Linköping Studies in Science and Technology
Dissertation No. 1882

Coding to cure

**NMR and thermodynamic software applied to
congenital heart disease research**

Markus Niklasson

Linköping University
Department of Physics, Chemistry and Biology
SE-58183 Linköping, Sweden

Linköping 2017

© Markus Niklasson, 2017

Published articles have been reprinted with permission from the publishers.

Cover: The artwork shows a sinus rhythm with a prolonged QT interval.

Markus Niklasson

Coding to cure

NMR and thermodynamic software applied to congenital heart disease research

ISBN: 978-91-7685-449-5

ISSN: 0345-7524

Linköping Studies in Science and Technology

Dissertation No. 1882

Typeset using L^AT_EX

Printed by LiU-Tryck, Linköping 2017

Till Mor och Far

Wikipedia is the best thing ever.
Anyone in the world can write anything they want about any subject.
So you know you are getting the best possible information.

Michael Scott
The Office

Principal supervisor

Associate professor Patrik Lundström

Department of Physics, Chemistry and Biology, Linköping University

Co-supervisors

Associate professor Malin Lindqvist Appell

Department of Medical and Health Sciences, Linköping University

Associate professor Björn Wallner

Department of Physics, Chemistry and Biology, Linköping University

Assistant professor Katja Petzold

Department of Medical Biochemistry and Biophysics, Karolinska Institutet

Opponent

Docent Anders Malmendal

Center for Molecular Protein Science, Lund University

Examination committee

Professor Lena Mäler

Department of Biochemistry and Biophysics, Stockholm University

Associate professor Christofer Lendel

Department of Chemistry, KTH Royal Institute of Technology

Associate professor Ann-Christin Brorsson

Department of Physics, Chemistry and Biology, Linköping University

Abstract

Regardless of scientific field computers have become pivotal tools for data analysis and the field of structural biology is not an exception. Here, computers are the main tools used for tasks including structural calculations of proteins, spectral analysis of nuclear magnetic resonance (NMR) spectroscopy data and fitting mathematical models to data. As results reported in papers heavily rely on software and scripts it is of key importance that the employed computational methods are robust and yield reliable results. However, as many scientific fields are niched and possess a small potential user base the task to develop necessary software often falls on researchers themselves. This can cause divergence when comparing data analyzed by different measures or by using subpar methods. Therein lies the importance of development of accurate computational methods that can be employed by the scientific community.

The main theme of this thesis is software development applied to structural biology, with the purpose to aid research in this scientific field by speeding up the process of data analysis as well as to ensure that acquired data is properly analyzed. Among the original results of this thesis are three user-friendly software:

COMPASS - a resonance assignment software for NMR spectroscopy data capable of analyzing chemical shifts and providing the user with suggestions to potential resonance assignments, based on a meticulous database comparison.

CDpal - a curve fitting software used to fit thermal and chemical denaturation data of proteins acquired by circular dichroism (CD) spectroscopy or fluorescence spectroscopy.

PINT - a line shape fitting and downstream analysis software for NMR spectroscopy data, designed with the important purpose to easily and accurately fit peaks in NMR spectra and extract parameters such as relaxation rates, intensities and volumes of peaks.

This thesis also describes a study performed on variants of the life essential regulatory protein calmodulin that have been associated with the congenital life threatening heart disease long QT syndrome (LQTS). The study provided novel insights revealing that all variants are distinct from the wild type in regards to structure and dynamics on a detailed level; the presented results are useful for the interpretation of results from protein interaction studies. The underlying research of this paper makes use of all three developed software, which validates that all developed methods fulfil a scientific purpose and are capable of producing solid results.

Populärvetenskaplig sammanfattning

Inom strukturbologi arbetar man bland annat med att karaktärisera proteiners struktur och dynamik. Att undersöka hur ett protein ser ut och uppför sig är viktigt för att förstå sjukdomsprocesser och för att utveckla läkemedel. Vid denna typ av forskning är datorer oersättliga då de används för att analysera data och utföra komplicerade beräkningar. Beroende på vilka metoder som används inom specialiserad forskning kan det finnas en brist på programvara som utför de analyser som förutsätts för att erhålla resultat. Till följd utav detta faller uppgiften att utveckla mjukvara ofta på forskaren själv, vilket kan ge konsekvenser om data inte analyseras på ett korrekt sätt och sedan används vid jämförelser med andra resultat och som grund till slutsatser. Det är därmed viktigt att robusta och tillförlitliga metoder utvecklas och sprids inom den vetenskapliga världen.

Arbetet som presenteras i denna avhandling har fokuserats på utveckling av användarvänlig mjukvara som kan användas inom strukturbologins ramar för att ge noggranna och tillförlitliga resultat. Avhandlingen beskriver fyra olika forskningsprojekt, av vilka tre är mjukvaror. Den första mjukvaran som beskrivs är COMPASS, som har utvecklats med syftet att tillordna signaler i spektra framtagna med kärnmagnetisk resonansspektroskopi (NMR spektroskopi). Genom att korrekt tillordna NMR signaler kan man dra slutsatser på atomnivå i spektra från andra NMR experiment; om signaler däremot är felaktigt tillordnade kan de slutsatser som dras få stora konsekvenser. Det är därmed viktigt att resultaten från denna tillordningsprocess är exakta.

Den andra mjukvaran som beskrivs är CDpal, vars syfte är att passa termodynamisk data till matematiska funktioner och extrahera parametrar som beskriver ett proteins stabilitet. Denna typ av data erhålls då ett veckat protein veckas upp genom tillförseln av värme eller tillsatsen av en denaturerande kemikalie. För en vanlig forskare är passning av termodynamisk data till matematiska modeller oftast en komplicerad process, vilket kan leda till att uppgiften

utförs på ett förenklat sätt som ger undermåliga resultat. CDpal passar termodynamisk data på ett korrekt sätt utan att göra processen krånglig för en vanlig användare.

Den tredje och sista mjukvaran som beskrivs i denna avhandling är PINT som används till att passa signaler i NMR spektra till matematiska funktioner samt extrahera parametrar som beskriver proteiners strukturella och dynamiska egenskaper. Då NMR experiment kan utformas på många olika sätt för att studera proteiner och dataanalysen oftast är komplicerad har PINT utvecklats för att kunna anpassas till analys av en mångfald av principiellt olika NMR experiment.

Samtliga mjukvaror som har utvecklats under arbetet på denna avhandling tillämpades sedan i ett forskningsprojekt där varianter av det kalciumbindande proteinet kalmodulin studerades. Dessa varianter har associerats med den medfödda dödliga hjärtsjukdomen långt QT-syndrom (LQTS) som kan leda till hjärtstillestånd vid bland annat stress och fysisk aktivitet. I vår grundforskning presenteras skillnader och likheter i både detaljerad struktur och dynamik mellan dessa varianter och friskt kalmodulin. Resultaten visar att de olika kalmodulinvarianterna skiljer sig från varandra på detaljnivå och bör orsaka LQTS på olika sätt.

List of publications

Paper I

Fast and Accurate Resonance Assignment of Small-to-Large Proteins by Combining Automated and Manual approaches

Markus Niklasson, Alexandra Ahlner, Cecilia Andresen, Joseph A. Marsh and Patrik Lundström.

PLOS Computational Biology, 2015; 11(1): e1004022

Digital illustration by MN was chosen as journal cover image.

Paper II

Robust and Convenient Analysis of Protein Thermal and Chemical Stability

Markus Niklasson, Cecilia Andresen, Sara Helander, Marie G.L. Roth, Anna Zimdahl Kahlin, Malin Lindqvist-Appell, Lars-Göran Mårtensson and Patrik Lundström.

Protein Science, 2015; 24(12):2055-2062

Digital illustration by MN was chosen as journal cover image.

Paper III

Comprehensive analysis of NMR data using advanced line shape fitting

Markus Niklasson, Renee Otten, Alexandra Ahlner, Cecilia Andresen, Judith Schlagnitweit, Katja Petzold and Patrik Lundström.

Journal of Biomolecular NMR, 2017; <https://doi.org/10.1007/s10858-017-0141-6>

Paper IV

Calmodulin variants of long QT syndrome: the same but different

Markus Niklasson, Cecilia Andresen, Christine Dyrager and Patrik Lundström.

Submitted

Publication not included in the thesis

Biophysical characterization of the calmodulin-like domain of *Plasmodium falciparum* calcium dependent protein kinase 3

Cecilia Andresen, Markus Niklasson, Sofie Cassman Eklöf, Björn Wallner and Patrik Lundström.

PLOS ONE, 2017; 12(7): e0181721

Contribution report

Paper I - (1st author)

MN conceived the idea and developed the majority of the software. PL contributed with code for chemical shift analysis and CA provided experimental data for analysis and the construction of chemical shift databases. JAM provided code for calculation of secondary structure propensity and AA developed the "Label" module. All authors tested and evaluated the software. MN analyzed the majority of data and wrote the paper together with PL with contributions from AA.

Paper II - (1st author)

The project was conceived by MN and PL. MN designed and developed the majority of the software and PL contributed with code for curve fitting routines. CA, SH, MGLR, AZK, MLA and LGM provided data used for the performance evaluation of the software. MN and PL analyzed the majority of data with contributions from CA. MN wrote the paper together with PL.

Paper III - (1st author)

The core idea of PINT was already designed and published when MN started working on its successor. MN identified areas in need of improvement and refinement to appeal to a wider user base. MN designed and developed the majority of the upgraded version and PL contributed with code for line shape fitting and downstream analysis. RO contributed with code for multi-threading and optimization. All authors contributed with data for analysis. MN, RO, CA, AA and PL provided experimental data for analysis and performed thorough tests of the software with contributions from JS and KP. MN wrote the paper together with PL.

Paper IV - (1st author)

The experiments were conceived and designed by PL. MN purified all protein samples and performed the majority of experiments with contributions from CA. PL assigned most of the resonances and analyzed the methyl ^{13}C CPMG experiments. CD performed and analyzed the homology modeling. MN analyzed the majority of data and wrote the paper together with PL with contributions from CA and CD.

Conference contributions

During work of this thesis I attended and presented original scientific research at the following international conferences:

Markus Niklasson, Alexandra Ahlner, Cecilia Andresen, Joseph A. Marsh and Patrik Lundström.

COMPASS: A Software for Accurate and Rapid Backbone Resonance Assignment of Proteins

EMBO Workshop: Magnetic resonance for cellular structural biology, **2014**, Grosseto, Italy.

Markus Niklasson, Alexandra Ahlner, Cecilia Andresen, Joseph A. Marsh and Patrik Lundström.

COMPASS: A Software for Accurate and Rapid Backbone Resonance Assignment of Proteins

18th Annual Conference of the Swedish Structural Biology Network, **2014**, Tällberg, Sweden. **Received Suraj Manrao's travel award for scientific poster.**

Markus Niklasson, Renee Otten, Alexandra Ahlner and Patrik Lundström.

mtPINT: Ultrafast integration of peak intensities in NMR spectra

19th Annual Conference of the Swedish Structural Biology Network, **2015**, Tällberg, Sweden.

Markus Niklasson, Renee Otten, Cecilia Andresen, Alexandra Ahlner and Patrik Lundström.

PINT: The ultimate tool for integration and analysis of peak volumes in NMR spectra

ICMRBS: The 27th International Conference on Magnetic Resonance in Biological Systems, **2016**, Kyoto, Japan.

Acknowledgments

During my PhD studies I have gotten the privilege to work with and alongside kind and brilliant minds. My heartfelt thanks to everyone I have gotten to know at the chemistry division of IFM.

Patrik Lundström, tack för att du tidigt under min grundutbildning öppnade mina ögon för NMR spektroskopi. Ditt engagemang och din pedagogiska förmåga är beundransvärda egenskaper. Den frihet och tillit du har gett mig inom forskningen har fått mig att utvecklas och älska mitt arbete.

Malin Lindqvist Appell och **Björn Wallner**, tack för samarbetet vi har haft och för ert jobb som mina bihandledare.

Katja Petzold, I cannot thank you enough for saving the end of my PhD studies! You are truly an amazing person and scientist. Thank you for the collaboration and inspiring meetings.

Cissi, en mer genomgod människa har jag aldrig träffat. Tack för att du sprider glädje, värme och kunskap. Jag kommer minnas alla stunder vi har dundrat hårdrock på labbet tillsammans.

Alexandra, tack för alla glada stunder och polariserade diskussioner vi har haft tillsammans! Det har varit ett sant nöje att få dela kontor med dig och att alltid ha någon att bolla idéer med.

Tack till alla övriga medlemmar i Lundströms forskargrupp jag lärt känna genom åren, **Ivana**, **Emma**, **Lisa***2, **Sara**, **Sofie**, **Sofia**, **Bitta**, **David**, **Alexander**, **Tove** och **Patricia**.

Johan Dahlén, tack för att jag har fått lära känna dig och fått äran och tilliten att undervisa på labb i dina och Elkes kurser. Din pedagogiska förmåga och optimism gör dig till en inspirationskälla.

Tack till **Joseph, Sara, Marie, Anna, Lasse, Bisse, Renee** och **Judith** för de samarbeten vi har haft tillsammans.

Tack till **Magda, Elke, Helena** och **Maria L** för konversationer, stöd och problemlösning i samband med undervisning.

Thanks to all friends I have gotten the pleasure to know before and during my PhD studies, **Ivana, Emma, Lisa, Madhan, Claudio, Amélie, Emily, Tove, Vishnu, Meri, Marie, Alexander, Marcus, Ottilia, Mattias, Simone, Sara** and **Misi**.

Tack till min älskade **Julia** för all din kärlek, allt ditt stöd och för att du är du och låter mig vara jag!

Tack till min älskade familj! Utan er hade jag inte strävat efter eller lyckats med detta!

Abbreviations and concepts

B₀ - Static magnetic field with z-axis orientation

CaM - calmodulin

CaMBD - calmodulin binding domain

CDI - calcium dependent inactivation

COMPASS - COMputer-aided Matching and Peak ASSignment

CPMG RD - Carr-Purcell-Meiboom-Gill relaxation dispersion

CPVT - Catecholaminergic polymorphic ventricular tachycardia

CSA - chemical shift anisotropy

CSP - chemical shift perturbation

FID - Free Induction Decay

HSQC - heteronuclear single quantum coherence

LQTS - long QT syndrome

LTCC - L-type calcium channel

NMR - nuclear magnetic resonance

NOE - nuclear Overhauser effect

PINT - Peak INTegration

RyR2 - Ryanodine receptor 2

SDF - spectral density function

SSP - secondary structure propensity

Table of contents

Abstract	vii
Populärvetenskaplig sammanfattning	ix
List of publications	xi
Contribution report	xiii
Conference contributions	xvi
Acknowledgments	xix
Abbreviations and concepts	xix
Table of contents	xxi
Introduction	1
Proteins	3
Protein structure	3
Protein dynamics	5
Thermodynamics of protein denaturation	6
Assaying protein stability	8
The EF-hand protein calmodulin	11
Long QT syndrome and calmodulin	12
NMR spectroscopy	17
The fundamentals	18
The chemical shift	19
Resonance assignment	20
Nuclear magnetic relaxation	23

Relaxation mechanisms	25
Chemical exchange	27
The nuclear Overhauser effect	28
The spectral density function	29
Studying protein dynamics with NMR	30
R_1 experiment	31
R_2 and $R_{1\rho}$ experiments	31
The CPMG relaxation dispersion experiment	32
Heteronuclear NOE	34
Results and discussion	35
Paper I: COMPASS	36
Motivation	36
How COMPASS works	37
How COMPASS performs	41
Paper II: CDpal	42
Motivation	43
Curve fitting	43
How CDpal works	45
How CDpal performs	47
Paper III: PINT	47
Motivation	49
How PINT works	49
How PINT performs	53
Paper IV: Calmodulin variants of LQTS	54
Motivation	54
The variants impact on the biophysical properties of CaM	55
Conclusions and future perspectives	59
References	65
Papers	76
Paper I	76
Paper II	91
Paper III	101
Paper IV	113

Introduction

---

*It takes a certain amount of courage to tackle very hard problems in science, I now realise.
You don't know what the timescale of your work will be: decades or only a few years.
Or your approach may be fatally flawed and doomed to fail.
Or you could get scooped just as you are finalising your work.
It is very stressful.*

Venkatraman Ramakrishnan

The scientific community is result oriented. Solid and reproducible results solving research questions and supporting postulations are the end goals and also what is presented in published manuscripts. Authors report on findings and the employed methods of their projects, but the underlying labor is an invisible aspect. Experiments and data analysis may not succeed at the first try, or even at all, and a research project can thus be described as an iterative process.

In NMR spectroscopy, the primarily employed method in this thesis, data analysis is often distinct and can be confusing. Furthermore, it has a steep learning curve as data from different types of experiments cannot be analyzed in the same way. The purposes of the thesis project were to develop and validate computational tools applied to the field of structural biology with NMR spectroscopy as a principal method. The source of motivation was to aid scientists and academic students by developing robust and accurate methods that increase throughput by speeding up the process of data analysis, which often is time-consuming and demanding yet pivotal for the end results. To attain the goals, bottlenecks of the data analysis processes had to be identified and studied in order to improve upon them. It was important that all developed tools were freely available and could be utilized by non-experts as well as to appeal to advanced users in terms of design, functionality and data output. The developed tools were applied to a specific project with the aim to study the structural and dynamical properties of long QT syndrome (LQTS) associated

calmodulin (CaM) variants.

The primary research questions of this thesis are:

1. How can the important but time-consuming resonance assignment process in protein NMR spectroscopy yield accurate results faster than the available alternatives?
(Paper I)
2. How can curve fitting and evaluation of protein denaturation data become a non-complex task for a common user without resorting to subpar methods? (Paper II)
3. Can a platform to analyze NMR spectroscopy data be designed to yield accurate and reproducible results while being user-friendly, streamlined, intuitive and applicable to distinct types of experimental data? (Paper III)
4. Can the differences and similarities in structure and dynamics of CaM variants associated with LQTS explain the phenotype? (Paper IV)

The following chapters will briefly describe proteins and NMR spectroscopy theory useful for understanding the presented papers. This is followed by a chapter motivating the research of each individual paper and summarizing the original results. The thesis ends with concluding remarks and a collection of the published papers.

Proteins

Harker: "Dr. Pauling, how do you manage to have so many good ideas?"

Pauling: "Oh! I just have lots of ideas, and throw away the bad ones."

Excerpt from a letter written by David Harker to Linus Pauling,
the discoverer of the α -helix and β -sheet protein structures.

Proteins are macromolecules that are essential for life. In structural biology they are a common subject for studies, often because of their involvement in diseases. When structural biologists study proteins they are often interested in answering questions such as: *What does the protein look like? How does the protein behave? Which targets does it interact with?* To answer these questions protein structure and dynamics are investigated and interaction studies with potential targets are performed. By studying proteins on a detailed level we can understand their behavior and develop potential drugs to cure or alleviate the burden of diseases.

— Protein structure —

The structure of a protein is an important pillar of its biophysical properties and functionality and it can be subcategorized into primary, secondary, tertiary and quaternary structure, Figure 1. Proteins consist of amino acids successively linked together in long sequences through peptide bonds of a carboxyl and an amine group. The linked amino acids construct the protein backbone consisting of a repeating pattern of amide nitrogen (N^H), C^α and C^O . The amino acid sequence starts at the N-terminal and ends at the C-terminal of the protein, which corresponds to the amine and carboxyl group of the first and last amino acids, respectively. This sequential pattern of amino acids for the entire protein is referred to as the primary structure.

The secondary structure of a protein is characterized by local structural patterns known as α -helices and β -strands; regions lacking these patterns are referred to as random coils or loops. These secondary structure elements are formed through hydrogen bonding between the backbone carbonyl and amide groups of different peptide bonds. In α -helices the described hydrogen bonding is sequential and occurs between amino acid i and $i+4$ with the amino acid side chains directed outwards from the helix. The hydrogen bonding of β -strands occurs between different strands, forming β -sheets that are either parallel or anti-parallel. Biophysical differences of amino acids are determined by the side

Figure 1: **A** The primary structure consists of linked amino acids, here depicted as spheres with one letter code. **B** The secondary structure of a protein is defined by local structures. **C** The protein tertiary structure describes the fold of the protein, here shown for a subunit of human hemoglobin. **D** The quaternary structure is the resulting structure from the association of multiple protein units. Here, four subunits of oxygen bound human hemoglobin, shown in **C**, form the quaternary structure. Panels **C** and **D** were rendered in PyMOL utilizing PDB accession code 1GZX[1].

chains, positioned at the backbone C^α , and the 20 common amino acids can be subgrouped in regard to their side chain properties. Common subgroups include positively (Arg, His and Lys) and negatively charged (Asp and Glu) side chains and polar (Ser, Thr, Cys, Tyr, Asn and Gln) and non-polar (Gly, Ala, Val, Leu, Ile, Pro, Phe, Trp and Met) side chains. The polar amino acids are often located at the surface of the protein due to their hydrophilic nature and, conversely, the non-

polar amino acids are hydrophobic and often situated in the protein core. The backbone will contort as a result of the side chain properties and consequently enable hydrogen bonding within α -helices or planarize β -strands. The degree of contortion is defined through the dihedral angles ϕ , ψ and ω , which are the right handed rotations around the N^H-C^α , $C^\alpha-C^O$ and C^O-N^H bonds, respectively. The dihedral angles of the protein backbone and the side chain interactions give rise to the tertiary structure, a three dimensional shape of the protein. Dihedral angles can be used to predict secondary structure elements and are helpful in combination with other constraints, such as homologous sequences and intramolecular distances, when predicting the protein tertiary structure. If multiple polypeptide chains associate, the resulting structure is called a quaternary structure.

— Protein dynamics —

It is not uncommon to represent a protein using the ball-and-stick model. In fact, it is a useful model despite a huge shortcoming: proteins are not static entities. Proteins are dynamic on different time scales, ranging from atomic vibrations and bond-vector fluctuations of nuclei, through molecular tumbling to protein folding and conformational changes, Figure 2. Through its dynamics a protein can facilitate processes like ligand binding, regulation, transportation and catalysis. It is thus important to study protein dynamics when investigating the function of a protein.

Figure 2: Protein dynamics can be categorized into different time scales, ranging from fast to slow motions, with relevance to different biophysical processes.

X-ray crystallography has for long been a common method of choice for structural determination of proteins but it has recently seen worthy competition from cryo-electron microscopy (cryo-EM) [2]. Both methods enable high resolution studies of proteins in a solid state but cryo-EM enables studies of the protein in different conformations because protein samples are flash-frozen; in X-ray

crystallography the protein is crystallized in a single conformation. Even though both methods can be used to probe protein dynamics, NMR spectroscopy has several advantages for this purpose. NMR spectroscopy experiments can be designed to study protein dynamics on different time scales. Site specific fast dynamics (ps) like NH bond vector fluctuations and molecular tumbling (ps-ns) can be investigated through model-free analysis [3–5]. Slower dynamics (μ s-ms), probed by experiments such as $R_{1\rho}$ and CPMG relaxation dispersion, give information on conformational changes, where regions of the protein exchange with an excited state; this information can be used to study ligand binding, folding, catalysis and allosteric effects [6–8]. Very slow dynamics (ms-days) can be probed through hydrogen-deuterium exchange and report on structural transitions [9]. Whereas other methods such as circular dichroism (CD) and fluorescence spectroscopy can be used to study protein dynamics for the total sample, NMR spectroscopy reveals information on a per nucleus basis. Currently there is no method other than NMR spectroscopy that can provide this structural and dynamical information on an atomic level.

— Thermodynamics of protein denaturation

In its native state, a protein adopts a specific structural ensemble defined by non-covalent intra- and intermolecular forces. These stabilizing forces are hydrogen bonds, ionic bonds, van der Waal's forces and the hydrophobic effect; the first three forces are attractive forces and the latter is the end result from the ordering of water around non-polar nuclei of the protein. By heating (or chilling, for some proteins) the sample or adding a chaotropic agent, the stabilizing forces are disrupted and the protein starts to unfold; this process is called denaturation. The thermodynamical parameters describing protein folding and unfolding are the constituents of Gibbs free energy (G): enthalpy (H), entropy (S) and heat capacity (C_p). Gibbs free energy is a measure that describes if a reaction is spontaneous, $\Delta G < 0$, or not, $\Delta G > 0$. At constant temperature, it can be expressed in terms of the change in enthalpy, ΔH , and entropy, ΔS , according to equation 1. Gibbs free energy can also be expressed to describe the equilibrium between the folded (F) and unfolded (U) states of the protein as seen in equation 2, where R is the universal gas constant, T is the absolute temperature and K_{eq} is the equilibrium constant defined by $K_{eq} = [U]/[F]$. The change in enthalpy, $\Delta H > 0$ for endothermic heat denaturation, equation 3, gives information on the heat taken up by the protein. The change in entropy, ΔS , equation 4, gives information on

whether the system becomes more disordered, $\Delta S > 0$, or ordered, $\Delta S < 0$, for a reaction.

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

$$\Delta G_0 = -RT \ln K_{eq} \quad (2)$$

$$\Delta H = \Delta H_0 + \int \Delta C_p dT \quad (3)$$

$$\Delta S = \Delta S_0 + \int \frac{\Delta C_p}{T} dT \quad (4)$$

As seen from equation 1, for a reaction to be energetically favorable the system must either move towards greater entropy, be exothermic or possess a combination of increased enthalpy and entropy that yields $\Delta G < 0$. For protein denaturation, the most common case is the latter where an uptake of energy is needed to disrupt intramolecular forces and expose hydrophobic surfaces to the solvent. The increase in entropy is accomplished through increased flexibility of the protein in the unfolded state as well as a larger ensemble of possible structural conformations to adopt.

A thermodynamical parameter that makes thermoanalysis difficult, yet important for the calculation of H, S and G, if it is temperature dependent, is the heat capacity, C_p . Heat capacity is the potential of the sample to take up heat energy and can be seen as energy spent for processes other than to increase the temperature of the sample [10]. The change in heat capacity for a protein undergoing denaturation is $\Delta C_p > 0$, if the denaturation is reversible and aggregation does not occur. The heat capacity can be separated into terms of contributions from protein-protein interactions (hydrogen bonding, hydrophobic interactions, salt bridges etc.) and hydration according to equation 5 [11].

$$\Delta C_p = \Delta C_p^{protein-protein} + \Delta C_p^{hydration} \quad (5)$$

The positive net change in heat capacity upon protein denaturation has been a subject of research for many years and this phenomenon can be seen as counterintuitive if $\Delta C_p^{protein-protein}$ would contribute significantly to the change in heat capacity. In this regard, the denaturation would yield a negative change in heat capacity as inter- and intramolecular forces are decreased when comparing a denatured state to a native state of a protein. However, early research showed that the change in heat capacity roughly correlates to the difference in accessible surface area (ΔASA) and size of the protein, and that variance can

be explained by an overestimation of the denatured state of the protein [12]. Furthermore, hydration of exposed areas during denaturation has been proven to be a major contributor to the change in heat capacity [13]. Interestingly, for irreversible protein denaturation, where protein aggregation occurs, the change in heat capacity is negative. Even though the thermodynamics and pathways of protein aggregation are unclear, hydration effects intuitively explain the negative change in heat capacity during this process and was implicated when studying the aggregation of insulin [14].

— Assaying protein stability —

Common techniques for quantification of thermodynamical parameters are the calorimetric techniques isothermal titration calorimetry (ITC) and differential scanning calorimetry (DSC). The primary uses of these methods are interaction studies for ITC and protein stability assays for DSC. NMR spectroscopy can also be applied to various studies of protein stability. Structural changes of the protein at atomic resolution can be monitored while altering pH, temperature or adding denaturants (Urea/GdnHCl). The useful hydrogen deuterium exchange experiment can be employed to study folding, unfolding and hydrophobic surfaces as the change in signal intensities due to solvent exchange can be monitored. Other simple and commonly employed techniques for assaying protein thermal or chemical stability are CD spectroscopy and fluorescence spectroscopy, respectively. In both these techniques the change in signal at a specified wavelength is monitored throughout the denaturation process. For CD spectroscopy, depending on the secondary structure content of the protein, these wavelengths are preferably set to the extremums 222 nm for α -helices or 217 nm for β -strands [15]. In fluorescence spectroscopy, the fluorescence of a sensitive probe, such as tryptophan or an added ligand, is monitored during denaturation. The resulting denaturation profile can be fitted to an appropriate model and evaluated to yield thermodynamical parameters.

Protein denaturation can be described with a simple two-state transition model where the first state is the protein in its native state (N) and the second state is fully denatured (D) protein. This process can either be reversible, $N \rightleftharpoons D$, or irreversible, $N \rightarrow D$, depending on the biophysical properties of the protein. The two-state transition model is the most common model for protein denaturation and the probed signal can be described by a sigmoid function when it is plotted against changes in temperature or concentration of a denaturant. The melting temperature, T_m , or concentration of denaturant, C_m , needed to reach the midpoint of the denaturation process are the most commonly used

Figure 3: A model depicting the process of reversible protein denaturation through an intermediate state.

parameters to describe protein stability. At equilibrium for a two-state transition model, T_m and C_m are the points during denaturation where the protein equally populates the native and denatured states, $[N] = [D]$, which gives $\Delta G_0 = 0$ and $\Delta H = T_m \Delta S$ through equations 1 and 2. In some cases, the protein will denature through an intermediate state (I). Initially intermediate states were described using the "molten globule" model [16]. In this state the protein will lack tertiary structure but still have secondary structure and a weak hydrophobic core. Intermediates are still described loosely with the term "molten globule", even though this model is not sufficient to describe all intermediate states [17]. Nevertheless, the denaturation of a protein through an intermediate state can be described using a three-state transition model, $N \rightleftharpoons I \rightleftharpoons D$, Figure 3, that also can be either reversible or not. For protein denaturation through intermediate states, multiple T_m and C_m values can be extracted, one for each transition. More complex models describing denaturation of multimers can also be employed, such examples include a dimer denaturing through either $N_2 \rightleftharpoons I_2 \rightleftharpoons 2D$ or $N_2 \rightleftharpoons 2I \rightleftharpoons 2D$. It can however be difficult to distinguish and characterize the correct type of model in these cases as a more complex model always can be used to describe a simpler model; this is a big dilemma when evaluating data from fitted models in general. This issue can be avoided by performing calorimetric measurements, not overinterpreting or overfitting the data and ensuring near identical conditions during experiments where comparison of different systems, such as protein variants, is the main aim.

Assuming a two-state transition model, the measured signal, y , can be described using the fractional populations of the native, f_N , and denatured, f_D , states:

$$y(\xi) = f_N(a_N + b_N\xi) + f_D(a_D + b_D\xi) \quad (6)$$

Figure 4: Simulated data of protein thermal denaturation following the three-state transition model $N \rightleftharpoons I \rightleftharpoons D$. The temperature is increased and the monitored signal changes as the protein denatures. The solid line represents the best fitted curve, where N, I and D indicates the plateau of the native, intermediate and denatured state, respectively. The thermal melting points for the $N \rightleftharpoons I$ (T_m) and $I \rightleftharpoons D$ ($T_{m,2}$) transitions have been extracted from the fitted data.

where $(a_x + b_x \xi)$ describes the signal of state x at temperature or concentration of denaturant ξ . The fractional populations are related to the equilibrium constant K_{eq} , given by equation 2, through $f_N = 1/(1 + K_{eq})$ and $f_D = K_{eq}/(1 + K_{eq})$. By fitting the denaturation profile to these equations ΔH_m and T_m (or C_m) can be extracted; an example showing data fitted to a three-state transition model is shown in Figure 4. The downside to this approach is that the thermodynamical parameters extracted from fits of denaturation profiles not necessarily represent the true values. Thermodynamical parameters of a fitted curve that are not true, yet give a good fit, are called apparent parameters. The enthalpy extracted from curve fitted data, often referred to as the van't Hoff enthalpy, is only equal to the true enthalpy if it is based on calculations using the true populations in the denatured and native states [18]. By comparing van't Hoff enthalpies to true enthalpies acquired through calorimetric measurements the extracted parameters from the fitted curve as well as the choice of model for the denaturation (two-state, three-state etc.) can be validated.

— The EF-hand protein calmodulin —

The term EF-hand protein comes from the study of parvalbumin by Kretsinger and Nockolds, where they denoted its six helices A, B, C, D, E and F and symbolized the CD and EF regions of the protein through a pair of right hands [19]. Canonical EF-hand proteins have a repeating pattern of a helix-loop-helix motif, where each motif is referred to as an EF-hand. Each EF-hand can coordinate a Ca^{2+} ion using seven ligands and upon doing so the protein tertiary structure changes from a closed conformation, with parallel α -helices, to an open conformation, with perpendicular α -helices, Figure 5. The seven Ca^{2+} ligands are situated in the loop and its adjoining helix of the EF-hand at amino acid positions I, III, V, VII, IX and XII. Position I, III and V bind Ca^{2+} using a side chain oxygen, position VII uses its backbone carbonyl oxygen, position IX binds a water molecule that in turn coordinates the Ca^{2+} ion and the bidentate binding position XII uses its side chain oxygens.

Figure 5: The C-lobe of CaM has two EF-hands capable of binding one Ca^{2+} ion each. **A** In the Ca^{2+} free state of CaM, also called the closed state, the α -helices of the EF-hands are parallel to each other. **B** When CaM binds calcium the α -helices will undergo conformational change and adopt an open conformation, orienting the α -helices of the EF-hands perpendicular to each other. In this open state CaM can bind target molecules. The arrows in both panels highlight the orientation of the α -helices for each EF-hand. Panels **A** and **B** were rendered in PyMOL utilizing PDB accession codes 1CMF[20] and 1FW4[21], respectively.

Calmodulin (CaM) is a highly conserved, ubiquitous and life essential EF-hand protein serving a purpose as a Ca^{2+} modulating sensor with a reg-

ulatory role in several biological processes including ion channel assembly/disassembly, gating of ion channels and muscle contraction. Its importance is reflected in being encoded by three different genes (CALM 1-3) located on three different chromosomes [22]. CaM consists of four EF-hands connected with short linkers and can bind a total of four Ca^{2+} ions with micromolar affinity. The different EF-hands of CaM have different Ca^{2+} affinities with EF-hand III and IV being the high affinity sites and Ca^{2+} binding is positive cooperative both within and between EF-hands [23]. In its Ca^{2+} bound state, illustrated in Figure 5B, CaM reorients its α -helices forming an aromatic cluster and exposing hydrophobic methionines that constitute binding pockets recognizing hundreds of different target proteins [20]. Among these targets are important myocyte targets including voltage-gated $\text{Na}_v1.5$, $\text{K}_v7.1$, $\text{Ca}_v1.2$ and RyR2 ion channels, responsible for the cellular influx and efflux of ions. However, the binding of CaM to targets is diverse as it is known that Ca^{2+} free CaM can preassociate to targets [24]. One study hypothesized that at low Ca^{2+} levels CaM is tethered to the the important L-type Ca^{2+} channels (LTCC) by binding to the cytoplasmic tail of the ion channel with both the N-lobe and the C-lobe; as the C-lobe binds Ca^{2+} it enables binding to the IQ motif of the ion channel [25].

— Long QT syndrome and calmodulin —

As many cases of sudden cardiac arrests go unresolved, CaM has found an emerging role in the causality as variants involved with the congenital heart diseases long QT syndrome (LQTS) and catecholaminergic polymorphic ventricular tachycardia (CPVT) have been discovered [26–30]. LQTS, first described by Jervell and Lange-Nielsen in 1957 [31], is a rare heart condition, affecting 1 in 2000 individuals, that causes a prolongation of the QT interval of the sinus rhythm [32], Figure 6. In the worst case scenario, the phenotype can lead to cardiac arrest with death or brain damage as consequences; the common triggers are gene specific and include stress, exercise and even sleep [33]. 75% of LQTS patients have a mutation in a gene encoding the pore forming α subunit of either the slow or rapid potassium channels $\text{K}_v7.1$ and $\text{K}_v11.1$, respectively, or the sodium channel $\text{Na}_v1.5$ [32]. Both $\text{K}_v7.1$ and $\text{K}_v11.1$ are voltage-gated channels responsible for efflux of potassium during the repolarization phase of the action potential; $\text{Na}_v1.5$ is also a voltage-gated channel but responsible for the influx of sodium ions during the depolarization phase. $\text{K}_v7.1$, $\text{K}_v11.1$ and $\text{Na}_v1.5$ are encoded by the KCNQ1, KCNH2 and SCN5A genes, respectively, and patients with mutations in these genes are diagnosed with LQT1, LQT2 or LQT3, respectively. LQT1 and LQT2 are different from LQT3 as increased activity of the sympathetic

Figure 6: A sinus wave of a healthy person and that of an individual afflicted by LQTS. The QT interval is prolonged as the repolarization (T wave) is delayed.

nervous system (e.g. during stress and exercise) triggers the phenotype for both LQT1 and LQT2 patients, whereas cardiac events for LQT3 patients are triggered during rest [34]. Consequently, the efficiency of treatment also differs between LQTS patients as β -blocker therapy is effective for LQT1 and LQT2 patients, with symptoms triggered by adrenergic stimuli, but not for LQT3 patients [34]. CaM has been shown to bind to the C-terminus of $K_v7.1$, contributing to both channel assembly and regulation of the $K_v7.1$ current [35]. Mutations in the IQ binding motif of the *KCNQ1* gene has been shown to impair CaM binding and the functionality of $K_v7.1$, this study also showed that CaM interacts with $K_v7.1$ in both presence and absence of Ca^{2+} [36]; there is no study showing direct CaM regulation of $K_v11.1$. $Na_v1.5$ has two sites that can bind CaM, one IQ motif located in the C-terminus and a CaM binding domain (CaMBD) located in a helix linker adjacent to a three amino acid motif (IFM) responsible for the gate inactivation of $Na_v1.5$. At low cytosolic Ca^{2+} levels CaM will bind to the IQ motif using its C-lobe. As Ca^{2+} levels rise CaM can bind either binding site of $Na_v1.5$, whereas the N-lobe of CaM can bind the IQ motif. By binding the IQ motif with the N-lobe and the CaMBD with the C-lobe, CaM will reduce the ability of $Na_v1.5$ to inactivate, thus keeping the gate open for sodium influx and extending the depolarization period [37]. The effect a mutation in the *SCN5A* gene has depends on whether it increases or decreases CaMBD affinity to CaM.

In 2013, a study was presented that revealed three novel CaM variants that were associated with LQTS (D95V, D129G and F141L) [27]. These variants are en-

coded by CALM1 (CaM^{D129G} and CaM^{F141L}) and CALM2 (CaM^{D95V}) genes, which interestingly are the genes that have the lowest relative expression of mRNA in myocyte cells, with CALM3 being responsible for almost twice the collective expression of mRNA produced from CALM1 and CALM2 [27]. Even so, the mutations are autosomal dominant, meaning that a single mutated gene out of six is sufficient to cause the phenotype. In CaM, D95 and D129 are located in positions III and I of the Ca²⁺ binding loops, respectively, and F141 is situated next to the bidentate binding position XII. Furthermore, F141 is a constituent of the aromatic cluster formed upon Ca²⁺ binding. The Ca²⁺ affinities of the variants were reduced to different degrees (wild type > F141L > D95V > D129G) and it was predicted that this would affect their regulatory functionality. To elucidate the mechanisms of the disease in respect to CaM, several interaction studies with the CaM variants were performed over the years following their discovery. It was found that these CaM variants had no significant effect on the I_{Na} current from the Na_v1.5 channels but impaired Ca²⁺ dependent inactivation (CDI) of L-type Ca²⁺ channels (LTCC) [38], responsible for regulating the Ca²⁺ influx during the action potential. A similar study published a few months prior also demonstrated that the CDI was impaired due to the mutations and clearly showed that the trend of the elevated surge of Ca²⁺ into the myocytes followed the order of decreased Ca²⁺ affinity for the CaM variants [39]. Furthermore, the study revealed that the mutants in Ca²⁺ free state could bind LTCC channels at least as well as wild type CaM. A recent study focused on the F141L variant and showed that the mutation did not affect the I_{Ks} current caused by the K_v7.1 channels but impaired CDI of LTCC [40]. As CaM binds several targets in its Ca²⁺ free state, it was suggested that pre-associated CaM^{F141L} could compete with wild type CaM for the modulation of LTCCs even though it exists in an unfavorable ratio.

CPVT is, like LQTS, a cardiac channelopathy where 50% to 65% of the patients have mutations in the ryanodine receptor 2 (RyR2) [41]. RyR2 is an ion channel, located in cardiac myocytes, responsible for driving muscle contraction by releasing Ca²⁺ from the endoplasmic and sarcoplasmic reticulum [42]. This ion channel will open and close in response to Ca²⁺ and CaM is a known regulator that can associate to RyR2 at both low and high Ca²⁺ levels with both the N- and C-lobe [43]. However, CaM only has an inhibiting effect on RyR2, which is peculiar as CaM activates the isoform RyR1 at low Ca²⁺ levels. Interaction studies of RyR2 and LQTS associated CaM variants (D95V, D129G and F141L) showed that CaM^{F141L} interacts with the CaM binding domain of RyR2 in a structurally different way than the wild type but still have inhibitory effects at low and high Ca²⁺ levels, whereas the inhibitory effects of CaM^{D95V} and CaM^{D129G} were markedly decreased [44, 45]. From this it is evident that the underlying mecha-

nisms of the idiopathies LQTS and CPVT are complex and could possibly be connected when variant CaM is involved. Also, given the promiscuity of CaM, there may be many different phenotypic pathways to LQTS. The performed studies on these novel CaM variants have been focused on the determination of Ca^{2+} affinities of the N- and C-lobes and interactions with target myocyte candidates. To provide evidence from another perspective we performed a detailed study on protein structure and dynamics for these CaM variants, presented in paper IV of this thesis.

NMR spectroscopy

As the attendants slowly rolled him into the cylindrical bore of the MRI instrument, Rabi found himself surrounded by a reflecting surface.

In that surface he saw a distorted image of himself.

"It was eerie," Rabi told me later. "I saw myself in that machine.

I never thought my work would come to this."

Excerpt from "Rabi, Scientist and Citizen" by John S. Rigden.

The author recalls a conversation with Isidor Isaac Rabi.

The origin of Nuclear Magnetic Resonance (NMR) spectroscopy dates back to late 1930's when Isidor Isaac Rabi *et al.* presented a method to re-orient the magnetization of lithium and chlorine in a LiCl beam [46], later awarded with the Nobel Prize in Physics in 1944. In 1945-1946, Felix Bloch and Edward Mills Purcell independently and simultaneously presented research on this effect in water and solid paraffin [47, 48], respectively, and later received a joint Nobel Prize in Physics in 1952. The reported findings mark the start of a scientific paradigm shift.

Today the NMR phenomenon is frequently used in medicine (MRI and metabolomics) and chemistry (structural and dynamical characterization of molecules, identification and assessment of purity). Important to this thesis is protein NMR spectroscopy which allows for studies of soluble proteins at atomic resolution. The method is applicable to study protein structure, protein-ligand interactions and protein dynamics at different time scales. No other scientific method allows for studying protein dynamics at this resolution, which makes NMR spectroscopy an invaluable tool for biophysical characterization studies of proteins. The main disadvantage of the method is poor sensitivity that necessitates high concentration and quality of the studied protein samples. Keeping the native state of proteins in high concentration is often cumbersome, especially as buffers used in NMR spectroscopy must meet certain requirements. The poor sensitivity of NMR spectroscopy also puts an upper size limit

on proteins; the largest protein successfully studied with NMR spectroscopy to date is the monomeric 82 kDa enzyme malate synthase G [49]. The development of new methods, such as sparse sampling [50] and the use of perdeuteration for resonance assignments [51], has tremendously reduced experiment acquisition times and pushed the size limit upwards for NMR experiments, respectively, and are testament to the importance of improving the methodology. The theory explained in this chapter partly follows that of Keeler [52].

— The fundamentals —

NMR spectroscopy can be briefly explained as applying radio frequency pulses, sampling frequencies in a given range, in specific manners to a sample in a strong magnetic field and detecting the induced current in a coil surrounding the sample over time. Processing the acquired signal with Fourier transformation will convert the signal from the time domain to the frequency domain and thus yield an NMR spectrum that can be analyzed.

NMR spectroscopy exploits that different nuclei have different properties, namely nuclear spins, which are used to design experiments and distinguish signals in recorded spectra. The spin is decided by the nucleus composition of protons and neutrons and in order for a nucleus to be NMR active, the spin quantum number, I , which can be described as angular momentum, must be non-zero. The lowest and preferred spin quantum number that a nucleus can possess is $\frac{1}{2}$. For $I = \frac{1}{2}$ nuclei, there are two possible spin states $m = \pm\frac{1}{2}$, called spin-up and spin-down, as $m = \{-I, -I+1, \dots, I-1, I\}$. A spinning nucleus will generate a magnetic field and from this we can define the magnetic moment, μ , through:

$$\mu = \gamma I \quad (7)$$

where γ is the gyromagnetic ratio, unique to different types of nuclei. For a sample, i.e. a protein sample, the net sum of all nuclear magnetic fields will be averaged to zero as the molecules tumble and adopt random orientations. However, when a strong external magnetic field, B_0 , is applied the nuclear spins will either align with or against it according to the Boltzmann distribution:

$$\frac{N_\alpha}{N_\beta} = e^{\Delta E/(k_B T)} \quad (8)$$

where N_α and N_β are the number of spins in the higher and lower energy states, respectively, k_B is the Boltzmann constant, T is the temperature and ΔE is the energy difference between the different states. The lower energy state is slightly favorable and will give rise to a net magnetization vector, directed along B_0 , with

energy proportional to the strength of the magnetic field, $E = -\mu \cdot B_0$. The nuclei will precess about B_0 with a frequency known as the Larmor frequency:

$$\nu_0 = \frac{\gamma B_0}{2\pi} \quad (9)$$

By applying a radio frequency pulse the net magnetization is tilted away from

Figure 7: A 90° radio frequency pulse has perturbed the net magnetization vector, initially aligned with B_0 (z-axis), bringing it into the transverse plane (XY plane). The net magnetization vector will precess about B_0 in the shape of a cone with the Larmor frequency until thermal equilibrium is reached through the phenomenon called relaxation.

alignment with the external magnetic field and will precess in the shape of a cone, Figure 7, with the Larmor frequency around B_0 until thermal equilibrium has been reestablished through the phenomenon known as relaxation (*vide infra*). During precession an induced current is detected in a coil, reporting on all nuclei in the sample; this signal is called the free induction decay (FID). By Fourier transforming the signal arising from the precession throughout the relaxation process, the Larmor frequencies can be recovered.

— The chemical shift —

Electrons surrounding nuclei will give rise to local induced magnetic fields, B_{ind} , opposing the external magnetic field, B_0 . As a result, a nucleus will only experience the difference of $B_0 - B_{ind}$, which will be different for every nucleus in the protein depending on the local electron density. The resulting Larmor frequency

can thus be expressed as:

$$\nu = \frac{\gamma B_0(1 - \sigma)}{2\pi} \quad (10)$$

where σ is the shielding factor. Consequently, Larmor frequencies for a type of nucleus in a molecule can differ substantially from each other due to this shielding effect. The Larmor frequency unit is not practical to use; instead NMR spectroscopists convert and report the Larmor frequencies as chemical shifts in the unit parts per million (ppm), δ . The chemical shift is calculated by comparing the measured Larmor frequency, ν , to that of a reference compound, ν_{ref} , as seen in equation 11.

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} 10^6 \text{ ppm} \quad (11)$$

A benefit from reporting the chemical shift instead of Larmor frequencies is that experiments acquired utilizing different magnetic field strengths can be directly compared.

In protein NMR spectroscopy, frequently analyzed types of nuclei with distinguishable chemical shifts are ^{15}N , $^1\text{H}^N$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}^\text{O}$. While it would be convenient to be able to predict these chemical shifts it can be very difficult. If the secondary structure is known, different ranges for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts can be expected, but large deviations can occur [53]. SPARTA+ is a useful tool for chemical shift prediction that not only considers protein secondary structure but also includes analysis of effects due to side-chain conformations, hydrogen bonding and more [54]. $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts are among the easiest chemical shifts to estimate and distinguish for amino acids as these nuclei are mostly affected by the backbone conformation. ^{15}N and $^{13}\text{C}^\text{O}$ chemical shifts are the hardest to calculate as they are affected by the adjoining amino acid, side chain composition and hydrogen bonding [55]. $^1\text{H}^N$ chemical shifts are also hard to calculate, being affected by a sum of different effects, for example, the ring current effect and hydrogen bonding. To be able to pinpoint the chemical shifts of each nucleus in a spectrum NMR spectroscopists employ a method known as resonance assignment.

— Resonance assignment —

Assigning the resonances of the protein backbone is essential for studying a protein with NMR spectroscopy. It is thus typically one of the first performed tasks when novel studies of a protein are conducted. As the number of observed signals in an NMR spectrum roughly is proportional to the size of the protein, one-dimensional or even two-dimensional NMR experiments are not sufficient

to extract the needed data for resonance assignment. The underlying reason is spectral crowding, where peaks coalesce and consequently cause difficulties when distinguishing the chemical shifts. In 1990 the classic work by Mitsuhiro Ikura, Lewis E. Kay and Ad Bax introduced 3D NMR spectroscopy for sequential assignment of the protein backbone [56]. The α -helical protein calmodulin was too difficult to assign using prior 2D methods as signals were heavily overlapped. By exploiting the through-bond J couplings, J_{CC} and J_{CN} , spectra could be resolved into three dimensions, correlating ^1H , ^{15}N and ^{13}C nuclei; this was truly ground breaking as the entire backbone of calmodulin now could be sequentially assigned. In the following years several additional triple resonance experiments were introduced [57–60] and collectively enabled studies of larger proteins. The triple resonance experiments call for an additional requirement of the protein sample, ^{15}N and ^{13}C isotopic labeling of backbone nuclei, which increases production cost. In the case of studying mid-sized to larger proteins, that often suffer from sensitivity issues due to relaxation properties, the employment of transverse relaxation-optimized spectroscopy (TROSY) variants of the three-dimensional experiments and deuteration can increase sensitivity [61–63].

A commonly employed battery of three-dimensional experiments for resonance assignment include HNCO, HN(CA)CO, CBCA(CO)NH and HNCACB, with nuclei within parentheses not being detected but used for transferring magnetization during the experiments. HNCO detects correlations between the chemical shifts of $^{15}\text{N}_i$, H_i^{N} and $^{13}\text{C}_{i-1}^{\text{O}}$ and similarly CBCA(CO)NH detects $^{15}\text{N}_i$, H_i^{N} , $^{13}\text{C}_{i-1}^{\alpha}$ and $^{13}\text{C}_{i-1}^{\beta}$ chemical shifts. HN(CA)CO and HNCACB detects correlations between the same chemical shifts as HNCO and CBCA(CO)NH, respectively, and additionally those of $^{13}\text{C}_i^{\text{O}}$, $^{13}\text{C}_i^{\beta}$ and $^{13}\text{C}_i^{\alpha}$. Combining these experiments, it is thus possible to distinguish between i and $i - 1$ type of chemical shifts and sequentially connect the signals for the entire protein, barring proline residues that lack the backbone amide proton. To assign proline residues NOE data of H^{δ} and H^{α} can be used as well as methods that correlate H^{δ} and H^{α} chemical shifts to the preceding residue [64, 65].

A central problem in resonance assignment is to solve ambiguities arising from amino acids having similar chemical shifts. As an example, the individual $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\beta}$ chemical shifts of Glu, Gln, His, Arg, Cys and Trp may be similar and could possibly be assigned to multiple positions in the protein sequence. This problem is resolved if their chemical shifts can be sequentially connected to construct a fragment with an unambiguous amino acid pattern. Resonance assignment becomes increasingly more difficult if several chemical shifts of an amino acid are missing, due to line broadening. This is because the missing

Figure 8: The figure shows the comparison of heteronuclear ^{15}N - ^1H NOE data (hetNOE) for the C-domain of calmodulin in the presence of Ca^{2+} when using assignments (left panel) and not (right panel). When using assignments, the hetNOE data show clear trends and can be compared to the protein topology (on top of the left panel) to distinguish flexible regions of the protein (low hetNOE values) from more rigid regions (high hetNOE values). The right panel shows valid hetNOE data but as assignments are missing the data is presented in a random order without indicating any trends and no specific conclusions can be drawn from this data.

chemical shifts cannot be used for sequential assignment and that the risk of ambiguity increases when assignments are based on few chemical shifts. If the protein is still stable and the protein dynamics do not increase with temperature, a possible remedy is to record the experiments at increased temperatures, which would increase molecular tumbling and suppress relaxation causing the line broadening.

Backbone resonance assignments are so important that results from other experiments would be rendered virtually useless without them. They pinpoint the exact amino acid of a protein and provide information used in other NMR experiments to locate regions of dynamics on different time scales, conformational change during ligand binding, environmental changes during solvent exchange and more. The assigned chemical shifts can also be used to predict a secondary structure [66, 67] and are useful constraints in protein structure modeling [68, 69]. The importance of accurately assigning resonances should by extension become evident as any conclusion drawn from other experiments, using erroneous resonance assignments, would be incorrect. A prime example of both the importance of assignments and their correctness is shown in Figure 8.

— Nuclear magnetic relaxation —

After his and Purcell's initial research that both had successfully detected magnetic resonance, Felix Bloch published a classical mathematical model known as the Bloch equations to describe the time dependence of the individual Cartesian components constituting the net magnetization [70]; it was later improved upon by Harden M. McConnell to include effects arising from chemical exchange (*vide infra*) [71]. Bloch introduced the term *relaxation* to describe the return of net magnetization to thermal equilibrium. He divided the relaxation process into longitudinal relaxation, a build-up of magnetization along B_0 , and transverse relaxation, the decay of transverse magnetization perpendicular to B_0 , and recognized that transverse relaxation could be several orders of magnitude faster than longitudinal relaxation.

After the net magnetization vector is perturbed by a radio frequency pulse it will eventually relax back to thermal equilibrium. For both longitudinal and transverse relaxation molecular motions will generate random fluctuating magnetic fields resulting in energy transfer between different spins and the spins and the lattice. Longitudinal relaxation is most sensitive to motions at the Larmor frequency whereas transverse relaxation is sensitive to all random motions but to different degrees. Just like the radio frequency pulse, sampling Larmor frequencies, causes transitions between energy levels for the entire sample, local magnetic fields with similar frequencies will cause random transitions of local regions in the molecule and can thus be seen as locally applied radio frequency pulses. This explains longitudinal relaxation but only partly explains transverse relaxation. Transverse relaxation is much more complex and is in essence the loss of coherence over time. After the application of a radio frequency pulse, bringing the net magnetization to the transverse plane, the spins will precess about B_0 with Larmor frequency. The spins have small differences in Larmor frequencies, which will cause them to dephase during time τ , but the spins can be refocused by applying a 180° pulse and waiting time τ again. However, in addition to the energy transitions described earlier, local differences in the magnetic field, chemical exchange (*vide infra*) and collisions between molecules will cause an irreversible loss of coherence and will subsequently reduce the net magnetization vector in the transverse plane; this is transverse relaxation.

To describe longitudinal and transverse relaxation their respective time constants T_1 and T_2 , often reported as rate constants $R_1 = T_1^{-1}$ and $R_2 = T_2^{-1}$, are used. T_1 and T_2 describe the build-up, Figure 9, and decay, Figure 10, of magnetization, respectively, and as transverse relaxation is affected by the same mechanisms as longitudinal relaxation plus additional factors T_1 is often longer

than T_2 . The T_1 and T_2 relaxation times are experimentally important as T_1 determines the time that we must wait for thermal equilibrium to reestablish before applying another pulse sequence, often set to $5T_1$, and T_2 determines the sensitivity of the experiment, reflected in the resulting line width of a peak that is equal to R_2/π at half height.

Figure 9: During longitudinal relaxation thermal equilibrium is reestablished according to the Boltzmann distribution. Here longitudinal relaxation is exemplified for an inversion recovery experiment where the net magnetization vector M_z immediately following a 180° pulse returns to thermal equilibrium over time.

Figure 10: For transverse relaxation the net magnetization vector in the transverse plane, M_{xy} , will decrease over time when the spins, making up the net magnetization, irreversibly lose coherence, represented by the arrows in the bottom panels.

The return to equilibrium for longitudinal relaxation, for magnetization in the transverse plane, is described by:

$$M_z(t) = M_0(1 - e^{-t/T_1}) \quad (12)$$

where M_0 is the equilibrium magnetization vector and T_1 describes the time elapsed for 63% of the original magnetization aligned with B_0 to be recovered. The magnetization vector, M_{xy} , aligned with the transverse plane decreases exponentially as a function of time according to:

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2} \quad (13)$$

where T_2 describes the time elapsed when only 37% of the transverse magnetization vector following the initial radio frequency pulse remains.

— Relaxation mechanisms —

As previously mentioned, relaxation occurs partly due to random fluctuating magnetic fields generated by molecular motion. This relaxation can be subdivided into different relaxation mechanisms, with different efficiency and relevance to protein NMR spectroscopy. In protein NMR spectroscopy, where large molecules with spin $\frac{1}{2}$ nuclei (often) are studied, the most efficient relaxation mechanisms are dipole-dipole interaction and chemical shift anisotropy (CSA). These and the relaxation mechanisms scalar coupling, spin rotation and quadrupolar coupling are briefly discussed in this section.

Dipole-dipole interaction

The importance of dipole-dipole relaxation was recognized early [72, 73]. Dipole-dipole interactions occur between different nuclei and are sensitive to internuclear distance and orientation of the internuclear vector in respect to the external magnetic field. Each nucleus with a magnetic moment will generate a local fluctuating magnetic field that will be experienced by nearby nuclei. As the molecule tumbles, the local magnetic fields will constantly be reoriented causing relaxation. The efficiency of the dipole-dipole relaxation is in essence determined by $\gamma_I^2 \gamma_S^2 / r_{IS}^6$, where γ is the gyromagnetic ratio and r_{IS} is the internuclear distance. From this it is evident that dipole-dipole relaxation for a proton-proton pair is approximately 100 times more efficient than for an ^{15}N -proton pair and that the distance between the nuclei has high impact on efficiency.

Chemical shift anisotropy

The chemical shift is a tensor and its components report on the chemical shift

relative to the direction of B_0 :

$$\delta = \begin{bmatrix} \delta_{XX} & \delta_{XY} & \delta_{XZ} \\ \delta_{YX} & \delta_{YY} & \delta_{YZ} \\ \delta_{ZX} & \delta_{ZY} & \delta_{ZZ} \end{bmatrix} \quad (14)$$

The average of the diagonal elements δ_{XX} , δ_{YY} and δ_{ZZ} equates to the reported chemical shift and is referred to as the isotropic chemical shift when the values of all diagonal element are equal:

$$\delta_{iso} = \frac{\delta_{XX} + \delta_{YY} + \delta_{ZZ}}{3} \quad (15)$$

Nuclei in large molecules like proteins rarely have isotropic chemical shifts due to constant tumbling and structural asymmetry. According to Haeberlen notation [74], δ_{ZZ} refers to the component with largest deviation from the isotropic value according to:

$$|\delta_{ZZ} - \delta_{iso}| \geq |\delta_{XX} - \delta_{iso}| \geq |\delta_{YY} - \delta_{iso}| \quad (16)$$

$|\delta_{ZZ} - \delta_{iso}|$ is thus a measure of CSA, δ_{CSA} , and if the chemical shift is anisotropic it will contribute to transverse relaxation. It is also seen that CSA correlates with the chemical shift and will thus have a high contribution to relaxation for nuclei such as $^{13}\text{C}^O$ and ^{15}N . CSA scales with B_0^2 and becomes a large source of relaxation at higher magnetic fields. For this reason it can be useful to record certain experiments, such as HNC0 and HN(CA)CO, at lower magnetic fields, e.g. 500 MHz, to increase the sensitivity.

Scalar coupling

The scalar coupling, also known as J coupling or indirect dipole dipole coupling, is a through-bond interaction where spins of adjacent nuclei polarize each other [75, 76]. This coupling is exploited to transfer magnetization between nuclei and correlating them but can also be a source of relaxation, namely scalar relaxation. Scalar relaxation is divided into two categories, scalar relaxation of the first and second kind, and it requires that affected nuclei are scalar coupled. Scalar relaxation of the first kind happens because of fast chemical exchange and scalar relaxation of the second kind happens when a coupled nucleus has a fast T_1 relaxation [77].

Spin rotation

Spin rotation is a relaxation mechanism only relevant for smaller molecules and in gas-phase, where magnetic fields generated from molecular rotation relax

nearby nuclei. Relaxation through spin rotation is enhanced when molecules collide, making spin rotation the only relaxation mechanism that increases in efficiency with temperature [78].

Quadrupolar coupling

Quadrupolar coupling is a dominating relaxation mechanism but only active for nuclei with $I > \frac{1}{2}$. The charge distribution in these nuclei is not symmetric and neither is the rotation; this will introduce an electric field gradient that will cause relaxation. In protein NMR spectroscopy it is relevant for deuterated protein samples and ^{14}N nuclei, as $I = 1$.

— Chemical exchange —

Proteins are dynamic and can adopt different conformations, as needed for processes like motility and target recognition. In protein NMR spectroscopy, conformational exchange can for example be directly seen in spectra during titration experiments and in results from the CPMG RD experiment (*vide infra*). It is an important process to study as it reports on structural changes in response to added ligand, potential binding sites and regions of flexibility. The underlying reason to the spectral changes in chemical shifts, peak intensities and linewidths, is chemical exchange.

Consider the simplest case where a nucleus in a protein can exist in two different states, A and B; chemical exchange is the intermittent change between these states. The states are chemically different and will thus yield different chemical shifts and relax with different rates. The exchange rate, k_{ex} , and populations of state A and B, p_A and p_B , are defined as:

$$k_{ex} = k_A + k_B \quad (17b)$$

$$p_A = \frac{k_B}{k_A + k_B} \quad p_B = \frac{k_A}{k_A + k_B} \quad (17c)$$

where the ground state A has the higher population, p_A , and the excited state B has the lower population, p_B . The magnitude of the exchange rate characterizes the chemical exchange as slow, intermediate or fast and will affect peaks in the spectrum differently. For slow exchange, $k_{ex} \ll |\Delta\omega|$, where $|\Delta\omega|$ is the absolute difference in angular frequency between states A and B, the nucleus will yield two separate signals in the NMR spectrum and for fast exchange,

$k_{ex} \gg |\Delta\omega|$, a single signal with a population weighted average will show in the NMR spectrum. Intermediate exchange, $k_{ex} \approx |\Delta\omega|$, is the regime between slow and fast exchange and the signal in the spectrum depends on the populations of the different states [79]. The signal can thus resemble either slow or fast exchange or be a mixture of these regimes; peaks with intermediate exchange are often extensively line broadened. The contribution to the transverse relaxation rate from chemical exchange depends on the exchange rate as well. Peaks of slowly exchanging nuclei are exchange broadened due to loss of coherence caused by altered populations; for fast exchanging nuclei in the free precession limit the exchange broadening is described by:

$$R_{ex} = \frac{p_A p_B \Delta\omega^2}{k_{ex}} \quad (18)$$

For fast chemical exchange, R_{ex} thus scales quadratically with the magnetic field strength, as $\Delta\omega \propto B_0$; for slow chemical exchange R_{ex} is not dependent on the magnetic field strength.

— The nuclear Overhauser effect —

In 1953 Albert W. Overhauser proposed a method to polarize nuclei in metals by irradiating electron spins [80], this method came to be known as the Overhauser effect. The method was developed into irradiating nuclear spins instead of electron spins [81] and was named the nuclear Overhauser effect (NOE). NOE is today applied to studies of inter- and intra-molecular interactions in proteins and one of the most widely used adaptations is the 2D heteronuclear NOE experiment (*vide infra*) [82] that enables studies of fast motions (ps - ns). To describe NOE, consider a nucleus, A , with two spin states, α and β , interacting with another nucleus X , also with two spin states, through dipole-dipole interaction. Saturating the spin states of nucleus A , making the populations of spin states α and β equal, will cause the nucleus to relax back to equilibrium populations. During dipole-dipole relaxation changes in spin states of nucleus A are transferred to nucleus X . The spin populations and thus intensities of nucleus B will either increase or decrease as a result, depending on the dominating relaxation pathway, Figure 11. The ω_2 pathway is favorable for small and fast tumbling molecules whereas for larger molecules, such as proteins, ω_0 is primarily used. This phenomenon can for example be used to locate regions of flexibility and yield important through space constraints used in protein structure modeling.

Figure 11: An energy diagram showing the different transitions for dipole-dipole relaxation of a two-spin system with spin states α and β . When nucleus A is saturated it will relax back to equilibrium using pathways ω_0 , ω_1 and ω_2 . $\omega_{1,A}$ transitions do not affect the spin of nucleus X , but ω_0 and ω_2 do. ω_0 , the dominating pathway for dipole-dipole relaxation in proteins, will cause a decrease in signal intensity of nucleus X as α_X will transition to β_X when $\beta_A \rightarrow \alpha_A$. The double quantum transition using the ω_2 pathway will conversely increase the polarization of nucleus B , yielding an increase in signal intensity.

— The spectral density function —

Molecules in solution tumble and collide. To describe their orientation after a period of time, τ , has passed the time correlation function, $G(\tau)$, can be used:

$$G(\tau) = G(0)e^{-|\tau|/\tau_c} \quad (19)$$

where the correlation time, τ_c , is defined as the time it takes for a molecule to rotate by one radian; a short τ_c equates to fast tumbling molecules and vice versa. The spectral density function (SDF) is the Fourier transform of the time correlation function and the real part describes the efficiency of angular frequencies to cause relaxation. In 1982 Giovanni Lipari and Attila Szabo presented the widely adopted Model-Free formalism, which is used to represent the motions of a molecule through an order parameter, S^2 , describing molecular rigidity and correlation times describing molecular tumbling, τ_c , and internal motions, τ_e [3, 4]. It is assumed that global and intramolecular motions are time-scale separated from other motions. With Model-Free, acquired relaxation rates and NOEs can be fitted to extract Model-Free parameters or the other way around. This is achieved by expressing the SDF, $J(\omega)$, as:

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S^2) \tau}{1 + \omega^2 \tau^2} \right] \quad (20a)$$

$$\frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e} \quad (20b)$$

where ω is the angular frequency. The relaxation rates R_1 and R_2 and NOE for an ^{15}N nucleus dipolar coupled to a proton and experiencing CSA and chemical exchange can be expressed using SDF as:

$$R_1 = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2J(\omega_N) \quad (21a)$$

$$R_2 = 0.5d^2[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{6}[3J(\omega_N) + 4J(0)] + R_{ex} \quad (21b)$$

$$NOE = 1 + (\gamma_H/\gamma_N)d^2[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]/R_1 \quad (21c)$$

where γ is the gyromagnetic ratio and d^2 and c^2 are the squared dipolar and CSA coupling constants, respectively, as defined by:

$$d^2 = \frac{\hbar^2 \gamma_H^2 \gamma_N^2}{4r_{NH}^6} \quad (22a)$$

$$c^2 = \frac{1}{3} \gamma_N^2 B_0^2 (\sigma_{\parallel} - \sigma_{\perp})^2 \quad (22b)$$

where $\hbar$ is Planck's reduced constant, r is the internuclear distance and $\sigma_{\parallel}$ and $\sigma_{\perp}$ are the parallel and perpendicular components of the chemical shift tensor. From equations 20a and 21 it can be seen that $R_1 \propto J(\omega_N)$, $R_2 \propto J(0)$ and $NOE \propto J(\omega_H)$, as $\omega_H \gg \omega_N$ and $J(0) \gg J(\omega_N) \gg J(\omega_H)$; proportionality does not apply in the extreme narrowing limit where τ_c is small, but only when $\omega^2 \tau_c^2 \gg 1$. Furthermore it is seen that dipolar and CSA contributions to relaxation rates are larger for R_2 than R_1 , due to the $J(0)$ contribution.

— Studying protein dynamics with NMR —

As previously discussed, protein dynamics can be divided into different time scales and NMR spectroscopy experiments can be designed to study different types of dynamical processes. Important to this thesis are ^{15}N relaxation experiments studying the mobility of the protein backbone. The following sections will briefly explain the employed experiments.

— R_1 experiment —

The two principal ways of measuring R_1 are by saturation and inversion recovery experiments. In the saturation recovery experiment magnetization, M_z , is tipped into the transverse plane and recovers to thermal equilibrium as described by equation 12. This experiment measures R_1 with a dynamic range based on the interval $[0, M_z]$, that interval of the inversion recovery experiment is $[-M_z, M_z]$ as a 180° pulse inverts the magnetization; this makes the inversion recovery experiment superior to the saturation recovery experiment. The build-up along B_0 for the inversion recovery experiment is described by:

$$M_z(t) = M_0(1 - 2e^{-t/T_1}) \quad (23)$$

where M_0 is the equilibrium magnetization, Figure 9.

— R_2 and $R_{1\rho}$ experiments —

In 1950 Erwin Hahn presented a technique to measure relaxation times, called *spin echo* [83]. The principle was to apply two 90° pulses, separated by a short delay τ , and detect the FID after another delay of length τ . Shortly thereafter Herman Carr and Edward Purcell improved the spin echo by replacing the second 90° pulse with a 180° pulse as well as enabling measurement of R_2 without restarting the pulse sequence from equilibrium [84]. This was achieved by continuously applying the refocusing 180° pulse n times, $90_x^\circ - (\tau - 180_x^\circ - \tau)_n$, and measuring the decay in between. Meiboom and Gill improved the spin echo method of Carr and Purcell by changing the phase of the 180° pulses by 90° in the transverse plane relative to the phase of the initial 90° pulse, $90_x^\circ - (\tau - 180_y^\circ - \tau)_n$; this resulted in that effects arising from pulse imperfections did not accumulate [85]. This experiment is today known as the CPMG experiment and it is a useful experiment to measure R_2 . A more robust way of measuring R_2 is to use the $R_{1\rho}$ experiment [86], where magnetization is spin-locked to an effective magnetic field, ω_{eff} , making the angle θ with B_0 , Figure 12. Magnetization will build up aligned with ω_{eff} with the rate $R_{1\rho}$ defined by:

$$R_{1\rho} = R_1 \cos^2(\theta) + R_2 \sin^2(\theta) \quad (24)$$

This equation is used to fit the acquired data and if R_1 rates have priorly been measured they can be directly inputted during fitting.

Figure 12: In the $R_{1\rho}$ experiment the laboratory frame is rotated with the angle θ forming an effective field, ω_{eff} , as defined by the strength of the applied field, ω_1 , and the carrier offset, Ω . The new frame is kept in place by applying continuous wave irradiation, also known as a spin-lock. Just like in the laboratory frame where magnetization builds up along B_0 and decays perpendicular to B_0 , magnetization will build up along ω_{eff} and decay perpendicular to ω_{eff} with the rates $R_{1\rho}$ and $R_{2\rho}$, respectively.

— The CPMG relaxation dispersion experiment —

The Carr-Purcell-Meiboom-Gill [84, 85] relaxation dispersion experiment (CPMG RD) [87] is designed to study slow conformational exchange on the μs -ms time scale where nuclei exchange between a ground state and a lowly populated state, or states, referred to as the excited state. As nuclei in different states may relax with different rates, R_2 must be described as:

$$R_{2,eff} = R_{2,0} + R_{ex} \quad (25)$$

where $R_{2,eff}$, the efficient R_2 , is the observed transverse relaxation rate, $R_{2,0}$ is the intrinsic transverse relaxation rate as defined by dipolar couplings, CSA, molecular tumbling and bond vector fluctuations and R_{ex} is the contribution to the transverse relaxation rate due to chemical exchange. By using the CPMG RD experiment dynamic amino acids can be distinguished from those with no chemical exchange and both $R_{2,0}$ and R_{ex} can be estimated, given that the exchange is detectable.

In the CPMG RD pulse sequence a 90° pulse flips the net magnetization into the transverse plane. This is followed by an echo train that refocuses the net magnetization, $(\tau - 180^\circ - \tau)_n$, with the delay τ and n number of repetitions. The

Figure 13: ^{15}N CPMG RD profiles of a dynamic amino acid (left) experiencing chemical exchange and that of an amino acid with no chemical exchange on the probed time scale (right).

experiment is repeated while altering the number of refocusing pulses during a constant time and detecting the decays of the resulting magnetization.

The intensity of the signal will decay as a function of the constant time, t , and is related to $R_{2,eff}$ according to:

$$I(\nu_{CPMG}) = I_0 \cdot e^{-R_{2,eff}t} \quad (26)$$

where I_0 is the reference intensity for $t = 0$ and $I(\nu_{CPMG})$ is the resulting intensity for the frequency, ν , of spin echoes. The R_{ex} contribution to $R_{2,eff}$ will decrease as ν_{CPMG} increases and when plotting $R_{2,eff}$ against ν_{CPMG} it will appear as exponential decays for exchanging nuclei and a constant value for non-exchanging nuclei, Fig 13. When these dispersions are fitted against an appropriate model the exchange rate, k_{ex} , the excited state population and difference in resonance frequency between the ground and excited state, $|\omega|$, can be estimated. One model appropriate for most time scales, except for very slow processes, is the corrected Carver-Richards equations [88, 89]:

$$R_{2,eff} = \frac{1}{2} \left(R_{2A}^0 + R_{2B}^0 + k_{ex} - 2\nu_{CPMG} \cosh^{-1} \left(D_+ \cosh(\eta_+) - D_- \cos(\eta_-) \right) \right) \quad (27a)$$

$$D_{\pm} = \frac{1}{2} \left(\pm 1 + \frac{\psi + 2\Delta\omega^2}{(\psi^2 + \zeta^2)^{1/2}} \right) \quad (27b)$$

$$\eta_{\pm} = 2^{-3/2} \frac{1}{\nu_{CPMG}} \left(\pm \psi + \sqrt{\psi^2 + \zeta^2} \right)^{1/2} \quad (27c)$$

$$\psi = \left(R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex} \right)^2 - \Delta\omega^2 + 4p_A p_B k_{ex}^2 \quad (27d)$$

$$\zeta = 2\Delta\omega \left(R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex} \right) \quad (27e)$$

For slower time scales, where the ground state population is much larger than the excited state population, $p_A \gg p_B$, a model proposed by Tollinger et al.[87] is more suitable:

$$R_{2,eff} = R_{2A}^0 + k_A - k_A \frac{\sin(\Delta\omega\tau_{CPMG})}{\Delta\omega\tau_{CPMG}} \quad (28)$$

where $2\tau_{CPMG}$ is the delay between the 180° pulses in the echo train. If the conformational exchange is faster or slower than what CPMG RD can detect it can be studied using the NMR experiments $R_{1\rho}$ RD and CEST, respectively.

— Heteronuclear NOE —

The ^1H - ^{15}N heteronuclear NOE experiment is useful for characterizing regions of a protein as flexible or rigid and its observable, hetNOE, is sensitive to motions faster than molecular tumbling. The experiment is recorded with and without saturating pulses and the results are used to calculate hetNOE. When the spin of a nucleus (S) is saturated spins of any dipolar coupled nuclei (I) will be affected by the transitions that S undergoes; this is the NOE effect as previously described. NOE is proportional to $J(\omega_H)$ and from this it can be concluded that for flexible regions of the molecule, low S^2 , the right term of equation 20a dominates; internal motions faster than molecular tumbling, $\tau_e < \tau_c$, will thus affect the observable.

When calculating hetNOE, the intensities of the experiments with, I , and without saturation, I_0 , are compared:

$$hetNOE = \frac{I}{I_0} \quad (29)$$

In protein NMR spectroscopy, flexible regions of the protein can be identified by plotting hetNOE against the amino acid sequence, as previously illustrated in the left panel of Figure 8. The hetNOE values can also be used together with R_1 and R_2 relaxation rates in Model-Free analysis.

Results and discussion

I think it is fair to say that personal computers have become the most empowering tool we have ever created. They are tools of communication, they are tools of creativity, and they can be shaped by their user.

Bill Gates

We humans are the self-proclaimed smartest species alive but when it comes to handling numbers we will never be able to compete with the processing power of computers. Computers have become pivotal technology in our society and can be found everywhere. In science, when it comes to data acquisition and analysis, computers are the primary tools. Therein lies the importance of software development. As scientific fields become niched, it is common that users must resort to writing their own scripts, programs and protocols to be able to analyze their data. A potential drawback to this is that differences in how the data is analyzed may cause inconsistencies when comparing results acquired by other measures. It is thus important to employ methods that can yield reproducible and accurate results.

This thesis reports on four papers, three of which present developed software for usage in the field of structural biology. The software were developed with the intention to aid research in this field as well as to encourage and teach novice scientists how to analyze their acquired data. The fourth paper describes an NMR spectroscopy study of CaM variants associated with the possibly fatal heart disease, LQTS. This paper can be seen as a testament to the results of the other three papers as the produced software have been used to analyze the collected data; hence fulfilling their collective aim of contributing to studies in the field of structural biology.

The following sections will describe and discuss the motivations and results of the different research projects of this thesis. For the three software projects, pseudocode of used algorithms will be presented and discussed as well.

— Paper I: COMPASS —

The time consuming process of assigning resonances in NMR spectra is an imperative task when studying a protein with NMR spectroscopy. If the assignments are missing or, even worse, are incorrect, results from NMR spectroscopy data of other experiments are rendered erroneous and misleading. NMR spectroscopists use different means to assign their proteins, which explains the wide array of protocols available today. While traditionalists opt for reliable yet time consuming manual approaches others choose automated protocols that are extremely efficient but sacrifice the important accuracy for speed. The loss of accuracy stems from that several amino acids have similar chemical shifts and that chemical shifts can deviate markedly from consensus values [90].

The first paper of this thesis describes the software COMPASS, short for COnputer-aided Matching and Peak ASSignment. COMPASS is a semi-automated resonance assignment software, developed to speed up labor intensive phases of the resonance assignment process as well as to perform in-depth analyses of the input data to yield reliable results that can be interpreted by the user.

— Motivation —

The initial source of motivation for COMPASS was to decrease the time spent manually typing assignments. Assuming a full battery of standard resonance assignment experiments, HSQC, HNCACB, CBCACONH, HN(CA)CO and HNCO [56, 59, 60], meticulous assignment of a mere 100 amino acid long protein would equate to typing approximately 700 labels; not counting typos, temporary assignments and corrections of wrongly made assignments. This combined with the intermittent analysis of chemical shifts and browsing of spectra drastically increase the time spent assigning NMR spectra. For this busy work to be alleviated a simple script typing assignments would work, but not without also performing the chemical shift analysis as the resonances must be specifically assigned. Automated and semi-automated software performing this task, with different degrees of accuracy, already existed [91–94], but the user control was either removed or complicated. COMPASS was thus designed to function as a middle ground to the manual and automated approaches, by performing the tedious typing tasks for the user and performing the chemical shift analysis while letting the user monitor the resonance assignment process to guarantee accuracy.

— How COMPASS works —

The basic principle of using COMPASS is to provide the software with ^{15}N , ^1H , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}^\text{O}$ chemical shift data, perform an automated analysis of the data and proceed to assign the resonances aided by results from the analysis. All tasks performed by COMPASS can be subdivided into four distinct modules named '*Convert*', '*Label*', '*Analyze*' and '*Assign*'. The '*Convert*' module is used for file preparation; as input data need a specific format we implemented routines to convert data to and from compatible formats. The user may convert peak lists containing chemical shift data and protein sequence data from different sources. COMPASS was designed to analyze peak lists containing chemical shift data, which means that users initially must pick peaks in their spectra using other software and provide the resulting peak lists to COMPASS. The '*Label*' module was designed to temporarily label and group the chemical shifts into 'spin systems'; these spin systems contain the chemical shifts of amino acid i and $i-1$. The spin system grouping is necessary in order to distinguish and compare chemical shifts during analysis. The '*Label*' module performs this task by using thresholds of chemical shift deviations during the formation of the spin systems. As this task is especially difficult, even during manual assignment, for overlapping regions in the spectra warnings indicating suspected overlaps, as well as other issues, are presented to the user after using the '*Label*' module. These warnings must be manually resolved, but this process does generally not consume much time.

After file preparation is complete and spin systems have been labeled the '*Analyze*' module is used. This module is the brain of COMPASS and it performs an in-depth analysis of the chemical shifts and generates data files needed for assigning the protein. The '*Analyze*' module can be described with Algorithm 1, where SSE is short for secondary structure element to indicate either α -helix, β -strand or random coil. All possible combinations of the three secondary structure elements are considered as the chemical shifts of $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}^\text{O}$ all depend on the dihedral angles [53]. The '*Analyze*' module make use of user set deviations for $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}^\text{O}$ chemical shifts, to control the degree of possible mismatches caused by chemical shifts of tentative assignments being unoptimized. A higher deviation will thus increase the number of wrongly generated fragments due to the ambiguous character of chemical shifts for several amino acids, but also ensure that more correctly generated fragments are included. It might instinctively seem like a bad tradeoff to include more erroneous fragments to gain few correctly generated fragments, but in COMPASS it is often very simple to evaluate the correctness of the fragments. This is done by

---

Algorithm 1 The main algorithm of the ‘*Analyze*’ module.

```

1: Read chemical shift data
2: Store spin systems as two residue long fragments
3: while chemical shifts of type  $i - 1$  can be paired to those of type  $i$  do
4:   if fragment length  $< 10$  then
5:     Extend the fragment with another fragment
6:   end if
7: end while
8: for each fragment and all positions in the amino acid sequence do
9:   for all SSE combinations of the fragment length do
10:    Compare the chemical shifts of the fragment to a database
11:    Calculate  $\chi^2$  for the fragment
12:    if  $\chi^2$  is among the top 20 lowest  $\chi^2$  values then
13:      Store the results of the fragment
14:    end if
15:  end for
16: end for
17: Write the top 20 results

```

---

evaluating a target function:

$$\chi^2 = \frac{10 \sum_{i=1}^m \sum_{j=1}^{n_i} (\omega_{ref,i}^j - \omega_{obs,i}^j)^2}{\sum_{i=1}^m n_i} \quad (30)$$

, where the inner sum is a sum of squared residuals for n number of chemical shifts with j types of chemical shifts, ω , and the outer sums span the fragment length, m . The χ^2 value in equation 30 has been weighted with $10/\sum_{i=1}^m n_i$, with the purpose to punish the value of χ^2 if chemical shifts are missing during analysis. It should be emphasized that the χ^2 value is easier to interpret for longer fragments, as it is very unlikely to get low scores for long and mismatched fragments. The drawback of generating longer fragments is the drastic increase in execution time. If we assume that we are analyzing two fragments of five and ten amino acids in length and that no chemical shifts are missing, COMPASS needs to make $3^5 = 243$ and $3^{10} = 59049$ iterations, respectively, to cover all possible combinations of secondary structure elements for a single segment of the amino acid sequence. For each iteration COMPASS will import database chemical shifts and calculate a χ^2 value. If this process was done for a small protein with a 100 long amino acid sequence it would mean over 5 million iterations for fully analyzing the fragment of ten amino acids in length, compared to the 20 000 iterations for the fragment containing five amino acids. Keeping in mind that the number of calculations of equation 30 for each iteration also increases with fragment length it is evident that fragment length has a huge impact on execution time. COMPASS uses the chemical shift database RefDB

**** Index ****	***** Residue type *****	***** Residue No *****	*** SSE ***	** χ^2 **
3285.1	ARG-GLY-HIS-SER-VAL	82-86	loop-loop-strd-strd-strd	5.02758
3285.2	ARG-GLY-HIS-SER-VAL	82-86	strd-loop-strd-strd-strd	5.09881
3285.3	ARG-GLY-HIS-SER-VAL	82-86	loop-loop-strd-loop-strd	5.85371
3285.4	ARG-GLY-HIS-SER-VAL	82-86	strd-loop-strd-loop-strd	5.92493
3285.5	ARG-GLY-HIS-SER-VAL	82-86	loop-loop-strd-strd-loop	6.23687
3285.6	ARG-GLY-HIS-SER-VAL	82-86	loop-strd-strd-strd-strd	6.23707
3285.7	ARG-GLY-HIS-SER-VAL	82-86	strd-loop-strd-strd-loop	6.30809
3285.8	ARG-GLY-HIS-SER-VAL	82-86	strd-strd-strd-strd-strd	6.30829
3285.9	ARG-GLY-HIS-SER-VAL	82-86	loop-loop-strd-loop-loop	7.06299
3285.10	ARG-GLY-HIS-SER-VAL	82-86	loop-strd-strd-loop-strd	7.06319
3285.11	ARG-GLY-HIS-SER-VAL	82-86	strd-loop-strd-loop-loop	7.13422
3285.12	ARG-GLY-HIS-SER-VAL	82-86	strd-strd-strd-loop-strd	7.13442
3285.13	ARG-GLY-HIS-SER-VAL	82-86	loop-helix-strd-strd-strd	7.33146
3285.14	ARG-GLY-HIS-SER-VAL	82-86	strd-helix-strd-strd-strd	7.40268
3285.15	ARG-GLY-HIS-SER-VAL	82-86	loop-strd-strd-strd-loop	7.44636
3285.16	ARG-GLY-HIS-SER-VAL	82-86	strd-strd-strd-strd-loop	7.51758
3285.17	ARG-GLY-HIS-SER-VAL	82-86	loop-loop-loop-strd-strd	7.66594
3285.18	ARG-GLY-HIS-SER-VAL	82-86	strd-loop-loop-strd-strd	7.73716
3285.19	ARG-GLY-HIS-SER-VAL	82-86	loop-helix-strd-loop-strd	8.15758
3285.20	ARG-GLY-HIS-SER-VAL	82-86	strd-helix-strd-loop-strd	8.22881

Peak name	1020N-HN(i-1)	1020N-HN	1090N-HN	1240N-HN	980N-HN
	Residue 1	Residue 2	Residue 3	Residue 4	Residue 5
13C α	55.351	45.906	55.5125	58.273	61.676
13C β	30.919	0	31.994	64.683	35.008
13C0	175.389	173.825	174.346	173.143	176.144

Figure 14: The top 20 results for each fragment outputted by the 'Analyze' module are presented to the user in the 'Assign' module. To assign the fragment to one of the results the user only has to click the unique fragment index (blue color) to the left. A single click for this particular fragment equates to typing 30 labels. Highlighted in green is a fragment with SSE correlating to that of a supplied file based on the crystal structure of the protein. All 20 results indicate that a single assignment, R82-V86, is most likely. This together with the low χ^2 values, a long fragment size and the fact that no chemical shifts are missing makes this assignment extremely probable.

[96] when evaluating the observed chemical shifts and a feature that is unique to COMPASS is to compensate for effects on the chemical shifts caused by expressing the protein using deuterated glucose (perdeuteration) or D₂O (partial deuteration). For perdeuteration the database was constructed by comparing chemical shifts from perdeuterated and protonated thioredoxin domain of human Grx3 [97] and for partial deuteration the database was constructed from similar measurements of partially deuterated proteins.

Finally the last module 'Assign' does the work that was the initial motivation of COMPASS, namely automatic typing of assignments. All results generated by the 'Analyze' module are presented to the user through data files, separated by the different fragment lengths and the minimum number of chemical shifts in each spin system, 1-3, that have been used to pair spin systems together. A typical result generated by COMPASS is displayed in Figure 14. The recommended

Figure 15: The near complete assignment of SAP97 PDZ-2 [95] in COMPASS. The results are continuously visualized in terms of completion as well as a calculated SSP. Highlighted in gray in the SSP plot is the secondary structure according to a provided data file. This file can, for instance, be based on crystal structure or the structure of a mutant or wild type and enables a direct comparison of the assigned chemical shift data. Green regions in the amino acid sequence represent fully assigned residues and yellow regions represent either assignments of only type *i* or *i-1* chemical shifts.

approach is to analyze long fragments with all chemical shift types, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}^\text{O}$, used for fragment generation first, as χ^2 values are easily interpreted for these fragments. By clicking indices of results that are evaluated to be correct, COMPASS will automatically assign the spin systems accordingly. Fragment results are presented in order of increasing χ^2 values and as a fragment is assigned all fragments that consequently cannot be correct are removed from the result browser. This too motivates using a higher chemical shift deviation during usage of the 'Analysis' module, as the user does not need to sift through every erroneous result. The shrinking lists of results will ultimately be empty or consist of ambiguous assignments, spin systems with missing chemical shifts and erroneous spin systems. As the remaining spin systems may need manual assignment, COMPASS has an included interface that lets the user specify an

assignment for a spin system without having to type the labels. Throughout the assignment process, the graphical user interface (GUI) will present feedback in form of a calculated secondary structure propensity (SSP) [67] for made assignments as well as the progress towards completion, Figure 15. The final results outputted by COMPASS are assigned peak lists that have been continuously updated throughout the assignment process.

— How COMPASS performs —

After implementing COMPASS as a standard protocol for resonance assignment of the protein backbone at Linköping University, for both research and teaching purposes, the assignment process is often finished within a days work depending on protein size. It is difficult to benchmark the speed of using COMPASS, as it approaches the resonance assignment task in a way not comparable to other available methods. Naturally, the speed of COMPASS cannot be compared to that of fully automated software, unless the time spent checking and correcting errors in the automatically assigned peak lists is accounted for. One aspect that can be benchmarked is the completeness of resonance assignments, meaning the number of correctly assigned peaks out of the total number of peaks in the spectra. We benchmarked the completeness of COMPASS against four unique alternative software, PINE Server [91], AutoAssign [92], EZ-ASSIGN [94] and SAGA [93], the former two being fully automated and the latter two needing manual intervention. We analyzed three different proteins using all software, a small protein (SAP97 PDZ-2 [95]) and two large proteins (TPMT*1 [98] and EphB2 JMS-KD [99]). Using only the results provided by the software as ground for assignments, all tests showed that COMPASS could assign the proteins to higher completeness than the alternatives, Figure 16. The automated and semi-automated assignment software PINE server and EZ-ASSIGN, respectively, tied for second place. We also performed additional benchmark tests where COMPASS was compared to manual assignment of the same proteins as in Figure 16 and the additional proteins Abp1p SH3 domain [100], CaM E140Q Tr2C [101] and CDPK3 CLD(B) [102]. The purpose of these tests were to investigate how many peaks that could be additionally assigned by manual analysis and the criterion was the same as before, only the results provided by COMPASS could be used as ground for assignments. For Abp1p SH3, SAP97 PDZ-2 and CDPK3 CLD(B) manual analysis did not yield additional assignments, but for the three remaining proteins a handful of extra peaks could be assigned. The difference is explained by ambiguities arising from missing chemical shifts that could not be assigned using the top 20 results from COMPASS. If COMPASS included the

Figure 16: When COMPASS was benchmarked against automatic (PINE Server and AutoAssign) and semi-automatic (EZ-ASSIGN and SAGA) assignment software, it outperformed the alternatives for all analyzed proteins in terms of completeness of resonance assignment. The completeness is based on the fraction of correctly assigned peaks out of the total number of available peaks in all analyzed spectra.

option to reanalyze a fragment, taking already made assignments into account, this issue would likely have been resolved. When it comes to speed COMPASS significantly outperforms the manual approach. The main advantage of using COMPASS, concluded from the benchmark tests, is thus that its results yield higher completeness than fully automated and alternative semi-automated protocols while being an efficient alternative to manual assignment.

— Paper II: CDpal —

The second paper describes the software CDpal (Circular Dichroism pal or Chemical Denaturation pal), developed with the aim to simplify the process of curve fitting thermal and chemical protein denaturation data and to extract thermodynamical parameters from the fits. CDpal achieves fast and reproducible curve fitting, which are the main advantages of using it. If a signal is being monitored as a function of temperature or concentration of added denaturant, it can be fitted in CDpal regardless of the acquisition method, making CDpal applicable for data originating from both CD and fluorescence spectroscopy.

— Motivation —

The source of motivation for CDpal was that we identified a need for a software capable of accurately fitting thermal and chemical denaturation data of proteins to different models. The available alternatives were often software distributed by manufacturers of spectrometers or graphical plotting tools such as GraphPad (<http://www.graphpad.com>), Psi-Plot (<http://www.polysoftware.com/aboutpsi.htm>), Origin (<http://www.originlab.com>), and SigmaPlot (<http://www.systat.com>). These alternatives either fitted the data in a simplified manner or even not at all. It is easy to understand the underlying reasons to this, as curve fitting is not a trivial task, but necessary in order to achieve accuracy of fitted parameters. To simplify curve fitting, making it a simple process for any user regardless of mathematical background, we developed CDpal. We aimed for a user-friendly design and functionality that let the user fit their data correctly, often automatically with the click of a button, thus making it so simple that no user would refrain from properly fitting their data. To understand how CDpal works and also relevant to Paper III describing a peak integration software, the next section will briefly explain the process and importance of curve fitting.

— Curve fitting —

A common task in many scientific fields is fitting data to a mathematical function based on physics and to extract the parameters describing the fit, this is known as curve fitting. In biophysical characterization of proteins, curve fitting yields important parameters describing, for example, protein stability, dynamical properties and molecular interactions. In biochemistry, most curve fitted functions are nonlinear and for this nonlinear least squares fitting [103] is used. Nonlinear least squares fitting is a type of curve fitting in which model parameters, that minimize a χ^2 value, must be calculated through iterations. For a dataset containing m number of data points that can be fitted to a function, $f(x; \beta_1, \beta_2, \dots, \beta_n)$, depending on n number of parameters, β , χ^2 is calculated through:

$$\chi^2 = \sum_{i=1}^m r_i^2 = \sum_{i=1}^m \frac{[y_i - f(x_i; \beta_1, \beta_2, \dots, \beta_n)]^2}{\sigma^2} \quad (31)$$

where r_i is the residual for the i :th data point and σ is the uncertainty. By choosing to minimize χ^2 , negative and positive deviations from the model function are treated equally and any increase in χ^2 , resulting from iteration with different β parameters, is construed as a step away from a minimum.

Figure 17: When minimizing χ^2 with the Levenberg-Marquardt algorithm, the start parameters decide at which minimum the minimization will converge. By providing reasonable estimates as start parameters the best fit, as defined by the global minimum, can be attained.

A useful algorithm for solving non-linear equations is the Levenberg-Marquardt algorithm [104, 105]. The Levenberg-Marquardt algorithm minimizes χ^2 by using initial β guesses and alternating between the steepest descent and the Gauss-Newton method. In the steepest descent method larger steps are taken, $\beta + \delta$, in the direction of a negative gradient; it is a good method if the solution is far from minimum. When the solution is close to a minimum the Levenberg-Marquardt algorithm instead shifts from the steepest descent method towards the Gauss-Newton method where a region is evaluated by assuming that the function is quadratic. A drawback to the Levenberg-Marquardt algorithm is that it is not guaranteed to find the global minimum and it thus necessitates good initial guesses of the start parameters for the curve fitting to converge at a global minimum, Figure 17. To simplify the process of finding a global minimum at the cost of speed, a grid search algorithm systematically testing different start values within a given range can be employed. The process of iteratively calculating gradients for all data points fitted to a given function and inverting matrices can for complicated models and large datasets be performance heavy even for computers, which is why the direction of software development is headed towards utilizing the threading capabilities of the GPU (Graphics Processing Unit) to perform these calculations [106].

How CDpal works

CDpal is divided into two different modules, one that fits thermal denaturation data and one that fits chemical denaturation data. The modules are similar with the exception of having different models used for curve fitting. To start a session in CDpal either thermal or chemical denaturation data is imported. CDpal can be customized to import data from a wide array of different formats. The user can modify the number of header lines, which columns make up the probed signal and increasing temperature or concentration of denaturant and search for a word that initiates data import. During data import, CDpal will offer the user to normalize the data, making it easier to perform comparisons of different sets of data. The imported data is at this stage ready to be fitted to any of the four implemented models: $N \rightleftharpoons D$, $N \rightleftharpoons I \rightleftharpoons D$, $N_2 \rightleftharpoons 2I \rightleftharpoons 2D$ or $N_2 \rightleftharpoons I_2 \rightleftharpoons 2D$. Unless a three state model, accounting for dimerization, is suspected, the data can be automatically fitted to either $N \rightleftharpoons D$ or $N \rightleftharpoons I \rightleftharpoons D$ or both models simultaneously; the automatic fitting of all data sets is described in Algorithm 2. All four models

Figure 18: The result of fitting thermal denaturation data of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) from *Oryctolagus cuniculus* (European rabbit) to a two-state model in CDpal.

can be fitted by manually estimating the start parameters. This task can seem a little daunting to a common user as the choice of start parameters are important to reach convergence at the global minimum, as seen in Figure 17. For this purpose CDpal have added functionality to perform linear regression of the native, intermediate and denatured states, which is simply performed by highlighting

---

Algorithm 2 Automatic curve fitting algorithm used in CDpal.

```
1: for all data sets do
2:   for all types of fitted models do
3:     Find the point(s) of inflection in the data set
4:     if thermal denaturation data is fitted then
5:       Initialize  $T_m$  to the inflection point
6:       Initialize  $\Delta H$  and  $\Delta C_p$  to standard values
7:     else
8:       /*Chemical denaturation data is fitted*/
9:       Initialize  $C_m$  to the inflection point
10:      Initialize  $m$  to a standard value
11:    end if
12:    Detect the native, intermediate and denatured states by calculating
    derivatives
13:    Perform linear regression on the native, intermediate and denatured
    states
14:    Use results of linear regression as start parameters
15:    Calculate  $\chi^2$ 
16:    while curve fitting has not converged and not reached maximum num-
    ber of iterations do
17:      Take a step  $(\beta + \delta\beta)$  in a descending direction of the function defined
      by the parameters
18:      Perform Gauss-Jordan elimination
19:      if convergence condition has been fulfilled then
20:        Calculate covariance matrix and curvature matrix
21:        Break out of while loop
22:      end if
23:      Evaluate  $\chi^2(\beta + \delta\beta)$ 
24:      if  $\chi^2(\beta + \delta\beta) \geq \chi^2$  then
25:        Increase step size
26:      else
27:         $\chi^2 = \chi^2(\beta + \delta\beta)$ 
28:         $\beta = \beta + \delta\beta$ 
29:        Decrease step size
30:      end if
31:    end while
32:    Calculate the variance using jackknife resampling
33:    Write data to files
34:    Plot fitted curves and display calculated parameters
35:  end for
36: end for
```

---

data points estimated to belong to one of the different states and clicking a corresponding button. All that is left to estimate are the melting temperatures or the midpoint concentration of denaturant, which are easily estimated by inspecting the denaturation data, and values for enthalpy (thermal denaturation) or the m value (chemical denaturation). To estimate changes in enthalpy the protein size can be used [107] and the range of m values is often $m \in [2, 30]$ kJ mol⁻¹ M⁻¹ for GdHCl or a factor two or three lower for urea [108]. Even if ΔH_m and m values can be hard to estimate their correctness rarely causes a problem during curve fitting as long as the other start parameters are properly estimated. The results outputted from CDpal are fitted curves and the final parameters yielding the different fits, Figure 18.

— How CDpal performs —

The process of analyzing the denaturation data is streamlined and more often than not, a session in CDpal is conducted within minutes, thanks to the included option to automatically fit all data. The results from CDpal are not by any means more accurate than that of any other software using the same set of equations derived from the laws of thermodynamics. It is however more convenient to use than the alternatives and it has a clear and concise purpose. Speaking of results being accurate or not is a pitfall as fitted parameters can be apparent, meaning that parameters that necessarily are not true can give a better fit to the experimental data than the true parameters. This is something that the user always should keep in mind, especially to avoid overinterpretation of the results and overfitting the data. The most accurate way to achieve correct thermodynamical parameters are to complement the method of choice with other methods and evaluate the results. Also, a really important factor when comparing data from different systems is to make sure that the experiment is conducted under the same conditions and at equilibrium.

— Paper III: PINT —

The analysis of NMR spectroscopy data is often very complicated and has likely deterred many potential users. The analysis process can be described with the following steps:

1. Process the acquired data.
2. Prepare the data for a specific type of analysis.
3. Perform the analysis.

4. Evaluate the results.

Even though the analysis process can be summarized in four steps, each step can vary substantially for different types of experimental data. For instance, examples of factors that cause divergence when processing NMR spectroscopy data are spectral dimensionality (1D, 2D, 3D etc.), sampling method (uniform or non-uniform), order of arrayed parameters and choice of processing software (different input and output formats). Differences in data preparation are primarily due to requirements on input data formats and the need for additional files when employing a script or software. As NMR spectroscopy has many areas of application, both the analysis and evaluation can involve distinct processes such as curve fitting, structural calculations, titration analysis, resonance assignment and preparation of plots and figures. Common software used to visualize spectra, pick peaks, fit peaks or extract relaxation parameters include SPARKY [109, 110], NMRPipe [111], FuDa (<http://www.biochem.ucl.ac.uk/hansen/fuda/>) and CCPN [112]. These software perform well and are widely used, but often have to be complemented with additional routines to fully analyze or evaluate the data. It is not feasible nor relevant to have a single software capable of performing all types of different analyses. But for the types of analyses that do share a similar methodology a common streamlined platform for analysis would be preferred.

Certain experiments for the analysis of protein dynamics with NMR spectroscopy have a high degree of similarity in the methodology from start to finish. A common battery of relaxation experiments include the measurement of R_1 , R_2 , hetNOE and conformational exchange (CPMG RD). The procedure of analyzing all of these experiments consists of an initial optimization of resonance assignments, followed by line shape fitting and downstream analysis. To extract relaxation parameters from the experiments R_1 , R_2 and CPMG RD, the data is fitted to decaying exponential functions and the heteronuclear NOE is calculated from the fitted peak intensities. It can be useful to fit the CPMG RD experiment to the corrected Carver-Richards equations [88, 89] as it is capable of fitting dispersions with fast or slow dynamics. Barring differences in models used for fitting data in the downstream analysis, all steps of the analysis can be performed in a similar fashion even though the results or the interpretations of the results are distinct. The third paper describes the software PINT (Peak INTEgration) developed for this purpose. PINT fits line shapes and performs consequential downstream analysis of NMR spectroscopy data and has been developed to function as a streamlined standalone application.

Motivation

PINT is based on the version published in 2013 [113], that originally was developed to fit peaks in NMR spectra and extract relaxation parameters. The software performed with high standards but had several recognized drawbacks. These drawbacks were:

1. PINT was restricted to Linux and macOS systems.
2. PINT was shell based.
3. PINT only utilized a single core of the processor during peak integration.
4. Additional software were required to visualize and evaluate the data.
5. The input data was restricted to spectra processed with NMRpipe.

The upgraded version had the motivation to eliminate all of these drawbacks as well as to include new features. As a result, PINT is now a standalone cross-platform software, available for Windows as well as Linux and macOS. It has gotten a comprehensive GUI that consequently does not require shell scripts to use. Multi-threading and memory mapped reading of spectra have been implemented to increase performance and PINT is now compatible with TopSpin processed spectra. Furthermore, there is no longer a need to use additional software for data visualization, evaluation of results and plotting of graphs as all this is performed from within the application. The streamlined layout of PINT is designed to be intuitive and to aid the user as much as possible; information on relevant functionality is found everywhere in the form of messages, a user manual and video tutorials.

How PINT works

The functionality of PINT is subdivided into seven different tabs, presented to the user in the same order as the suggested workflow, these are: '*Project files*', '*Spectra*', '*Integration*', '*3D fits*', '*2D fits*', '*Parameters*' and '*Plots*'. These tabs correspond to steps 2-4 (data preparation, analysis and evaluation) as previously described, Figure 19. A standard session in PINT starts with data preparation; spectra, peak lists and additional files such as files used for defining the experimental settings are imported in the '*Project files*' tab. The spectra and resonance assignments are then visualized in the '*Spectra*' tab, with the option to view the spectrum in 2D or 3D graphics. The purpose of the '*Spectra*' tab is to optimize peak positions, which is done either by manual repositioning of peak labels or using a peak centering routine, and peak radii considered for line shape fitting.

Figure 19: The seven different tabs subdividing the functionality of PINT. PINT covers data preparation, analysis and result evaluation but does not accept raw data formats and can thus not be used for the initial processing step.

Once the data has been prepared work is continued in the '*Integration*' tab, the heart and brain of PINT. PINT is in essence a curve fitting software for NMR spectroscopy data and to ensure versatility the line shape fitting and downstream analysis are customized by adding one-liner syntax to a parameter browser. The syntax is described in detail in the user manual and the GUI continuously provides descriptions of usage and purpose of the written or highlighted syntax. By altering just a few lines of syntax, every fitting routine available in PINT can be used and the analysis can be tailored in detail both locally, for a subset of peaks, and globally, for all peaks subjected to analysis. For instance, to fit CPMG RD data to the corrected Carver-Richards equations the necessary lines of syntax that need to be entered are:

```
-fitCPMG #Fit CPMG RD data
-array ncyc #Array N/2 number of refocusing pulses
-timeT2 0.03 #The constant time relaxation delay
```

Similarly, here is the syntax entered to fit a hydrogen-deuterium exchange experiment:

```
-fitExpOffset #Fit a decaying exponential with an offset
-array time #Array the total elapsed time
```


After entering the preferred syntax and ordering PINT to start parsing the inputted parameters, PINT will perform the line shape fitting and optional downstream analysis. The algorithm responsible for this, Algorithm 3, is very similar to parts of Algorithm 2, as both CDpal and PINT utilize the Levenberg-Marquardt algorithm for curve fitting. The four employed steps of line shape fitting seen

---

Algorithm 3 Line shape fitting algorithm used in PINT.

```

1: Parse syntax describing the analysis to perform
2: Import experimental parameters from headers of spectra
3: Read peak list
4: if PINT should automatically detect overlaps then
5:   Calculate overlaps using integration settings
6: end if
7: if any errors have been encountered then
8:   Report errors and abort the algorithm
9: end if
10: for all planes or spectra to be analyzed do
11:   Read data using memory mapping
12:   if spectra are processed with TopSpin then
13:     Convert and store the data to a compatible format
14:   end if
15:   Calculate spectral noise
16:   for all groups of peaks to analyze do
17:     Store spectral data as defined by peak radii
18:   end for
19: end for
20: for all groups of peaks do
21:   Initialize a worker for multi-threading purposes
22: end for
23: while threads have work to do do
24:   Start threads
25:   Fix peak positions
26:   //Curve fitting is described in Algorithm 2 (lines 14-29)
27:   Fit a single plane
28:   Copy data from previous fit and let peak positions float
29:   Fit a single plane again
30:   Fix positions and line widths to those acquired from previous fit
31:   Fit all planes
32:   Copy data from previous fit and let all parameters float
33:   Fit all planes
34:   Write plot files
35:   Calculate volumes and errors
36:   Perform downstream fitting
37:   Write results
38:   Terminate thread
39: end while

```

---

in Algorithm 3 were implemented to increase robustness. As the Levenberg-Marquardt algorithm only finds a local minimum and relies on good estimates of input parameters, it is much faster to initially fix parameters and use the results in consequential steps. By the final curve fitting step, the start parameters should be really close to the true parameters and would thus cause the line shape fitting to quickly converge when all parameters are allowed to float.

After the line shape fitting has finished the user may start evaluating the data using the '3D fits', '2D fits', 'Parameters' and 'Plots' tabs. The '3D fits' tab will draw plots of experimental and fitted data together with a difference map for easy evaluation of the line shape fitting quality, Figure 20. Peaks that have unsatisfactory fits are spotted easily and, if necessary, the input parameters of these peaks may be altered before performing another integration. The line shape fitting process is an iterative process and for all but the simplest of spectra the peak integration parameters must be optimized over several integrations. The first integration often reveals if the default Gaussian line shape is suited to represent the majority of peaks in the spectrum. If the results in general are unsatisfactory, the user may try any other available line shape: Lorentzian, Galore (a linear superposition of Gaussian and Lorentzian line shapes) or Voigt (a convolution of Gaussian and Lorentzian line shapes).

Figure 20: The results of the line shape fitting is evaluated through superimposed mesh plots, black and white for experimental and fitted data, respectively, together with a difference map. In this example two overlapping peaks have been fitted successfully, as indicated by the green regions in the difference map, to a gaussian line shape.

In ‘*2D fits*’ decays and fitted dispersions are visualized together with a residual plot, provided that the fitted line shape data has been subjected to a downstream analysis. The ‘*Parameters*’ tab contains a table of all fitted parameters and can be used to selectively plot any parameter against the amino acid sequence; this plot will be rendered in the ‘*Plots*’ tab. All plots rendered in PINT can be modified and exported.

— How PINT performs —

The performance of PINT has been drastically improved with the release of its successor by the addition of multi-threading. This means that instead of integrating a single group of peaks, PINT can integrate multiple groups at once as defined by the CPU and an optimal thread count, thus utilizing the full power of the CPU. Another added feature that has significantly sped up integration times is the addition of memory mapped reading of spectra. Memory mapping is a useful technique, best applied to larger files such as spectra due to its large overhead, which increases the speed of reading files by reading large chunks of the file and storing it as virtual memory before parsing; parsing this mapped memory is faster than sequentially reading the file.

PINT was benchmarked by fitting peaks to line shapes and performing downstream analysis of pseudo-3D data including ^{15}N R_1 , ^{15}N $R_{1\rho}$, ^{15}N CPMG RD, heteronuclear ^{15}N - ^1H NOE and CEST data. The benchmark resulted in calculated rates varying between 32-387 analyzed peaks per second. The calculated rates were based on data from small to large proteins and included fitting of all planes, which of course improved the rates as all planes are fitted simultaneously for every peak as seen in line 31 of Algorithm 3. The fastest analysis with the duration of 3 seconds was of an ^{15}N R_1 experiment of a small protein, with 58 mostly well-resolved peaks and 20 planes to be analyzed (387 peaks per second). Compared to a slow analysis of a large protein with many overlaps, 189 peaks and 70 planes, the duration was 4 minutes and 45 seconds (46 peaks per second). It is important to note that the duration for the integration heavily depends on factors such as initial peak positions, spectral resolution, signal-to-noise ratio and the total number of peaks, planes and overlaps. The rate determining factor is more often than not the group with the highest number of overlapped peaks. This is exemplified when comparing the above ^{15}N R_1 experiment with a 3 second duration to the same analysis of a protein with similar size. For this ^{15}N R_1 analysis, 66 peaks in 25 planes with the largest overlap consisting of seven peaks, the analysis took 44 seconds (38 peaks per second), a rate approximately reduced by a factor of ten. The reason why the number of overlaps heavily affects inte-

gration times lies within the number of calculations made during minimization of χ^2 using the $\mathcal{O}(n^3)$ Levenberg-Marquardt algorithm. The speed of the line shape fitting and downstream analysis is not the only important factor when it comes to time savings using PINT instead of alternative software. As PINT is a standalone application its streamlined workflow has a substantial impact on throughput. For example, the user does not have to manually generate plots or tables after fitting and analyzing peaks as these are automatically generated and presented within PINT.

The quantitative results of the line shape fitting was compared to that of Sparky [109, 110] and the results correlated very well except for a handful of outliers due to overlapped peaks. The results of PINT indicated that the line shape fitting of the outliers were of a quality comparable to the bulk of the analyzed peaks. As PINT is unique in its visualization of integrated data, it is hard to evaluate the results of Sparky in a similar manner for these outliers.

— Paper IV: Calmodulin variants of LQTS —

As previously described, several CaM variants have been implicated in the congenital heart disease LQTS. Paper IV describes a detailed analysis of structural and dynamical properties of three of the novel protein variants [27] involved with the disease, probed by NMR spectroscopy. The study was performed on the C-lobe of CaM (C-CaM) for the wild type, D95V, D129G and F141L variants. In the study we acquired and assigned resonances of sparse sampled 3D-spectra [114, 115] for all variants and the wild type in both apo and holo states. To investigate the dynamics we acquired and analyzed ^{15}N R_1 , ^{15}N $R_{1\rho}$, $[^1\text{H}]\text{-}^{15}\text{N}$ NOE, ^{15}N CPMG and ^{13}C CPMG for $^{13}\text{C}^\epsilon$ methyl groups of methionines [84–86, 116–118].

— Motivation —

The importance of CaM cannot be denied, which is why mutations of this highly conserved protein are interesting. Studies on the discovered variants have been focused on the effect mutated CaM has when interacting with target myocytes [38, 39, 44, 45, 119–121], which is extremely important to elucidate the phenotypic pathways. However, not a single study has been conducted to study the effects the mutations have on detailed CaM structure and dynamics. Moreover, interaction studies using NMR spectroscopy do not study changes in labeled CaM but that of the labeled target. Do CaM variants still have a similar fold as the wild type? What is the structural and dynamical response to the presence of Ca^{2+} ? Are the apo and/or holo states of variant CaM different from those of

the wild type? These are important questions to answer for fully understanding their involvement in LQTS; these questions motivated paper IV.

— The variants impact on the biophysical properties of CaM —

As previously known, C-CaM Ca^{2+} affinities of the variants are markedly decreased when compared to the wild type, leaving that of N-CaM (N-domain of CaM) virtually unaffected [27, 120]. Using NMR spectroscopy, there are multiple ways to probe Ca^{2+} binding of CaM, of which the simplest probes are seen directly in the ^{15}N - ^1H HSQC spectra. Upon addition of Ca^{2+} , large downfield chemical shift perturbations (CSPs) of the $^{15}\text{N}^H$ chemical shifts of I100 and V136, situated in position VIII of the Ca^{2+} binding motif of the EF-hand, are expected as the backbone carbonyl of position VII directly coordinates Ca^{2+} , which results in a deshielding effect of I100 and V136, Figure 21. For EF-hands I and III the resulting downfield shift is in the range of +14-16 ppm and +4-6 ppm for that of EF-hands II and IV [122]. The larger CSP for EF-hands I and III is explained by the side-chain reorientation of isoleucine from a *trans* to *gauche*⁻ conformation. Using this probe it was found that neither C-CaM^{D95V} or C-CaM^{D129G} bound Ca^{2+} strongly in their mutated EF-hand. Furthermore, C-CaM^{D129G} failed to bind Ca^{2+} strongly in the unmutated EF-hand as well. C-CaM^{F141L} is most similar to the wild type, binding Ca^{2+} with both EF-hands. Another probe indicative of Ca^{2+} binding is the $^1\text{H}^N$ chemical shift of glycines in position VI of the EF-hand Ca^{2+} binding motif. Hydrogen bonding with the side chain of position I upon Ca^{2+} coordination causes a significant downfield CSP of the $^1\text{H}^N$ chemical shift. The results of using this probe were in agreement with the results described above, as absent $^1\text{H}^N$ chemical shifts >10.2 ppm were construed as weak Ca^{2+} binding. Note that the mutation of C-CaM^{D129G} is situated in position I of EF-hand IV and thus natively lack a side chain capable of hydrogen bonding with G134.

In the paper, we used chemical shifts in two distinct ways: to investigate conformational exchange upon Ca^{2+} addition and to compare the variants to wild type CaM, Figure 22. We calculated compound chemical shifts by weighting the $^1\text{H}^N$ and $^{15}\text{N}^H$ chemical shifts according to:

$$\omega_{comp} = [(\omega_{HN})^2 + (\omega_N/5)^2]^{0.5} \quad (32)$$

We used the compound chemical shift, ω_{comp} , to calculate $\text{CSP} = \omega_{comp,holo} - \omega_{comp,apo}$. We noted a marked decrease in CSP, reporting on conformational exchange, located in the EF-hand III for C-CaM^{D95V} and both EF-hands of C-CaM^{D129G}, also the hinge region between EF-hands III-IV had lower CSPs for both of these CaM variants. CSPs of C-CaM^{F141L} were near identical to those

Figure 21: I100 and V136 are two excellent probes that indicate Ca^{2+} binding of EF-hand III and IV, respectively. The I100 and V136 downfield $^{15}\text{N}^{\text{H}}$ CSPs of +13.810 ppm and +5.796 ppm, respectively, for wild type C-CaM indicate that both EF-hands bind Ca^{2+} .

of the wild type, which is not surprising as its Ca^{2+} affinity resembles that of the wild type indicating similar structure. To compare the different variants to the wild type we correlated the compound chemical shifts of C-CaM in apo and holo states. This gave further insight and strongly indicated that C-CaM^{F141L} has similar structure to that of wild type CaM in both apo and holo states, whereas C-CaM^{D129G} differs the most of the studied variants.

When studying protein dynamics we estimated correlation times by first calculating the R_2/R_1 ratios and continuously trimming the values with two standard deviations until no values were removed; this was done to exclude residues with chemical exchange. Using $[^1\text{H}]\text{-}^{15}\text{N}$ hetNOE data the ratios were further trimmed by removing residues with $\text{hetNOE} \leq 0.65$, this removed large contributions from fast internal motions to the SDF and allowed τ_c to be estimated by fitting the resulting data to a simplified model [82], with the assumption that

Figure 22: Examples of wild type C-CaM and C-CaM^{D129G} chemical shift comparisons. Both panels make use of compound chemical shifts of $^{15}\text{N}^H$ and $^1\text{H}^N$ nuclei. To the left, the wild type (black) shows large CSPs as a result of conformational exchange upon the addition of Ca^{2+} . For C-CaM^{D129G} (gray), Ca^{2+} addition does not induce large conformational changes. To the right, the apo (hollow) and holo (black) compound chemical shifts have been correlated, showing that C-CaM^{D129G} has a degree of similarity to the apo state of wild type C-CaM but deviates substantially from the holo state.

rotational diffusion was isotropic, using:

$$T_1/T_2 \sim \left[d^2 \{ J'(\omega_H - \omega_N) + 3J'(\omega_N) + 6J'(\omega_H + \omega_N) \} + c^2 J'(\omega_N) \right] / \left[0.5d^2 \{ 4J'(0) + J'(\omega_H - \omega_N) + 3J'(\omega_N) + 6J'(\omega_H) + 6J'(\omega_H + \omega_N) \} + \frac{c^2}{6} \{ 3J'(\omega_N) + 4J'(0) \} \right] \quad (33a)$$

$$J'(\omega) = \frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} \quad (33b)$$

From this we saw similar correlation times for the wild type and variants, $\langle \tau_c \rangle = 5.3 \pm 0.2$ ns and 6.3 ± 0.5 ns for the apo and holo states, respectively; this shows that all variants tumble as monomers. C-CaM is expected to tumble more slowly when the α -helices of each EF-hand are reoriented to a perpendicular conformation in response to Ca^{2+} . It is also expected that C-CaM becomes more rigid when coordinating Ca^{2+} , which was seen for wild type C-CaM and C-CaM^{F141L}; C-CaM^{D95V} was more flexible in the mutated EF-hand and C-CaM^{D129G} was flexible in both EF-hands. Slow dynamical data showed an overall increase in dynamics for all CaM variants, most notably for CaM^{D95V}

and CaM^{D129G}. In fact, the dynamics of C-CaM^{D129G} in its apo state were too fast to be detected by ¹⁵N CPMG RD and were measured using ¹⁵N R_{1ρ} RD yielding $k_{ex} = 26000 \pm 900 \text{ s}^{-1}$, that of apo wild type CaM was $k_{ex} = 2860 \pm 50 \text{ s}^{-1}$. The slow dynamics of apo CaM^{F141L} were different from the wild type indicating destabilization, even though the structural data showed a high degree of similarity. To further investigate why C-CaM^{F141L} may be involved with LQTS we investigated the dynamics of the methyl groups of the methionines forming the substrate binding pocket [20]. The investigation yielded that C-CaM^{F141L} side chain dynamics of methionines were significantly increased from that of the wild type in both apo and holo states. Furthermore, through homology modeling it was found that the formation of the binding pocket was absent for C-CaM^{F141L}.

Collectively our results show that C-CaM^{D95V} fails to bind Ca²⁺ in the low affinity site of EF-hand III, where the mutation is located, C-CaM^{D129G} only binds Ca²⁺ transiently not introducing any major structural changes in doing so and C-CaM^{F141L} binds Ca²⁺ in close analogy to the wild type but does not form a pocket essential for ligand binding in its Ca²⁺-bound state. Preassociation cannot be excluded as an important factor in the phenotypic pathway of LQTS and it has previously been proposed that apo CaM binds the ryanodine receptor and LTCCs [39, 43, 123]. However, it has been shown that for these three variants preassociation to Ca_v1.2 is still possible [39]. From our results it was thus suggested that apo C-CaM variants still may preassociate to myocyte targets but not perform the task as a modulating Ca²⁺ sensor properly and thus affecting the ability of the ion channels to properly regulate the influx and efflux of ions to the cells. C-CaM^{D95V} and C-CaM^{D129G} may cause LQTS by not being able to respond properly to Ca²⁺ whereas for C-CaM^{F141L} the likely involvement in LQTS is failure to recognize targets or having a lower target affinity, even though the response to Ca²⁺ may be sufficient to trigger the conformational switch.

Conclusions and future perspectives

The scientific man does not aim at an immediate result.

He does not expect that his advanced ideas will be readily taken up.

His work is like that of the planter - for the future.

His duty is to lay the foundation for those who are to come, and point the way.

He lives and labors and hopes.

Nikola Tesla

Three of the papers presented in this thesis describe developed software that can be used for streamlined data analysis of proteins and the results presented in the fourth paper are generated by the software and describes a biophysical characterization study of protein variants of CaM.

COMPASS was my first conceived software and project. The project started during my master's thesis when I during manual assignment of a 28 kDa protein realized that the vast majority of the time spent assigning the protein was spent on typing labels for each assignment and browsing through spectra. By writing a small script that automatically generated and analyzed fragments of paired resonances I noticed that my efficiency was markedly increased. The work on COMPASS continued into my PhD studies and by the end of the project the small script had been extended to become a GUI application. Now at the end of my time as a PhD student I have identified several features that could be implemented and improve COMPASS as a result of my personal development as a scientist and programmer. Programmatically speaking, I would like to redo several algorithms to become more robust and efficient and to include features such as a spectrum viewer, strip plot functionality and allow for more types of experiments to be used for input data. Several of these features are already implemented in PINT so it would be easy to implement them in COMPASS. There are already excellent software for similar purposes, but as COMPASS is unique in its performance it would benefit from these additions and a long term vision

of mine is even to bundle COMPASS together with PINT to create a suite of NMR spectroscopy software. One scientific improvement to be added is to be able to distinguish chemical shifts of reduced (free) and oxidized (disulfide bonded) cysteines as well as prolines in cis and trans conformations. Currently COMPASS utilize statistics representing cysteines in a reduced state but as $^{13}\text{C}^\alpha$ and especially $^{13}\text{C}^\beta$ chemical shifts are sensitive to the existence or absence of a disulfide bond [124] this could easily be implemented. The same approach is valid for prolines where $^{13}\text{C}^\beta$ chemical shifts are sensitive to cis/trans conformations [125]. Further potential improvements include to allow for easier detection of ambiguities through intercomparison of fragments and to update the chemical shift database and perhaps even make the database connect directly to the Biological Magnetic Resonance Data Bank (<http://www.bmrb.wisc.edu/>). To include analysis of data and spectra from additional experiments such as NOESY is also a potential improvement.

Even though there are several features that can be added or altered, the software in its current state fills the purpose for which it was developed. COMPASS has continuously been used in both our research and teaching and has drastically sped up the time spent for assigning proteins.

The software CDpal is very useful and has made the task of curve fitting less cumbersome for users. Being automated and robust have made CDpal a useful tool in academic research as well as a method for teaching students protein thermodynamics. The implemented algorithm for automatic curve fitting has proven to be extremely robust as I have yet to encounter a data set where curve fitting did not properly converge. For thermal denaturation, the algorithm uses a constant transition enthalpy ($\text{N} \rightleftharpoons \text{D}$) of 350 kJ mol^{-1} as an initial guess regardless of input data, the inflection point(s) is/are used to estimate midpoint(s) of the denaturation without regard to data noise and the five leading and trailing data points of the data set define the native and denatured states, respectively, through linear regression. Collectively this approach makes up for excellent start parameters for curve fitting thermal denaturation data; the algorithm uses a similar approach when fitting chemical denaturation data.

When writing this sentence, CDpal has been adopted in teaching as well as research, with good prospects for the future. Currently, CDpal fulfils the purpose it was designed for, namely streamlined, rapid and accurate analysis of protein thermal and chemical stability. More complicated models for curve fitting, describing the denaturation of oligomers, are potential software updates that can be implemented by any advanced user. A planned improvement to CDpal is to include fitting of the first derivative of supplied thermal denaturation data. This

can be a useful resort in cases where the standard models cannot properly fit a data set. It is hard to determine a reliable thermal denaturation temperature for highly stable proteins where the denaturation curve does not reach the plateau for the denatured state. A tentative plan for the future is thus to add support for fitting thermal denaturation curves of samples with added denaturant. By recording multiple experiments with increasing concentrations of denaturant, the T_m can potentially be determined through extrapolation.

In my opinion, the software PINT is the pinnacle of this thesis. It has the potential to appeal to both advanced and inexperienced users in the NMR spectroscopy field. Even prior the publication of the paper, describing the new version of PINT, it had been adopted in several labs. Even if speed often is of importance, the most important factor that constitute the performance of PINT is the results. It was of key concern to present the results to the user in a transparent way, meaning that users themselves must be allowed to inspect the quality of the fit and not blindly trust quantities outputted by PINT. To achieve this we presented the data in multiple ways. First we presented a quality rating of each fitted peak to guide the eye of the user when trying to spot poorly fitted peaks. This was followed by rendering superimposed plots of experimental and fitted data together with a difference map to spot regions with errors. We also added the possibility to individually plot peaks in overlapped regions to increase transparency. Finally we presented the extracted parameters, including line widths, optimized peak positions, intensities, volumes and statistics for the fit. Collectively, this makes up for an excellent way of efficiently and accurately evaluating the analyzed data. Among the improvements made to reach out to a wider user base was to turn PINT into a cross-platform application. This turned out to be a wise decision as the Windows version of PINT, previously not supported, currently make up for 50% of the total number of downloads.

Effort has been put into making the software user-friendly, intuitive and well-documented and it is hard to estimate the impact this will have on the number of users and their experience with PINT, but hopefully it will not be difficult to get information on any question that may arise during usage. When letting novice NMR spectroscopists, not involved with the production of PINT, try the software they quickly got used to the workflow and managed to produce satisfying results with little guidance; this was truly a good sign. Even so, it is virtually impossible to create a foolproof and error-free application of this magnitude and any reported feedback will be helpful to continuously improve PINT.

One of the difficulties when planning and designing PINT was to create an open-ended platform that allowed for the addition of new features. As new

additions have been continuously added throughout development, I believe that this have been achieved. Useful future improvements include the addition of tailored support for the integration of titration data and even 3D spectra. The support to further analyze downstream analysis results within PINT is another potential improvement.

The fourth paper describing a biophysical characterization of three novel CaM variants associated with LQTS using NMR spectroscopy adds important insight into their structural and dynamical properties. The main focus of papers published on this subject are interaction studies with candidate targets. These studies are very important and relevant in order to understand the phenotype, finding the correct treatment and for drug design. Our study contributes to the research in a different but equally important way, by providing detailed information on a per nucleus basis. This is ultimately very important for drug design and to increase the understanding of results from interaction studies, which often report on averaged results for the total sample. Even though it is easy to hypothesize that a mutation in the consensus pattern of an EF-hand of CaM will impact its Ca^{2+} binding affinity and through methods as fluorescence spectroscopy measure affinities for respective N- or C-lobe, it cannot and must not be assumed that the altered affinity is only due to one EF-hand not responding to calcium as binding is known to be positive cooperative. As seen for both D95V and D129G, the EF-hands of the C-lobe are structurally affected by the single mutations and neither EF-hand's response to Ca^{2+} is comparable to that of the wild type. This evidence could not have been provided without utilizing a higher resolution method such as NMR spectroscopy.

When analyzing the data presented in this paper all resulting software of Paper I-III have been used. Although not presented in the results of the final paper, CDpal (Paper II) was used during an early phase of the project to validate previously published data. The current research on the involvement of CaM in LQTS has revealed novel variants and investigated myocyte target interactions. An interesting potential study for the future is to monitor titrations of isotopically labeled full-length CaM to target myocytes in presence and absence of Ca^{2+} and investigate if variants bind the targets the same way as the wild type. It would also be interesting to elucidate if the functionality of the N-domain is intact for variants with C-domain mutations. Based on recent research [38–40], I find that the most promising candidate myocyte target to subject to further studies together with the CaM variants is the $\text{Ca}_v1.2$ channel.

I conclude my work on this thesis with great hope and anticipation that the published methods become parts of something bigger and that our CaM study is a building block in finding a treatment to LQTS and CPVT caused by CaM variants.

References

---

1. Paoli, M., Liddington, R., Tame, J., Wilkinson, A. & Dodson, G. Crystal Structure of T State Haemoglobin with Oxygen Bound At All Four Haems. *J. Mol. Biol.* **256**, 775–792 (1996).
2. Callaway, E. The revolution will not be crystallized. *Nature* **525**, 172 (2015).
3. Lipari, G. & Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).
4. Lipari, G. & Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **104**, 4559–4570 (1982).
5. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C. & Gronenborn, A. M. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **112**, 4989–4991 (1990).
6. Malmendal, A., Evenäs, J., Forsén, S. & Akke, M. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. *J. Mol. Biol.* **293**, 883–899 (1999).
7. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590 (2004).
8. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).

9. Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Mol. Biol.* **7**, 740–743 (2000).
10. Cooper, A. Protein Heat Capacity: An Anomaly that Maybe Never Was. *J. Phys. Chem. Lett.* **1**, 3298–3304 (2010).
11. Prabhu, N. V. & Sharp, K. A. Heat Capacity in Proteins. *Annu. Rev. Phys. Chem.* **56**, 521–48 (2005).
12. Myers, J. K., Pace, N. C. & Scholtz, M. J. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).
13. Velicelebi, G. & Sturtevant, J. M. Thermodynamics of the denaturation of lysozyme in alcohol-water mixtures. *Biochemistry* **18**, 1180–1186 (1979).
14. Dzwolak, W., Ravindra, R. & Winter, R. Hydration and structure—the two sides of the insulin aggregation process. *Phys. Chem. Chem. Phys.* **6**, 1938–1943 (2004).
15. Greenfield, N. J. & Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).
16. Ohgushi, M. & Wada, A. 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett.* **164**, 21–24 (1983).
17. Fink, A. L., Oberg, K. A. & Seshadri, S. Discrete intermediates versus molten globule models for protein folding: characterization of partially folded intermediates of apomyoglobin. *Fold. Des.* **3**, 19–25 (1998).
18. Freire, E. [6] Thermal denaturation methods in the study of protein folding. *Methods Enzymol.* **259**, 144–168 (1995).
19. Kretsinger, R. H. & Nockolds, C. E. Carp Muscle Calcium-binding Protein. *J. Biol. Chem.* **248**, 3313–3326 (1973).
20. Finn, B. E., Evenäs, J., Drakenberg, T., Waltho, J. P., Thulin, E. & Forsén, S. Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Mol. Biol.* **2**, 777–783 (1995).
21. Olsson, L.-L. & Sjölin, L. Structure of Escherichia coli fragment TR2C from calmodulin to 1.7 Å resolution. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **57**, 664–669 (2001).
22. Toutenhoofd, S. L., Foletti, D., Wicki, R., Rhyner, J. A., Garcia, F., Tolon, R. & Strehler, E. E. Characterization of the human CALM2 calmodulin gene and comparison of the transcriptional activity of CALM1, CALM2 and CALMS. *Cell Calcium* **23**, 323–338 (1998).

23. Linse, S., Helmersson, A. & Forsén, S. Calcium binding to calmodulin and its globular domains. *J. Biol. Chem.* **266**, 8050–8054 (1991).
24. Tidow, H. & Nissen, P. Structural diversity of calmodulin binding to its target sites. *FEBS J.* **280**, 5551–5565 (2013).
25. Pitt, G. S., Zühlke, R. D., Hudmon, A., Schulman, H., Reuter, H. & Tsien, R. W. Molecular basis of calmodulin tethering and Ca²⁺-dependent inactivation of L-type Ca²⁺ channels. *J. Biol. Chem.* **276**, 30794–30802 (2001).
26. Nyegaard, M., Overgaard, M. T., Søndergaard, M. T., Vranas, M., Behr, E. R., Hildebrandt, L. L., Lund, J., Hedley, P. L., Camm, A. J., Wettrell, G., Fosdal, I., Christiansen, M. & Børglum, A. D. Mutations in Calmodulin Cause Ventricular Tachycardia and Sudden Cardiac Death. *Am. J. Hum. Genet.* **91**, 703–712 (2012).
27. Crotti, L., Johnson, C. N., Graf, E., Ferrari, G. M. D., Cuneo, B. F., Ovadia, M., Papagiannis, J., Feldkamp, M. D., Rathi, S. G., Kunic, J. D., Pedrazzini, M., Wieland, T., Lichtner, P., Beckmann, B.-M., Clark, T., Shaffer, C., Benson, D. W., Kääh, S., Meitinger, T., Strom, T. M., Chazin, W. J., Schwartz, P. J. & Alfred L. George, J. Calmodulin Mutations Associated with Recurrent Cardiac Arrest in Infants. *Circulation* **127**, 1009–1017 (2013).
28. Marsman, R. F., Barc, J., Beekman, L., Alders, M., Dooijes, D., van den Wijngaard, A., Ratbi, I., Sefiani, A., Bhuiyan, Z. A., Wilde, A. A. & Bezzina, C. R. A Mutation in CALM1 Encoding Calmodulin in Familial Idiopathic Ventricular Fibrillation in Childhood and Adolescence. *J. Am. Coll. Cardiol.* **63**, 259–266 (2014).
29. Makita, N., Yagihara, N., Crotti, L., Johnson, C. N., Beckmann, B.-M., Roh, M. S., Shigemizu, D., Lichtner, P., Ishikawa, T., Aiba, T., Homfray, T., Behr, E. R., Klug, D., Denjoy, I., Mastantuono, E., Theisen, D., Tsunoda, T., Satake, W., Toda, T., Nakagawa, H., Tsuji, Y., Tsuchiya, T., Yamamoto, H., Miyamoto, Y., Endo, N., Kimura, A., Ozaki, K., Motomura, H., Suda, K., Tanaka, T., Schwartz, P. J., Meitinger, T., Kääh, S., Guicheney, P., Shimizu, W., Bhuiyan, Z. A., Watanabe, H., Chazin, W. J. & George, A. L. Novel Calmodulin (CALM2) Mutations Associated with Congenital Arrhythmia Susceptibility. *Circ. Cardiovasc. Genet.* **7**, 466–474 (2014).
30. Pipilas, D. C., Johnson, C. N., Webster, G., Schlaepfer, J., Fellmann, F., Sekarski, N., Wren, L. M., Ogorodnik, K. V., Chazin, D. M., Chazin, W. J., Crotti, L., Bhuiyan, Z. A. & Jr, A. L. G. Novel calmodulin mutations associated with congenital long QT syndrome affect calcium current in human cardiomyocytes. *Heart Rhythm* **13**, 2012–2019 (2016).

31. Jervell, A. & Lange-Nielsen, F. Congenital deaf-mutism, functional heart disease with prolongation of the Q-T interval, and sudden death. *Am. Heart. J.* **54**, 59–68 (1957).
32. Tester, D. J. & Ackerman, M. J. Genetics of long QT syndrome. *Methodist DeBakey Cardiovasc J* **10**, 29–33 (2014).
33. Schwartz, P. J., Priori, S. G., Spazzolini, C., Moss, A. J., Vincent, G. M., Napolitano, C., Denjoy, I., Guicheney, P., Breithardt, G., Keating, M. T., Towbin, J. A., Beggs, A. H., Brink, P., Wilde, A. A. M., Toivonen, L., Zareba, W., Robinson, J. L., Timothy, K. W., Corfield, V., Wattanasirichaigoon, D., Corbett, C., Haverkamp, W., Schulze-Bahr, E., Lehmann, M. H., Schwartz, K., Coumel, P. & Bloise, R. Genotype-phenotype correlation in the long-QT syndrome: gene-specific triggers for life-threatening arrhythmias. *Circulation* **103**, 89–95 (2001).
34. Moss, A. J. & Kass, R. S. Long QT syndrome: from channels to cardiac arrhythmias. *J. Clin. Invest.* **115**, 2018–2024 (2005).
35. Ghosh, S., Nunziato, D. A. & Pitt, G. S. KCNQ1 assembly and function is blocked by long-QT syndrome mutations that disrupt interaction with calmodulin. *Circ. Res.* **98**, 1048–1054 (2006).
36. Shamgar, L., Ma, L., Schmitt, N., Haitin, Y., Peretz, A., Wiener, R., Hirsch, J., Pongs, O. & Attali, B. Calmodulin Is Essential for Cardiac IKs Channel Gating and Assembly. *Circ. Res.* **98**, 1055–1063 (2006).
37. Sorensen, A. B., Søndergaard, M. T. & Overgaard, M. T. Calmodulin in a Heartbeat. *FEBS J.* **280**, 5511–5532. issn: 1742–4658 (2013).
38. Yin, G., Hassan, F., Haroun, A. R., Murphy, L. L., Crotti, L., Schwartz, P. J., George, A. L. & Satin, J. Arrhythmogenic Calmodulin Mutations Disrupt Intracellular Cardiomyocyte Ca²⁺ Regulation by Distinct Mechanisms. *J. Am. Heart Assoc.* **3**, e000996 (2014).
39. Limpitkul, W. B., Dick, I. E., Joshi-Mukherjee, R., Overgaard, M. T., Jr., A. L. G. & Yue, D. T. Calmodulin mutations associated with long QT syndrome prevent inactivation of cardiac L-type Ca²⁺ currents and promote proarrhythmic behavior in ventricular myocytes. *J. Mol. Cell. Cardiol.* **74**, 115 – 124 (2014).
40. Rocchetti, M., Sala, L., Dreizehnter, L., Crotti, L., Sinnecker, D., Mura, M., Pane, L. S., Altomare, C., Torre, E., Mostacciuolo, G., Severi, S., Porta, A., De Ferrari, G. M., George, A. L. Jr, Schwartz, P. J., Gneccchi, M., Moretti, A. & Zaza, A. Elucidating arrhythmogenic mechanisms of long-QT syndrome

- CALM1-F142L mutation in patient-specific induced pluripotent stem cell-derived cardiomyocytes. *Cardiovasc. Res.* **113**, 531–541 (2017).
41. Tester, D. J., Kopplin, L. J., Will, M. L. & Ackerman, M. J. Spectrum and prevalence of cardiac ryanodine receptor (RyR2) mutations in a cohort of unrelated patients referred explicitly for long QT syndrome genetic testing. *Heart Rhythm* **2**, 1099–1105 (2005).
  42. Petegem, F. V. Ryanodine Receptors: Structure and Function. *J. Biol. Chem.* **287**, 31624–31632 (2012).
  43. Lau, K., Chan, M. M. Y. & Van Petegem, F. Lobe-Specific Calmodulin Binding to Different Ryanodine Receptor Isoforms. *Biochemistry* **53**, 932–946 (2014).
  44. Søndergaard, M. T., Liu, Y., Larsen, K. T., Nani, A., Tian, X., Holt, C., Wang, R., Wimmer, R., Petegem, F. V., Fill, M., Chen, S. R. W. & Overgaard, M. T. The Arrhythmogenic Calmodulin p.Phe142Leu Mutation Impairs C-domain Ca²⁺ Binding but Not Calmodulin-dependent Inhibition of the Cardiac Ryanodine Receptor. *J. Biol. Chem.* **292**, 1385–1395 (2017).
  45. Søndergaard, M. T., Tian, X., Liu, Y., Wang, R., Chazin, W. J., Chen, S. R. W. & Overgaard, M. T. Arrhythmogenic Calmodulin Mutations Affect the Activation and Termination of Cardiac Ryanodine Receptor-mediated Ca²⁺ Release. *J. Biol. Chem.* **290**, 26151–26162 (2015).
  46. Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. A New Method of Measuring Nuclear Magnetic Moment. *Phys. Rev.* **53**, 318 (1938).
  47. Bloch, F., Hansen, W. W. & Packard, M. Nuclear Induction. *Phys. Rev.* **69**, 127 (1946).
  48. Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69**, 37–38 (1946).
  49. Tugarinov, V., Choy, W.-Y., Orekhov, V. Y. & Kay, L. E. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. USA* **102**, 622–627 (2005).
  50. Rovnyak, D., Frueh, D. P., Sastry, M., Sun, Z.-Y. J., Stern, A. S., Hoch, J. C. & Wagner, G. Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J. Mag. Res.* **170**, 15–21 (2004).

51. Venters, R. A., II, B. T. F., Fierke, C. A. & Spicer, L. D. Characterizing the Use of Perdeuteration in NMR Studies of Large Proteins: ^{13}C , ^{15}N and ^1H Assignments of Human Carbonic Anhydrase II. *J. Mol. Biol.* **264**, 1101–1116 (1996).
52. Keeler, J. *Understanding NMR spectroscopy* (John Wiley & Sons, 2011).
53. Spera, S. & Bax, A. Empirical correlation between protein backbone conformation and C.alpha. and C.beta. ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**, 5490–5492 (1991).
54. Shen, Y. & Bax, A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **48**, 13–22 (2010).
55. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
56. Ikura, M., Kay, L. E. & Bax, A. A novel approach for sequential assignment of proton, carbon- ^{13} , and nitrogen- ^{15} spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659–4667 (1990).
57. Bax, A. & Ikura, M. An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α -carbon of the preceding residue in uniformly $^{15}\text{N}/^{13}\text{C}$ enriched proteins. *J. Biomol. NMR* **1**, 99–104 (1991).
58. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ^{15}N , ^{13}C , ^2H Labeled Proteins with High Sensitivity. *J. Am. Chem. Soc.* **116**, 11655–11666 (1994).
59. Grzesiek, S. & Bax, A. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293 (1992).
60. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Mag. Res.* **99**, 201–207 (1992).
61. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371 (1997).

62. Salzmänn, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **121**, 844–848 (1999).
63. Salzmänn, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. USA* **95**, 13585–13590 (1998).
64. Olejniczak, E. T. & Fesik, S. W. Two-Dimensional Nuclear Magnetic Resonance Method for Identifying the H. α ./C. α . Signals of Amino Acid Residues Preceding Proline. *J. Am. Chem. Soc.* **116**, 2215–2216 (1994).
65. Bottomley, M., Macias, M., Liu, Z. & Sattler, M. A novel NMR experiment for the sequential assignment of proline residues and proline stretches in $^{13}\text{C}/^{15}\text{N}$ -labeled proteins. *J. Biomol. NMR* **13**, 381–385 (1999).
66. Hafsa, N. E., Arndt, D. & Wishart, D. S. CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Res.* **43**, W370–W377 (2015).
67. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: Implications for fibrillation. *Protein Sci.* **15**, 2795–2804 (2006).
68. Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperski, T., Montelione, G. T., Baker, D. & Bax, A. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci.* **105**, 4685–4690 (2008).
69. Shen, Y., Vernon, R., Baker, D. & Bax, A. De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* **43**, 63–78 (2009).
70. Bloch, F. Nuclear Induction. *Phys. Rev.* **70**, 460–474 (1946).
71. McConnell, H. M. Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* **28**, 430 (1958).
72. Bloembergen, N., Purcell, E. M. & Pound, R. V. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Phys. Rev.* **73**, 679–712 (1948).
73. Solomon, I. Relaxation Processes in a System of Two Spins. *Phys. Rev.* **99**, 559–565 (1955).

74. Harris, R. K., Becker, E. D., De Menezes, S. M. C., Granger, P., Hoffman, R. E. & Zilm, K. W. Further Conventions for NMR Shielding and Chemical Shifts (IUPAC Recommendations 2008). *Magn. Reson. Chem.* **46**, 582–598 (2008).
75. Proctor, W. G. & Yu, F. C. On the Nuclear Magnetic Moments of Several Stable Isotopes. *Phys. Rev.* **81**, 20–30 (1951).
76. Gutowsky, H. S., McCall, D. W. & Slichter, C. P. Coupling among Nuclear Magnetic Dipoles in Molecules. *Phys. Rev.* **84**, 589–590 (1951).
77. Bruch, M. *NMR spectroscopy techniques* 125 (CRC Press, 1996).
78. Mason, J. *Multinuclear NMR* 150 (Springer Science & Business Media, 2012).
79. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814**, 942–968 (2011).
80. Overhauser, A. W. Polarization of Nuclei in Metals. *Phys. Rev.* **92**, 411–415 (1953).
81. Carver, T. R. & Slichter, C. P. Polarization of Nuclear Spins in Metals. *Phys. Rev.* **92**, 212–213 (1953).
82. Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979 (1989).
83. Hahn, E. L. Spin Echoes. *Phys. Rev.* **80**, 580–594 (1950).
84. Carr, H. Y. & Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **94**, 630–638 (1954).
85. Meiboom, S. & Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
86. Korzhnev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. A. & Kay, L. E. An NMR Experiment for the Accurate Measurement of Heteronuclear Spin-Lock Relaxation Rates. *J. Am. Chem. Soc.* **124**, 10743–10753 (2002).
87. Tollinger, M., Skrynnikov, N. R., Mulder, F. A. A., Forman-Kay, J. D. & Kay, L. E. Slow Dynamics in Folded and Unfolded States of an SH3 Domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).
88. Carver, J. P. & Richards, R. E. A general two-site solution for the chemical exchange produced dependence of T₂ upon the carr-Purcell pulse separation. *J. Magn. Reson.* **6**, 89–105 (1969).

89. Davis, D. G., Perlman, M. E. & London, R. E. Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1 ρ and T2 (CPMG) methods. *J. Magn. Reson., Ser B* **104**, 266–275 (1994).
90. Frueh, D. P. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **78**, 47–75 (2014).
91. Bahrami, A., Assadi, A. H., Markley, J. L. & Eghbalnia, H. R. Probabilistic Interaction Network of Evidence Algorithm and its Application to Complete Labeling of Peak Lists from Protein NMR Spectroscopy. *PLOS Comput. Biol.* **5**, 1–15 (2009).
92. Zimmerman, D. E., Kulikowski, C. A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., ya Chien, C., Powers, R. & Montelione, G. T. Automated analysis of protein NMR assignments using methods from artificial intelligence. *J. Mol. Biol.* **269**, 592–610 (1997).
93. Crippen, G. M., Rousaki, A., Revington, M., Zhang, Y. & Zuiderweg, E. R. P. SAGA: rapid automatic mainchain NMR assignment for large proteins. *J. Biomol. NMR* **46**, 281–298 (2010).
94. Zuiderweg, E. R. P., Bagai, I., Rossi, P. & Bertelsen, E. B. EZ-ASSIGN, a program for exhaustive NMR chemical shift assignments of large proteins from complete or incomplete triple-resonance data. *J. Biomol. NMR* **57**, 179–191 (2013).
95. Muller, B., Kistner, U., Veh, R., Cases-Langhoff, C., Becker, B., Gundelfinger, E. & Garner, C. Molecular characterization and spatial distribution of SAP97, a novel presynaptic protein homologous to SAP90 and the Drosophila discs-large tumor suppressor protein. *J. Neurosci.* **15**, 2354–2366 (1995).
96. Zhang, H., Neal, S. & Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25**, 173–195 (2003).
97. Witte, S., Villalba, M., Bi, K., Liu, Y., Isakov, N. & Altman, A. Inhibition of the c-Jun N-terminal kinase/AP-1 and NF- κ B pathways by PICOT, a novel protein kinase C-interacting protein with a thioredoxin homology domain. *J. Biol. Chem.* **275**, 1902–1909 (2000).
98. Woodson, L. C. & Weinshilboum, R. M. Human kidney thiopurine methyltransferase purification and biochemical properties. *Biochem. Pharmacol.* **32**, 819–826 (1983).
99. Chan, J & Watt, V. Eek and Erk, new members of the Eph subclass of receptor protein-tyrosine kinases. *Oncogene* **6**, 1057–1061 (1991).

100. Lila, T. & Drubin, D. G. Evidence for physical and functional interactions among two *Saccharomyces cerevisiae* SH3 domain proteins, an adenylyl cyclase-associated protein and the actin cytoskeleton. *Mol. Bio. Cell* **8**, 367–385 (1997).
101. Evenäs, J., Thulin, E., Malmendal, A., Forsén, S. & Carlström, G. NMR studies of the E140Q mutant of the carboxy-terminal domain of calmodulin reveal global conformational exchange in the Ca²⁺-saturated state. *Biochemistry* **36**, 3448–3457 (1997).
102. Li, J.-L., Baker, D. A. & Cox, L. S. Sexual stage-specific expression of a third calcium-dependent protein kinase from *Plasmodium falciparum*. *BBA-Gene Struct. Expr.* **1491**, 341–349 (2000).
103. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical recipes in C* (Cambridge Univ Press, 1982).
104. Levenberg, K. A Method for the Solution of Certain Non-linear Problems in Least Squares. *Quart. Appl. Math.* **2**, 164–168 (1944).
105. Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.* **11**, 431–441 (1963).
106. Zhu, X. & Zhang, D. Efficient Parallel Levenberg-Marquardt Model Fitting towards Real-Time Automated Parametric Imaging Microscopy. *PLOS ONE* **8**, 1–9 (2013).
107. Rees, D. C. & Robertson, A. D. Some thermodynamic implications for the thermostability of proteins. *Protein Sci.* **10**, 1187–1194 (2001).
108. Myers, J. K., Nick Pace, C & Martin Scholtz, J. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).
109. Goodard, T. D. & Kneller, D. G. Sparky 3. *University of California, San Francisco*.
110. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325 (2015).
111. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
112. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).

113. Ahlner, A., Carlsson, M., Jonsson, B.-H. & Lundström, P. PINT: a software for integration of peak volumes and extraction of relaxation rates. *J. Biomol. NMR* **56**, 191–202 (2013).
114. Orekhov, V. Y., Ibraghimov, I. & Billeter, M. Optimizing resolution in multidimensional NMR by three-way decomposition. *J. Biomol. NMR* **27**, 165–173 (2003).
115. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
116. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
117. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E. & Kay, L. E. Measurement of bond vector orientations in invisible excited states of proteins. *Proc. Natl. Acad. Sci. USA* **104**, 18473–18477 (2007).
118. Lundström, P., Vallurupalli, P., Religa, T. L., Dahlquist, F. W. & Kay, L. E. A single-quantum methyl ¹³C-relaxation dispersion experiment with improved sensitivity. *J. Biomol. NMR* **38**, 79–88 (2007).
119. Berchtold, M. W., Zacharias, T., Kulej, K., Wang, K., Torggler, R., Jespersen, T., Chen, J.-N., Larsen, M. R. & Cour, J. M. I. The Arrhythmogenic Calmodulin Mutation D129G Dysregulates Cell Growth, Calmodulin-dependent Kinase II Activity, and Cardiac Function in Zebrafish. *J. Biol. Chem.* **291**, 26636–26646 (2016).
120. Vassilakopoulou, V., Calver, B. L., Thanassoulas, A., Beck, K., Hu, H., Buntwal, L., Smith, A., Theodoridou, M., Kashir, J., Blayney, L., Livaniou, E., Nounesis, G., Lai, F. A. & Nomikos, M. Distinctive malfunctions of calmodulin mutations associated with heart RyR2-mediated arrhythmic disease. *Biochim. Biophys. Acta* **1850**, 2168–2176 (2015).
121. Hwang, H.-S., Nitu, F. R., Yang, Y., Walweel, K., Pereira, L., Johnson, C. N., Faggioni, M., Chazin, W. J., Laver, D., George, A. L., Cornea, R. L., Bers, D. M. & Knollmann, B. C. Divergent Regulation of RyR2 Calcium Release Channels by Arrhythmogenic Human Calmodulin Missense Mutants. *Circ. Res.* **114**, 1114–1124 (2014).

122. Biekofsky, R. R., Martin, S. R., Browne, J. P., Bayley, P. M. & Feeney, J. Ca²⁺ Coordination to Backbone Carbonyl Oxygen Atoms in Calmodulin and Other EF-Hand Proteins: ¹⁵N Chemical Shifts as Probes for Monitoring Individual-Site Ca²⁺ Coordination. *Biochemistry* **37**, 7617–7629 (1998).
123. Rodney, G. G., Moore, C. P., Williams, B. Y., Zhang, J.-Z., Krol, J., Pedersen, S. E. & Hamilton, S. L. Calcium Binding to Calmodulin Leads to an N-terminal Shift in Its Binding Site on the Ryanodine Receptor. *J. Biol. Chem.* **276**, 2069–2074 (2001).
124. Sharma, D. & Rajarathnam, K. ¹³C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* **18**, 165–171 (2000).
125. Shen, Y. & Bax, A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J. Biomol. NMR* **46**, 199–204 (2010).

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-142785>