–2246.
- [13] G. Wagner, A. Kumar, K. Wüthrich, *Eur J Biochem* **1981**, *114*, 375–384.
- [14] aP. C. Lauterbur, *Nature* **1973**, *242*, 190–191; bP. Mansfield, I. L. Pykett, P. G. Morris, R. E. Coupland, *Br J Radiol* **1978**, *51*, 921–922.
- [15] P. Rovó, *Solid State Nucl Magn Reson* **2020**, *108*, 101665.
- [16] B. A. Grzybowski, A. V. Ishchenko, C.-Y. Kim, G. Topalov, R. Chapman, D. W. Christianson, G. M. Whitesides, E. I. Shakhnovich, *Proc Natl Acad Sci U S A* **2002**, *99*, 1270–1273.
- [17] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res* **2000**, *28*, 235–242.
- [18] W. Kühlbrandt, *Science* **2014**, *343*, 1443–1444.
- [19] aJ. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589; bM. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, *Nucleic Acids Res* **2022**, *50*, D439–D444.

- [20] P. Aller, A. M. Orville, *Dynamic Structural Biology Experiments at XFEL or Synchrotron Sources*, Springer US, New York, **2021**.
- [21] D. K. Sasmal, L. E. Pulido, S. Kasal, J. Huang, *Nanoscale* **2016**, *8*, 19928–19944.
- [22] S. A. Hollingsworth, R. O. Dror, *Neuron* **2018**, *99*, 1129–1143.
- [23] A. D. Gossert, W. Jahnke, *Prog Nucl Magn Reson Spectrosc* **2016**, *97*, 82–125.
- [24] P. Csermely, R. Palotai, R. Nussinov, *Trends Biochem Sci* **2010**, *35*, 539–546.
- [25] H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **1991**, *254*, 1598–1603.
- [26] M. P. Williamson, *Prog Nucl Magn Reson Spectrosc* **2013**, *73*, 1–16.
- [27] Z. Bu, D. J. Callaway, *Adv Protein Chem Struct Biol* **2011**, *83*, 163–221.
- [28] R. Ishima, S. Bagby, *Protein Dynamics Revealed by CPMG Dispersion*, Springer, Cham, **2016**.
- [29] A. G. Palmer, III, *Chem Rev* **2004**, *104*, 3623–3640.
- [30] S. J. Ullrich, S. Hölper, C. Glaubitz, *J Biomol NMR* **2014**, *58*, 27–35.
- [31] D. Marion, D. F. Gauto, I. Ayala, K. Giandoreggio-Barranco, P. Schanda, *ChemPhysChem* **2019**, *20*, 276–284.
- [32] J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *J Chem Phys* **1979**, *71*, 4546–4553.
- [33] P. Vallurupalli, G. Bouvignies, L. E. Kay, *J Am Chem Soc* **2012**, *134*, 8148–8161.
- [34] A. G. Palmer, III, *J Magn Reson* **2014**, *241*, 3–17.
- [35] aJ. Iwahara, J. M. Wojciak, R. T. Clubb, *J Biomol NMR* **2001**, *19*, 231–241; bC. Zwahlen, P. Legault, S. J. F. Vincent, J. Greenblatt, R. Konrat, L. E. Kay, *J Am Chem Soc* **1997**, *119*, 6711–6721; cK. Ogura, H. Terasawa, F. Inagaki, *J Biomol NMR* **1996**, *8*, 492–498; dA. L. Breeze, *Prog Nucl Magn Reson Spectrosc* **2000**, *36*, 323–372.
- [36] R. Rovó, , Petra, C. Smith, D. Gauto, B. d. Groot, P. Schanda, R. Linser, *J Am Chem Soc* **2019**, *141*.
- [37] aD. F. Gauto, P. Macek, D. Malinverni, H. Fraga, M. Paloni, I. Sučec, A. Hessel, J. P. Bustamante, A. Barducci, P. Schanda, *Nat Commun* **2022**, *13*, 1927; bP. Ma, J. D. Haller, J. Zajakala, P. Macek, A. C. Sivertsen, D. Willbold, J. Boisbouvier, P. Schanda, *Angew Chem Int Ed* **2014**, *53*, 4312–4317; cV. Kurauskas, S. A. Izmailov, O. N. Rogacheva, A. Hessel, I. Ayala, J. Woodhouse, A. Shilova, Y. Xue, T. Yuwen, N. Coquelle, J.-P. Colletier, N. R. Skrynnikov, P. Schanda, *Nat Commun* **2017**, *8*, 145; dJ. M. Lamley, C. Öster, R. A. Stevens, J. R. Lewandowski, *Angew Chem Int Ed* **2015**, *54*, 15374–15378; eA. G. Palmer, F. Massi, *Chem Rev* **2006**, *106*, 1700–1719.
- [38] J. Kotschy, R. Linser, *Proton-detected solid-state NMR and its applications to membrane proteins*, IOP Publishing, **2020**.
- [39] B. Rowlinson, E. Crublet, R. Kerfah, M. J. Plevin, *Biochem Soc Trans* **2022**, *50*, 1555–1567.
- [40] aS. K. Vasa, H. Singh, K. Grohe, R. Linser, *Angew Chem Int Ed* **2019**, *58*, 5758–5762; bL. B. Andreas, K. Jaudzems, J. Stanek, D. Lalli, A. Bertarello, T. Le Marchand, D. Cala-De Paepe, S. Kotelovica, I. Akopjana, B. Knott, S. Wegner, F. Engelke, A. Lesage, L. Emsley, K. Tars, T. Herrmann, G. Pintacuda, *Proc Natl Acad Sci U S A* **2016**, *113*, 9187–9192.
- [41] aD. Mance, T. Sinnige, M. Kaplan, S. Narasimhan, M. Daniëls, K. Houben, M. Baldus, M. Weingarth, *Angew Chem Int Ed* **2015**, *54*, 15799–15803; bJ. Medeiros-Silva, D. Mance, M. Daniëls, S. Jekhmane, K. Houben, M. Baldus, M. Weingarth, *Angew Chem Int Ed* **2016**, *55*, 13606–13610.
- [42] D. M. LeMaster, *Prog Nucl Magn Reson Spectrosc* **1994**, *26*, 371–419.
- [43] S. Asami, P. Schmieder, B. Reif, *J Am Chem Soc* **2010**, *132*, 15133–15135.

- [44] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono, P. Güntert, *Nature* **2006**, *440*, 52-57.
- [45] aD. Sheppard, R. Sprangers, V. Tugarinov, *Prog Nucl Magn Reson Spectrosc* **2010**, *56*, 1-45; bT. L. Religa, L. E. Kay, *J Biomol NMR* **2010**, *47*, 163-169; cR. Sprangers, L. E. Kay, *Nature* **2007**, *445*, 618-622; dS. Schütz, R. Sprangers, *Prog Nucl Magn Reson Spectrosc* **2020**, *116*, 56-84.
- [46] K. Xue, R. Sarkar, Z. Tosner, D. Lalli, C. Motz, B. Koch, G. Pintacuda, B. Reif, *J Biomol NMR* **2019**, *73*, 625-631.
- [47] K. Xue, R. Sarkar, C. Motz, S. Asami, V. Decker, S. Wegner, Z. Tosner, B. Reif, *J Phys Chem* **2018**, *122*, 16437-16442.
- [48] aR. Sprangers, L. E. Kay, *J Am Chem Soc* **2007**, *129*, 12668-12669; bR. Sprangers, A. Velyvis, L. E. Kay, *Nat Methods* **2007**, *4*, 697-703.
- [49] aH. Steiner, B. H. Jonsson, S. Lindskog, *Eur J Biochem* **1975**, *59*, 253-259; bR. G. Khalifah, *J Biol Chem* **1971**, *246*, 2561-2573; cT. H. Maren, *Physiol Rev* **1967**, *47*, 595-781.
- [50] N. U. Meldrum, F. J. Roughton, *J Physiol* **1933**, *80*, 113-142.
- [51] aS. Lindskog, *Pharmacol Ther* **1997**, *74*, 1-20; bC. T. Supuran, *Curr Pharm Des* **2008**, *14*, 603-614.
- [52] A. Nocentini, C. T. Supuran, *Expert Opin Drug Discov* **2019**, *14*, 1175-1197.
- [53] A. Di Fiore, C. T. Supuran, A. Scaloni, G. De Simone, *J Enzyme Inhib Med Chem* **2020**, *35*, 1450-1461.
- [54] aB. Becker, W. H. Middleton, *AMA Arch Ophthalmol* **1955**, *54*, 187-192; bE. Masini, F. Carta, A. Scozzafava, C. T. Supuran, *Expert Opin Ther Pat* **2013**, *23*, 705-716.
- [55] aM. Aggarwal, B. Kondeti, R. McKenna, *Expert Opin Ther Pat* **2013**, *23*, 717-724; bS. D. Silberstein, E. Ben-Menachem, R. P. Shank, F. Wiegand, *Clin Ther* **2005**, *27*, 154-165.
- [56] A. Scozzafava, C. T. Supuran, F. Carta, *Expert Opin Ther Pat* **2013**, *23*, 725-735.
- [57] aC. T. Supuran, *Future Med Chem* **2011**, *3*, 1165-1180; bM. Benej, S. Pastorekova, J. Pastorek, *Subcell Biochem* **2014**, *75*, 199-219.
- [58] W. S. Sly, P. Y. Hu, *Annu Rev Biochem* **1995**, *64*, 375-401.
- [59] G. N. Shah, G. Bonapace, P. Y. Hu, P. Strisciuglio, W. S. Sly, *Hum Mutat* **2004**, *24*, 272.
- [60] aB. V. Alvarez, G. L. Vilas, J. R. Casey, *EMBO J* **2005**, *24*, 2499-2511; bD. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, *Nucleic Acids Res* **2018**, *46*, D1074-D1082.
- [61] A. Liljas, K. K. Kannan, P. C. Bergstén, I. Waara, K. Fridborg, B. Strandberg, U. Carlbom, L. Järup, S. Lövgren, M. Petef, *Nat New Biol* **1972**, *235*, 131-137.
- [62] aT. O. Wambo, L. Y. Chen, S. F. McHardy, A. T. Tsin, *Biophys Chem* **2016**, *214-215*, 54-60; bS. H. Koenig, R. D. Brown, 3rd, *Ann N Y Acad Sci* **1984**, *429*, 99-108.
- [63] G. Aronsson, L. G. Mårtensson, U. Carlsson, B. H. Jonsson, *Biochem* **1995**, *34*, 2153-2162.
- [64] D. Duda, C. Tu, M. Qian, P. Laipis, M. Agbandje-McKenna, D. N. Silverman, R. McKenna, *Biochem* **2001**, *40*, 1741-1748.
- [65] aS. K. Vasa, P. Rovó, R. Linser, *Acc Chem Res* **2018**, *51*, 1386-1395; bJ. Zheng, B. S. Avvaru, C. Tu, R. McKenna, D. N. Silverman, *Biochem* **2008**, *47*, 12028-12036.

- [66] aH. Singh, S. K. Vasa, H. Jangra, P. Rovó, C. Päslack, C. K. Das, H. Zipse, L. V. Schäfer, R. Linser, *J Am Chem Soc* **2019**, *141*, 19276–19288; bZ. Liang, Y. Xue, G. Behravan, B. H. Jonsson, S. Lindskog, *Eur J Biochem* **1993**, *211*, 821–827.
- [67] R. E. Tashian, *Bioessays* **1989**, *10*, 186–192.
- [68] aA. E. Eriksson, T. A. Jones, A. Liljas, *Proteins* **1988**, *4*, 274–282; bK. M. Merz, Jr., *J Mol Biol* **1990**, *214*, 799–802.
- [69] S. K. Vasa, H. Singh, P. Rovó, R. Linser, *J Phys Chem Lett* **2018**, *9*, 1307–1311.
- [70] H. Singh, C. K. Das, S. K. Vasa, K. Grohe, L. V. Schäfer, R. Linser, *Angew Chem Int Ed* **2020**, *59*, 22916–22921.
- [71] aL. L. Kiefer, C. A. Fierke, *Biochem* **1994**, *33*, 15233–15240; bL. L. Kiefer, J. F. Krebs, S. A. Paterno, C. A. Fierke, *Biochem* **1993**, *32*, 9896–9900.
- [72] G. Chen, X. Kong, D. Lu, J. Wu, Z. Liu, *Phys Chem Chem Phys* **2017**, *19*, 11690–11697.
- [73] D. K. Srivastava, K. M. Jude, A. L. Banerjee, M. Haldar, S. Manokaran, J. Kooren, S. Mallik, D. W. Christianson, *J Am Chem Soc* **2007**, *129*, 5528–5537.
- [74] S. M. Gould, D. S. Tawfik, *Biochem* **2005**, *44*, 5444–5452.
- [75] F. Briganti, S. Mangani, A. Scozzafava, G. Vernaglione, C. T. Supuran, *J Biol Inorg Chem* **1999**, *4*, 528–536.
- [76] B. S. Avvaru, S. A. Busby, M. J. Chalmers, P. R. Griffin, B. Venkatakrisnan, M. Agbandje-McKenna, D. N. Silverman, R. McKenna, *Biochem* **2009**, *48*, 7365–7372.
- [77] aS. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, *Nucleic Acids Res* **2021**, *49*, D1388–d1395; bS. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, *Nucleic Acids Res* **2023**, *51*, D1373–D1380.
- [78] C. T. Supuran, *J Enzyme Inhib Med Chem* **2016**, *31*, 345–360.
- [79] aK. Håkansson, M. Carlsson, L. A. Svensson, A. Liljas, *J Mol Biol* **1992**, *227*, 1192–1204; bS. Mangani, K. Håkansson, *Eur J Biochem* **1992**, *210*, 867–871.
- [80] aB. M. Jönsson, K. Håkansson, A. Liljas, *FEBS Lett* **1993**, *322*, 186–190; bS. K. Nair, D. W. Christianson, *Eur J Biochem* **1993**, *213*, 507–515; cV. Kumar, K. K. Kannan, P. Sathyamurthi, *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 731–738.
- [81] K. Håkansson, C. Briand, V. Zaitsev, Y. Xue, A. Liljas, *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 101–104.
- [82] S. Glöckner, K. Ngo, C. P. Sager, T. Hüfner-Wulsdorf, A. Heine, G. Klebe, *ACS Chem Biol* **2020**, *15*, 675–685.
- [83] R. Gaspari, C. Rechlin, A. Heine, G. Bottegoni, W. Rocchia, D. Schwarz, J. Bomke, H.-D. Gerber, G. Klebe, A. Cavalli, *J Med Chem* **2016**, *59*, 4245–4256.
- [84] S. Glöckner, Dissertation thesis, Philipps-Universität Marburg **2020**.
- [85] C. T. Supuran, *Nat Rev Drug Discov* **2008**, *7*, 168–181.
- [86] A. Scozzafava, F. Briganti, M. A. Ilies, C. T. Supuran, *J Med Chem* **2000**, *43*, 292–300.
- [87] J. Y. Winum, C. Temperini, K. El Cheikh, A. Innocenti, D. Vullo, S. Ciattini, J. L. Montero, A. Scozzafava, C. T. Supuran, *J Med Chem* **2006**, *49*, 7024–7031.
- [88] aF. Saczewski, J. Sławiński, A. Kornicka, Z. Brzozowski, E. Pomarnacka, A. Innocenti, A. Scozzafava, C. T. Supuran, *Bioorg Med Chem Lett* **2006**, *16*, 4846–4851; bG. De Simone, R. M. Vitale, A. Di Fiore, C. Pedone, A. Scozzafava, J. L. Montero, J. Y. Winum, C. T. Supuran, *J Med Chem* **2006**, *49*, 5544–5551.
- [89] G. T. H. Nguyen, T. N. Tran, M. N. Podgorski, S. G. Bell, C. T. Supuran, W. A. Donald, *ACS Cent Sci* **2019**, *5*, 308–318.

- [90] M. A. Pinard, C. D. Boone, B. D. Rife, C. T. Supuran, R. McKenna, *Bioorg Med Chem* **2013**, *21*, 7210–7215.
- [91] Q. Huang, E. Y. Rui, M. Cobbs, D. M. Dinh, H. J. Gukasyan, J. A. Lafontaine, S. Mehta, B. D. Patterson, D. A. Rewolinski, P. F. Richardson, M. P. Edwards, *J Med Chem* **2015**, *58*, 2821–2833.
- [92] aP. W. Snyder, J. Mecinovic, D. T. Moustakas, S. W. Thomas, 3rd, M. Harder, E. T. Mack, M. R. Lockett, A. Héroux, W. Sherman, G. M. Whitesides, *Proc Natl Acad Sci U S A* **2011**, *108*, 17889–17894; bV. Alterio, S. M. Monti, E. Truppo, C. Pedone, C. T. Supuran, G. De Simone, *Org Biomol Chem* **2010**, *8*, 3528–3533; cA. Liljas, *IUCrJ* **2018**, *5*, 4–5; dH. Waldmann, M. Koppitz, *Small Molecule - Protein Interactions*, 1st ed., Springer, Berlin, Heidelberg, **2003**.
- [93] S. Sri Wahyu Effendi, J.-Y. Lin, I.-S. Ng, *Bioresour Technol* **2022**, *363*, 127980.
- [94] A. Stein, D. Chen, N. V. Igareta, Y. Cotelte, J. G. Rebelein, T. R. Ward, *ACS Cent Sci* **2021**, *7*, 1874–1884.
- [95] J. Kotschy, B. Söldner, H. Singh, S. K. Vasa, R. Linser, *Angew Chem Int Ed* **2024**, *63*, e202313947.
- [96] C.-H. Chen, *Deuterium Bonding Versus Hydrogen Bonding*, Springer, Cham, **2022**.
- [97] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–221.
- [98] aZ. Sun, Q. Liu, G. Qu, Y. Feng, M. T. Reetz, *Chem Rev* **2019**, *119*, 1626–1665; bX. Guo, D. He, L. Huang, L. Liu, L. Liu, H. Yang, *Comput Theor Chem* **2012**, *995*, 17–23.
- [99] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486–501.
- [100] P. Jonasson, A. Kjellsson, I. Sethson, B.-H. Jonsson, *FEBS Lett* **1999**, *445*, 361–365.
- [101] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, M. Nilges, *Bioinformatics* **2007**, *23*, 381–382.
- [102] aT.-L. Hwang, S. Mori, A. J. Shaka, P. C. M. van Zijl, *J Am Chem Soc* **1997**, *119*, 6203–6204; bC. Eichmüller, N. R. Skrynnikov, *J Biomol NMR* **2005**, *32*, 281–293.
- [103] A. Mandal, J. C. Boatz, T. B. Wheeler, P. C. van der Wel, *J Biomol NMR* **2017**, *67*, 165–178.
- [104] aD. P. Frueh, A. Leed, H. Arthanari, A. Koglin, C. T. Walsh, G. Wagner, *J Biomol NMR* **2009**, *45*, 311–318; bS. H. Mishra, B. J. Harden, D. P. Frueh, *J Biomol NMR* **2014**, *60*, 265–274; cP. Rossi, Y. Xia, N. Khanra, G. Veglia, C. G. Kalodimos, *J Biomol NMR* **2016**, *66*, 259–271.
- [105] C. Amero, P. Schanda, M. A. Durá, I. Ayala, D. Marion, B. Franzetti, B. Brutscher, J. Boisbouvier, *J Am Chem Soc* **2009**, *131*, 3448–3449.
- [106] K. Grohe, E. Nimerovsky, H. Singh, S. K. Vasa, B. Söldner, B. Vögeli, C. M. Rienstra, R. Linser, *ChemComm* **2019**, *55*, 7899–7902.
- [107] R. Linser, B. Bardiaux, V. Higman, U. Fink, B. Reif, *J Am Chem Soc* **2011**, *133*, 5905–5912.
- [108] L. M. I. Koharudin, A. M. J. J. Bonvin, R. Kaptein, R. Boelens, *J Magn Reson* **2003**, *163*, 228–235.
- [109] C. Gerlach, M. Smolinski, H. Steuber, C. A. Sotriffer, A. Heine, D. G. Hangauer, G. Klebe, *Angew Chem Int Ed* **2007**, *46*, 8511–8514.

- [110] P. Paluch, R. Augustyniak, M.-L. Org, K. Vanatalu, A. Kaldma, A. Samoson, J. Stanek, *Front Mol Biosci* **2022**, *9*.
- [111] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular biology of the cell*, 4th ed., Garland Science Taylor & Francis Group, New York, **2002**.
- [112] F. Guerra, A. N. Bondar, *J Membr Biol* **2015**, *248*, 443-453.
- [113] aL. Fagerberg, K. Jonasson, G. von Heijne, M. Uhlén, L. Berglund, *Proteomics* **2010**, *10*, 1141-1149; bJ. P. Overington, B. Al-Lazikani, A. L. Hopkins, *Nat Rev Drug Discov* **2006**, *5*, 993-996.
- [114] P. A. Borea, S. Gessi, S. Merighi, F. Vincenzi, K. Varani, *Physiol Rev* **2018**, *98*, 1591-1625.
- [115] Y. Zhang, K. Wang, Z. Yu, *J Med Chem* **2020**, *63*, 15258-15278.
- [116] S. J. Singer, G. L. Nicolson, *Science* **1972**, *175*, 720-731.
- [117] K. Jacobson, E. D. Sheets, R. Simson, *Science* **1995**, *268*, 1441-1442.
- [118] aD. A. Brown, E. London, *J Biol Chem* **2000**, *275*, 17221-17224; bA. G. Lee, *Biochim Biophys Acta* **2003**, *1612*, 1-40.
- [119] S. Vanounou, D. Pines, E. Pines, A. H. Parola, I. Fishov, *Photochem Photobiol* **2002**, *76*, 1-11.
- [120] B. Lugtenberg, *Trends Biochem Sci* **1981**, *6*, 262-266.
- [121] Z. Zerrouk, S. Alexandre, C. Lafontaine, V. Norris, J. M. Valleton, *Colloids Surf B Biointerfaces* **2008**, *63*, 306-310.
- [122] K. S. Mineev, K. D. Nadezhdin, *Nanotechnol Rev* **2017**, *6*, 15-32.
- [123] J. M. Dörr, S. Scheidelaar, M. C. Koorengel, J. J. Dominguez, M. Schäfer, C. A. van Walree, J. A. Killian, *Eur Biophys J* **2016**, *45*, 3-21.
- [124] Y.-M. Li, M.-T. Lai, M. Xu, Q. Huang, J. DiMuzio-Mower, M. K. Sardana, X.-P. Shi, K.-C. Yin, J. A. Shafer, S. J. Gardell, *Proc Natl Acad Sci U S A* **2000**, *97*, 6138-6143.
- [125] P. C. Fraering, M. J. LaVoie, W. Ye, B. L. Ostaszewski, W. T. Kimberly, D. J. Selkoe, M. S. Wolfe, *Biochem* **2004**, *43*, 323-333.
- [126] aS. Urban, M. S. Wolfe, *Proc Natl Acad Sci U S A* **2005**, *102*, 1883-1888; bM. K. Lemberg, J. Menendez, A. Misik, M. Garcia, C. M. Koth, M. Freeman, *EMBO J* **2005**, *24*, 464-472.
- [127] aM. Freeman, *Annu Rev Genet* **2008**, *42*, 191-210; bM. A. Sturtevant, M. Roark, E. Bier, *Genes Dev* **1993**, *7*, 961-973.
- [128] E. V. Koonin, K. S. Makarova, I. B. Rogozin, L. Davidovic, M. C. Letellier, L. Pellegrini, *Genome Biol* **2003**, *4*, R19.
- [129] M. K. Lemberg, M. Freeman, *Genes Dev* **2007**, *17*, 1634-1646.
- [130] S. Düsterhöft, U. Künzel, M. Freeman, *Biochim Biophys Acta Mol Cell Res* **2017**, *1864*, 2200-2209.
- [131] A. Nejatfard, N. Wauer, S. Bhaduri, A. Conn, S. Gourkanti, N. Singh, T. Kuo, R. Kandel, R. E. Amaro, S. E. Neal, *Cell Rep* **2021**, *37*, 109840.
- [132] G. Jürgens, E. Wieschus, C. Nüsslein-Volhard, H. Kluding, *Wilehm Roux Arch Dev Biol* **1984**, *193*, 283-295.
- [133] aJ. R. Lee, S. Urban, C. F. Garvey, M. Freeman, *Cell* **2001**, *107*, 161-171; bS. Urban, J. R. Lee, M. Freeman, *Cell* **2001**, *107*, 173-182.
- [134] G. A. McQuibban, J. R. Lee, L. Zheng, M. Juusola, M. Freeman, *Curr Biol* **2006**, *16*, 982-989.
- [135] R. P. Baker, R. Wijetilaka, S. Urban, *PLoS Pathog* **2006**, *2*, e113.

- [136] E. W. Sayers, E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, S. T. Sherry, *Nucleic Acids Res* **2022**, *50*, D20-d26.
- [137] R. P. Baker, S. Urban, *Nat Chem Biol* **2012**, *8*, 759-768.
- [138] aY. Wang, Y. Zhang, Y. Ha, *Nature* **2006**, *444*, 179-180; bS. Cho, S. W. Dickey, S. Urban, *Mol Cell* **2016**, *61*, 329-340; cK. R. Vinothkumar, K. Strisovsky, A. Andreeva, Y. Christova, S. Verhelst, M. Freeman, *EMBO J* **2010**, *29*, 3797-3809; dA. Ben-Shem, D. Fass, E. Bibi, *Proc Natl Acad Sci U S A* **2007**, *104*, 462-466.
- [139] Z. Wu, N. Yan, L. Feng, A. Oberstein, H. Yan, R. P. Baker, L. Gu, P. D. Jeffrey, S. Urban, Y. Shi, *Nat Struct Mol Biol* **2006**, *13*, 1084-1091.
- [140] aC. Lazareno-Saez, E. Arutyunova, N. Coquelle, M. J. Lemieux, *J Mol Biol* **2013**, *425*, 1127-1142; bH. Ghasriani, J. K. C. Kwok, A. R. Sherratt, A. C. Y. Foo, T. Qureshi, N. K. Goto, *Biochem* **2014**, *53*, 5907-5915.
- [141] E. Arutyunova, P. Panwar, P. M. Skiba, N. Gale, M. W. Mak, M. J. Lemieux, *EMBO J* **2014**, *33*, 1869-1881.
- [142] A. R. Sherratt, D. R. Blais, H. Ghasriani, J. P. Pezacki, N. K. Goto, *Biochem* **2012**, *51*, 7794-7803.
- [143] aJ. D. A. Tyndall, T. Nall, D. P. Fairlie, *Chem Rev* **2005**, *105*, 973-1000; bS. J. Hubbard, *Biochim Biophys Acta Protein Struct Mol Enzymol* **1998**, *1382*, 191-206.
- [144] Y. Wang, Y. Ha, *Proc Natl Acad Sci U S A* **2007**, *104*, 2098-2102.
- [145] Y. Zhou, S. M. Moin, S. Urban, Y. Zhang, *Structure* **2012**, *20*, 1255-1263.
- [146] C. L. Brooks, C. Lazareno-Saez, J. S. Lamoureux, M. W. Mak, M. J. Lemieux, *J Mol Biol* **2011**, *407*, 687-697.
- [147] R. P. Baker, K. Young, L. Feng, Y. Shi, S. Urban, *Proc Natl Acad Sci U S A* **2007**, *104*, 8257-8262.
- [148] aY. Wang, S. Maegawa, Y. Akiyama, Y. Ha, *J Mol Biol* **2007**, *374*, 1104-1113; bA.-N. Bondar, C. del Val, S. H. White, *Structure* **2009**, *17*, 395-405.
- [149] aO. Engberg, D. Ulbricht, V. Döbel, V. Siebert, C. Frie, A. Penk, M. K. Lemberg, D. Huster, *Sci Adv* **2022**, *8*, eabq8303; bA.-N. Bondar, *J Phys Chem B* **2019**, *123*, 4159-4172.
- [150] A. C. Foo, B. G. Harvey, J. J. Metz, N. K. Goto, *Protein Sci* **2015**, *24*, 464-473.
- [151] K. R. Vinothkumar, *J Mol Biol* **2011**, *407*, 232-247.
- [152] S. M. Moin, S. Urban, *Elife* **2012**, *1*, e00173.
- [153] K. Strisovsky, H. J. Sharpe, M. Freeman, *Mol Cell* **2009**, *36*, 1048-1059.
- [154] S. Urban, D. Schlieper, M. Freeman, *Curr Biol* **2002**, *12*, 1507-1512.
- [155] L. G. Stevenson, K. Strisovsky, K. M. Clemmer, S. Bhatt, M. Freeman, P. N. Rather, *Proc Natl Acad Sci U S A* **2007**, *104*, 1003-1008.
- [156] S. Maegawa, K. Ito, Y. Akiyama, *Biochem* **2005**, *44*, 13543-13552.
- [157] L. Käll, A. Krogh, E. L. Sonnhammer, *J Mol Biol* **2004**, *338*, 1027-1036.
- [158] Y. Xue, S. Chowdhury, X. Liu, Y. Akiyama, J. Ellman, Y. Ha, *Biochem* **2012**, *51*, 3723-3731.
- [159] aJ. W. Harper, K. Hemmi, J. C. Powers, *Biochem* **1985**, *24*, 1831-1841; bJ. C. Powers, C. M. Kam, L. Narasimhan, J. Oleksyszyn, M. A. Hernandez, T. Ueda, *J Cell Biochem* **1989**, *39*, 33-46.
- [160] O. A. Pierrat, K. Strisovsky, Y. Christova, J. Large, K. Ansell, N. Bouloc, E. Smiljanic, M. Freeman, *ACS Chem Biol* **2011**, *6*, 325-335.

- [161] E. V. Wolf, A. Zeißler, O. Vosyka, E. Zeiler, S. Sieber, S. H. Verhelst, *PLoS One* **2013**, *8*, e72307.
- [162] S. Zoll, S. Stanchev, J. Began, J. Škerle, M. Lepšík, L. Peclinovská, P. Majer, K. Strisovsky, *EMBO J* **2014**, *33*, 2408-2421.
- [163] A. Tichá, S. Stanchev, K. R. Vinothkumar, D. C. Mikles, P. Pachi, J. Began, J. Škerle, K. Švehlová, M. T. N. Nguyen, S. H. L. Verhelst, D. C. Johnson, D. A. Bachovchin, M. Lepšík, P. Majer, K. Strisovsky, *Cell Chem Biol* **2017**, *24*, 1523-1536.e1524.
- [164] C. Shi, C. Öster, C. Bohg, L. Li, S. Lange, V. Chevelkov, A. Lange, *J Am Chem Soc* **2019**, *141*, 17314–17321.
- [165] S. Wagner, M. M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Högbom, K. J. van Wijk, D. J. Slotboom, J. O. Persson, J. W. de Gier, *Proc Natl Acad Sci U S A* **2008**, *105*, 14371-14376.
- [166] K. R. Andersen, N. C. Leksa, T. U. Schwartz, *Proteins* **2013**, *81*, 1857-1861.
- [167] J. Marley, M. Lu, C. Bracken, *J Biomol NMR* **2001**, *20*, 71-75.
- [168] A. Tichá, S. Stanchev, J. Škerle, J. Began, M. Ingr, K. Švehlová, L. Polovinkin, M. Růžička, L. Bednárová, R. Hadravová, E. Poláchová, P. Rampírová, J. Březinová, V. Kašička, P. Majer, K. Strisovsky, *J Biol Chem* **2017**, *292*, 2703-2713.
- [169] E. Poláchová, K. Bach, E. Heuten, S. Stanchev, A. Tichá, P. Lampe, P. Majer, T. Langer, M. K. Lemberg, K. Stříšovský, *J Med Chem* **2023**, *66*, 251-265.
- [170] S. W. Dickey, R. P. Baker, S. Cho, S. Urban, *Cell* **2013**, *155*, 1270-1281.
- [171] S. Cho, R. P. Baker, M. Ji, S. Urban, *Nat Struct Mol Biol* **2019**, *26*, 910-918.
- [172] R Core Team, R Foundation for Statistical Computing, Vienna, Austria, **2023**.
- [173] M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, D. F. Hochstrasser, *Methods Mol Biol* **1999**, *112*, 531-552.
- [174] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687-696.
- [175] S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, G. W. Vuister, *J Biomol NMR* **2016**, *66*, 111-124.
- [176] M. Bieri, P. R. Gooley, *BMC Bioinformatics* **2011**, *12*, 421.
- [177] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J Appl Crystallogr* **2007**, *40*, 658-674.
- [178] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J Comput Chem* **2004**, *25*, 1605-1612.
- [179] W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132.
- [180] S. N. Cohen, A. C. Chang, L. Hsu, *Proc Natl Acad Sci U S A* **1972**, *69*, 2110-2114.

List of publications

Kotschy, J., Söldner, B., Singh, H., Vasa, S., Linser, R., Microsecond timescale conformational dynamics of a small-molecule ligand within the active site of a protein, *Angew Chem Int Ed* 2023, e202313947, DOI: 10.1002/anie.202313947.

(and its german translation: Konformationeller Austausch eines Liganden in der Bindetasche eines Proteins auf der Mikrosekundenzeitskala)

Kotschy, J., Linser, R., 2020, "Proton-detected solid-state NMR and its applications to membrane proteins" in "Solid-state NMR - Applications in biomembrane structure", edited by Frances Separovic and Marc-Antoine Sani, pages 15-1 to 15-35, IOP Publishing, DOI: 10.1088/978-0-7503-2532-5ch15.

Pallach, R., Keupp, J., Terlinden, K., Frenzel-Beyme, L., Kloß, M., Machalica, A., **Kotschy, J.**, Vasa, S. K., Chater, P. A., Sternemann, C., Wharmby, M. T., Linser, R., Schmid, R., Henke, S., Frustrated flexibility in metal-organic frameworks, *Nature Comm* 12, 2021, DOI: 10.1038/s41467-021-24188-4.

Conferences and workshops

- 2019 Workshop “XIII ‘Manuel Rico’ NMR School”, Jaca (Spain)
- 2019 Workshop “Good Scientific Practice”, Dortmund
- 2019 Workshop “Tools for Structural Biology of Membrane Proteins”, Hamburg
- 2020 Workshop “Presentation Training for a Virtual Stage: Engaging the Listener in Your Talk”, online
- 2020 Workshop “Fast Forward” workshop, online
- 2020 Workshop “G-NMR school”, Göttingen
- 2020 Research visit at Strisovsky group, Prague (Czech Republic)
- 2021 IMRPS symposium, Dortmund,
contribution: poster presentation
- 2021 Workshop “Statistics for biologists”, Dortmund
- 2022 Conference “NMR a tool for biology”, Paris (France),
contribution: poster presentation
- 2022 Workshop “Scientific writing”, Dortmund
- 2022 Conference “ICMRBS”, Boston (USA),
contribution: poster presentation
- 2022 IMPRS-LM retreat 2022, Borken,
contribution: talk
- 2023 Conference “Tag der Chemie”, Dortmund,
contribution: talk
- 2023 Workshop “GMP”, online
- 2023 IMPRS symposium, Dortmund,
contribution: poster presentation
- 2023 Conference “PSB Symposium 2023”, Grenoble (France),
contribution: poster presentation
- 2023 BioNMR network NRW meeting, Aachen,
contribution: talk

Eidesstattliche Versicherung (Affidavit)

Kotschy, Julia

Name, Vorname

(Surname, first name)

Belehrung:

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

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

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

Die oben stehende Belehrung habe ich zur Kenntnis genommen.

Matrikel-Nr.

(Enrolment number)

Official notification:

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

The submission of a false affidavit is punishable.

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

I have taken note of the above official notification.

Ort, Datum

(Place, date)

Unterschrift

(Signature)

Titel der Dissertation:

(Title of the thesis):

“Structure and dynamics studies of the enzymes hCAII and GIpG via NMR spectroscopy”

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.

Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

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

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

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

Ort, Datum

(Place, date)

Unterschrift

(Signature)