Molecular recognition and dynamics in proteins studied by NMR

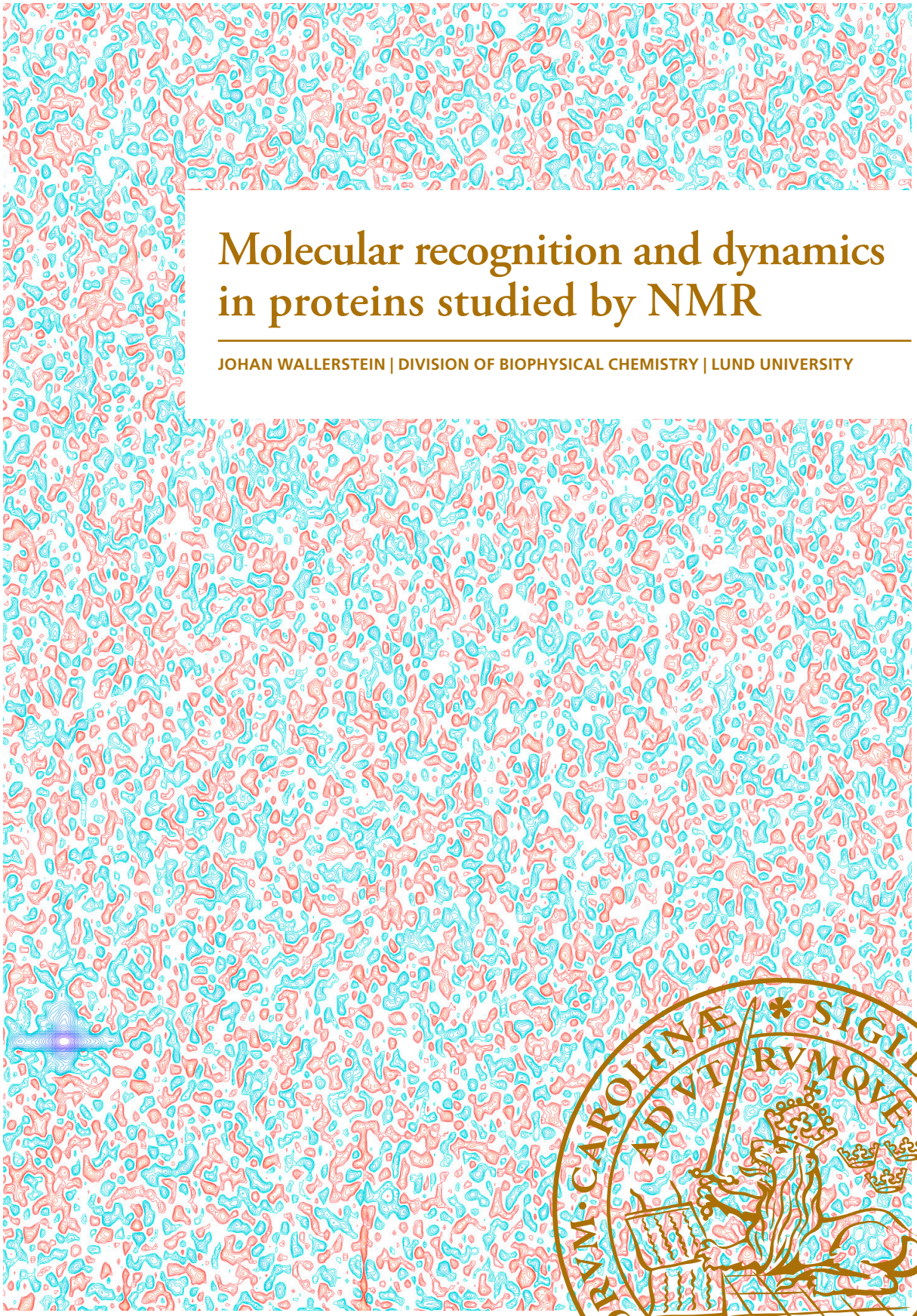
Wallerstein, Johan

2019

Document Version:
Publisher's PDF, also known as Version of record

Molecular recognition and dynamics in proteins studied by NMR

JOHAN WALLERSTEIN | DIVISION OF BIOPHYSICAL CHEMISTRY | LUND UNIVERSITY



Molecular recognition and dynamics in proteins studied by NMR

Johan Wallerstein

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Engineering (LTH) of Lund University, Sweden.
To be defended in lecture hall F at Kemicentrum.
Wednesday, 20th of November 2019 at 1.15 pm

Faculty opponent
Magnus Wolf-Watz
Department of Chemistry, Umeå University, Sweden

Organization
LUND UNIVERSITY
Biophysical Chemistry
Lund University
PO Box 124
SE-221 00 Lund, Sweden

Author
Johan Wallerstein

Document name
DOCTORAL DISSERTATION

Date of issue
2019-10-10

Title: Molecular recognition and dynamics in proteins studied by NMR

Abstract

Knowledge of dynamics in protein is very important in the description of protein function and molecular recognition. The thesis investigates protein dynamics on time-scales from milli- to sub-nanosecond, with focus on the latter, using NMR spin relaxation experiments on two proteins, the 138-residue carbohydrate recognition domain of galectin-3 (Gal3C) and the 56-residue B1 domain of bacterial protein G (PGB1). Fives studies are presented.

By measuring the exchange contribution to R2-relaxation, as a function of pH, using 13C-CPMG-experiments directed on the side-chain carbonyls of glutamic and aspartic acid in PGB1, we could determine site specific protonation rate constants k(on), and deprotonation rate constants k(off). A linear free-energy relationship between log k(on) and pKa is found, which provides information on the free-energy landscape of the protonation reaction, showing that the variability among residues in these rates arises from charge stabilization of the deprotonated state.

Dimethyl sulfoxide (DMSO) is often used for dissolving nonpolar ligands in protein-ligand studies. DMSO change the viscosity, which is proportional to rotational correlation time. The correlation time of PGB1 and Gal3C was determined for different DMSO-concentrations. Effects of minute additions of DMSO in samples used in spin-relaxation experiments were examined for the case of 2H-methyl side chain model-free studies. Chemical shift perturbations studies on apo-Gal3C show uniform changes of chemical shifts, indicating DMSO has a minor impact on the hydration layer.

15N-backbone and 2H-methyl side chain NMR order parameters (S^2) are determined for three Gal3C complexes using the Lipari-Szabo model-free formalism developed in the 80s. Minor changes in ligand structure generate differences in the conformational entropy. Specifically, the radial distribution of conformational entropy, with ligand in the centre reveals how entropy varies between consecutive shells within the protein. The study combines NMR relaxation with isothermal titration calorimetry (ITC), X-ray crystallography, and computational approaches.

The structure–thermodynamic relationship for halogen bonds C–X (X=F, Cl, Br, and I) between the ligand substituents and the backbone C=O of a glycine in Gal3C was studied with NMR, ITC, X-ray crystallography and computational approaches. The sigma-hole associated with the halogen bond affects the electron density of surrounding 15N nuclei in amides, and consequently the chemical shift of these nuclei. There is a correlation between ITC-determined enthalpy of binding vs amide chemical shift.

In Gal3C bound to natural substrate lactose, the chemical shifts changes of the 13C-E1 in histidine side chains are traced as a function of pH using 1H-13C HSQC-experiments. The experiment helps in determination of the tautomeric state in the proteins four histidines. The histidine in the binding site is in tautomer ND1, another is in tautomer NE2 and two are partly charged. Neutron diffraction studies at pH 7.5 confirms the NMR-studies.

Key words

Chemical exchange, Conformational entropy, Entropy, Galectin-3, Isothermal Titration Calorimetry, ITC, Model-free formalism, Molecular recognition, Nuclear Magnetic Resonance, NMR, PGB1, Protein dynamics, Protein–ligand interactions, Proton-transfer reactions, Spin relaxation, Viscosity

Language
English

ISBN
978-91-7422-694-2 (print)
978-91-7422-695-9 (e-version)

Number of pages: 212

Recipient's notes

Price

Molecular recognition and dynamics in proteins studied by NMR

Johan Wallerstein

LUND
UNIVERSITY

Front cover: “Coherence and noise”, by NMRpipe

Back cover: “The fifth element”, by Anonymous

Division of Biophysical Chemistry
Lund Institute of Technology (LTH)
Lund University, Lund, Sweden

ISBN 978-91-7422-694-2 (print)
ISBN 978-91-7422-695-9 (e-version)

Printed in Sweden by Media-Tryck, Lund University, Lund 2019

Media-Tryck is an environmentally certified and ISO 14001:2015 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

To the one who wants to read…

"We tend to fight the next war in the same way we fought the last one. We are prisoners of our own experience. Many of the things that we learned that worked in World War II were not applicable to the war in Vietnam."

– Samuel V. Wilson[1]

[1] From the 18 hours TV-Documentary, ***The Vietnam War*** (2017), written by *Geoffrey C. Ward* and directed by *Ken Burns* and *Lynn Novick*. The highly decorated general *Samuel V. Wilson* (1923-2017) coined the word "counter-insurgency", later known as COIN. In 2002, prior to the invasion of Iraq, Wilson expressed concern that the United States might repeat its mistakes from previous wars, now specifically referring to Vietnam.

Table of Contents

List of papers and author contributions

I. **Johan Wallerstein**, Ulrich Weininger, M. Ashhar I. Khan, Sara Linse, and Mikael Akke
(2015) "Site-Specific Protonation Kinetics of Acidic Side Chains in Proteins Determined by pH-Dependent Carboxyl 13C NMR Relaxation",
Journal of American Chemical Society

Contributions: The author performed all the experiments and the subsequent analysis of data, and aided in the purification of the protein. The author contributed in the writing of the manuscript and made a large majority of pictures and tables.

II. **Johan Wallerstein**, Mikael Akke
(2018) "Minute Additions of DMSO Affect Protein Dynamics Measurements by NMR Relaxation Experiments through Significant Changes in Solvent Viscosity",
ChemPhysChem

Contributions: The author performed the experiments and the subsequent analysis of data, and contributed in the writing the manuscript, and made all pictures.

III. **Johan Wallerstein**, Majda Misini Ignjatović, Rohit Kumar, Octav Caldararu, Kristoffer Peterson, Hakon Leffler, Derek T. Logan, Ulf J. Nilsson, Ulf Ryde and Mikael Akke
(2019) "Entropy–Entropy compensation between the conformational and solvent degrees of freedom finetunes affinity in ligand binding to galectin-3C",
Manuscript.

Contributions: The author performed the NMR- and ITC-experiments, and the subsequent analysis of data. The author contributed in the writing of the manuscript and made most pictures and tables.

IV. Maria Luisa Verteramo, Majda Misini Ignjatović, Rohit Kumar, **Johan Wallerstein**, Veronika Chadimová, Hakon Leffler, Fredrik Zetterberg, Derek T. Logan, Ulf Ryde, Mikael Akke and Ulf J. Nilsson
(2019) "Structural and thermodynamic studies on halogen-bond interactions in ligand–galectin-3 complexes: electrostatics, solvation and entropy effects", *Manuscript*.

Contributions: The author performed the NMR-experiment, and the subsequent analysis of data, and aided in design and analysis of the ITC-experiments. The author had minor contributions in the writing of the manuscript.

V. Francesco Manzoni, **Johan Wallerstein**, Tobias E. Schrader, Andreas Ostermann, Leighton Coates, Mikael Akke, Matthew P. Blakeley, Esko Oksanen, and Derek T. Logan
(2018) "Elucidation of Hydrogen Bonding Patterns in Ligand-Free, Lactose- and Glycerol-Bound Galectin-3C by Neutron Crystallography to Guide Drug Design",
Journal of Medicinal Chemistry

Contributions: The author performed the NMR-experiment and the subsequent analysis of data. The author had minor contributions in the writing of the manuscript.

Populärvetenskaplig sammanfattning

Att proteiner har med livet att göra är allmänt känt, men exakt på vilket sätt engagerar en stor del av världens naturvetenskapliga forskare. Doktorsavhandling studerar två aspekter av förhållandet ”liv–proteiner”, nämligen hur proteiner kommunicerar – så kallad molekylär igenkänning – och hur proteindynamik kan te sig. Dynamik, till exempel hur en viss aminosyra i protein rör sig, eller hur proteinet förflyttas i sin helhet genom vätskan, sammanlänkas ofta med funktion. Vidare är dynamik en förutsättning för kommunikation för det går rimligtvis inte att kommunicera utan att något rör sig, typiskt läppar eller händer vid mänsklig kommunikation.

Två proteiner – del av ett bakterieprotein och ett sockerbindande protein

Två proteiner har studerats, galektin-3, ett sockerbindande protein, samt en domän av protein G, som är ett bakterieprotein som binder antikroppar. Forskning har bedrivits vid Lunds Universitet under decennier på dessa proteiner, sedan 70-talet för protein G. Avhandlingen är därmed bara en pusselbit i kartläggningen. Karaktären på forskningen får anses vara så pass allmän – grundforskningskaraktär – att man kan hävda att slutsatserna är giltiga för i stort sett alla proteiner. Studierna på galektin-3 är emellertid även inriktade mot medicinsk tillämpning då läkemedelsföretag involverats.

NMR-spektroskopi kräver flytande helium

Proteindynamiken har huvudsakligen studerats med NMR-spektroskopi. NMR står för *nuclear magnetic resonance*, kärnmagnetisk resonans och för att förstå NMR-mätningar av proteindynamik måste vi kortfattat beskriva metoden.

NMR-experimentet inleds med att provet placeras i ett enormt starkt magnetfält på 10–20 tesla, att jämföra med jordens magnetfält som är en miljon gånger svagare (30–60 mikrotesla). För att uppnå så starka magnetfält krävs supraledning, som i sin tur kräver extremt låg temperatur ner mot absoluta nollpunkten (0 K, eller –273,15 °C). Därför omsluts magneten av flytande helium med en temperatur kring 4 K. Helium är en sinande resurs och det är av vikt att materialforskning uppfinner material som är supraledande vid högre temperaturer.

När proteinprovet placeras i magneten skapas energinivåskillnader (s k *Zeeman-effekt*) mellan spinnenergier i atomkärnan. Spinn är en kvantmekanisk egenskap och liknelsen små roterande *magnetiska dipoler,* eller leksakssnurra brukar användas, men lika ofta påpekas att det inte finns något som verkligen korresponderar mot spinn i vår ”klassiska värld”. Max Plancks *kvanthypotes* säger i alla fall att spinnenerginivåskillnaden ska vara proportionell mot frekvensen, och eftersom det är små skillnader (trots enorma

magnetfält) bestrålas NMR-provet med lågfrekvent elektromagnetisk strålning, mer precist radiovågor i intervallet 50–1000 MHz.

Halvtaligt spinn

Väte som är den vanligaste atomen i biologisk materia, har en kärna som är särskilt relevant för NMR eftersom 99,9% av allt väte har så kallat halvtaligt kärnspinn. Typen av spinn, halv, hel, 3/2 o s v bestäms av sammansättningen av protoner och neutroner. Av vikt i teorin kring halvtaligt spinn är att det ger upphov till två energinivåer.

Andra viktiga biomolekylatomer är kol (C), kväve (N) och syre (O) men bara C och N har isotoper med halvtaligt kärnspinn. 13C- och 15N-isotoperna har emellertid bara ca 1% naturlig förekomst. Genom att mata bakterier som producerar galektin och protein G domänen med näringsämnen inmärkta med 100% 13C och 100% 15N erhålles nästan 100% NMR-aktiva kol- och kvävekärnor i proteinerna. Metoden för framställning av isotopinmärkta proteiner uppstod på 90-talet och har revolutionerat proteinforskningen.

Resonanseffekt och kemiskt skift

Magnetfältet skapar således en polarisering, två energinivåer för atomkärnorna i proteinet, och nu kommer vi till en viktig del av experimentet för vid bestrålning med en radiovågfrekvens motsvarande *spinnenerginivåskillnaden* (kvanthypotesen) uppstår resonansfenomenet som gett NMR-metoden dess namn. Observera att radiovåg-*amplituden* inte behöver vara stor för att resonansfenomenet ska uppstå, det viktiga är frekvensen. Innebörden av resonans är just detta, en slags synkronisering av frekvenser, eller ”timing”. När radiosignalen stängs av fortsätter kärnspinnen sända samma frekvens, som klingar av under loppet av cirka en sekund. Mer korrekt beskrivet är att kärnspinnen sänder ut *aningen olika* frekvens, som sällan skiljer mer än 100 Hz. Skillnaden förklaras av att spinnfrekvensen beror på den kemiska strukturen, *elektrontätheten* kring atomen. Just att frekvensen som mäts är något skild från den frekvens som bestrålar provet är komplicerat och inte något vi går in på i en kort text, men den lilla skillnaden kallas *kemiskt skift* eftersom det har med ”kemin” att göra. Fysiker kom på namnet, möjligen då de inte förstod fenomenet. Tack vare kemiska skiftet har molekyler unika spektra. I synnerhet stora molekyler som proteiner har remarkabelt unika spektra.

Proteindynamik mäts med kärnspinnrelaxation

Signalens avklingning anses vara komplicerad att förstå än skapandet och kräver en något längre utläggning. *NMR-spinnrelaxation* är benämningen på avklingningsfenomenet som i en mening är begripligt eftersom det är en återgång till jämvikt. Relaxationen är den mätbara storhet som ger information om proteindynamik på atomnivå. Avklingning på cirka en sekund kan låta snabbt men i atomernas värld är det en ansenlig tidsrymd. Skälet att det går långsamt är att NMR-metoden inte medger *spontan* eller *stimulerad relaxation* från exciterat till grundtillstånd, vilket är regeln för t ex optisk spektroskopi. Einstein beskrev detta och förenklat beror avsaknad av den sortens relaxation inom NMR på radiovågornas låga frekvens. Nej, det som relaxerar kärnspinnen är egeninducerade radiofrekvenser, av samma storleksordning som skapade signalen. Dessa uppstår när kärnspinnen, som alltså kan ses som små magnetiska dipoler och följaktligen också har

ett magnetfält, kommer nära varandra. Dipol–dipol-magnetfälten interagerar stokastiskt, eller slumpmässigt. Även här kommer Einsteins teorier in eftersom han först förklarade *brownsk rörelse* vilken kan sägas beskriva *rotationsdiffusionen* som leder till dipol–dipol-fluktuationerna. Rotationsdiffusion kan liknas vid en plastboll som guppar i en bassäng, små slumpmässiga ”störningar” får bollen, sett över längre tid, att rotera flera varv, i alla riktningar.

Alltså, det bör nu gå att föreställa sig att proteinstorlek har betydelse för *fördelningen av frekvenser* som rotationsdiffusionen, via dipol–dipol-fluktuationen, ger upphov till och då är vi nära metodens essens eftersom det uppenbarligen finns samband mellan kärnspinnrelaxation och proteiners dynamik. Ett annat ord för frekvensfördelning är *spektraltäthet* och relaxationskonstanter kan matematiskt skrivas som en serie av spektraltätheter. Återigen har “timing” betydelse, för det råkar vara så att givet ett magnetfält (och dessa varierar mellan magneter) så är det bara en viss protein-storlek som skapar maximal relaxation. Detta illustreras i avhandlingen men är ett välkänt faktum.

Resultat

En mer ingående beskrivning, i frånvaro av matematik, om kärnspinnrelaxation bör nog inte göras i en förenklad sammanfattning. Jo en sak bör tilläggas, metoden är enormt kraftfull, och avhandlingen kan faktiskt visa på unika resultat. Till exempel har vi bestämt, på atomnivå, hur olika delar av galektin, särskilt de nära bindningsfickan, rör sig vid små ändringar av sockerarten. Detta åstadkoms genom att låta galektin binda till tre snarlika sockerarter och studera dynamikskillnader som då uppstår. Kärnspinn-relaxationen har även möjliggjort bestämning av hastighetskonstanter för vätejoner som hoppar av och på aminosyror i protein G-domänen. Det är en ny upptäckt och av visst värde för teoretisk modellering av proteindynamik. I framtiden kommer antagligen datorsimuleringar dominera proteindynamik-fältet, men experiment behövs likväl. För inte vill väl folk ta läkemedel som *enbart* simulerats fram?

Dynamik länkas till entropi och termodynamik

Vi har här beskrivit NMR-metoden och dess tillämpning på proteinforskning, men grunden för all fysikalisk kemi är termodynamik, om vilken Albert Einstein sagt ungefär (en tredje och sista Einstein-referens): ”Den klassiska termodynamiken är den enda fysikaliska teorin vilken jag är övertygad om att inte kommer kullkastas”. Proteiners dynamik kan relateras till frihetsgrader och detta begrepp är sammanlänkat med det kanske viktigaste begreppet inom termodynamiken – *entropin* (enhet: energi/temperatur). Fler frihetsgrader i ett system ökar entropin och den *totala* entropin strävar *alltid* efter ett maximum vid varje kemisk reaktion, följaktligen i alla livsprocesser. Som en motpol till entropin finns *energin*, det som upprätthåller ordningen genom att binda atomer samman och skapa strukturer. Dualismen entropi–energi är fundamental inom termo*dynamiken*.

Sammanfattningsvis*,* avhandlingens *specifika* bidrag är en ökad förståelse om proteiners dynamik, kommunikation och funktion, men det finns även ett *allmänt* bidrag, av många bedömare nog betraktat som försvinnande litet men likväl inte noll, nämligen ett bidrag som skapar aningen mer *ordning* och förståelse av begreppen energi och entropi.

Preface

Everything must come to an end. Some say the circle never ends. One day might come though, when nothing exists, and even the *idea of the circle* is gone. What we know for sure is that this small book marks the end of my doctoral studies, which I have enjoyed, for many reasons.

The book/thesis describes different ways to use *nuclear spin relaxation* measurements to get information about protein function and structure. It also describes the method *isothermal titration calorimetry* which gives you thermodynamic information from chemical reactions. These two methods have been used with the *specific* aim to characterize two (already quite well-studied) proteins, galectin-3 and the B1 domain of bacterial protein G. A *general aim* of my research and thesis can be described as – contribute with knowledge about biophysical processes in proteins. The following 60+ pages serve to introduce the research, explain the underlying theory, and provide cross-reference to the research papers which are *the actual* research contribution.

The supposed reader is a doctoral student in the same or a similar field, or anyone with interests in the topics. I should not deny that the outline is based on my interests, but the goal has been to put the research into a broader context. The reader judge if this succeeded. Three chapters are presented.

Thermodynamics is of the greatest relevance for a physical chemist. Some consider thermodynamics a supreme ideology in science, one of few that will exist forever: “the doctrine of energy and entropy” (Müller, 2007). It is acknowledged that a thorough understanding and description of biomolecular processes is not possible without involving thermodynamic concepts, such as free energy landscapes.

Nuclear Magnetic Resonance is the main experimental method during my doctoral studies. A very fascinating and versatile human invention discovered about 70 years ago which likely will persist as an analytical tool in chemistry the next 70 years.

Proteins are the models I have studied; without them I would have looked into a solution of water and salt. Structural biology is the research field that examines proteins (and other biological macromolecules) on atomic resolution and it contains some of the most exciting, as well as most complex stories in science. The fantasy of Nature seems to be endless. Molecular recognition, used in the thesis title, is a subdiscipline in structural biology and it is related to protein communication. Communication is always difficult.

These chapters can all be covered by the umbrella called *biophysical chemistry*, a subject that seems to cover so much of the collected human knowledge.

If you continue from here, I hope you find something interesting. If the topic does not seem to be fun, maybe you can give the back cover a few seconds? What do you see?

Thermodynamics

The terminology and concepts of thermodynamics are outlined in the context of biochemistry and the goal is to present it in a simple shape, based on how it has been taught the last 120 years. The thermodynamics related to protein chemistry is highlighted, e.g. protein–ligand interaction. Attention is given to the concept of entropy, due to its close connection to dynamics.

Two thermodynamic relationships particularly important for the thesis are the following: i) $\Delta G° = RT \ln K_d$, where $\Delta G°$ is the standard state Gibbs free energy of binding, K_d the equilibrium dissociation constant and R the gas constant (8.314 $\mathrm{Jmol^{-1}K^{-1}}$). ii) $\Delta S = R \sum_k ln\left[\left(1 - S^2_{\mathrm{A},k}\right)/\left(1 - S^2_{\mathrm{B},k}\right)\right]$ explain how the squared NMR order parameters S^2, for N number of residues can be treated as messengers of the conformational entropy difference (ΔS) between two protein states A and B. Isothermal titration calorimetry (ITC), a technique which directly gives thermodynamic information, is presented in this section and it makes use of equation i).

The main references for this chapter are physical chemistry textbooks (Chandler, 1987; Atkins and De Paula, 2006, 2010) and course material (Halle, 2016).

Internal energy and enthalpy

The internal energy E, of a *system* is the total kinetic and potential energy stored in all molecules in the system. Everything outside the system is called the *surrounding*. The system and the surrounding are sometimes referred to as the *total system*. E changes either by transfer of heat q, or work w:

$$\Delta E = w + q \qquad (1)$$

and the system cannot recognize any difference between heat and work. Heat is defined as positive if added to the system and work is positive if it is done on the system, i.e. opposite to the displacement in case of expansion work. The internal energy is a *state function* meaning that a change in its value is independent of the path between the two states. The infinitesimal change of a state function is an *exact differential*[2]. The equation can also be written $\Delta E = \delta w + \delta q$ where δ indicates inexact differentials. "For

[2] The line integral of an exact differential depends only on the endpoints of the path, but the line integral of an inexact differential depends on the path.

an *isolated system* the internal energy is a constant", this is one version of the first law of thermodynamics.[3]

Assuming the system volume to be constant (an isochoric system), we have $\Delta E = q_V$ where subscript V means constant volume. *Heat capacity*, $C = q/\Delta T$, where ΔT is change in temperature, is then defined as:

$$C_V = \frac{\Delta E}{\Delta T} \tag{2}$$

This is the isochoric heat capacity. Since most chemistry and especially the experiments related to this thesis, are assumed being done under constant pressure rather than constant volume, there is a need for reshaping of the concepts, and the *enthalpy*[4] H is defined:

$$H = E + pV \tag{3}$$

where p is pressure. Studying changes in enthalpy, we get:

$$\Delta H = \Delta E + p\Delta V \quad (\text{const.}\, p) \tag{4}$$

The enthalpy is derived from three state functions E, p and V; hence enthalpy is also a state function, and at constant pressure and when no work is done, we have:

$$\Delta H = q_p \tag{5}$$

If $\Delta H > 0$ we call the reaction *endothermic*, and the reverse $\Delta H < 0$ is called an *exothermic* reaction. ITC-experiments referred to in papers III and IV show exothermic reactions. The isobaric heat capacity is defined as:

$$C_p = \frac{\Delta H}{\Delta T} \tag{6}$$

Since most processes in biochemistry occur in liquids, the volume changes are very small, and we can assume $\Delta H \approx \Delta E$.

Strength of chemical bonds and protein folding

A chemical bond can be characterized by measuring its bond enthalpies, which can be thought of as the strength of the bond. Enthalpy changes related to strong covalent bonds (~ 300–1000 kJ/mol) rarely occur in biochemistry but protein–ligand interactions and the molecular recognition processes involve breaking and making of non-covalent bonds. The energy range for non-covalent bonds range from J/mol for very weak van der Waals interactions, up to ~200 kJ/mol for strong hydrogen bonds (van Holde, Johnson and Ho, 2006; Müller, Faeh and Diederich, 2007; *CRC Handbook of Chemistry and Physics*, 2018). The interaction strength can be compared with the proposed difference between a native and denatured globular protein under physiological conditions, also called the *conformational stability* (relates to ΔG of the folding process, ΔG will soon be introduced)

[3] The first formulation of equivalence of heat and work is ascribed to Benjamin Thompson (Count Rumford) in 1798 (Glasstone, 1947).

[4] The word enthalpy means 'heat inside' in Greek, an old term is *Gibbs heat function.*

which is 21–63 kJ/mol (Pace, 1990). This must be considered a small value in comparison with the total potential energy of the thousands of covalent and non-covalent bonds in proteins. This delicate balance between folded and unfolded state is a fascinating consequence of the evolution.

As an example, the B1 domain of protein G (PGB1), studied in papers I and II, has a quite low conformational stability but considered being a stable protein due to its considerably high melting temperature (see coming section 'Protein GB1'). Reasonable is to believe that the native state represent a local energy minima ("kinetic protein folding hypothesis") and not a global minima ("thermodynamic folding hypothesis" (Wetlaufer, 1973; Schulz and Schirmer, 1979). Many folding processes are also found to involve intermediate states (Löw *et al.*, 2008). Protein stability is also governed by solvent interactions, usually referred to as hydration-effect.

The ability to fine-tune biological processes, so that specific, but transient binding can be accomplished, has been important for survival of the organism. Conserved protein structures have not been an advantage. Protein evolution is indeed very interesting but not the main topic of this thesis.

Entropy

The concept of entropy is complicated, but still one should be able to explain it[5], especially since it is mentioned about 130 times in paper III. One short description could be: "Matter and energy tend to disperse". Thermodynamic concepts can be seen from macroscopic as well as molecular perspective and the macroscopic interpretation of entropy can be traced to Rudolf Clausius who in 1865 explained that entropy maximization is the ultimate goal of the physical universe (Leiter and Leiter, 2003). This continuous increase in total entropy is the driving force behind spontaneous changes.

"The more disorder[6] the higher entropy", is often used to explain entropy, making it logical that entropy for a pure substance increases in order: solid → liquid → gas. Considering that there is an increase in 'degrees of freedom' accompanied with an increase in entropy may aid understanding. Entropic penalty in the context of protein–ligand binding refers to a loss in mobility for either ligand, protein or solvent, which reduces the affinity.

The entropy concept is even used outside of thermodynamics. In information theory it is 'missing information' or 'uncertainty'. These thoughts were developed by Claude E. Shannon ("Mathematical theory of communication", 1948) and Edwin T. Jaynes. He preferred to see statistical thermodynamics as an inferential method and developed the

[5] "…one of the most subtle concepts of theoretical physics, or natural philosophy…" (Müller, 2007); "Entropy is one of those scientific technical terms, like *relativity, the uncertainty principle*, *evolution*, more recently *chaos*, that have from time to time captured the literary imagination." (Taylor, 2001)

[6] Arieh Ben-Naim has long argued for abandoning the description of entropy as degree of disorder (Ben-Naim, 2011). Atkins to some extent supports this (Atkins and De Paula, 2006), p 81.

maximum entropy principle for this purpose. "The expression of entropy has a deeper meaning, quite independent of thermodynamics" (Jaynes, 1957).

Change of entropy

Along any reversible path there exists an integrating factor T, leading to that dS becomes an exact differential:

$$\mathrm{d}S = \frac{\delta q_{rev}}{T} \tag{7}$$

Remember δq_{rev} refers to an inexact differential. By integration along a path we get that the change in entropy of the system, going from state A to B is $\Delta S = \int_A^B \delta q_{rev}/T$, where T refers to temperature of the surrounding. The subscript rev means the reaction is *reversible*, which means there is chemical equilibrium during the reaction. For other processes, so-called *irreversible*, we have $\Delta S > \int_A^B \delta q/T$.

There are usually two kinds of entropy changes, i) when temperature is changing, ii) phase transitions. The first is described $dq_{rev} = C \cdot dT$, where C can be either the isochoric (2) or isobaric (6) heat capacity (and we assume C is independent of temperature) and by this (7) can be rewritten as:

$$\Delta S = \int_A^B \frac{C dT}{T} = C \ln \frac{T_B}{T_A} \tag{8}$$

A higher final temperature (T_B) results in a positive ΔS. These entropy changes are of interest in biochemistry but are not studied specifically in this thesis.

We focus on ii), ΔS upon phase transitions. Phase transitions can be used to explain protein folding, since the folding process involves a disorder ↔ order transition (Schulz and Schirmer, 1979). Melting of ice is a well-known phase transition. Other examples are the ferromagnetic Ising spin systems (Chandler, 1987), droplet and micelle formation. Symmetry arguments can describe phase transitions; symmetry is a discontinuous variable and can only be changed in stepwise fashion.

How do we calculate entropies for (phase) transitions? We start with (7) and look at changes in the surrounding:

$$\Delta S_{sur} = \frac{q_{sur}}{T} \tag{9}$$

The heat here is assumed to arise solely form changes inside the system, i.e. the system loses heat q, and we set $-q = q_{sur}$, we get $\Delta S_{sur} = -q/T$. The surrounding has constant pressure, and we can make the identification $q = \Delta H$ via (6). We then write:

$$\Delta S_{sur} = \frac{-\Delta H}{T} \tag{10}$$

H, U and S without subscripts always refer to the system. Appropriate positioning of the system borders is not trivial.

We are now able to comment on the distinction between work and heat and how they are appreciated by the surrounding. Work necessarily leads to a translational, ordered motion of molecules, while transfer of heat stimulates random motions. The intimate relationship to randomness is one reason entropy often seems abstract.

Gibbs free energy and chemical equilibrium

It is problematic for the chemist that entropy of both the system and the surrounding are changing, since it is difficult to control events in the surrounding. By using (10) and the fact the total change of entropy (which we know has to increase) is the sum of the entropy changes in the system and in the surrounding:

$$\Delta S_{tot} = \Delta S + \Delta S_{sur} = \Delta S - \frac{\Delta H}{T} > 0 \qquad (11)$$

at constant T and p. By defining a new state function, *Gibbs free energy*, $G = H - TS$, (henceforth we shorten it Gibbs energy). We multiply (11) with $-T$ to get:

$$\begin{aligned} &\Delta H - T\Delta S < 0 \\ &\Leftrightarrow \\ &\Delta G < 0 \end{aligned} \qquad (12)$$

In (12) ΔH is a measure of the average potential energy between molecules, and $T\Delta S$ is a measure of the correlation between molecules.

From (11) and (12) we also get $\Delta G = -T\Delta S_{tot}$. So, the requirement for a spontaneous process to have an increase in total entropy (always true), can be restated as, there must be a decrease in Gibbs energy (under constant T and p). An advantage with (12) is that the system only, without involving the surrounding, provides all information to determine if the process is spontaneous. Of importance when studying protein interaction is to remember that changes in *total* entropy drive the reaction.

The reaction Gibbs free energy and the equilibrium constant

The Gibbs energy is extremely useful due to the connection with the equilibrium constant of a reaction. To explain this we invent *the reaction Gibbs energy* $\Delta_r G$, used for studying changes of the molar Gibbs energy, i.e. the change of *chemical potential* $\mu_j = \left(\partial G/\partial n_j\right)_{p,T,n_{i\neq j}}$ during the course of a reaction. Subscript refers to species j, the *reaction coordinate* ξ (units of mole) refers to extent (0–100%) of the reaction. Differential changes in the reaction, with stochiometric numbers ν_j is studied. ν_j is negative for reactants and positive for products and we have $dn_j = \nu_j d\xi$. Using this we write the differential form of Gibbs energy as:

$$dG = VdP - SdT + \sum_j \mu_j dn_j = \left(\sum_j \nu_j \mu_j\right) d\xi \tag{13}$$

for constant T and p. We define the reaction Gibbs energy as

$$\Delta_r G = \left(\frac{\partial G}{\partial \xi}\right)_{p,T} = \sum_j \nu_j \mu_j \tag{14}$$

Next, we use another (*sic*) definition of the chemical potential for an ideal solution:

$$\mu_j = \mu_j^o + RT \ln a_j \tag{15}$$

Where μ_j^o is the *standard chemical potential*, the chemical potential of pure substances j at 1M and a_j is the activity[7], which in solutions are concentration of j divided by 1M to remove units in the natural logarithm. We substitute this into (14):

$$\begin{aligned} \Delta_r G &= \sum_j \nu_j \mu_j^o + RT \sum_j \ln\left(a_j^{\nu_j}\right) \\ &= \Delta_r G° + RT \ln \prod_j a_j^{\nu_j} \end{aligned} \tag{16}$$

The product $\prod_j a_j^{\nu_j}$ is commonly known as the *reaction quotient* Q.

$$\Delta_r G = \Delta_r G° + RT \ln Q \tag{17}$$

At chemical equilibrium the change of Gibbs energy during the reaction should be zero, since the reaction does not go. The left side of (17) will become zero, and we substitute Q for the *equilibrium constant* K:

$$\Delta_r G° = -RT \ln K \tag{18}$$

EXAMPLE: Study of mixing to better understand entropy

Reactions for which $\Delta_r G° < 0$ are called *endergonic* and (18) is of great interest for papers III and IV. We analyse this in more detail, by using an example from a review covering protein–ligand interactions (Homans, 2007). In the review, the case of mixing of two substances A and B is studied: $A \overset{K_{mix}}{\Longleftrightarrow} B$, with $K_{mix} = [A]/[B]$. The Gibbs energy for the mixing reaction:

$$\Delta_r G = RT\{x_A \ln x_A + x_B \ln x_B\} \tag{19}$$

x_A and x_B are mole fraction of A and B, we have $x_A = 1 - x_B$. The free energy of the system will contain another part, $x_B \Delta G°$ defining the mole fraction of the pure parts of A and B. We sum these and get:

[7]With γ_j as the activity coefficient and x_j the molar ratio we have the activity $a_j = \gamma_j x_j$.
Also $a_j \to x_j$ and $\gamma_j \to 1$ as $x_j \to 0$

$$\Delta_r G = x_B \Delta_r G° + RT\{(1 - x_B)\ln(1 - x_B) + x_B \ln x_B\} \qquad (20)$$

$\Delta_r G$ is shown in Figure 1 as a blue line. The non-mixing and mixing contributions are indicated. Using derivatives of (20), the equilibrium point is given as:

$$\begin{aligned} 0 &= \frac{d\Delta_r G}{dx_B} = \Delta_r G° + RT\{-\ln(1 - x_B) + \ln x_B\} \\ &= \Delta_r G° + RT \ln \frac{x_B}{1 - x_B} \end{aligned} \qquad (21)$$

The logarithm contains K_{mix}, hence we have (18). Combining (12) and (18) gives the relationship used in all of the ITC-experiments:

$$\Delta_r H° - T\Delta_r S° = -RT \ln K_a \qquad (22)$$

Often the superscripts ° are dropped for convenience, as well as the subscript r after Δ.

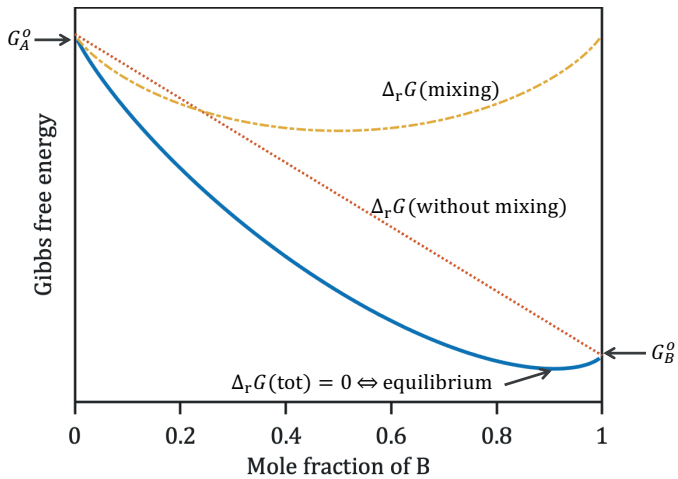


Figure 1. Entropy of mixing and chemical equilibrium. Blue line shows the change of Gibbs energy over the reaction for the ideal solution of two mixing substances A and B. "Mole fraction of B" on x-axis could as well be labelled "ξ" for reaction coordinate. The dotted red line shows if mixing would not be considered, the yellow dash-dotted line shows the mixing. The equilibrium point is given by the equilibrium constant *K* for the reaction.

For protein–ligand interaction, the association constant K_a is typically substituted for the dissociation constant $K_d = 1/K_a$, adjusting the equation for determining equilibrium constants with ITC to: $\Delta_r H° - T\Delta_r S° = RT \ln K_d$.

Protein–ligand interactions

The thermodynamics will be applied on a simple model – a protein (P) with a single binding site interacts with a ligand (L) to form a complexed protein (PL):

$$P + L \rightleftharpoons PL$$

The ligand can be anything, except the solvent, that binds to the protein, an ion, a small molecule, a hydrogen atom, or a protein. The terminology implies that the receptor molecule is larger than the ligand molecule. Chemical equilibrium is described with:

$$\Delta G° = RT \ln K_d\,; \quad K_d = \frac{[L][P]}{[PL]} \tag{23}$$

The lower the K_d-concentration, the higher the affinity. More insights into the dissociation constant is given by making use of the concept of saturation, defined (for the one-binding site) as $0 < \theta = [PL]/([P] + [PL]) < 1$. θ can be rearranged:

$$\theta = \frac{[L]}{K_d + [L]} \tag{24}$$

$[L]$ refers to free ligand concentraion, and the total ligand concentration is $[L]_{tot} = [L] + [PL]$. Using these concepts, we plot the respective concentrations. θ = 0.5 means $[L] = K_d$ (Figure 2).

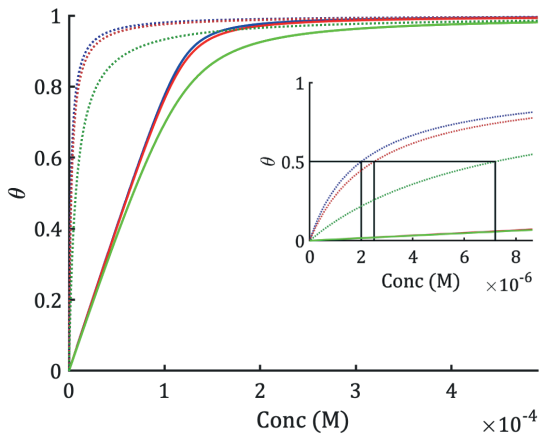


Figure 2 Calculation of saturation curves for a single-site binding equilibrium: $P + L \rightleftharpoons PL$. θ = 0.5 (50% saturation). The zoom-in picture is identical to a free ligand concentration [L] = K_d. The ligands are given the same K_d (and colours) as for the three Gal3C complexes in paper III. Dotted line is [L] and full line is $[L]_{tot}$.

Huge efforts are made within the scientific community to understand protein–ligand interactions and their decomposition into enthalpic and entropic components (Perozzo, Folkers and Scapozza, 2004; Fox *et al.*, 2018). Typically to achieve this is to make a series of homologous ligands (as in papers III and IV) – with small variation between them – or make small adjustment on the proteins by mutations (not done in the thesis) and study $\Delta\Delta G°$, $\Delta\Delta H°$ and $T\Delta\Delta S°$. In ligand-binding the debated concept of enthalpy-entropy compensation refers to $\Delta\Delta H° \approx T\Delta\Delta S°$ resulting in no net change in binding affinity (Chodera and Mobley, 2013). Indication of this is found in papers III and IV.

A structure-thermodynamic database, SCORPIO (Olsson *et al.*, 2008), built from hundreds of ITC-experiments on protein–ligand interaction reveals that the $\Delta H°$ provides the largest favourable contribution to binding. However, the database also reveals that *apolar surface area burial*, seen as an entropic effect, is the feature most correlated with $\Delta G°$.

Partitioning of the entropic contributions to binding

We dissect the binding free energy (22) as (Wand and Sharp, 2018):

$$\begin{aligned} &\Delta G = \Delta H - T\Delta S \\ &= \Delta H - T\{\Delta S_{\mathrm{conf,P}} + \Delta S_{\mathrm{conf,L}} + \Delta S_{\mathrm{sol}} + \Delta S_{\mathrm{r-t}} + \Delta S_{\mathrm{other}}\} \end{aligned} \tag{25}$$

The ΔH consists of interactions described in section ‘

Electrostatic interactions’ and is considered straightforward to investigate, though not saying it is a simple task. Paper V gives a detailed description of hydrogen bond patterns for Gal3C–ligand interactions.

The separation of the binding entropy is generally seen as a greater challenge, and the partitioning makes clear that the NMR-experiments (papers II and III) can only measure a (probably quite small) fraction of ΔS. NMR order parameters determine $\Delta S_{\mathrm{conf,P}}$, the conformational entropy changes upon binding for the protein. $\Delta S_{\mathrm{conf,L}}$ is the corresponding term for the ligand, and is not determined by NMR in this thesis but by collaborators performing molecular dynamics (MD) simulations (Ignjatovic, 2019). ΔS_{sol} relates to changes of degrees of freedom for the water molecules attached to L and P before binding. *Desolvation* is the process whereby water molecules bound to the proteins are released. This is studied by collaborators performing MD simulations (paper III) and neutron crystallography (paper V). $\Delta S_{\mathrm{r-t}}$ relates to rotational and translation degrees of freedom for L and P before binding, and finally $\Delta S_{\mathrm{other}}$ is unknown contribution. The partitioning of entropy as in (25) is one example of many possible ways to partition. (Fox *et al.*, 2018) presents another example.

Complex binding and cooperativity

Nature shows more complex examples than single site binding. Multiple binding sites are common (Ma *et al.*, 2002; Bellelli and Carey, 2017) and there can be *allosteric effects*, i.e. transfer of signal to remote sites on the protein upon contact with ligand. *Cooperativity* refers to communication between different binding sites (for the same ligand) at the protein. Positive cooperativity means binding at one site increases the affinity at another, negative means that binding at one site decreases affinity at another.

The first studies of cooperativity are attributed to A. V. Hill[8] who in 1913 analysed how gases of O_2 and CO bind to the protein haemoglobin. Hill could not explain binding curves

[8] The Englishman Archibald Vivian Hill (1886–1977) received a Nobel Prize in physiology, "for his discovery relating to the production of heat in the muscle". During the studies prior to the prize he used an instrument, “the Blix’s apparatus for thermosensory stimulation” developed

due to the effect of salt and CO_2 (Hill, 1913) and included an exponent n_H (later named *Hill-coefficient*) to (24):

$$\theta = \frac{[L]^{n_H}}{K_d + [L]^{n_H}} \tag{26}$$

If $n_H > 1$ we call it positive binding cooperativity, and $n_H < 1$ refers to negative binding cooperativity. Hill's conclusion reveals his ambiguity: "It is shown here that the theory [using n_H] is capable of including all the known facts in relation to CO- and O_2-dissociation curves" (Hill, 1913). n_H is an *ad hoc* assistant, used since it works and this is how it is applied in paper I.

pH-titration of acidic side chains

An example of protein–ligand interaction involving a proton as a ligand, is the acid-base reaction for titratable side chains. In paper I, glutamic acid (Glu) and aspartic acid (Asp) are titrated to study this equilibrium. In Paper V we titrate histidine (His). The titrations involve pH-adjustment while at the same time record the NMR chemical shifts.

Knowledge of protonation/deprotonation processes is essential for understanding protein function and theoretical chemists depend critically on knowledge of the overall protein charge for their simulations. For example, a theoretical investigation uses the information about protonation rates to understand protein folding rates processes. The rate of folding depends on the protonation state of some residues, and the proton-transfer is then a rate-limiting factor (Mizukami, Sakuma and Maki, 2016).

The acid-base equilibrium is written:

$$\mathrm{HA} \overset{K_a}{\Leftrightarrow} \mathrm{H^+ + A^-} \tag{27}$$

$$K_a = \frac{a_{H^+} \cdot a_{A^-}}{a_{HA}} = \frac{[\mathrm{H^+}][\mathrm{A^-}]}{[\mathrm{HA}]} \tag{28}$$

Second equal sign assumes ideal dilute solutions. For an acidic, negatively charged amino acid, we redefine the saturation explained by (26) as the fraction of charge by using the definitions $\mathrm{p}K_a = -\log_{10} K_a$ and $\mathrm{pH} = -\log_{10}[H^+]$:

$$\theta_{A^-} = \frac{1}{1 + 10^{n_H(pK_a - pH)}} \tag{29}$$

Figure 3 plots $-\theta_{A^-}$ vs pH with varying n_H (the ratio is identical to the charge).

by prof. Magnus Blix at Lund University (Norrsell, 2000). The instrument measured heat from contracting muscles. Blix also conducted studies on why humans cannot fly.

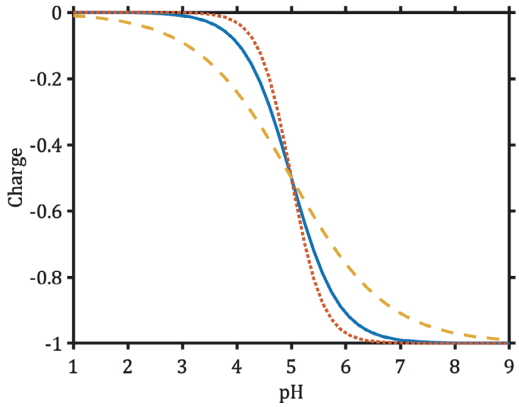


Figure 3. Charge of an amino acid with pK_a=5 as a function of pH. Blue line represents a non-cooperative case (n_H = 1), red dotted line is positive cooperativity (n_H > 1) and yellow dashed line is negative cooperativity (n_H < 1).

Equation (29) is a modified Henderson-Hasselbalch equation and a physicochemical interpretation of the Hill-coefficient is as a measure of the reduction of the charge–charge interaction due to salt screening (Lindman *et al.*, 2006).

NMR is the most suitable method for determination of pK_a-values in a solution. However it is important that the nuclear spins are close to the titrating site (Nielsen, 2008). This recommendation was followed in paper I, but was more complicated to apply in titration studies of histidine in paper V. Since protonation events occur in the so-called *fast chemical exchange regime* (explained in the NMR-chapter) the observed chemical shift δ_{obs} is the population averaged shift:

$$\delta_{obs} = \delta_{A^-}\theta_{A^-} + \delta_{HA}(1 - \theta_{A^-}) \qquad (30)$$

Titration of Histidine side chains

Histidine is a sophisticated amino acid and plays a critical role in the function of numerous proteins, much due to the pK_a-values of the imidazole ring, which are close to physiological pH. Because of this, histidine can be both an acid and a base at physiological conditions, making it an ideal switch or gatekeeper (Pelton *et al.*, 1993). Histidine is assumed to be involved in ~50% of all protein active sites (Hansen and Kay, 2014). Also, the ^{1}H chemical shift of histidine proton C2 made it suitable for NMR titration studies since it is far apart (> 8 ppm) from the other, hence well resolved[9]. (Bradbury and Scheraga, 1966; Markley, 1975). Paper V study the histidines in Gal3C to determine the protonation state.

[9] Could it be that Nature wants to give the scientists "easy access" to certain interesting objects?

Non-covalent interactions in proteins

Molecular recognition is typically a non-covalent event where the interacting molecules form a complex without creating covalent bond. Chemical bonding in proteins are ultimately explained by quantum mechanical calculations, which provide the theoretical chemist with a complete thermodynamic state of the protein. The experimental methods used in the thesis – NMR and ITC – usually do not measure and quantify individual bond energies but give qualitative measure of aggregate values of bond energies.

As mentioned in section 'Strength of chemical bonds and protein folding', chemists still struggle to understand the specific contributions from weak interactions to protein stability and protein folding. Ferreira de Freitas and Schapira investigated 11000 PDB-structures of ligand–proteins complexes, representing ca 750000 protein-ligand atomic interactions, in search for common structural pattern (Ferreira de Freitas and Schapira, 2017). Their aim was to improve understanding of how to optimize drug design. They found seven common interactions in their data: i) hydrophobic contacts (loosely defined as apolar···apolar interactions), ii) hydrogen bonds, iii) π-stacking, iv) weak long-range hydrogen bonds, v) salt bridge, vi) amide π-stacking between the backbone amides and aromatic rings of ligands, and vii) cation–π interactions. We only describe some of these in more detail. A comprehensive description of the non-covalent interactions is given in the main resources used for this section (van Holde, Johnson and Ho, 2006; Halle, 2016).

Electrostatic interactions

The electrostatic interactions can be either intra- or intermolecular, they depend on distance and on the polarizability, expressed as a *permittivity* ϵ, which is a measure of the degree to which the material is permeated by the electric field in which it is immersed. *Coulomb's law* gives the electric potential between to charges Z and z. $E_{\mathrm{el}} = Zze^2/4\pi\epsilon r$. The permittivity is expressed in terms of multiples $\epsilon = \epsilon_r\epsilon_0$, where $\epsilon_0 = 8.854 \times 10^{-12}C^2J^{-1}m^{-1}$ and is called *vacuum permittivity*, and ϵ_r is the *relative permittivity* (preferred name before the formerly used *dielectric constant*). r is the separation of the atomic centres.

Dipole-dipole interactions depends on slight separation of the charge distributions, giving rise to regions of $\delta+$ and $\delta-$ in the atom, and we call this a dipole moment. We describe it: $E_{\mathrm{d-d}} = f(\boldsymbol{\mu_1}, \boldsymbol{\mu_2})/r^3$, where $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ are the two interacting dipole moments and r is the separation between them. The most important dipole–dipole interaction in biology is the *hydrogen bond,* involving H and an electronegative atom, primarily O and N. Because of its importance it is considered being an interaction type of its own.

The *van der Waals interaction*[10] is an induced dipole–dipole interaction that occur between atoms without a permanent dipole moment due to fluctuating electric dipoles. The attraction – also called dispersion attraction or *London dispersion force* – appears because atom-1 causes tiny imbalance in charge distribution of atom-2, and is described

[10] Johannes Diderik van der Waals (1837-1923). Dutch scientist who received the Nobel Prize in Physics in 1910 for his equations of states for gases and liquids.

$E_{\mathrm{vdW,attr.}} = -A/r^6$, where A is derived from quantum mechanical variables, r is the separation between the centres of the atoms. Adding a repulsive part – a *hard sphere approximation* – finally leads to a combined van der Waals interaction potential: $E_{\mathrm{vdW}} = -A/r^6 + B/r^{12}$, where power of 12 is an empirical estimate. The minimum gives the *van der Waals radius* r_{vdW}, or the distance between atoms. In biological matter r_{vdW} is estimated to be 1 Å for aromatic H; 1.2 Å for aliphatic H; 1.5, 1.6 and 1.7 Å for O, N and C respectively. The van der Waals radii of bound halogen atoms to a primary alkyl are 1.4, 1.7, 1.8 and 2.0 Å for F, Cl, Br and I respectively (Bondi, 1964), of the *Halogen bond* described in Paper IV. The van der Waals interactions are omnipresent in protein solutions and exist even between hydrophobic molecules and water (Chandler, 2005).

Hydrogen bonds

As mentioned, the hydrogen bond (HB) is an electrostatic interaction but given a section of its own due to its importance. Already 70 years ago it became evident it was a main stabilizer in the secondary structures of helices and sheets in proteins (Pauling, Corey and Branson, 1951). No person is attributed to the discovery of the mechanism.

How come it is so important? A general answer would be that hydrogen is by far the most abundant element in the universe. Also, water is built from two hydrogens and one oxygen and in principle all biology is related to the water medium[11].

A HB is an interaction between a polarized *donor* (D–H) and a polarized non-bonding orbital in an *acceptor* A. When the difference in electronegativity between D and H is so large that H remains unshielded the HB can form (Jeffrey, 1997). The proteins are often carrying the donor atom in protein–ligand interactions. The N–H···O is the most common HB in proteins and the most common acceptors is glycine. The most common donors are glycine and aspartic acid (Ferreira de Freitas and Schapira, 2017). The β-sheets and α-helices are formed by intramolecular N–H···O=C interactions in the backbone between the amide hydrogens and carbonyl oxygens by nearby residues, this formation increases the π-bond character of the peptide C–N bond. “All potential HB donors and acceptors in proteins are satisfied a significant fraction of the time, either by intramolecular HBs or by HBs to solvent water.” (Fleming and Rose, 2005).

There are *intra-residue* HBs between the amide protons and carbonyl oxygens of the same residue (Newberry and Raines, 2016). HBs in the side chains govern biological functions by gating mechanism and molecular recognition (Fersht, 1987; Bondar and White, 2012).

Paper I shows how HBs are involved in determination of pK_a of acidic side chains, and paper V is a systematic analysis of HB-network in galectin-complexes using neutron crystallography, which is a method especially suitable for studies of HBs since the H is localised directly and not inferred by other atoms as in X-ray crystallography.

[11] One should not be too afraid of sometimes recapitulating facts (almost) every person knows.

Hydrophobic interaction and hydration

The HB partly explains the fascinating hydrophobic interaction observed when oily (non-polar) molecules are mixed into water. Walter Kauzmann first realized its significance in the 1950s when he hypothesized: "The hydrophobic bond is probably one of the more important factors involved in stabilizing the folded configuration in many native proteins" (Kauzmann, 1959). 'Interaction' or 'effect' is preferred before Kauzmann's 'bond'.

Theoretical studies clarified the mechanism and made a distinction between adding small and large non-polar solutes to water (Chandler, 2005). With small solutes, the water cavity of radius R does not disrupt the HB network of water molecules – the hydration layer – to large extent and as a consequence a cavity volume dependence of the Gibbs energy of solvation (ΔG_{sol}) appears, and we have $\Delta G_{\mathrm{sol}} \propto R^3$. Small refers to $R \lesssim 10$ Å. On the contrary, for large solutes or large cavity creation, the HB network necessarily is destroyed and an interface between the non-polar molecules and water molecules appears with the consequence of a solvated surface dependence of Gibbs energy of solvation: $\Delta G_{\mathrm{sol}} \propto R^2$. The point, i.e. radius R, where the R^3-dependence $\rightarrow$ R^2-dependence is where the ΔG_{sol} for the complex is lower than ΔG_{sol} for the individual components (Homans, 2007). This is the driving force of the hydrophobic effect.

In the context of protein–ligand interactions, a frequent hydrophobic interaction is when an aliphatic carbon in the protein and an aromatic carbon in the ligand connect to minimize the water accessible surface. This we find in Paper III where an alanine connects to the aromatic ring of the ligand.

Halogen bonds

A recently discovered non-covalent interactions is the halogen bond (Auffinger *et al.*, 2004), but an "ancestor" was discovered by Hassel already 50 years ago (Hassel, 1970).

Halogen bond, as it is used here, refers to a non-covalent directional C–X···Y interaction with X=Cl, Br, I and Y=O, N, S. The three heavier halogens adapt the $s^2p_x^2p_y^2p_z^1$ configuration, where the z-axis is along the C–X bond. The three unshared electron pairs cluster around the X, but the outermost region becomes positive when one of the X valence electrons is pulled into the σ-bond. This anisotropic distribution of electron density around X results in a "polar flattening" or a σ-*hole* (Clark *et al.*, 2007). The interaction is linear in the direction of the C–X bond. The bonding nucleophile is an electron-rich O, N or S, most often an oxygen atom (Paulini, Müller and Diederich, 2005; Scholfield *et al.*, 2017). F cannot form halogen bonds due to its higher electronegativity and lack of polarizability.

Paper IV is a combined ITC, NMR, X-ray, and computational study of a series of ligands forming halogen bonds.

Thermodynamic fingerprints of protein–ligand interactions using ITC

Calorimetry is a very old technique for measuring heat in a scientific manner and has been used for more than 200 years.[12] Modern microcalorimeters can titrate volumes of ~1–5 μL into cell volumes ~200 μL to monitor the interactions in the biomolecules.

The principles of ITC

Thermal analysis refers to measurements of time dependence of the sample temperature when the sample is perturbed (Hemminger and Sarge, 1998). The instrument used in this thesis, a MicroCal PEAQ-ITC (Malvern Panalytics), perform thermal analysis to achieve calorimetric data. Measurement is performed indirectly in the sense heat is picked up by a Peltier device and converted to an electric signal. A precise calibration of this conversion is vital to get the thermodynamic data correct.

The instrument measures the power ($\mu J/s$) required to maintain a constant temperature difference between the sample cell and the reference cell. The "injection peak area" is transformed to enthalpic heat for an assumed reversible reaction. An exothermic reaction results in negative peaks (less heat is needed for the reference cell), and an endothermic reaction results in positive peaks. The fitted parameters[13] are the ΔG converted to K_d-values and ΔH. From these ΔS is derived according to (22). The ability to decompose the binding into the three thermodynamic state functions is a unique feature. The limits for direct determination of affinities are between 10^{-2} to 10^{-9} M affinities. Complexes in paper III have $K_d \sim 10^{-6}$ and in paper IV $K_d \sim 10^{-7}$–10^{-6}.

Some prefer to explicitly distinguish the thermodynamic state functions given in ITC measurement from the absolute, by tagging with 'observed' as ΔG_{obs} (Ladbury, Klebe and Freire, 2010). We do not make this distinction.

Reproducibility and ITC

As always it is of important that data can be reproduced. An interlaboratory study revealed relative standard deviations of 20%, for the three thermodynamic parameters. This exceeds the individual precision estimates by a factor of 10 (Myszka *et al.*, 2003). The study suggested that concentration errors contribute most to the systematic error. Re-analysis of the data concludes that concentration errors in an ITC-experiment typically

[12] Again we credit the spirit in revolutionary France for its contribution to mankind since Antoine Lavoisier and Pierre Simon Laplace built the first calorimeter around 1780. Using it they investigated the distinctions between heat and temperature proposed by the Scot Joseph Black. Another hundred years later *another* Frenchman, Pierre-Eugene Marcelin Berthelot developed the first calorimeter which resemble modern instruments (Moncrief, 2000; Knight, 2002).

[13] Usually an ITC-measurement involves a series of ~15 injections separated by ~4 minutes.

are 10%, exceeding expected levels[14] and this is not revealed by the least-square analysis (Tellinghuisen and Chodera, 2011). Own estimations of concentration errors are ~4 % for the protein and ~1% for the ligand (paper III).

Data fitting has been improved recently and there exists software that is able to attribute individual errors to the injection peak, using noise-information in the inter-injection period (Keller *et al.*, 2012; Zhao, Piszczek and Schuck, 2015; Brautigam *et al.*, 2016).

ITC in drug discovery

Due to the unique features of the ITC-method, it has been used much in drug design the last decade. All kinds of protein interactions can be studied: protein–small molecule, enzyme–inhibitor, protein–protein, protein–DNA, protein–carbohydrate are some examples. Ten years ago ITC was suggested being a "hot tip" in drug discovery (Ladbury, Klebe and Freire, 2010) and "of enormous value in drug design" (Chaires, 2008). Recent publications emphasize that the output of global enthalpy and entropy parameters are crude measures at the screening stage, and the community seems to move towards a use of more theoretical methods, of course along with the established experimental methods (Geschwindner, Ulander and Johansson, 2015). Figure 4 illustrate the development of ITC in drug design and there seems to be a decline in interest the last four years.

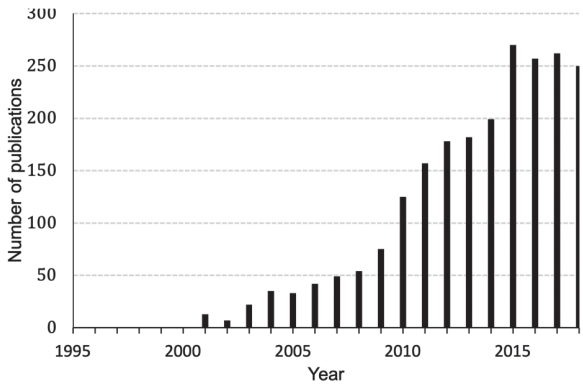


Figure 4 Publications per year. Using search phrase "isothermal titration calorimetry drug design" on web of science database (Clarivate Analytics).

Statistical thermodynamics

To understand one of the core concepts in this thesis – the NMR order parameter S^2 – it is of great value with some background in statistical thermodynamics, a theory based on probabilistic argument. Statistical thermodynamics connects the microscopic world of atoms and the macroscopic world as our senses perceive it. As such it might seem abstract

[14] This finding has indirect implication for paper II, since that paper argues that awareness of the DMSO-concentration is of importance for evaluation of the relaxation experiment.

and incomplete in a brief text as this. Motivation for presenting the theory in an experimental oriented thesis is that theory should guide the experiment.

The reader can skip this section and only consider the last equations, which are used for linking changes in free energy to NMR order parameters.

Most of the derivation is provided by John Slater[15] (Slater, 1939), together with few other books (Nash, 1966; Morowitz, 1978; Chandler, 1987; Atkins and De Paula, 2006).

Ensemble averages, microstates, and phase space

We write the observed value of an arbitrary property X as an *ensemble averages*:

$$X_{obs} = \sum_{\nu} P_\nu X_\nu = \langle X \rangle \tag{31}$$

where P_ν is the probability or fraction of time spent in *microstate*[16] ν. Ensemble refers to the assembly (of subsystem) of all possible microstates, which are all states consistent with the constraint with which we characterize the system macroscopically. More specifically a microstate can be defined by N coordinates and N momenta (for N/3 particles) and this makes up a 2N-dimensional *phase space*. There are different ensembles, based on constraints given to them, we will briefly discuss few ensembles and the consequences that follows.

We start with the so-called *microcanonical ensemble*, with constant energy, volume, and number of molecules for each subsystem in the ensemble. Specifically, we study ensemble averages in *stationary systems*, i.e. for which $\langle X \rangle = \langle X(t) \rangle$. We provide an example using a one-dimensional harmonic oscillator with energy gaps $\epsilon = \hbar\omega$ between each level. Three oscillators a, b and c share energy of 3ϵ, which gives rise to three different configurations (roman numbers) and ten microstates according to Figure 5.

[15] Productive American physicist and MIT-professor. Invented the Slater determinant. The first sentence of the book reads: "…unfortunate that physics and chemistry ever where separated […] A wide range of study is common to both subjects. The sooner we realize the better."

[16] Called "Complexion" by Boltzmann (Sharp and Matschinsky, 2015).

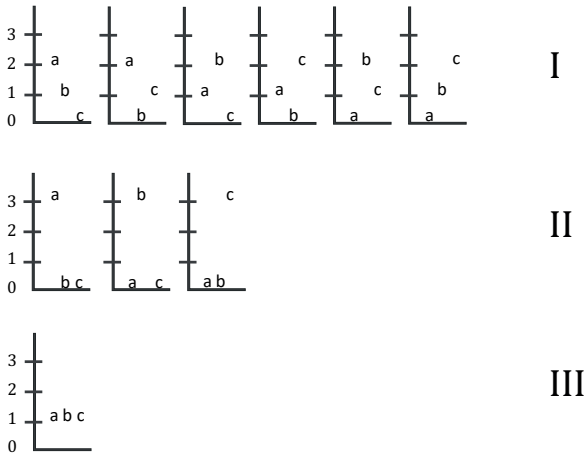


Figure 5. Graph with the number of configurations (I, II and III) and microstates (6, 3 and 1) for three harmonic oscillators (letter a, b and c) in a microcanonical ensemble.

P_ν for each of the ten states in the ensemble is $1/10$, however, the degeneracy (1 or none, 3 and 6) gives the probability of the each of the three configurations I, II and III as $6/10$, $3/10$ and $1/10$ summing up to 1. The degeneracy is called the weight, or *density of states* and given the symbol Ω[17]. Statistical mechanics postulates that all microstates are equally probable – with same *a priori* probability. This cannot be verified in an experiment[18].

The predominant configuration and the statistical entropy

The example is scaled up, from three oscillators to an infinite number (i.e. 6×10^{23}) leading to huge numbers of energy levels ϵ_N, and it turns out that of the immense number of microstates that appear in a macroscopic sample of molecules, an overwhelming majority comes from a small set of configurations, centred on and very similar to the *predominant configuration*. This configuration gives rise to the average value in (31). In Figure 5 the configuration I is the predominant. At equilibrium the probability of a macroscopic state ν is determined by the weight and we write $P_\nu = 1/\Omega$. Now we *postulate* that the entropy of the system can be written:

$$S = -k \sum_\nu P_\nu \cdot \ln P_\nu \tag{32}$$

Using $P_\nu = 1/\Omega$ and the fact $\sum_\nu P_\nu = 1$, we get:

$$S = k \cdot \ln \Omega \tag{33}$$

[17] "Permutabilitätmass" by Boltzmann (Sharp and Matschinsky, 2015).

[18] It is interesting that many of the important statements as this, or the hypothesis of a wave function Ψ in quantum mechanics, cannot be verified. We rely on them because they work.

which is the famous Boltzmann entropy equation, that can be said to connect Clausius macroscopic entropy S with the statistical microscopic Ω. We know that Clausius entropy should be maximum at equilibrium and to see if Ω result in this we differentiate (32):

$$dS = -k \sum_{\nu} (1 + \ln P_\nu) dP_\nu \tag{34}$$

Again using the fact $\sum_\nu P_\nu = 1$ we get $\sum_\nu dP_\nu = 0$ and first term in (34) vanishes; assuming the probability density is uniform we have that $\ln p_\nu$ is independent of ν and we remove it from the summation. We can now use the same arguments as before, $\sum_\nu dP_\nu = 0$, finally giving $dS = 0$ i.e. the entropy in (32) does not vary but are on a maximum, given a uniform probability density.

The canonical ensemble and the partition function

In the previous example all systems in the ensemble have the same energy (N, V and E are common for the systems). The distribution of largest entropy in such a case, is that in which the systems are distributed uniformly through phase space. All P_ν's of the energy E_ν is the same, and all P_ν's with energy deviating from E_ν are zero. This is not realistic for a system at thermal equilibrium, where the energy must be allowed to fluctuate – transferred between systems – and instead we set N, V and T as common for the systems.

The ensemble that represents the possible states of a system in thermal equilibrium with a heat bath at a fixed temperature is called the *canonical ensemble*. Mathematically the following conditions can be formulated:

$$\begin{aligned} &\text{i) } dS = -k \sum_{\nu} \ln P_\nu \, dP_\nu = 0; \\ &\text{ii) } d\langle E \rangle = \sum_{\nu} E_\nu \, dP_\nu = 0; \text{ conservation of energy} \\ &\text{iii) } \textstyle\sum_\nu dP_\nu = 0; \text{ conservation of probabilty;} \end{aligned} \tag{35}$$

Equation i) rephrases (34). These conditions must hold simultaneously, and the optimization problem is solved using Lagrange multipliers. The only way to satisfy i)–iii) for arbitrary values of dP_ν:s, is a linear combination of the coefficients of dP_ν and this leads to:

$$\ln P_\nu = \alpha - \beta E_\nu \tag{36}$$

Substitution of this into the three equations in (35) will prove that we found the solution. Using $\sum_\nu P_\nu = 1$ and noting $P_\nu = e^{\ln P_\nu}$, we get $\alpha = -\ln Z$ where

$$Z = \sum_{\nu} e^{-\beta E_\nu} \tag{37}$$

is the canonical *partition function*. What about the interpretation of β? It is not trivial, but we set $\beta = 1/kT$ and assume there is something called a *Boltzmann distribution* with probabilities for a state given as:

$$P_\nu = \frac{e^{-\beta E_\nu}}{\mathrm{Z}} \tag{38}$$

A preparation of the system according to this gives an ensemble in equilibrium. A completely random mixture in this system corresponds to $\beta = 0$ or $T \to \infty$ and we get $P_\nu = 1/\Omega$, as for the microcanonical ensemble. Using (38) together with the statistical definition of entropy in (32) gives:

$$dS = -k \sum_\nu \ln P_\nu \, dP_\nu = k\beta \sum_\nu E_\nu \, dP_\nu \tag{39}$$

$\sum_\nu E_\nu \, dP_\nu$ corresponds to the heat δq. This is perhaps more easily seen if we study a process slow enough to preserve equilibrium, for which we can differentiate the energy:

$$\mathrm{d}\langle E \rangle = \sum_\nu (E_\nu \, dp_\nu + p_\nu dE_\nu) \tag{40}$$

The term $\sum_\nu p_\nu \, dE_\nu$ in (40) is due to a shift of the energy levels, caused by external perturbation, hence by work δw. Using the first law (1) in combination with (39) we obtain: $\sum_\nu E_\nu \, dp_\nu = \delta q$, or $\sum_\nu E_\nu \, dp_\nu = \delta q_{rev}$ to be exact. We then see that our definition of β agrees with $\mathrm{d}S = \delta q_{rev}/T$, which is the thermodynamic definition of entropy.

The Boltzmann probability distribution in (38) has huge implications for chemistry.

Applications: Helmholtz free energy

Using (38), we can derive the Helmholtz (free) energy, defined as:

$$A = E - TS \tag{41}$$

which we rewrite as:

$$\begin{aligned} A &= \sum_\nu P_\nu E_\nu + kT \sum_\nu P_\nu \ln P_\nu \\ &= \sum_\nu \frac{e^{-\beta E_\nu}}{\mathrm{Z}} E_\nu + kT \sum_\nu \frac{e^{-\beta E_\nu}}{\mathrm{Z}} (-\beta E_\nu - \ln Z) \\ &= -kT \ln Z \end{aligned} \tag{42}$$

As mentioned in the section ‘Internal energy and enthalpy’ the two forms on free energy can be equated for normal experimental conditions in life science. In the NMR chapter we show how this form of energy can be applied, which is the done in paper III.

Transport phenomena

Transport phenomena refers to transfer of mass, heat and momentum and they have in common that there is a change of a thermodynamic variables over time. Kinetics is related to transport processes and is dealt with in paper I. These non-stationary processes are harder to treat than processes at equilibrium and they are typically explained in books with the word 'Nonequilibrium' or 'irreversible' in the title (Eu, 2004; Demirel, 2014). The formal treatment using partial differential equations is beyond the scope of the thesis.

A cornerstone is *the fluctuation-dissipation theorem* developed by Albert Einstein during his studies of Brownian motion and by Lars Onsager (Onsager, 1931). It was explained by Callen and Welton in 1951 (Chandler, 1987) and the theorem connects relaxation in a non-equilibrium system with the spontaneous microscopic dynamics of the equilibrium system. "The fluctuation-dissipation theorem tells us that wherever there is damping there must be fluctuations." (van Kampen, 2007). Small deviation from equilibrium are studied and the so-called *linear response theory*[19] is applied.

The following sections on viscosity and diffusion are, if not otherwise stated, referring to Ira Levine's textbook in physical chemistry (Levine, 2009).

Viscosity

Fluids that flow more easily than others are said to have low viscosity, η. Honey is an example of a fluid with high viscosity under unstressed conditions. In paper II we mixed DMSO and water and studied with NMR how the DMSO induced viscosity changes had implications for NMR-parameters. A simple explanation of viscosity starts from a distribution of velocities in a fluid which gives rise to a transport of momentum. Assuming a one-dimensional laminar flow in x-direction, a frictional force F_x appear:

$$F_x = -\eta \mathcal{A} \frac{dv_x}{dy} \tag{43}$$

v_x refers to velocity in x-direction. $\mathcal{A}$ is the area of an imaginary surface drawn between parallel sheets with different velocities v_x as shown in Figure 6, and η is called the (kinematic) viscosity coefficient. Force per area in this context is the same as *shear stress*.

[19] Linear response theory is a broad term for the study of systems weakly displaced from equilibrium. "[The *response function*] relates the general behavior of a non-equilibrium system in the linear regime to the correlations between spontaneous fluctuations at different times as they occur in the equilibrium system" (Chandler, 1987)

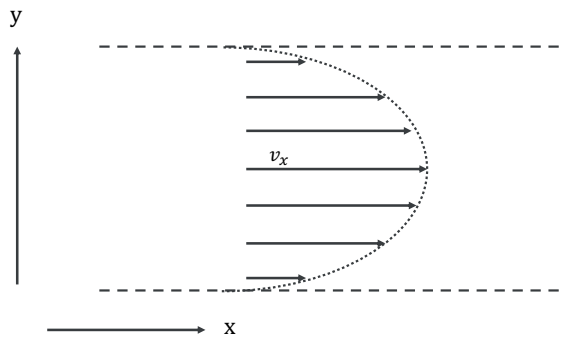


Figure 6. Velocity profile of a fluid exemplifying laminar flow in one dimension. Arrows in the profile show speed at that position. This idealized case holds for Newtonian fluids, i.e. where η independent of shear stress.

When two imaginary adjacent sheet slides against each other, each of them exerts a frictional resistive force on the other, and this is the viscosity. The frictional force is from the slower-moving fluid (short arrows) on the faster moving (long arrows). A prerequisite for (43) to hold is that the velocity is small, otherwise there is turbulent flow.

The SI units for η are N m^2 s or Pa s. (43). For a solid sphere of radius R moving in a fluid, a gradient of speed develops in the fluid surrounding the sphere. Stokes' found the force exerted on the flowing sphere is proportional to the speed v. $F_x = 6\pi\eta R v_x$, where $6\pi\eta R$ is the *friction coefficient* and symbols have their ordinary meaning.

Brownian motion, diffusion, and Stokes-Einstein equation

Thermally agitated solvent molecules undergo random collisions with any particle (molecule or colloid) immersed in the solution. In the absence of other forces, these collisions cause the particle to undergo a random walk called *Brownian motion* (Figure 7), with a mean-square displacement along all three axis (here exemplified in with x) that increases linearly with time t:

$$\langle(\Delta x)^2\rangle = 2Dt \tag{44}$$

D is the *translational diffusion coefficient*, which for a protein in water at room temperature is $\sim 10^{-10}$ m^{2}s^{-1} (Price, 2009). This relationship was discovered by Einstein when he studied colloidal particles. He also assumed that each particle has an average kinetic energy of $1.5 \cdot kT$ and based on this he concluded: [20]

$$\langle(\Delta x)^2\rangle = \frac{2kTt}{F_x} \tag{45}$$

F is the force exerted on the colloidal particles. If the particles are spheres, we reuse the expression for the force found by Stokes, and get the *Stokes-Einstein equation* connecting the viscosity and diffusion coefficients:

[20] …after all he was Albert Einstein.

$$D = \frac{kT}{6\pi\eta R} \tag{46}$$

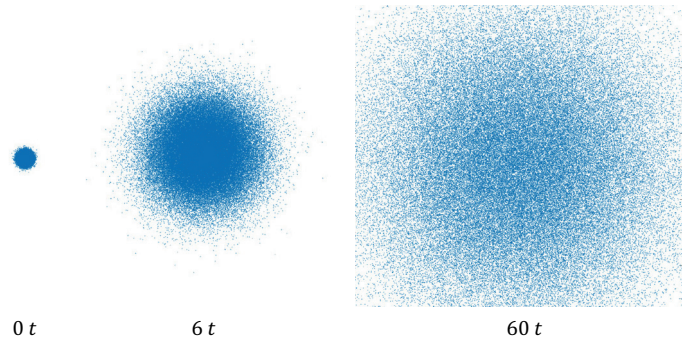


Figure 7. Simulation of two-dimensional translational Brownian motion for 100 000 particles. At time: 0, 6 t and 60 t; t refers to an arbitrary time step.

The rotational Brownian motion for spheres or ellipsoids, analogous to the translational Brownian motion in Figure 7, has large implications for NMR since it leads to random modulation of the magnetic field that drives spin relaxation (Favro, 1960; Woessner, 1962; Hubbard, 1972). The friction coefficient for *rotational diffusion* is $8\pi\eta R^3$, resulting in the rotational diffusion coefficient: $D_{\mathrm{rot}} = kT/8\pi\eta R^3$ (Korzhnev *et al.*, 2001). The *isotropic rotational correlation time* $\tau_c = 1/6D_{\mathrm{rot}}$ and combining the last two equations we have for a spherical protein:

$$\tau_{\mathrm{c}} = \frac{\eta 4\pi R^3}{3kT} = \frac{\eta V}{3kT} \tag{47}$$

R is the effective hydrodynamic radius and, V is the volume which scales linearly with mass so for proteins τ_{c} is approximately 1 ns per 2.6 kDa (Rule and Hitchens, 2006).

NMR

Nuclear Magnetic Resonance is a central experimental method in chemistry, also is called *the central science*. The interest and demand for the NMR-method and the NMR-technology is large in many branches of science. The technique of magnetic resonance in solids and liquids was discovered[21] 1946 by two American research groups led by Purcell and Bloch, more or less independently, and because of that both received the Nobel prize (Bloch, Hansen and Packard, 1946; Purcell, Torrey and Pound, 1946). One of the first reviews on the topic, ten years later concluded:[22] "The resonant techniques provide incisive probes, often to allow a direct question about molecular structure or environment to be answered in more direct fashion than by other means." (Wertz, 1955). The main reason for the success of NMR is exactly this, its unique ability to provide detailed information about molecular processes and structures at atomic resolution. The close connection to the theory of quantum mechanics, which underlies all principles of NMR, is part of the success.

This thesis focuses on protein NMR which involves structural as well as dynamics studies of proteins. It is considered a demanding NMR-specialisation since proteins are large.

When sources are not cited this presentation of the incredibly large NMR-field is based on few comprehensive textbooks (Harris, 1983; Hore, 1995; van de Ven, 1995; Rule and Hitchens, 2006; Cavanagh *et al.*, 2007; Chary and Govil, 2008; Levitt, 2008).

Details on NMR instrumentation and data processing are not presented.

Protein dynamics and NMR – motional regimes

The applications of protein NMR might interest many readers more than the method itself. In any case it can be fruitful to introduce applications before theory. This is done briefly with aid of two recent reviews (Charlier, Cousin and Ferrage, 2016; Kovermann, Rogne and Wolf-Watz, 2016). The didactic must be sacrificed with this approach in the sense concepts presented in this section will be explained in later sections.

NMR is more universally applicable to motional studies than other techniques in chemistry and the range of time scales that can be covered is enormous, from picoseconds

[21] Isidor Isaac Rabi is credited for the discovery of nuclear magnetic resonance. Using molecular beams he found that NMR more accurately could determine magnetic moments (Rabi *et al.*, 1938).

[22] Another insightful comment by John Wertz in the 1955-review is: "…[NMR] represents a worthy challenge to any chemist or physicist…"

to seconds, and the understanding of protein dynamics and function today would not be on the same level without the NMR-relaxation studies. The research field protein dynamics took off around year 1990, much due to the development of triple-resonance probes and ^{13}C and ^{15}N isotope labelling of proteins.

Kovermann et al go through important methodological developments and advancements in technology over the years. Among these are the Lipari-Szabo model-free approach (1982) and the development of pulse sequences for determination of heteronuclear relaxation constants by Kay et al (1989), which made global analysis of protein dynamics achieveable. Introduction of transverse relaxation optimized spectroscopy (TROSY) (Pervushin *et al.*, 1997) paved the way for NMR studies on larger protein systems (size > 30 kDa). The continuous development of coherence transfer techniques and high field magnets in the GHz-range along with invention of cryo-probes, of course have been important for increased sensitivity. Integration of NMR with other methods becomes more common. Typically, the other methods are MD-simulation and crystallography. This is also the case for three of the studies presented here (papers III-V).

Charlier et al emphasize the usefulness of protein-NMR for determination of the conformational space of a protein (also called 'energy-landscape') and explain the theory behind the use of NMR for studies of reorientational and translational dynamics. The chemical exchange motional regimes are covered by different NMR-methods, as exchange spectroscopy also called ZZ-exchange, (k_{ex}=1–50 s^{-1}), CPMG-relaxation dispersion (k_{ex}=10^2–10^3 s^{-1}) and relaxation in the rotating frame (k_{ex}=10^3–10^5 s^{-1}). A recent method, chemical exchange saturation transfer (CEST[23]) investigate low populated states of the protein indirectly by studying the exchange with visible states (Vallurupalli, Bouvignies and Kay, 2012). Relaxometry refers to measurement of relaxation rates by continuous change of the magnetic field (in range of μT to 1T) and the method enables determination of exchange rates in the time-regime k_{ex}=10^5–10^8 s^{-1} which typically is difficult to probe in NMR. Exchange rates of internal water molecules in BPTI has been determined (in Lund) using this method (Persson and Halle, 2008).

Figure 8 gives an overview of some of the protein dynamic time scales and NMR-methods used for studying them.

NMR vs X-ray crystallography

Within structural biology and studies of proteins, the NMR method is often seen as complementary to X-ray crystallography and comparisons between them are made. As of spring 2019 there were 13000 NMR PDB-structures and 140 000 X-ray PDB-structures (Takeuchi, Baskaran and Arthanari, 2019) and due to the impressing speed and efficiency with which high-resolution X-ray structures of proteins are produced, one has to reconsider the role for NMR-methods in protein studies. Takeuchi et al emphasize the need for NMR to study protein dynamics: "NMR often provides the missing pose that is

[23] The so-called saturation transfer methodology used for CEST was developed in Lund 60 year ago (Forsén and Hoffman, 1963). CEST is useful in MRI and enables imaging of low concentration compounds.

critical to tell a whole story. This is especially true for modular or multi-domain proteins, where orchestrated structural rearrangement is an essential part of the function." (Takeuchi, Baskaran and Arthanari, 2019).

Another aspect considered an advantage to NMR is that a large fraction of proteins appears to be intrinsically disordered proteins (IDPs) or has large unfolded segments. Since these do not adopt unique three-dimensional structures in solution but fluctuate rapidly over an ensemble of conformations crystallography cannot provide precise data (Dyson and Wright, 2001). The IDP-concept challenge the Structure–Function paradigm and complements this with a Disorder–Function relationship (Habchi *et al.*, 2014).

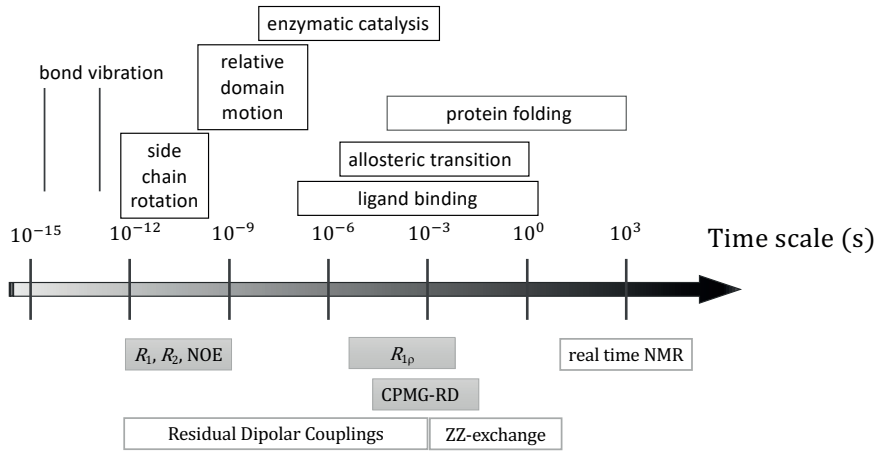


Figure 8. Time scales around room temperature for some of the dynamic processes in proteins (above the arrow) and NMR-methods (below the arrow) that can sense the respective processes. Adopted from several sources (McCammon and Harvey, 1987; Palmer, 2004; Halle, 2016; Kovermann, Rogne and Wolf-Watz, 2016). Grey boxes are methods used in the thesis

Angular momentum and nuclear spin

Magnetic resonance is found in magnetic systems that possess both magnetic moments and intrinsic angular momentum and the nuclear magnetic resonance phenomenon appears when the atomic nuclei has either odd number of protons or neutrons or both (e.g. ${}^{1}_{1}\mathrm{H}$, ${}^{2}_{1}\mathrm{H}$, ${}^{13}_{6}\mathrm{C}$, ${}^{14}_{7}\mathrm{N}$, ${}^{31}_{15}\mathrm{P}$). If the numbers of protons and neutrons both are even there is intrinsic angular moment (e.g. ${}^{12}_{6}\mathrm{C}$, ${}^{16}_{8}\mathrm{O}$). The intrinsic angular momentum in NMR is so called spin and here denoted I. Spin is often described as a purely quantum mechanical phenomenon that has no correspondence in the classical world. Malcolm Levitt, who is an authority in the field of spin dynamics writes:

"Spin is a highly abstract concept, which may never be entirely 'grasped' beyond knowing how to manipulate the quantum mechanical equations. Geometrical arguments can never tell the whole truth, because the human mind is probably incapable of grasping the entire content of quantum mechanics" (Levitt, 2008) and we must accept that a system of nuclei with spin angular momentum I give rise to a magnetic moment μ:

$$\hat{\mu} = \gamma \hat{I} \tag{48}$$

The word 'spin' in this thesis refers only to nuclear spin. Hat over the symbol refers to quantum mechanical operators and γ is the gyromagnetic ratio that can be written:

$$\gamma = \frac{e}{2m_p} g_n = \frac{g_n \mu_B}{\hbar} \tag{49}$$

e is the charge, m_p the proton mass, the nuclear magneton and g_n the g-factor of the nucleon. The most important spin-half nuclei in biomolecular NMR with their respective natural abundance and gyromagnetic ratios in MHz^{-1} are ^{1}H (99.89%, 42.58), ^{13}C (1.07%, 10.71) and ^{15}N (0.36%, 4.31), data from (*CRC Handbook of Chemistry and Physics*, 2018)

By applying a magnetic field to a nuclear state with spin I the degeneracy is broken into $2I+1$ levels. Most nuclei in biomolecular NMR have half-integer spin, hence there are two energy levels for spin quantum numbers $m = +1/2$ and $m = -1/2$. Nuclei with $I > 1/2$ are known as quadrupolar nuclei and deuterium (^{2}H with $I = 1$) is the only example in this thesis. The energy separation is the scalar product $E = -\boldsymbol{\mu} \cdot \boldsymbol{B_0}$, where $\boldsymbol{\mu}$ is the magnetic moment vector and $\boldsymbol{B_0}$ the magnetic field vector. Subscript 0 refers to static time-independent field. The z-axis of the strong magnetic field is defined to coincide with the direction of the magnetic field and the energy becomes $E = -\mu_z B_0$. Eq. (48) gives for spin-1/2: $\mu_z = \gamma I_z$, where $I_z = m\hbar$ and $\Delta m = \pm 1$ and the resonance condition becomes:

$$\Delta E = \hbar \gamma B_0 \tag{50}$$

This splitting of the spin energy levels is called the *nuclear Zeeman effect* (Figure 9).[24]

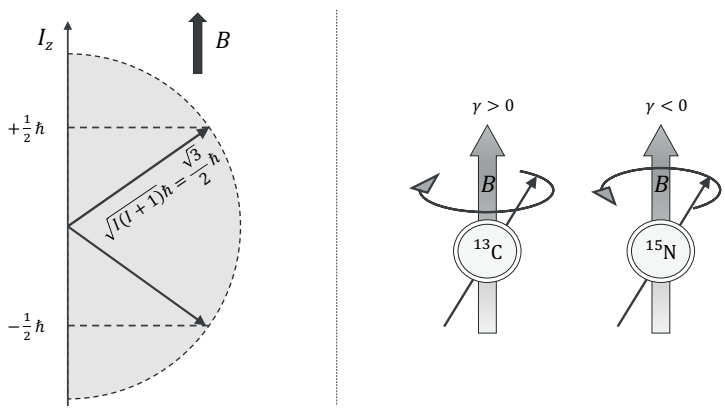


Figure 9 Left: Space quantization of spin 1/2 particle. Right: The sign of the gyromagnetic ratio, γ, determines the sense of precession. Nuclei such as ^{1}H and ^{13}C are said to have clockwise precession while ^{15}N has counterclockwise precession. Right-hand side image inspired by 'Russel39' [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)]

[24] From Pieter Zeeman (1865-1943), dutch physicist sharing a Nobel prize because he discovered the phenomena, and interestingly Wertz suggested in the first NMR-review that "Zeeman-level spectroscopy" would be an appropriate name (Wertz, 1955).

In the classical view the nuclei are small rotating magnet moments, doing precessional motions (Figure 9). The precession frequency[25] ν_0 of the electromagnetic radiation sent out from the NMR-sample is:

$$\nu_0 = -\frac{\gamma B_0}{2\pi} \leftrightarrow \omega_0 = -\gamma B_0 \tag{51}$$

Almost equal sign could be more right since the magnetic field at the nucleus differs slightly from the static B_0-field due to the local chemistry around the nucleus.

The populations in the two energy levels are distributed according to the Boltzmann distribution (38) and this results in a tiny spin polarization; hence the energy is also very low, so low frequency electromagnetic waves are used in NMR-experiments.

For example, for protons placed in a 14.1 T magnet (600 MHz) at 301 K, like conditions for the experiments in the thesis we have:

$$\begin{aligned} \frac{n_\beta}{n_\alpha} &= e^{\frac{-\Delta E}{kT}} = e^{\frac{-\hbar\gamma_H B_0}{kT}} \\ &= e^{\frac{-(1.055\cdot10^{-34}\ \mathrm{Js})\cdot(267.5\cdot10^{6}\ \mathrm{T^{-1}s^{-1}})\cdot(14.1\ \mathrm{T})}{(1.38\cdot10^{-23}\ \mathrm{JK^{-1}})\cdot(301\ \mathrm{K})}} = 0.99990 \end{aligned} \tag{52}$$

so there are 9999 spins in state β ($m = -1/2$, 'spin down', high energy) for every 10000 spins in state α ($m = +1/2$, 'spin up' – low energy) and this difference makes the signal. This results in a very weak signal in general and the search for acceptable signal to noise ratio has always been a challenge in NMR. Since the proton has the highest γ of all nuclei and high natural abundance, NMR-experiments involving other isotopes results in an even worse starting point seen from an experimenter's perspective.

Nuclear spin interactions and NMR parameters

The NMR-spectrum is an ultimate fingerprint of the molecules under study. All features of the spectrum, as number of peaks, their interrelationship, the peak shape, and intensities are explained by the nuclear spin interactions, also called spin couplings. A spin interaction is defined as something that modifies the energy levels of a spin. In this section the important interactions – for the thesis – are presented, these are: *the chemical shift*, *the scalar or J-coupling*, *the dipolar coupling* and *the quadrupolar coupling*.

Chemical shift

The most central NMR-parameter is the chemical shift giving each nucleus its characteristic frequency due to the molecular ("chemical") surrounding and because of its importance we give it little more attention than the others. The chemical shift is

[25] Larmor frequency is another term for precession frequency from the Irish physicist Joseph Larmor (1857-1942).

extremely sensitive to the electron density and it is a consequence of the interaction between the B_0-field and the electrons, and of the electrons with the nuclei. It was discovered 1950, just few years after the NMR-method was discovered that the ^{14}N-shifts for the ammonium ion differed depending on counter ion (Proctor and Yu, 1950).

Equation (51) is restated as $\omega_{\mathrm{loc}} = -\gamma B_{\mathrm{loc}}$, where loc refers to local angular frequency (in radians) and local magnetic field. $B_{\mathrm{loc}} = (1-\sigma)B_0$, where σ is the shielding constant. The shielding constant σ is not suitable as a measure of the shift since absolute shifts are rarely needed. Because of this a relative measure of the shielding – the chemical shift – is invented. This is defined as the dimensionless parameter:

$$\delta = \frac{\nu - \nu_{\mathrm{ref}}}{\nu_{\mathrm{ref}}} = \frac{(1-\sigma) - (1-\sigma_{\mathrm{ref}})}{1-\sigma_{\mathrm{ref}}} \approx 10^6 \cdot (\sigma_{\mathrm{ref}} - \sigma) \qquad (53)$$

Almost equal sign is chosen since $\sigma_{\mathrm{ref}} \ll 1$, and ref stands for a calibration standard compound, which is DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) in aqueous solution for protein NMR. δ is given in ppm (parts per million) and it is a deshielding constant – it increases with decreasing σ.

Shielding anisotropy

The general description of the shielding is as a tensor with anisotropic properties and its time average over all orientations – what is seen in the spectra due to rapid random tumbling of molecules in liquids – is called the *isotropic chemical shift*, defined as:

$$\bar{\sigma} = \frac{\sigma_{xx} + \sigma_{yy} + \sigma_{zz}}{3} \qquad (54)$$

The subscripts refer to the tensor diagonal elements and the right-hand side is the average trace of the shielding tensor. The anisotropy $\Delta\sigma$ and the asymmetry η are defined as:

$$\Delta\sigma = \sigma_{zz} - \frac{(\sigma_{xx} + \sigma_{yy})}{2}; \quad \eta = \frac{3(\sigma_{xx} - \sigma_{yy})}{2\Delta\sigma} \qquad (55)$$

Examples of σ-tensors, with their respective isotropic shift, anisotropy and asymmetry are shown in Figure 10. When referring to anisotropy of δ we say *chemical shift anisotropy* (CSA). The CSA originates from the anisotropic electron density in the magnetic field. In proteins CSA are ~170 ppm for ^{15}N (NH) and ~40 ppm ^{13}C (CH).

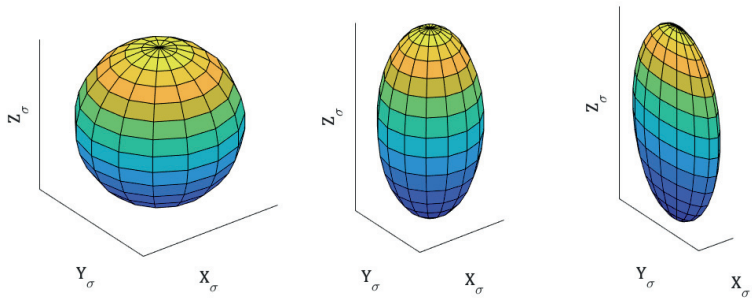


Figure 10. The ellipsoid for the σ-tensor in the principal axis system of the tensor with Zσ aligned with the B0-field. The coordinates {Xσ, Yσ, Zσ} are {1,1,1}; {1,1,2} and {1,3,5} from left to right and the this result in respective isotropic shift, anisotropy and asymmetry: {1.0, 0, 0} ; {1.33, 1, 0} ; {3, 3, –1}.

Contributions to nuclear shielding

A useful decomposition of the shielding constant is (Hore, 1995; Chary and Govil, 2008):

$$\sigma = \sigma_d + \sigma_p + \sigma_n + \sigma_{\text{other}} \tag{56}$$

where d represent local diamagnetic, p local paramagnetic and n neighbouring group anisotropy. 'Other' refers to all other sources as electric fields, solvent effects or van der Waals interactions. In the section 'pH-titration of acidic side chains' we explained how the electric field influence the shielding of the nuclei and how it can be used for determination of pK_a-values.

It is a formidable task to make *ab initio* calculations of the magnetic shielding tensor and relativistic effects are needed to account for (Kaupp, 2004; Olejniczak *et al.*, 2012). However, the basic non-relativistic theory holds well for spin-1/2 nuclei involved in protein NMR studies.

Local diamagnetic and paramagnetic term

Lamb showed that for a single free atom with a spherical symmetric electron density, (i.e. electrons in S-state only) the shielding factor is (Lamb, 1941):

$$\sigma_d = \frac{4\pi e^2}{3mc^2}\int_0^\infty r\rho(r)dr \tag{57}$$

where $\rho(r)$ is the radial electron density, and other symbols have their usual denotation. The induced magnetic field based on (57) was determined as $B_{\text{ind}} = -0.319 \times 10^{-4} Z^{\frac{4}{3}} B_0$, where Z is the mass number, hence for protons, $B_{\text{ind}} \approx 10^{-4} \times B_0$. The calculations by Lamb explains σ_d. In a diamagnetic material, the magnetization induced by the external field (B_0) is opposing that field.

σ_p, of opposite sign to σ_d, cannot be explained classically, and it corrects for effects from electrons outside the s-shell, especially the presence of p and d electrons determine the magnitude of σ_p.

Ramsey[26] outlined the theory of the chemical shift by combining the effect from σ_d and σ_p: "they [σ_d and σ_p] at least partially and perhaps completely explain the chemical effect that has been reported by various observers in measurements of nuclear moments" (Ramsey, 1950a, 1950b). The *temperature-independent paramagnetic term* is an extended name for σ_p, to distinguish it from the temperature-dependent term that results from unpaired electrons and are of much greater magnitude (Becker, 2000).

Three factors determine σ_p. i) An inverse relationship with the energy gap $< 1/\Delta E >$ between ground and excited states; ii) an inverse relationship to the average cube distance $\langle 1/r^3 \rangle$ between the nucleus and the orbitals of electrons (Jameson and Gutowsky, 1964); iii) the asymmetry in distributions of electrons between p-orbitals near the nucleus.

The considerably small chemical shift range of 10 ppm for protons, is explained by the *few electrons – less shielding* leading to a small σ_d. There is *little influence of p-orbitals and large energy gap* between ground and excited states leading to a small σ_p as well. The smaller shielding makes proton shifts to an inferior indicator of trends in chemical shift changes. For larger atoms such as C, N and F, with p- and d-orbitals, the energy gap is smaller and because of this, the σ_p creates a larger dispersion, or range of chemical shifts for these nuclei. The size of the atoms decreases in a period from left to right, and the inverse cube distance effect leads to higher impact of σ_p for atoms to the right in the periodic system.

The chemical shift ranges are around 200 ppm for ^{13}C and 900 ppm for both ^{15}N and ^{19}F.

Neighbouring group anisotropy

The circulation of electrons in a benzene rings in a strong magnetic field will serve as an example of *neighbouring group anisotropy,* σ_n in (56).

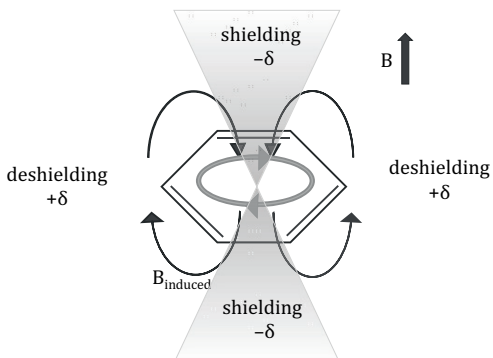


Figure 11. Schematic presentation of shielding and deshielding zones around benzene ring due to ring current of delocalized π-electrons (grey circle) when the plane of the benzene molecule is perpendicular to the magnetic field lines. Ring current induced chemical shift changes are discussed in paper IV.

[26] Norman F. Ramsey developed the theory as (the only at that time) graduate student with Isidor Isaac Rabi at Columbia 1937-49, he moved to Harvard 1949. (Ramsey's Nobel Lecture 1989)

When an aromatic compound – for example the ligands used in papers III and IV – is placed in an external magnetic field (B_0), the delocalized π-electrons circulate around the benzene ring and generate a current. This ring current induces a local secondary magnetic field with maximum when the plane of the benzene ring is perpendicular to the magnetic field lines (Figure 11). The induced field opposes B_0 in the middle of the ring (above and below) leading to shielded regions, but reinforces B_0 in the transverse plane, hence deshielding the aryl protons and neighbouring nuclei.

Intramolecular protons inside extended aromatic compounds (e.g. annulene) has been shown to have very shielded protons at –3 ppm (Hore, 1995). The magnetic anisotropy of carbon double bonds has been assumed to be like that of aromatic rings.

Random coil chemical shift of peptides and proteins

Figure 12 shows a two-dimensional ^{1}H–^{15}N HSQC spectrum of Gal3C. Chemical shifts in proteins are partitioned into so-called *random coil chemical shift* and the conformation-dependent *secondary chemical shifts*. Random coil shifts are defined as the characteristic chemical shifts of amino acid residues in short – usually tri- and tetra-peptide, disordered polymers (Richarz and Wüthrich, 1978; Wishart *et al.*, 1995).

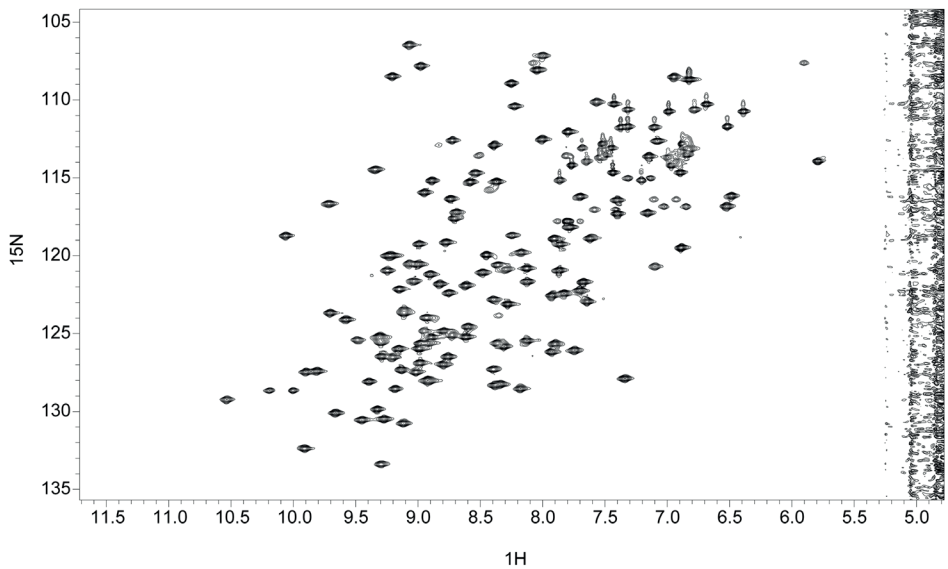


Figure 12 ^{1}H-^{15}N-HSQC spectrum of one of the Gal3C-complexes (M-ligand) from paper III shows the correlation peaks of the NH amide bonds.

The secondary chemical shift is the difference between the observed and the random-coil shifts, these usually have non-covalent structural or dynamic information (Wishart and Case, 2002). Especially for some isotopes the secondary chemical shift points at specific protein secondary structure (Spera and Bax, 1991). Since twenty years, backbone torsion angle can be predicted based on chemical shift (Cornilescu, Delaglio and Bax, 1999).

Another step was taken when chemical shifts, in conjunction with Monte Carlo assembly-based structure determination protocol, were used to determine protein structures only based on the chemical shift data (Shen *et al.*, 2008). In protein–ligand interaction studies, the small changes of the chemical shift due to different ligands are typically studied, and it is called *chemical shift perturbation* (CSP) studies (Williamson, 2013). CSP of ^{1}H–^{15}N HSQC-spectra is a useful and straightforward tool, used in all the papers in this thesis.

Scalar coupling

The multiplet splitting of peaks in liquid state spectra arise because each spin interacts with its neighbour through a mechanism called scalar or J-coupling[27]. This interaction between two spins S and I is given by a (scalar product) Hamiltonian: $\hat{H}_J = 2\pi J \hat{I}\hat{S} = 2\pi J(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y + \hat{I}_z\hat{S}_z)$, which contribute to a splitting of S into $2I + 1$ lines and of I to $2S + 1$ lines. For spin 1/2-particles this leads to 2 lines. The mechanism is mediated by bonding electrons. When the J-coupling extends over more than one bond it is called a long-range coupling. The number of bonds over which the coupling is active is indicated as a superscript ^{2}J or ^{3}J. Typical values are 125–250 Hz (^{13}C, ^{1}H); 30–80 Hz (^{13}C, ^{13}C); 70–110 Hz (^{15}N, ^{1}H); 2–20 Hz (^{15}N, ^{13}C). The ^{2}J and ^{3}J have a magnitude of about –20–20 Hz, where minus sign refer to few cases of ^{2}J (van de Ven, 1995). J-couplings are expressed in Hertz since they are independent of the magnetic field.

The existence of J-couplings is used heavily when magnetization is transferred via NMR pulse sequences. The *weak-coupling* limit, which give rise to so-called first-order spectra refers to the condition $|\nu_I - \nu_S|/|J_{IS}| \gg 1$. If $|\nu_I - \nu_S| \approx |J_{IS}|$ the multiplets appear to lean toward one another. The magnitude of the ^{3}J couplings can also provide information on the magnitude of torsion angles which provide direct structural information (Karplus, 1959).

Dipolar coupling

The dipolar interaction summarizes the energy relationship between two magnetic moment. It is also called *direct dipole–dipole coupling*[28] to emphasize it is not mediated by electrons but instead occurs directly between nuclear spins, it can be described by the Hamiltonian (in frequency units):

$$\hat{H}_{dd} = \frac{\mu_0}{4\pi} \cdot \frac{\gamma_I \gamma_S \hbar}{r^3} \left[\hat{I} \cdot \hat{S} - \frac{3\left(\hat{I} \cdot \bar{r}\right)\left(\hat{S} \cdot \bar{r}\right)}{r^2} \right] \qquad (58)$$

The vector $\bar{r}$ points from spin I to S and makes an angle θ with the orientation of B_0. The factor in front of the brackets is the *dipolar coupling constant* and μ_0 ($4\pi \times 10^{-7}$H m^{-1}) is the permeability of free space. The interaction is tensorial in nature, meaning it depends on the relative spatial orientation of the two dipoles. It turns out that for isotropic fluids

[27] Also called indirect coupling or spin-spin coupling

[28] Also called through-space dipole–dipole coupling

the expression in the brackets averages to zero and due to this, the dipolar coupling does not cause any splitting for NMR-spectra in liquids.

Quadrupolar coupling

The quadrupolar interaction couples the electrical quadrupolar moment of a nucleus with its nuclear magnetic moment. The theory for this interaction is beyond the scope of the thesis. Briefly we have no *electric* energy terms that depend on the orientation or internal structure of the nucleus, for the spin one half particles, but for deuterons (^{2}H), with spin-one, used in the methyl spin-relaxation studies in Paper III and referred to in Paper II, the electric energy term appears. The quadrupolar coupling is strong, resulting in heavy damping of the relaxation rates in these studies.

Spin relaxation

The NMR spectrum is mainly determined by the chemical shift, the J-coupling and exchange effects. The timescale of the disappearance of the peaks, the processes called relaxation, also has consequences for the spectrum since it affects the intensity.

Relaxation is caused by microscopic fluctuations in the spin interactions. Therefore, the relaxation rates have information on the average dynamical behaviour of individual spins and their surroundings, hence the rates supply knowledge of dynamic processes in molecules, as well as structure.

A complete outline of protein NMR spin relaxation theory involves stressful quantum mechanical operator theory at a level above the goal of this presentation. The following descriptions and reviews on spin relaxation theory are used (Kowalewski, 1990; van de Ven, 1995; Korzhnev *et al.*, 2001; Luginbuhl and Wüthrich, 2002; Jarymowycz and Stone, 2006; Cavanagh *et al.*, 2007; Levitt, 2008; Nicholas *et al.*, 2010; Grzesiek, 2017).

Classical description, Bloch equations and rotating frame

The classical picture of NMR starts by studying the equation for the magnetic spinning top. Let $\boldsymbol{M} = \boldsymbol{M}(t) = \{M_x(t), M_y(t), M_z(t)\}$ be the nuclear magnetization in cartesian coordinates. The precession of the magnetization (without taking chemical shielding into account) is described by:

$$\frac{d}{dt}\boldsymbol{M} = \gamma \boldsymbol{M} \times \boldsymbol{B} \tag{59}$$

which according to the nature of the vector product says that the change in magnetization, $d\boldsymbol{M}$ is perpendicular to the plane spanned by $\boldsymbol{M}$ and $\boldsymbol{B}$. Because of this "magnetic force" the spin precess. According to (59) the magnetization keeps precessing forever at frequency $|\gamma B_0|$. Felix Bloch solved the problem with eternal precession by adding damping constants R_1 and R_2 and he also assumed an initial magnetization M_0 at thermal equilibrium (Bloch, 1946):

$$\frac{d}{dt}\boldsymbol{M} = \gamma \boldsymbol{M} \times \boldsymbol{B} - \boldsymbol{R}(\boldsymbol{M} - \boldsymbol{M_0});$$
$$\boldsymbol{R} = \begin{bmatrix} R_2 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & R_1 \end{bmatrix}; \quad \boldsymbol{M_0} = \begin{bmatrix} M_{0,x} \\ M_{0,y} \\ M_{0,z} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dfrac{CB_0}{T} \end{bmatrix} \tag{60}$$

which are the Bloch equations. R_1 is the *longitudinal* or *spin-lattice* relaxation rate constant and R_2 is the *transverse* or *spin-spin* relaxation rate constant and $C = N\mu_0\mu/k$ is the Curie constant, N the number of spins and other symbols have their usual meaning.

Very often in NMR a *rotating reference frame* is implemented, by introducing a weak variable magnetic field $B_1(t)$ with frequency ω_{rf}, resulting in change of coordinates to ones which rotate in the xy plane at offset frequency $\Omega = \omega_0 - \omega_{\mathrm{rf}}$. The rotating frame is indicated by superscript $\boldsymbol{r}$ in (60), as $(d/dt)\boldsymbol{M^r} = \gamma \boldsymbol{M^r} \times \boldsymbol{B^r}$. The transformation might be more easily understood by studying each component:

$$\begin{aligned} \frac{dM_x}{dt} &= -R_2 M_x + (\omega_0 - \omega_{\mathrm{rf}}) M_y \\ \frac{dM_y}{dt} &= -R_2 M_y - (\omega_0 - \omega_{\mathrm{rf}}) M_x \\ \frac{dM_z}{dt} &= R_1 (M_0 - M_z) \end{aligned} \tag{61}$$

and by introducing $M^+ = M_x + iM_y$, we get the magnetization in the rotating frame immediately after a 90° y-pulse as $M^+ = \exp\{-t(i\Omega + R_2)\}$. $\boldsymbol{B}^r$ is called *the effective field in the rotating frame*.

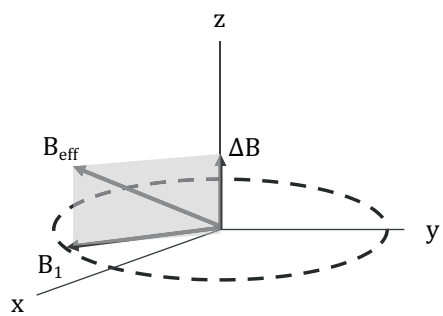


Figure 13. In the rotating frame the static B_0 field is reduced to a residual field of size comparable to B_1. The vector sum B_{eff} is the effective field in the rotating frame.

The original, real world, coordinate frame, is called the laboratory frame. In the rotating frame the z-component of the effective field: $\Delta B = \Omega/\gamma = B_0 - \omega_{\mathrm{rf}}/\gamma$ determines the offset frequency. This residual static field is comparable to B_1 when transmitter frequency

is close to ω_0. Figure 13 show the relationship between B_1, ΔB and $B_{\text{eff}} = \sqrt{B_1^2 + \Delta B^2}$ and Figure 14 shows development of the magnetization with and without an offset.

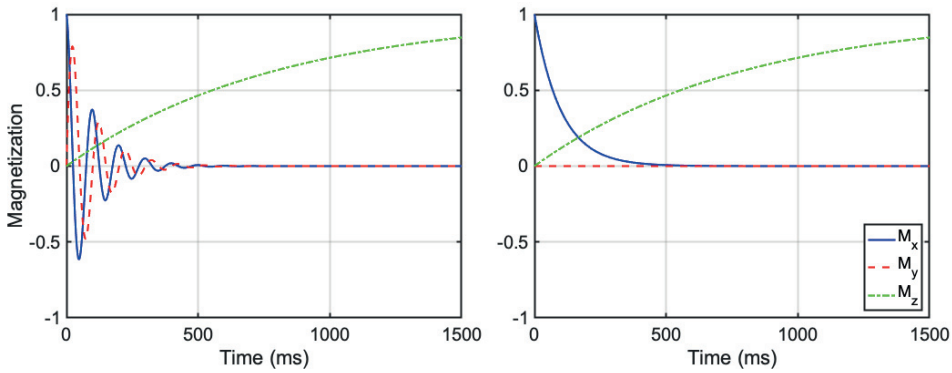


Figure 14. Simulation of equations (61) for magnetization after a 90° y-pulse with R_1 = 1.25 s^{-1} (green dot-dashed line) and R_2 = 10 s^{-1}, (blue line in right image) typical for small to medium sized proteins in the magnetic fields commonly used. Left picture shows Ω = 5 Hz and right shows Ω = 0 Hz, i.e. on-resonance with observer at M_y = 0.

Relaxation processes – T_1- and T_2-relaxation

The relaxation to thermodynamic equilibrium and the Boltzmann distribution (52) is called T_1-relaxation ($T_1 = 1/R_1$) and involves energy transfer from excited state to the surroundings (lattice). By introducing the B_1-field we tilt the magnetization from the z-axis to the xy-plane and when the B_1-field is switched off, the tilted magnetization returns to the original Boltzmann distribution in a process described by a time constant T_1.

The T_2-relaxation ($T_2 = 1/R_2$) occurs without any exchange of energy with the surrounding but is governed by the stochastic movements due to Brownian diffusion in the liquid. It is a measure of the speed for the frequencies of individual nuclei to get "out of sync". T_2-relaxation is generally said to consist of three contributions. i) *Inhomogeneous magnetic field* results in different precession frequency depending on sample location. This is a systematic effect. Stochastic molecular movements give rise to magnetic field fluctuations and these are ii) the magnetic field *fluctuations due to the dipolar coupling* between spins, and iii) the *fluctuations due to the CSA* (Figure 10 and Figure 11). Of these two stochastic fluctuation contributions to transverse relaxation, the dipolar is most important. [29]

[29] Spin-spin relaxation is a synonym to transverse relaxation because of this.

Redfield semi-classical relaxation theory

This section describes the relaxation theory for spin-1/2 particles as outlined by Alfred Redfield in the 1950s[30] (Redfield, 1957). His ideas starts with Wangsness-Bloch theory[31] (Wangsness and Bloch, 1953) but Redfield find their method insufficient, especially he concludes: "if the thermal bath is a complicated thing, like a liquid, it will be impossible to solve its Hamiltonian.". Redfield notes that in the famous BPP-paper (Bloembergen, Purcell and Pound, 1948) the effect of the thermal bath can be estimated from the classical mean-square size of its interaction, and the rate of the statistical fluctuations of this interaction caused by magnetic moments of surrounding molecules, and the word Semi-classical can be traced to this assumption. The surrounding is treated classically and the spin system with quantum mechanics. Redfield introduces important concepts into relaxation theory such as *density matrix*, *coherence*, *correlation time,* and *spectral density*. The outline here does not carefully describe these concepts. Also, the *not so interested* reader can jump the sections 'Density matrix' and 'Liouville-von Neumann equation' without losing track of the thesis' message. The motivation for still having these sections are that they derive the very important relaxation parameters.

The outline is adopted from James Keeler's simplified compilation (Keeler, 2002) with aid from other sources (van de Ven, 1995; Cavanagh *et al.*, 2007).

Density matrix

The starting point is the density matrix and the density operator. We use the so-called Dirac bracket (*bra*-c-*ket*) notation. The two-level spin 1/2 system is a superposition of two basis states, the *ket* (row vector) is:

$$|\psi\rangle = c_\alpha|\alpha\rangle + c_\beta|\beta\rangle \tag{62}$$

Where α and β represent the basis vectors and c are complex numbers having information of phase and probability of the basis states. There is a corresponding complex conjugate *bra* (column vector). The expectation value, the experimental mean value, of an operator $\hat{A}$ – usually a Hamiltonian operator or an angular momentum operator – is:

$$\langle\hat{A}\rangle = \langle\psi|\hat{A}|\psi\rangle = \left(\langle\alpha|c_\alpha^* + \langle\beta|c_\beta^*\right)|\hat{A}|\left(c_\alpha|\alpha\rangle + c_\beta|\beta\rangle\right) \tag{63}$$

The operator is also a matrix, which is always Hermitian since the eigenvalues must be real, and we can write it as:

[30] The seminal NMR-paper is published in IBM Journal, founded the same year. He worked for IBM at Watson Research Laboratory studying nuclear and electronic spin magnetic resonance and low-temperature physics. It is unlikely that companies today accept that their employees devote themselves to non-profit, basic research as Alfred Redfield did.

[31] Redfield theory is also called *Bloch-Redfield-Wangsness* theory, but there is no consensus since the name *Abragam-Redfield* theory is also used (Nodet and Abergel, 2007).

$$\boldsymbol{A} = \begin{bmatrix} \langle\alpha|\hat{A}|\alpha\rangle & \langle\alpha|\hat{A}|\beta\rangle \\ \langle\beta|\hat{A}|\alpha\rangle & \langle\beta|\hat{A}|\beta\rangle \end{bmatrix} = \begin{bmatrix} A_{\alpha\alpha} & A_{\alpha\beta} \\ A_{\beta\alpha} & A_{\beta\beta} \end{bmatrix} \tag{64}$$

Using this the expectation value can also be written as:

$$\langle\hat{A}\rangle = [c_\alpha^* \quad c_\beta^*] \begin{bmatrix} A_{\alpha\alpha} & A_{\alpha\beta} \\ A_{\beta\alpha} & A_{\beta\beta} \end{bmatrix} \begin{bmatrix} c_\alpha \\ c_\beta \end{bmatrix} \tag{65}$$

Removing the matrix $\boldsymbol{A}$ gives the square of the length of ψ, or the normalization factor:

$$\langle\psi|\psi\rangle = [\mathrm{c}_\alpha^* \quad \mathrm{c}_\beta^*] \begin{bmatrix} \mathrm{c}_\alpha \\ \mathrm{c}_\beta \end{bmatrix} = |\mathrm{c}_\alpha|^2 + \left|\mathrm{c}_\beta\right|^2 = 1 \tag{66}$$

We define an operator $\hat{\sigma}$ and a corresponding matrix $\boldsymbol{\sigma}$[32] as:

$$\hat{\sigma} \equiv |\psi\rangle\langle\psi| = \begin{bmatrix} c_\alpha \\ c_\beta \end{bmatrix} [c_\alpha^* \quad c_\beta^*] = \begin{bmatrix} c_\alpha c_\alpha^* & c_\alpha c_\beta^* \\ c_\beta c_\alpha^* & c_\beta c_\beta^* \end{bmatrix} = \boldsymbol{\sigma} \tag{67}$$

this is the *density operator* and the *density matrix*. The dimension of the operators, and the number of elements in the matrix for a spin $1/2$ system with N spins is 4^N. The space is called the Liouville space in contrast to the Hilbert space (dimension 2^N) that builds up the wave function ψ. The trace operation of the density matrix times the operator matrix equals the expectation value:

$$\langle\hat{A}\rangle = \mathrm{Tr}(\boldsymbol{\sigma A}) = \mathrm{Tr}(\boldsymbol{A\sigma}) \tag{68}$$

which holds for a *pure* quantum state. There is only one initial quantum state described in (62): $|\psi\rangle = c_\alpha|\alpha\rangle + c_\beta|\beta\rangle$. A *mixed* quantum state is a statistical mixture of many wave function, here exemplified as a combination of two spin quantum states as $\Psi = p_1|\psi_1\rangle + p_2|\psi_2\rangle$. The probabilities p_i should not be confused with the c, but p_i are classical probabilities (sums to 1) describing the ratio or the weight of the two quantum states. The *ensemble average expectation value* (of N expectation values as (65)) is:

$$\overline{\langle\hat{A}\rangle} = \sum_k^N p_k \langle\hat{A}\rangle_\mathrm{k} \tag{69}$$

which is simplified (for a two-state system) as:

$$\overline{\langle\hat{A}\rangle} = \mathrm{Tr}(\bar{\boldsymbol{\sigma}}\boldsymbol{A}) = \mathrm{Tr}\,\overline{\begin{bmatrix} c_\alpha c_\alpha^* & c_\alpha c_\beta^* \\ c_\beta c_\alpha^* & c_\beta c_\beta^* \end{bmatrix}} \times \begin{bmatrix} A_{\alpha\alpha} & A_{\alpha\beta} \\ A_{\beta\alpha} & A_{\beta\beta} \end{bmatrix} \tag{70}$$

which is identical to (68) except for the added overbar, which denotes an ensemble average. Hereafter the overbar is skipped even though we work with mixed spin states.

[32] σ here is not related to the shielding tensor. Also, equation (67) is an example of where the positioning of equality and definition signs is not trivial. QM tend to interweave concepts.

Coherence

The density matrix helps in explaining one of the most important concepts in NMR, that of *coherence*. The diagonal density matrix elements in (67) represent the populations, and the off-diagonal the coherences. The c's are complex numbers as: $c_\alpha = |c_\alpha|e^{i\phi_\alpha}$, and $c_\beta = |c_\beta|e^{i\phi_\beta}$, and we rewrite the off-diagonal elements for the mixed two-state in (70):

$$\begin{aligned}(+1)&: \overline{c_\alpha c_\beta^*} = \overline{|c_\alpha||c_\beta|e^{\imath(\phi_\alpha-\phi_\beta)}} \\ (-1)&: \overline{c_\beta c_\alpha^*} = \overline{|c_\alpha||c_\beta|e^{\imath(\phi_\beta-\phi_\alpha)}}\end{aligned} \tag{71}$$

$(+1)$ and (-1) stands for coherence order and $(+1)$ comes from $\Delta m = (+1/2) - (-1/2) = +1$. For an NMR-experiment that starts from with a 90° RF-pulse on a longitudinal equilibrium magnetization I_z, a *phase relationship* between the states, $|\alpha\rangle$ and $|\beta\rangle$ is created and the ensemble average for the off-diagonal elements in (71) are non-zero. There is "a coherent superposition between the two states" (Cavanagh *et al.*, 2007). The goal for a pulse sequence (see section 'Pulse sequences for two- and multidimensional NMR') is to "treat" the density matrix in such a way that this phase relationship is kept through the sequence, and we end up with some kind of coherence of interest during the acquisition part. This is not a trivial task, and sometimes the spectra show only the background noise, absent of any sort of coherence or correlation.

Liouville-von Neumann equation

The density matrix has a central role in the Redfield theory. We can describe the time derivative of the density operator in units of $\hbar$ (so-called Planck units) as:

$$\frac{d\hat{\sigma}}{dt} = -i\left[\hat{H}, \hat{\sigma}\right] \tag{72}$$

This is the *Liouville-von Neumann equation*. The Hamiltonian $\hat{H} = \hat{H}_0 + \hat{H}_1(t)$ consists of a static part $\hat{H}_0$ and a random time dependent part $\hat{H}_1(t)$. We have mentioned that relaxation is due to stochastic fluctuations, hence our interest lies in the random part and by transforming it to $\hat{H}_1^T(t)$ via:

$$\hat{H}_1^T(t) = e^{i\hat{H}_0 t}\hat{H}_1(t)e^{-i\hat{H}_0 t} \tag{73}$$

we deal only with the random operator:

$$\frac{d\hat{\sigma}^T(t)}{dt} = -i\left[\hat{H}_1^T(t), \hat{\sigma}^T(t)\right] \tag{74}$$

We call this the *interaction representation*. Equation (74) can be solved (Nicholas *et al.*, 2010 gives a complete solution) to second order in the perturbation to give:

$$\frac{d\hat{\sigma}^T(t)}{dt} = -\int_0^\infty \overline{\left[\hat{H}_1^T(t), \left[\hat{H}_1^T(t-\tau), \hat{\sigma}^T(t)\right]\right]}\, d\tau \tag{75}$$

Assumptions behind the derivation of the time-derivative of $\hat{\sigma}^T(t)$ are:

- $\hat{\sigma}^T(t)$ and $\hat{H}_1^T(t)$ are uncorrelated.
- Time scale: $\tau_c \ll t \ll \frac{1}{R}$, where τ_c is the correlation time of the stochastic Hamiltonian driving the relaxation, R is the relaxation rate.
- The replacement $\hat{\sigma}^T(t) \rightarrow \hat{\sigma}^T(t) - \sigma_{eq}$, where σ_{eq} is the equilibrium density operator, for $\hat{\sigma}^T(t)$ to not relax to 0, but to the Boltzmann energy distribution.

To make better use of the master equation the $\hat{H}_1^T(t)$ is expressed as:

$$\hat{H}_1^T(t) = \sum_q \hat{F}^q(t)\,\hat{A}^q \tag{76}$$

where $\hat{F}^q$ are *random spatial functions*. $\hat{A}^q$ are *spin operators*, each with a characteristic frequency ω_p^q under the $\hat{H}_0$ expressed as:

$$e^{i\hat{H}_0 t}\hat{A}^{q,T}e^{-i\hat{H}_0 t} = \sum_p e^{i\omega_p^q t}\,\hat{A}^q \tag{77}$$

We give an example of $\hat{A}^q$ in the case of dipolar relaxation for an *IS* two spin system (e.g. I ^{1}H and S). We have the following two-spin operators: I_zS_z, I^-S^+, I^+S^-, I_zS^+, I^+S_z and I^+S^+, with corresponding ω_p^q: 0, $\omega_S - \omega_I$, $\omega_I - \omega_S$, ω_S, ω_I and $\omega_I + \omega_S$. Since $\hat{A}^q$ is Hermitian there are equal numbers of $\hat{A}^{-q}$, with a negative frequency according to $\omega_p^{-q} = -\omega_p^q$. The same holds for the spatial functions $\hat{F}^q$, which also have a corresponding $\hat{F}^{-q}$.

The two-spin operators are direct products of single-spin spin operators (e.g. $I_zS_z = I_z \otimes S_z$) and the product operator methodology for designing pulse sequences is also based on direct products (Sørensen *et al.*, 1983).

Combining (75) and (77) yields, after extensive calculations (see Cavanagh et al., 2007, chapter 5) the *master equation* for calculation of spin relaxation rates:

$$\frac{d\hat{\sigma}^T(t)}{dt} = -\frac{1}{2}\sum_{p,q,p',q'}\left[\hat{A}_p^q,\left[\hat{A}_{p'}^{q'},\hat{\sigma}^T(t)\right]\right] \times J(\omega) \times e^{i\left(\omega_p^q+\omega_{p'}^{q'}\right)t} \tag{78}$$

J is the spectral density[33], which is the Fourier transform of the time correlation function: $J\left(\omega_p^q\right) = \int_{-\infty}^{\infty} G(\tau) \times e^{-i\omega_p^q\tau}\,(\tau)d\tau$, where $G(\tau) = \overline{F^q(t)F^{q'}(t-\tau)}$ is the time correlation function of the spatial functions.

$G(\tau)$ is expectation values of products of F^q at different times and it is large for small τ and goes towards zero for large τ. $G(\tau)$ is difficult to display since knowledge of the dynamics is needed to calculate it. Examples of $G(\tau)$ (confusingly called $C(\tau)$) are shown in section 'Spectral density modelling and generalized order parameters'.

[33] Originally, 'Fourier intensities at frequency ω' were used for $J(\omega)$.

A simplification of (78) can be made by assuming that no cross-correlation occurs. Then, the only pair of q and q' that results in anything else than $G = 0$ are $q' = -q$. Another effect of neglecting cross-correlation is that the exponential in (78) becomes equal to unity since $\omega_p^{q'} = \omega_p^{-q} = -\omega_p^q$. The simplification means we only look at time evolution of the secular terms, a so-called *secular approximation.* We restate the master equation:

$$\frac{d\hat{\sigma}^T(t)}{dt} = -\frac{1}{2}\sum_{p,q}\left[\hat{A}_p^q, \left[\hat{A}_{-p}^{-q}, \hat{\sigma}^T(t)\right]\right] \times J\left(\omega_p^q\right) \tag{79}$$

To save space a *relaxation super-operator* can be defined as:

$$\hat{\Gamma} = \frac{1}{2}\sum_{p,q}\left[\hat{A}_p^q, \left[\hat{A}_{-p}^{-q}, \right]\right] \times J\left(\omega_p^q\right) \tag{80}$$

To better understand (79) we expand $\sigma(t)$ in terms of Liouville space basis operators $\hat{B}$. For a spin-1/2 system these are either the Cartesian basis operator set $\{\hat{1}, \hat{I}_x, \hat{I}_y, \hat{I}_z\}$ or spherical basis operator[34] set $\{\hat{I}_-, \hat{I}_+, \hat{I}^\alpha, \hat{I}^\beta\}$. The density operator can be represented as a linear combination of a complete set of $\hat{B}$s:

$$\sigma(t) = \sum_k^N b_k(t)\hat{B}_k \tag{81}$$

where $b_k(t)$ are complex coefficients. Orthogonality in operator space means:

$$\mathrm{Tr}\{\hat{B}_j^\dagger \hat{B}_k\} = \delta_{jk}\langle \hat{B}_k | \hat{B}_k \rangle \tag{82}$$

Normalization is accomplished via division by $\beta = \langle \hat{B}_k | \hat{B}_k \rangle$. From this one finally computes the matrix elements c_{jk} which represent the rate constant for the transfer of magnetization from operator $\hat{B}_j$ to $\hat{B}_k$ as:

$$c_{jk} = \frac{1}{\beta}\mathrm{Tr}\{\hat{B}_j | \hat{\Gamma} | \hat{B}_k\} \tag{83}$$

The relaxation supermatrix $\boldsymbol{R}$ (not same $\boldsymbol{R}$ as in the Bloch-equations (60)), built from c_{jk} is block diagonal when the secular approximation is used. Cavanagh *et al.* 2007, Chapter 5 explains adjustments needed outside of the secular approximation and the resulting change in appearance of (83). The matrix allows calculation of relaxation rates from multiple mechanisms, e.g. dipolar coupling, CSA.

We try to summarize this theory dense section in few words. Fast molecular processes on pico- to nanosecond time scale modulate the dipole–dipole interaction and the CSA. The magnitudes of the relaxation rates for elements of the spin density operator depend on i)

[34] $\hat{I}_-$ and $\hat{I}_+$ can be seen as the –1 and +1 coherences in (71).

the variance of the local magnetic fields induced by the time-varying Hamiltonians and ii) the time correlation function. This is the content of the master equation (79).

Spectral densities and heteronuclear relaxation rates

Using (79) and summing the contribution from dipole–dipole interaction ($R_{1,2}^{DD}$) and CSA ($R_{1,2}^{CSA}$) for a heteronuclei X (^{13}C in paper I, ^{15}N in papers II and III), in pair with ^{1}H, results in that relaxation rates can be written as linear combinations of spectral densities.

The longitudinal relaxation rate $R_1 = R_1^{DD}+R_1^{CSA}$ becomes:

$$R_1 = \frac{1}{4}d^2[J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X)] + c^2 J(\omega_X) \tag{84}$$

The transverse relaxation rate $R_2 = R_2^{DD}+R_2^{CSA} + R_{\text{ex}}$ becomes:

$$\begin{aligned} R_2 = {} & \frac{1}{8}d^2[4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H + \omega_X)] \\ & + \frac{1}{6}c^2[4J(0) + 3J(\omega_X) + R_{\text{ex}} \end{aligned} \tag{85}$$

where $d = (\mu_0 \hbar \gamma_X \gamma_H / 4\pi) \cdot \langle 1/r_{XH}^3 \rangle$; $c = \Delta\sigma \cdot \omega_X/\sqrt{3}$; μ_0 is the permeability of free space, r_{XH} is the distance between the X and ^{1}H spins; $\Delta\sigma$ is the chemical shielding anisotropy (assuming symmetric chemical shift tensor) from (55). R_{ex} is exchange contribution to R_2 and will be discussed later. Additionally, we define *the {^{1}H-X} nuclear Overhauser effect (NOE)* or cross-relaxation rate constant σ_{XH} (Overhauser, 1953) as:

$$\begin{aligned} \frac{I_{SS}}{I_{eq}} &= NOE = 1 + \frac{1}{4R_1}d^2\left(\frac{\gamma_X}{\gamma_H}\right)[6J(\omega_H + \omega_X) + J(\omega_H - \omega_X)] \\ \sigma_{XH} &= \frac{1}{4}d^2[6J(\omega_H + \omega_X) + J(\omega_H - \omega_X)] \end{aligned} \tag{86}$$

I_{SS} is the proton saturation steady state intensity, and I_{eq} is the equilibrium state intensity.

Figure 15 shows the spectral densities used in equations (84) – (86) for three τ_c-values, for an isotropic diffusion tensor at 500 MHz and with X = ^{15}N. In Figure 16 the R_1 and R_2 are shown for two magnetic field strength over the motion range 40 MHz – 1 GHz.

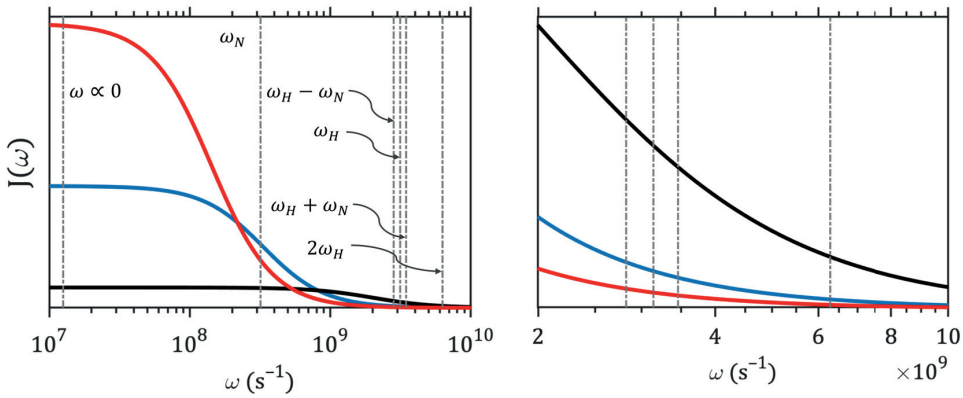


Figure 15. J(ω) at 500 MHz (11.7 T) for τ_c of 7 ns (red), 3 ns (blue) and 0.5 ns (black). These corresponds to MW of approximately 16–17 kDa (e.g. Gal3C), 6–7 kDa (e.g. PGB1) and a 0.5 kDa = 500 g/mol molecule (e.g. ligands used in paper III). The vertical bars shows position of the respective frequencies found in the expressions for R_1, R_2 and NOE, the 0-frequency bar corresponds to J(ω) ≈ J(0). The right-hand side picture zooms on the region of the high-frequency spectral densities. The dominance of J(0) becomes clear. R_1 and R_2 using these J(ω) are shown on a log-log plot in Figure 16.

A striking difference between the rate constants is that R_1 shows a maximum while R_2 does not. One way to explain the continuous increase in R_2 for larger proteins (longer τ_c) is that the dipolar interaction is efficient under longer time, leading to more loss of signal. This set a limit for protein size that can be measured with NMR and currently it is around 50 kDa (Charlier, Cousin and Ferrage, 2016).

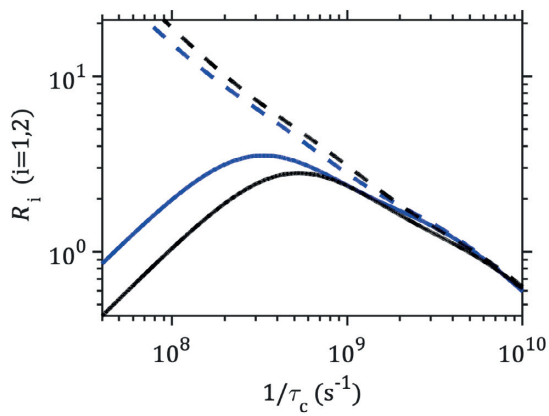


Figure 16 R_1 (solid line) and R_2 (dashed line) as a function of $1/\tau_c$, i.e. the frequency of the motion as described in (84) and (85) with dipolar and CSA-contributions. 500 MHz, 11.7 T (blue) and 800 MHz, 18.7 T (black) Higher ω_0 implies R_1-maximum at higher frequency. x-axis from 40 MHz to 1 GHz, slow motions are defined to be at the left of R_1-max and fast to the right. The R_2 is growing with slower motions and coincide with R_1 for fast motions.

The corresponding expressions for relaxation rates for deuterons, which are measured in paper III and simulated in paper II, are presented in paper II.

Chemical exchange

The relaxation rates computed by the Redfield theory are determined by stochastic processes on a time scale of pico- to nanosecond. We turn to stochastic processes on slower time scale – micro- to milliseconds – resulting in spin relaxation through modulation of isotropic chemical shifts. It is called *chemical exchange* and it is a powerful probe for characterizing the kinetics, thermodynamics, and conformational changes in proteins. Chemical exchange was recognized early in the development of NMR: "Processes inducing transitions include spin-lattice and spin-spin relaxation; also, a similar effect would be produced by an actual chemical exchange of atoms between two molecular species." (Gutowsky, McCall and Slichter, 1953). The authors modified the Bloch-equation to consider the kinetics, an approach further studied by McConnell. The equations were consequently called the Bloch-McConnell equations (McConnell, 1957). Chemical exchange is well described in reviews and text books (Sandström, 1982; Helgstrand, Härd and Allard, 2000; Bain, 2003; Woessner, 2007; Palmer, 2014; Sauerwein and Hansen, 2015).

The model used here to try to explain chemical exchange is an amino acid side chain undergoing conformational dynamics. One of the spins in the side chain is in a magnetic environment A, and moves due to conformational changes on a time scale of milliseconds to another magnetic environment B. The reaction for the spin is:

$$A \underset{k_\mathrm{b}}{\overset{k_\mathrm{f}}{\rightleftharpoons}} B \tag{87}$$

with forward (k_f) and backward (k_b) rates. The example is slightly different from the proton transfer reaction described in paper I. So the Bloch equations (60) are adjusted by adding the rate matrix $\boldsymbol{K}$:

$$\sum_{j=1}^{N} \frac{d}{dt}\boldsymbol{M}_j = \sum_{j=1}^{N} \{\gamma[\boldsymbol{M} \times \boldsymbol{B}]_j - \boldsymbol{R}(\boldsymbol{M}_j - \boldsymbol{M}_0) + \sum_{k=1}^{N} K_{jk}\boldsymbol{M}_k\} \tag{88}$$

Equation (87) is reused for every j species in the reaction, hence there are N number of M_x, M_y and M_z, and the rate matrix $\boldsymbol{K}$ is of dimension N. In our case $N = 2, j = A, B$ and the chemical reaction is written as coupled differential equations:

$$\begin{bmatrix} \frac{d[A]}{dt} \\ \frac{d[B]}{dt} \end{bmatrix} = \begin{bmatrix} -k_\mathrm{f} & k_\mathrm{b} \\ k_\mathrm{f} & -k_\mathrm{b} \end{bmatrix} \begin{bmatrix} [A](t) \\ [B](t) \end{bmatrix} \tag{89}$$

At equilibrium we have $d[A]/dt = d[B]/dt = 0$, hence $-k_\mathrm{f}[A]_\mathrm{eq} + k_\mathrm{b}[B]_\mathrm{eq} = k_\mathrm{f}[A]_\mathrm{eq} - k_\mathrm{b}[B]_\mathrm{eq}$ and $[A]_\mathrm{eq}/[B]_\mathrm{eq} = k_\mathrm{b}/k_\mathrm{f}$. The equilibrium fractional population of A is: $p_A = [A]_\mathrm{eq}/([B]_\mathrm{eq} + [A]_\mathrm{eq}) = k_\mathrm{b}/(k_\mathrm{f} + k_\mathrm{b}) = k_\mathrm{b}/k_\mathrm{ex}$, with $p_A + p_B = 1$. *The*

exchange rate is defined as $k_{\mathrm{ex}} = k_{\mathrm{f}} + k_{\mathrm{b}}$, which is the average "jump frequency" between position A and B. Scalar couplings were not considered here.

We study the transverse magnetization to see how R_2 is affected during chemical exchange. By transforming (88) to the rotating frame and look at the x- and y-components (the transverse), we get the Bloch-McConnell equations for transverse relaxation:

$$\frac{d\boldsymbol{M}^{+}(t)}{dt} = (i\boldsymbol{\Omega} - \boldsymbol{R} + \boldsymbol{K})\boldsymbol{M}^{+}(t) \quad \Leftrightarrow$$

$$\begin{bmatrix} \dfrac{dM_A^{+}(t)}{dt} \\ \dfrac{dM_B^{+}(t)}{dt} \end{bmatrix} = \begin{bmatrix} i\Omega_A - R_{2,A} - k_{\mathrm{f}} & k_{\mathrm{b}} \\ k_{\mathrm{f}} & i\Omega_B - R_{2,B} - k_{\mathrm{b}} \end{bmatrix} \begin{bmatrix} M_A^{+}(t) \\ M_B^{+}(t) \end{bmatrix} \tag{90}$$

In order to solve the system via eigenvalue decomposition we introduce $\boldsymbol{T} = i\boldsymbol{\Omega} - \boldsymbol{R} + \boldsymbol{K}$, and then have $\boldsymbol{M}^{+}(t) = e^{-\boldsymbol{T}t}\boldsymbol{M}^{+}(0) = \boldsymbol{U}^{-1}e^{-\boldsymbol{D}t}\boldsymbol{U}$ where diagonal matrix $\boldsymbol{D} = \boldsymbol{URU^{-1}}$, and $\boldsymbol{U}$ is a unitary transformation matrix. The solution, a 2×2 matrix $\boldsymbol{A}$, covers the time evolution and consists of lengthy and convoluted matrix elements a_{ij} (Chap. 5 in Cavanagh *et al,* 2007). By Fourier transformation of $M_A^{+}(t) + M_B^{+}(t) = a_{11}M_A^{+}(0) + a_{12}M_B^{+}(0) + a_{21}M_A^{+}(0) + a_{22}M_B^{+}(0)$, the free induction decay (FID) becomes as displayed in Figure 17 for different ratios of $\Delta\omega/k_{\mathrm{ex}}$.

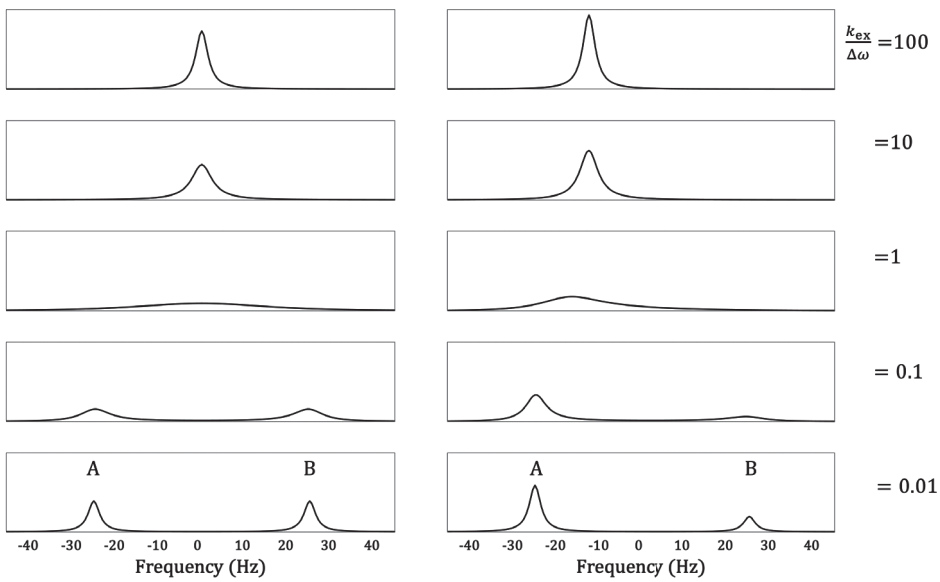


Figure 17. Two-site chemical exchange for symmetric (left – 50% of both populations) and skewed (right – 75% of A and 25% of B) cases. Five different exchange regimes are shown, where $k_{ex}/\Delta\omega = 100$ corresponds to fast, $k_{ex}/\Delta\omega = 1$ to intermediate and $k_{ex}/\Delta\omega = 0.01$ to slow regime. Calculations are based on (90) with input parameters –25 Hz and +25 Hz frequency for the respective sites A and B, hence $\Delta\omega = 2\pi 50$. $R_{2,0} = 10\ \mathrm{s}^{-1}$ for both peaks.

Chemical exchange time-regimes

There are three regimes based on the ratio $\Delta\omega/k_{\mathrm{ex}}$ where $\Delta\omega = |\Omega_A - \Omega_B|$. Slow $k_{\mathrm{ex}} \ll \Delta\omega$, intermediate $k_{\mathrm{ex}} \approx \Delta\omega$ and fast $k_{\mathrm{ex}} \gg \Delta\omega$. The terms slow, intermediate and fast, refer to the frequency of the reaction and Levitt argues it would be better to call the regimes 'infrequent' to 'frequent exchange' (Levitt, 2008).

In the slow to intermediate regime the movement of spins between position A and B we have the relationship $k_{\mathrm{ex}} \propto R_{\mathrm{ex}}$, a so-called *motional broadening*. Strictly this holds for the symmetric case, while the peaks in the general case broaden differently with $R_{\mathrm{ex},A} = k_{\mathrm{f}}$ and $R_{\mathrm{ex},A} = k_{\mathrm{b}}$. While going into the intermediate to fast regime one might be surprised to find that the exchange contribution to the relaxation rates goes in opposite direction and we see a *motional narrowing* of the two uppermost diagrams, $1/k_{\mathrm{ex}} \propto R_{\mathrm{ex}}$. This is accounted for by the frequent exchange between sites A and B resulting in that phase cannot accumulate at each site, but the resonance frequency, as well as the intrinsic relaxation are population-weighted averages, $\bar{\Omega} = p_A\Omega_A + p_A\Omega_A$ and $\overline{R_2} = p_A R_{2,A} + p_B R_{2,B}$ and we have $R_{\mathrm{ex}} = p_A p_B \Delta\omega^2/k_{\mathrm{ex}}$, which is used in paper I.

Paper IV gives examples of strong binding ligands, resulting in slow exchange processes.

To move between time regimes, one typically change the temperature, however this is easier to do for small organic molecules than for temperature sensitive large proteins. In protein NMR, the exchange effect can be modulated by varying pressure, buffer, or degree of ligand saturation.

The exchange rate constants $k(T)$ are related to the activation energies E_a through the Arrhenius equation $k(T) = \mathbb{A} \cdot \exp\{-E_a/\, N_A kT\}$ where N_A is the Avogadro constant and $\mathbb{A}$ is the pre-exponential factor which can be seen as a frequency of attempts to overcome the energy barrier.

Spin relaxation reports on protein dynamics

We turn to the application of the theory. ^{15}N relaxation is the most used technique for studying dynamic in protein NMR, it provides general hydrodynamic information (global correlation times), fast internal motion, and can detect slow motions on milli- to microsecond time scale. Before the relaxation experiments the NMR-spectroscopist must link the resonance peaks to the molecular structure and many two- and three-dimensional assignment NMR-pulse sequences are invented for this purpose.

Pulse sequences for two- and multidimensional NMR

Pulse sequences are series of radio-frequency pulses spaced in a specific way. All modern pulse sequences utilize the NMR Fourier transform (FT) technique developed by Richard Ernst and Weston Anderson in the 1960s (Ernst and Anderson, 1966; Douty, 2016). The two-dimensional NMR spectroscopy – with two frequency dimensions and one intensity – was also developed by Ernst but inspired by an idea proposed by Jean Jeener (Jeener,

1971).[35] He suggested a 2D FT-experiment consisting of two $\pi/2$ pulses, with a variable time t_1 between, and with a time t_2 measuring the time elapsed after the second pulse (Figure 18a). Ernst and co-workers recognized the power of the idea, explained it and because of this they are credited for it (Aue, Bartholdi and Ernst, 1976). The experiment is called COSY (Correlated spectroscopy[36]) and the Fourier transformation of a two-dimensional signal from time to frequency domain can be written:

$$S(\omega_1, \omega_2) = \int_0^\infty dt_1 e^{-i\omega_1 t_1} \int_0^\infty dt_2 e^{-i\omega_2 t_2}\, s(t_1, t_2) \tag{91}$$

There are an extensive number of multi-dimensional experiment for assignment, structural and dynamical studies of proteins (Sattler, Schleucher and Griesinger, 1999; Kanelis, Forman-Kay and Kay, 2001). They gained in popularity due to the invention of uniform isotopic enrichment with ^{13}C and ^{15}N. Better probes, invention of pulsed gradients, and advanced water suppression techniques aided in this development (Bax, 2011). The complexity of large proteins systems, and intrinsically disordered proteins, leads to even more spectral overlap and new techniques has to be invented (Frueh, 2014). Non-uniform sampling during acquisition and 5-dimensional pulse sequences are some of the latest inventions (Pustovalova, Mayzel and Orekhov, 2018).

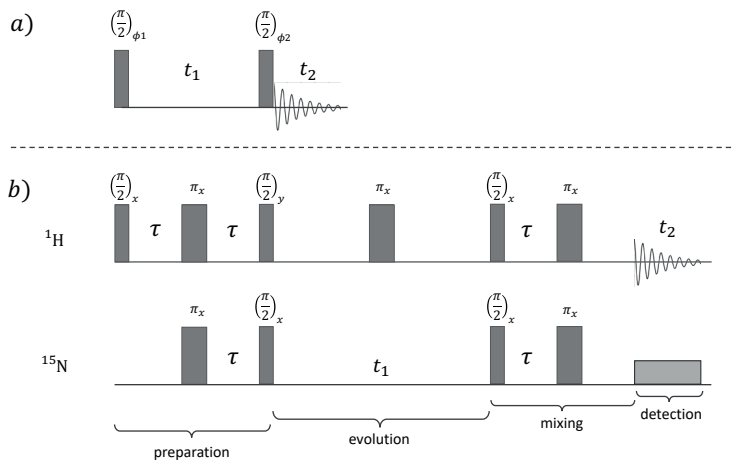


Figure 18. a) Schematic description of the simplest possible 2D experiment with an adjustable time t_1 and phases ϕ1 and ϕ2 of the 90°-pulses. b) A HSQC pulse sequence. The top set of pulses are applied to the protons and the lower set of the pulses are applied to the heteronuclear spins (^{15}N in this illustration) via a separate radio-frequency channel. Narrow bars correspond to 90°-pulses and wider bars to 180°-pulses. The delay τ is set to 1/4J in an INEPT-block. The ^{1}H at $t_1/2$, and the grey ^{15}N decoupling box during acquisition, are optional.

[35] The years around 1970 were pivotal for the NMR-technology, with the foundation of 2D and FT-spectroscopy. During the period Paul Lauterbur proposed the use of gradients to map three dimensional objects. The technology led to Magnetic resonance imaging (Keller, 2019).

[36] 'COrrelated SpectroscopY' is uninformative. Information sacrificed to achieve a fun name.

Common themes in multidimensional pulse sequences

Almost all protein NMR pulse sequences consist of four building blocks: i) preparation, ii) evolution, iii) mixing and iv) detection, as shown for the much-used HSQC-sequence (Heteronuclear Single Quantum Coherence) (Bodenhausen and Ruben, 1980) in Figure 18b. The HSQC is built on INEPT-schemes (Morris and Freeman, 1979) which result in the $I_x \rightarrow I_z S_y$, i.e. magnetization is transferred from in-phase to anti-phase.

Of importance is implementation of proper *coherence selection*. This is the process whereby the signal is systematically treated by varying the phases of RF pulses and receiver to select specific coherences and compensate for imperfections to cancel artefacts. It is done by using *phase cycles* and *field gradients* (Bodenhausen, Kogler and Ernst, 1984).

Most of the assignment experiments performed for thesis (especially for paper III) uses sequential assignment. The method combines information from triple resonance pulse sequences such as CBCANNH and CBCA(CO)NNH (Grzesiek and Bax, 1992a, 1992b). CBCANNH correlates NH group with the Cα and Cβ of its own residue (strong correlation) and of the residue preceding (weak correlation). CBCA(CO)NNH only correlates the NH group to the preceding Cα and Cβ. Assignment of the backbone can be followed up with assignment of the side chains. This is often carried out via so-called TOCSY-transfer (Total correlation spectroscopy) sequences such as HCCH-TOCSY.

The increased R_2 with protein size (Figure 16) leads to loss of sensitivity in pulse sequences during defocusing and refocusing delay required for the INEPT-transfer. Invention of the TROSY (transverse relaxation optimized spectroscopy) assists by using the interference between the CSA and dipole-dipole relaxation mechanisms and this can lead to large decrease of the R_2. TROSY-experiments are not used for this thesis.

Measuring relaxation rates

The methods for measuring the backbone and side chains R_1, R_2 and NOE in equations (84) – (86) for ^{15}N spins are well documented (Kay, Torchia and Bax, 1989; Stone *et al.*, 1992; Farrow *et al.*, 1994; Ferrage, 2012). Detailed description of the pulse sequences is given in (Korzhnev *et al.*, 2001; Cavanagh *et al.*, 2007). The pulse sequences are based on the HSQC sequence in Figure 18b with additional features, as reverse INEPT blocks in the mixing part for sensitivity enhancement by a factor of $\sim\sqrt{2}$. This is called PEP-technology (preservation of equivalent pathways) (Palmer *et al.*, 1991; Cavanagh and Rance, 1993). Additional gradients (PFG – pulsed field gradients) for coherence selection make it to a PFG-PEP-HSQC pulse sequence (Kay, Keifer and Saarinen, 1992).

The relaxation delays used for experiments in papers I–III are implemented as a CPMG-echo trains (Carr and Purcell, 1954; Meiboom and Gill, 1958) for R_2 and as an inversion recovery type block for R_1 (Vold *et al.*, 1968). The R_1 is a difference experiment as described in (Sklenář, Torchia and Bax, 1987) to get a single exponential decay. The pulse sequences for R_1 and R_2 are based on (Farrow *et al.*, 1994).

No further experimental details of pulse sequences are presented but here. The interested reader are referred to papers I–III.

Measuring methyl side-chain relaxation rates

Experiments are performed for determination of methyl-side chain order parameters, by studying methyl group deuterons in uniformly ^{15}N- and ^{13}C-labelled and fractionally (60%) ^{2}H-labelled proteins. Methyl groups are evenly spread over the protein and there are many of them, which make them an interesting target (J. Hajduk *et al.*, 2000). The experiments specifically select for the isotopomer $^{13}CH_2D$ and since relaxation of (spin-1) deuterons is heavily dominated by the quadrupolar interaction is utilized (Muhandiram *et al.*, 1995; Millet *et al.*, 2002; Skrynnikov, Millet and Kay, 2002). A drawback is the lower sensitivity (compared to ^{15}N-relaxation sequences), a consequence of the faster relaxation. However, five independently relaxing elements of the density matrix results in five unique relaxation rates, which gives better determined spectral densities. Paper II and III study methyl-side chain relaxation.

Data processing

The data processing pipeline for our relaxation studies involves: Measurements → Spectral processing using NMRPipe (Delaglio *et al.*, 1995) → Peak assignment using CcpNmr analysis (Vranken *et al.*, 2005) → Peak intensity determination using PINT (Ahlner *et al.*, 2013). Further data fitting, e.g. of exponential decaying functions has been done either in CcpNmr, PINT or in Matlab (MathWorks). In cases of spectral density fitting of ^{15}N-relaxation rates, either 'relax' (paper III) (d'Auvergne and Gooley, 2008a, 2008b) or 'ROTDIF' (paper II) (Walker, Varadan and Fushman, 2004) has been used.

Spectral density modelling and generalized order parameters

The method for analysing internal motions in proteins in two of the presented papers (papers II and III) is based on the model-free approach developed 40 years ago (Halle and Wennerström, 1981; Lipari and Szabo, 1982a, 1982b). Using several relaxation rates, preferably from different magnetic fields, the model-free method can determine different kinds of correlation times τ_i and a so-called *generalized order parameter*[37] S^2.

We first reiterate the connection between relaxation rates and dynamics in protein NMR in liquids. Each observable relaxation rate/process (as R_1, R_2 and NOE) involves transitions between quantized magnetic energy levels. The transitions depend on magnetic fields oscillating at the transition frequencies. Thus, the relaxation rates for a specific nucleus in a protein are determined by the likelihood this nucleus experience appropriate oscillating magnetic fields. The appropriate oscillations are created by stochastic fluctuations of magnetic nuclei relative to each other, or relative to the B_0 field.

[37] The generalized order parameter $\mathbb{S}$ is most of the time *squared*, $\mathbb{S}^2$, hence the correct word is *squared generalized order parameter*. There is also a normal, usual order parameter S as well. The generalized reduces to the usual order parameter when the motion is axially symmetric. This is explained in the original article (Lipari and Szabo, 1982a):
$\mathbb{S} = \langle 0.5(3\cos^2(\theta) - 1)\rangle = S$, where θ is the angle of the diffusion cone. All of $\mathbb{S}, \mathbb{S}^2, S$ and S^2 might, somewhat misleadlingly, be called 'order parameter'.

Consequently, interpretation of relaxation data gives information of molecular motion in the protein. This is exactly what the model-free approach exploits.

Simplified outline of model-free

We start with a bond vector connecting two nuclei (Clarkson, 2007). Figure 19a is a schematic picture of a NH bond vector in a protein that tumble isotropically as well as having internal motions. The translational motions of the bond, sideways and forward–backwards, are not considered. A critical assumption in the model-free analysis are that the global and the internal motions occur on different time scale and are statistically independent. Luckily for large proteins, at normal temperature this holds most of the time (Halle, 2009). The global correlation time τ_c, depends on the tumbling speed and for the isotropic case, the time correlation look like an exponential decay, i.e. $C_g(t) = \exp(-t/\tau_c)$ which is exemplified in Figure 19b for two proteins with $\tau_c \approx 7$ ns (e.g. Gal3C) and $\tau_c \approx 3$ ns (e.g. PGB1). The flexibility is modelled by a generalized order parameter S^2, in the range $0 \leq S^2 \leq 1$, where 1 is rigid and 0 means no motional constraints. $C(t)$ goes to 0 in Figure 19b. We now decouple the tumbling, imaging we can remove it, and let the internal motion only determine $C(t)$, which now refuse to decay to zero. Two cases are studied, i) a much-constrained internal motion leads to little decay (e.g. to 0.8), ii) a less constrained internal motions leads to much decay (e.g. to 0.3). Often the NH bond vector motions (15N–1H amide bond but could as well be a 13C–1H methyl axis bond vector) are described as a diffusion-in-a-cone. Mathematically the internal motion with constraints is modelled as a correlation function $C_i(t) = S^2 + (1 - S^2) \cdot \exp(-t/\tau_i)$, where τ_i stands for internal correlation time (set to 0.1 ns) and Figure 19c show case i) $S^2 = 0.8$ and ii) $S^2 = 0.3$. Mathematically we combine the two motions as:

$$C_T(\mathrm{t}) = C_g(t) \cdot C_i(t) = \left\{e^{-\frac{t}{\tau_c}}\right\}\left\{S^2 + (1 - S^2)e^{-\frac{t}{\tau_i}}\right\}; \quad \frac{1}{\tau_T} = \frac{1}{\tau_c} + \frac{1}{\tau_i} \qquad (92)$$

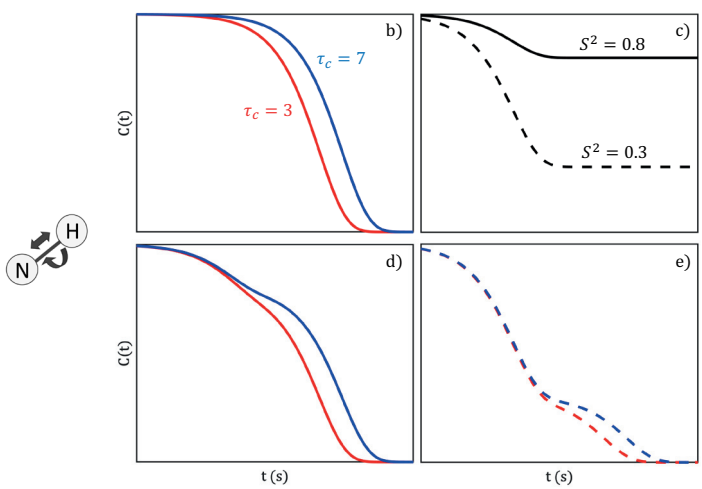


Figure 19. a) NH bond vector showing rotational as well as vibrational motions, and a schematic representation of the correlation functions. b) isotropic tumbling with two characteristic global correlation times (τ_c) of 3 ns (red) and 7 ns (blue). c) "frozen" global motions, with *only* internal motions, for S^2=0.8 (solid line) and S^2=0.3 (dashed line). Combinations of the two motional regimes in b) and c), for the two proteins with d) S^2=0.8 and e) S^2=0.3.

The result is shown for $S^2 = 0.3$ (Figure 19d) and $S^2 = 0.8$ (Figure 19e). C_T and τ_T refers to total correlation function and total correlation time, respectively. Note that for the shorter τ_c (red line) it is difficult to distinguish the "step" separating the motional modes. The Fourier transform of (92), i.e. the spectral density looks like:

$$J(\omega) = \frac{S^2\tau_c}{1+\omega^2\tau_c^2} + \frac{(1-S^2)\tau_T}{1+\omega^2\tau_T^2} \tag{93}$$

In the outline here we have not used any scaling factors.

The term model-free is used since apparently no assumptions are made about the nature of the movement. Another name is 'two-step function' (Halle and Wennerström, 1981).

The theory was extended to include motions on two different timescales (Clore *et al.*, 1990) in which the faster of the motions is parameterized by S_f^2 and τ_f and the slower by S_s^2 and τ_s, where we have $S^2 = S_f^2 S_s^2$.

Complications with the model-free formalism

The formalism is powerful, yet its founder – Lipari and Szabo – wisely restated a fundamental principle in deductive science: "models cannot be proven; they can only be eliminated" (Lipari and Szabo, 1982a). While this is true for all models, we specify few methodological drawbacks with the model-free: i) the assumption of motional regime decorrelation, i.e. $\tau_c \gg \tau_i$ obviously limits on the method. Fluctuations with frequency close to, or slower than τ_c cannot be sensed. ii) No information of the directions (e.g. up-down or sideways) of motions are provided. iii) An overestimation of the calculated entropy is provided since each individual bond in a protein is assumed to be independent (see also comments related to equation (95)). iv) Typically, in the software used, the NH bond length is fixed (1.02 Å) as well as all CSA of the ^{15}N-backbone (–172 ppm), based on CSA estimations in Ribonuclease H (Kroenke, Rance and Palmer, 1999). Average CSA of –160 ppm has also been suggested (Fushman and Cowburn, 2001). Ideally residue specific CSA should be used. Also, if exchange exist the software relax by default fit fast exchange, leading to a quadratic field strength dependence of R_{ex}, which may not be valid for real data (Millet *et al.*, 2000).

Diffusion tensor

So far, all examples assume an isotropic rotational diffusion tensor, but anisotropic diffusion is more likely since very few proteins are completely spherical. Anisotropic diffusion means the protein tumble faster in certain directions. The correlation functions (92) will be described by a sum of exponentials and so the spectral density equation (93) is rearranged as a sum:

$$J(\omega) = \frac{2}{5}\sum_{j=1}^{5} A_j \left(\frac{S^2\tau_{c,j}}{1+\omega^2\tau_{c,j}^2} + \frac{(1-S^2)\tau_{T,j}}{1+\omega^2\tau_{T,j}^2} \right) \tag{94}$$

The five A_j depends on the direction cosines of the NH vector and can be found in (Tjandra *et al.*, 1995). We have used the correct scaling factor of 2/5. Theory for diffusion tensors are explained carefully in this review article (Korzhnev *et al.*, 2001)

Linking order parameters and entropy

As mentioned in the section 'Statistical thermodynamics', that theory provides us with a powerful approach for connecting macroscopic properties, such as the Helmholtz free energy, A in equation (42) to the molecular events. The connection between S^2 and the conformational entropy (of importance especially for paper III) is described in (Akke, Brüschweiler and Palmer, 1993). A shortened version of the arguments goes as follows.

Helmholtz free energy (equated here with Gibbs energy) is $A = -kT \ln Z$, where Z is the canonical partition function (37) which is decomposed as:

$$Z = Z_U Z_O = Z_U \prod_{j=1}^{N} z_j \tag{95}$$

U refers to 'unknown contribution' to free energy and O to 'observed fluctuations', z_j are partition functions for the N individual fluctuating bonds (e.g. $N = 116$ in paper III). Since the z_j in (95) are assumed to be independent, the total contribution to free energy is overestimated, also note that we cannot be known that Z_U is uncorrelated with Z_O.

Adjustment to a reference energy level ($E_{j,\mathrm{ref}}$) leads to:

$$z_j \to \tilde{z}_j = \sum_{\nu} e^{\frac{-\Delta E_{j,\nu}}{kT}} \; ; \; \Delta E_{j,\nu} = E_{j,\nu} - E_{j,\mathrm{ref}} \tag{96}$$

$E_{j,\nu}$ is the energy for the j:th bond vector in the state (bond orientation) ν. The relationship with $\tilde{z}_j$ and the S^2 is:

$$S_j^2 = \frac{4\pi}{5} \sum_{m=-2}^{2} \left| \sum_{\nu} \frac{e^{\frac{-\Delta E_{j,\nu}}{kT}}}{\tilde{z}_j} \cdot Y_2^m(\theta_{j,\nu}, \phi_{j,\nu}) \right|^2 \tag{97}$$

$Y_2^m(\theta_{j,\nu}, \phi_{j,\nu})$ are spherical harmonics, described in most quantum mechanics textbooks, and $\theta_{j,\nu}$ and $\phi_{j,\nu}$ define the orientation of bond vector j for state ν.

Hence (97) link the S^2 to the probability distribution of orientations for the NH bond vector and from that further derivation leads to the important relationship:

$$\Delta G = -kT \sum_{j=1}^{N} \ln\left(\frac{1 - S_{j,\mathrm{final}}^2}{1 - S_{j,\mathrm{initial}}^2}\right) \tag{98}$$

which is the conformational entropy contribution to free energy upon a change from initial to final state, which as well can be two different protein complexes. For one mole the entropy change becomes $\Delta S = R \sum_k \ln\left[\left(1 - S_{\mathrm{A},k}^2\right)/\left(1 - S_{\mathrm{B},k}^2\right)\right]$.

The assumptions rely on that no “dramatic” changes have occurred during the reaction. Several different models for the bond vector movement have been investigated (Yang and Kay, 1996) and one of the conclusions made was “the conformational entropy versus order parameter profile calculated for backbone […] bond vectors […] is extremely well approximated by the profile obtained assuming a *diffusion-in-a-cone model*.”

The determination of conformational entropy changes for side-chain methyl-axis dynamics is in principle the same but presumably more advanced due to more degrees of freedom. A dictionary for calculation of entropy for the side chain of each amino acid type has been created for this purpose (Li and Brüschweiler, 2009). Paper III uses this.

The next and last theory chapter preceding the research, is about the protein systems and it has no equations. It then appears – to the author at least – as the last equation (98) connects with the derivation of the Gibbs free energy in first initial equations (1) – (12).

Protein systems

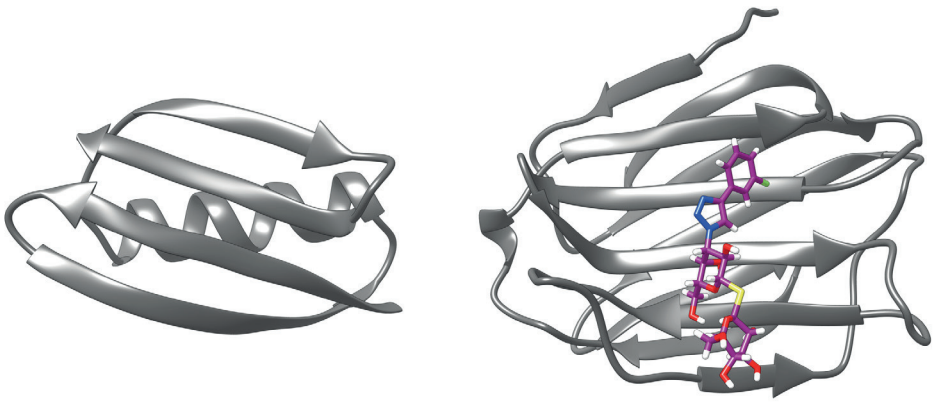


Figure 20. The two protein models studied in the thesis, none of them is an intact protein but a fragment of larger proteins, this to simplify the investigations. To the left, the 56 residue PGB1-QDD, a domain of streptococci protein G. To the right Gal3C, the 138-residue carbohydrate binding domain (CRD) of the 250-residue human galectin-3. The coloured ligand is the so-called M-ligand in Paper III. Image made in Chimera (Pettersen *et al.*, 2004).

Protein GB1

Papers I and II study the small 56 residue domain (B1-domain, 6.5 kDa) (Figure 20) of a large (58–65 kDa depending on strain) bacterial *Streptococcus* surface protein called protein G[38]. One of its siblings, Protein A, was found in the late 1950s. Protein A and Protein G belong to a group of immunoglobulin (Ig) binding bacterial proteins named with letters A, H, G, M or L. The Ig-binding activity of protein G was discovered in Lund in the 1970s (Kronvall, 1973) and the protein was given the name a decade later (Björck and Kronvall, 1984). Soon thereafter Protein G was found to bind to several human plasma proteins (Kronvall, 2013). The B1-domain[39] (PGB1) is one of three consecutive domains.

Ig's are essential components of the human immune system, which main function is to separate self from non-self. Ig's are large, around 150 kDa, and partitioned into a mobile *a*nti-gen *b*inding part, 'Fab', and a less flexible hence *c*rystallizable part, 'Fc', to which the majority of the Ig-binding bacterial proteins interact (Wikström, 1995). Protein G

[38] Putting 'protein' to the name of a protein is nowadays a tautology, but perhaps not at the time of the finding in the beginning of the protein science era.

[39] B are exchanged for a C, calling the domains C1, C2 and C3 in early publications.

belongs to the class of bacterial surface protein evolved for adhesion to both Fab and Fc and thereby weakening the hosts immune system.

Protein G consists of repetitive, stable and thermoresistant Ig-binding domains each of which can fold and function as separate units. PGB1 has reversible melting to 87° C. The topology and the extensive hydrogen-bonding network with a tightly packed and buried hydrophobic core, is claimed to be the cause for the thermal stability, which also relates to a low ΔC_P(folding) (Gronenborn *et al.*, 1991; Alexander *et al.*, 1992).

The PGB1 structure, revealed by NMR, consists of a single α-helix packed against a four-stranded β-sheet (Gronenborn *et al.*, 1991). The interaction with the IgG Fab region is mediated via an IgG β-strand which bind to the second β-strand of PGB1. The interaction PGB1 – Ig Fc-region, are eased by charged and polar side chains in the α-helix and the third β-strand. This was confirmed by NMR (Gronenborn and Clore, 1993) and crystallography (Sauer-Eriksson *et al.*, 1995).

PGB1 is considered being an excellent model for protein NMR research and there are many NMR studies on the protein. Four examples follow. Covariance studies of coupled (synchronized) motion using 10 mutants of PGB1 (Mayer *et al.*, 2003). Development of magic-angle spinning solid-state NMR methods for protein research (Trent Franks *et al.*, 2005). Relationship between structure and chemical shift (Tomlinson *et al.*, 2010). PGB1 as crowding agent (Bille *et al.*, 2019).

The model studied in the thesis have the mutations T2Q, N8D, N37D, hence the name PGB1-QDD. Threonine at position 2 in wild type is mutated to reduce N-terminal cleavage. A side effect of that mutation is an increased deamidation of the new glutamine, and to reduce that, two residues with amide side chains are mutated, N8 and N37.

Some of the information presented here is retrieved from two, not easily accessible resources, a thesis and a small conference proceeding (Wikström, 1995; Kronvall, 2013).

Galectin-3C

The carbohydrate recognition domain (CRD) of galectin-3 (Figure 20) is a thoroughly studied protein. It is the C-terminal domain of the human protein galectin-3 and consists of 138 residues (residue 113–250, 15.8 kDa). Galectin-3 (30 kDa) belongs to the large lectin family. Lectin is the generic term for carbohydrate binding proteins, i.e. proteins which bind to carbohydrates, glycoproteins, or glycolipids. These have numerous functions in biological recognition (Leffler *et al.*, 2002; Dumic, Dabelic and Flögel, 2006).

Galectins, fifteen are found, are defined by evolutionary conserved amino acid sequences and by recognition of β-galactoside structures. They are divided into three groups, the *prototype* containing one CRD, the *tandem-repeat* with two distinct but homologous CRDs, separated by a non-conserved linker sequence of 70 amino acids, and the *chimera type* with one member, galectin-3. The CRD connects to a proline- and glycine-rich N-terminal domain involved in oligomerization, through formation of a coiled-coil structure.

Galectin-3 has been found to be involved in cell–cell and cell–extracellular matrix adhesion, cell growth and differentiation, and apoptosis, angiogenesis, immune reactions, tumour growth and metastasis (Yang, Hsu and Liu, 1996; Newlaczyl and Yu, 2011). Because of this it receives attention as a potential drug target. Galectin expression is altered in malignant vs normal tissues and galectin expression can serve as an independent prognostic marker (Thijssen *et al.*, 2015). In a recent publication involvement in Alzheimer's disease and amyloid formation is shown (Boza-Serrano *et al.*, 2019).

The natural substrate of galectin-3 are β-galactosides, e.g. lactose, and these are polar molecules, hence the binding site is considerably polar with residues as arginine, histidine, glutamic acid, tryptophan, and asparagine. The drug candidates synthesized by the collaborators, are based on the lactose structure or a similar thiogalactoside. Recently two theses were presented that describes synthesis and the molecular basis of the galectin-ligand interactions for papers III–V (Peterson, 2018; Verteramo, 2019).

The Gal3C-structure was solved using X-ray crystallography (Seetharaman *et al.*, 1998) and it consists of two anti-parallel β-sheets with 5 and 6 β-strands respectivily (Figure 20). The complete galectin-3 has shown a tendency to form dimers or oligomers at higher concentration upon ligand binding (Lepur *et al.*, 2012). This is one of the reasons for using the CRD-domain, which also from an NMR-perspective is easier to study than the twice as large intact protein.

Results and conclusions

Comments to the papers are provided here. The section 'List of papers and author contributions' might be of interest in connection to this chapter, since focus here is on the author's contribution to the respective project.

This chapter is not supposed to reproduce the article abstracts, even if some information from them naturally are repeated (without citation). Here, somewhat more speculation is allowed, and few details which are not included in the articles are presented. Hopefully the most important parts are selected for a publication but a selection procedure, what to include, can always be debated. Article pictures are referenced as: Paper I: Fig. 3.

Paper I: proton transfer in PGB1

Proton transfer plays a critical role in many biological processes, such as proton pumping across membranes and enzyme catalysis. It has key roles in energy conversion processes in biological systems, including photosynthesis and respiration. Most enzymes use acid-base catalysis, where the rates of proton transfer can be rate limiting for the overall reaction. The question we tried to answer in this study was: *How fast are protons going on and off the acidic residues?* A general answer would be: k_{off} (deprotonation rate) ~ 0.1–3×10^6 s^{-1} and k_{on} (protonation rate) ~ 0.6–300×10^9 M s^{-1}.

The proton affinity is described by the pK_a-value and there are two kinds of acidic side-chains in proteins. PGB1-QDD had 12 of these. By changing pH and simultaneously measure the (fast) exchange contribution to relaxation, $R_{ex} = p_A p_B \Delta\omega / k_{ex}$, using a CPMG-experiment we could determine the rates in the pseudo first-order reaction, since the chemical shifts informed about the populations (p_A, and p_B via the pK_a) of the protonated and deprotonated states as well as the chemical shift difference $\Delta\omega$ between the states (see section 'Chemical exchange'). pH-adjustments were made using 21 separate and unbuffered NMR-samples by adding an acid (HCl) or a base (NaOH).

Measurement of proton transfer rates in proteins has not been done in this direct way before. The pulse sequence correlates the side chain ^{13}CO chemical shift with the neighbouring ^{1}Hβ/γ to yield 2D spectra (Hansen and Kay, 2011), thereby we could study the shifts close to the protonation site.

A linear free energy relationship appears between the on-rate, $\log(k_{\mathrm{on}})$ and the pK_a. More specifically a *positive* linear correlation exists between $\Delta_r G \propto -\log K$ and $\Delta_r^{\ddagger} G \propto -\log k_{\mathrm{on}}$ for the on-reaction:

$$R\text{—}COO^- + H^+ \overset{K}{\rightarrow} R\text{—}COOH; \quad K = 1/K_a \tag{99}$$

The more favourable this reaction is, the higher are the on-rates (Figure 21). This is presented in Paper I: Fig. 3 for $\log k_{\mathrm{on}}$ vs $-\log K_a$ [$= -\log(1/\mathrm{K})$]. The schematic Figure 21 below instead shows $-\log k_{\mathrm{on}}$ vs $-\log K$.

A linear-free energy relationship is not expected for the k_{off} since there are no collisions when the protons leave the side chains.

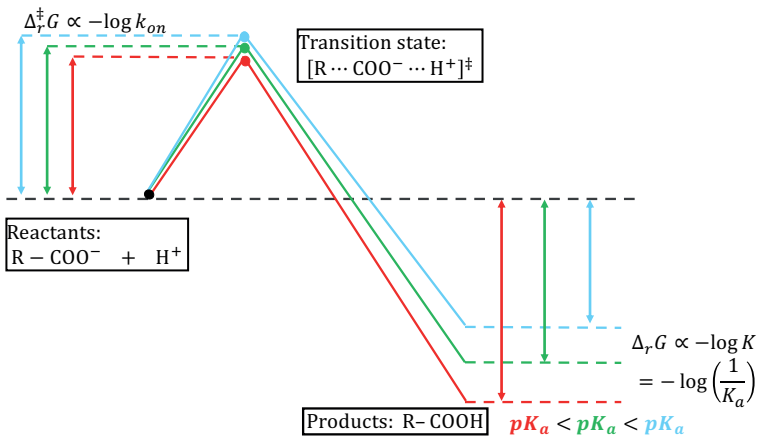


Figure 21. Schematic linear-free energy relationship, showing a reduction of the transition state barrier, due to higher k_{on}, as the reaction Gibbs free energy becomes more favourable. The reaction is described in the text. Note that $-\log(K) = -\log(1/K_a) = -\mathrm{p}K_a$.

Hydrogen bonding (see section 'Hydrogen bonds') likely explains the extreme values of k_{on} and pK_a since the four lowest pK_a: Asp22, Asp46, Asp47, and Glu56 are HB acceptors while the highest (by far) pK_a for Asp37 is a HB donor.

It could be interesting to apply the method used here, on other titrating residues (Lys, Arg and His) and possibly on advanced protein complexes.

The study was supposed to be completed with a rotating-frame spin-lock $R_{1\rho}$-experiments of the slowest exchange processes discovered in the study, for residue Asp37, pK_a = 6.5. Paper I: Fig. S6 shows a flat dispersion profile, indicating that the exchange process is faster than ~5000 s^{-1} (it was expected to be ~ 83000 s^{-1} for the sample at pH 6.7). In Lund we used a 11.7 T (500 MHz) magnet, and the idea was to exploit the higher spin-lock field strengths at Swedish NMR center using the cryogenically cooled probe heads as done previously (Ban *et al.*, 2012). The experiments failed.

Glu15 is an interesting outlier and shows an unexplained deviation between relaxation determined pK_a=4.6, and chemical shift determined pK_a=4.2 (Paper 1: Fig S4).

Correction paper I: In eq. 2, the 'tan'-function should be a 'tanh'-function.

Paper II: viscosity change of DMSO affects NMR spin relaxation

We investigated an effect expected from the theory outlined in the section 'Viscosity'. From theory we anticipated an increase in viscosity (η) due to the addition of DMSO (Dimethyl sulfoxide, $(CH_3)_2$–S=O) to water. The answer was positive. We also investigated, in detail, the consequences of this effect for certain NMR relaxation studies of protein-ligand binding.

One purpose with this control study was to create awareness of this effect and of its implication for NMR spin relaxation experiments in water–DMSO mixtures. Another was to assure the DMSO did not influence the Gal3C studies, for example by altering hydration of protein to a large extent. Many Gal3C articles use ligands dissolved in DMSO (Diehl *et al.*, 2010; Peterson *et al.*, 2018; Kumar, Ignjatović, *et al.*, 2019; Verteramo *et al.*, 2019). The reason is that DMSO is a non-toxic, good solvent for almost all organic molecules due to the relative permittivity ϵ_r=46, which is between that of water and that of non-polar solvents. The concentration of DMSO when used as solubilizer is low, 1–5% (v/v) and it is well-known to increase viscosity (η) (Cowie and Toporowski, 1961). The approximate expression for the rotational correlation time τ_c in (47) tells that η varies linearly with τ_c. However, η varies in a non-linear fashion with the DMSO-concentration (Cowie and Toporowski, 1961).

We varied the DMSO concentrations for two proteins, PGB1 (6.5 kDa) and the apo form of Gal3C (15.5 kDa). Since Galectin-3 is a drug target, it was of interest to know if DMSO affects the hydration layer. We studied the chemical shift perturbation due to DMSO, and the perturbations were small and uniform over the protein surface (Paper III: Fig. 1).

The rotational correlation time was determined using the ROTDIF-software (Walker, Varadan and Fushman, 2004) with ^{15}N R_2/R_1-ratios and the corresponding pdb-file. The R_2/R_1-ratio is a good approximation, independent of fast internal motions and of the CSA. It therefore provides a good measure for the rate at which each NH vector reorients with global tumbling τ_c (Kay, Torchia and Bax, 1989). The approach involves using the ratios from rigid NH sites only (~95% of collected data).

A quote from the paper: "The resulting values of $\tau_{c,DMSO}/\tau_{c,0}$ are 1.13, 1.05 and 1.025 for 5%, 3%, 2% and 1% (v/v) DMSO. We note that as a rule of thumb the percent increase in τ_c compared to a sample without DMSO, equals the v/v percentage of DMSO multiplied by a factor 2.5."

To perform model-free fitting of methyl side chains using ^{2}H-relaxation (paper III) τ_c is used as input parameter, usually measured with ^{15}N-backbone-relaxation experiments. How does this relate to paper II? If it is a ligand–protein study, and if DMSO is used on one of the two samples, or in varying concentrations for the ^{15}N- and ^{2}H-relaxation experiments, the DMSO-concentrations matters.

The error in the fitted methyl-axis S^2 were estimated by creating synthetic relaxation data (for four of the five relaxation rates) using an array of τ_c (reflecting 0 to 11% DMSO (v/v)) and S^2 with values between 0 and 1 in steps of 0.02. The synthetic data were used

to (re)fit the S^2, assuming a fixed τ_c which reflects a 4% (v/v) addition of DMSO (in the ^{15}N-backbone-relaxation experiment). Paper II: Fig. 4 shows that this may result in a misfit S^2 if the actual DMSO-concentration is for example 0%, or 8%. Figure 22 illustrates the workflow.

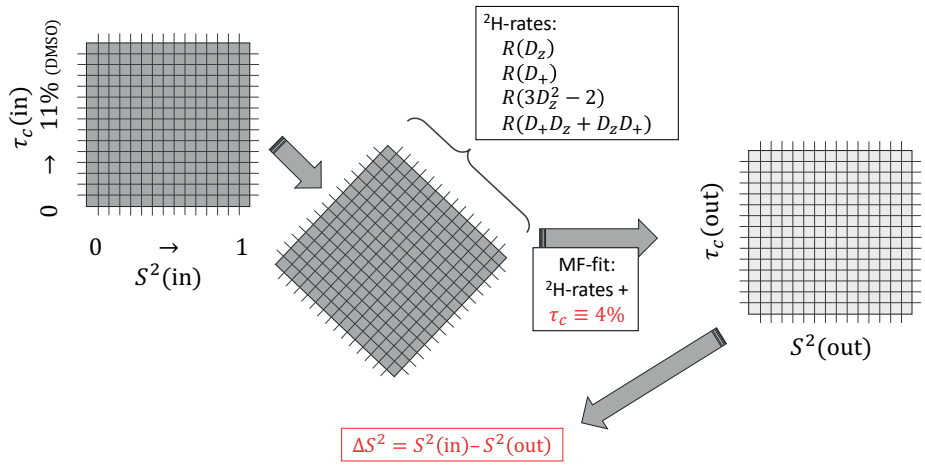


Figure 22. Workflow for the optimization (using MATLAB's 'fmincon'-function) leading to Paper II: Fig 4.

Correction paper II: Fig 4 caption: "(C) τ_c=14.2 ns", should be "(C) τ_c=15.8 ns"

Paper III: changes in conformational entropy in ligand bound galectin-3C

The study resembles two previous studies, where Gal3C is mixed with ligands followed by thorough NMR model-free analysis, along with MD-simulation, ITC and X-ray crystallography studies (Diehl *et al.*, 2010; Verteramo *et al.*, 2019). The ability of the NMR method to capture dynamics on different time scales is useful to understand conformational motions in proteins. The dynamics information is naturally suppressed in the crystallized proteins.

Three thiogalactoside ligands, differing with respect to the position of a fluorine in the ortho-, meta- or para-position on a benzene ring (called O, M and P, MW 503 g/mol) are attached to Gal3C (Paper III: Fig. 1, Fig. 3 and Fig. S1). The ligands resemble those in a recent study, not using NMR (Kumar, Misini Ignjatović, *et al.*, 2019), where the glucose is substituted for S–toluene. The substitution in this study to glucose, served to make the ligands water-soluble. Attempts were made to solubilize the S–toluene ligands in mixtures of DMSO and PEG, for NMR- and ITC-experiments, without success. In the NMR- and ITC-experiments it is of value to have few additives, and the choice of glucose-substituted ligands contribute to less systematic errors.

The ITC-experiments reveal, as expected, that the binding is enthalpically driven (Paper III: Fig. 2). The entropic contribution to binding is similar for the three complexes. An enthalpy-entropy compensation is found over the ligand series.

By running several kinds of ^{15}N- and ^{2}H spin relaxation experiments, the conformational entropy is studied, and it reveals subtle but significant differences between the complexes, especially the ^{15}N data report this. Affinity seems to show positive correlation with (backbone) mobility close to the binding site (Paper III: Fig. 7). A new metric – *a radial distribution of conformational entropy* – is implemented.

Comments to the discrepancy between NMR and MD-data

When comparing results from two methods the researcher normally prefer consistency, i.e. results should be similar, within statistical error. It is important to acknowledge that even if two independent methods give the same results, this do not infer the results are correct. Also, when results are in conflict, an opportunity of method improvement appears – either one method is wrong or both. The synergistic relationship between simulations and NMR is necessarily based on continuous evaluations of a large mass of inconsistencies found over the years when comparing the two methods (Sharp, 2019). In this work we have used standard methods in the respective fields to measure and compute the conformational entropy, and more work needs to be done to understand these standard methods. We guess on the causes for the different results for MD-simulations based on dihedral angle histogramming, and the NMR-studies, in their respective estimation of conformational entropy. Three differing outcomes are specially studied here.

i) One discrepancy is that NMR backbone S^2 shows the bound state higher in conformational entropy (more flexible), while MD shows the opposite (Paper III: Fig. S12). The same discrepancy between methods is seen in (Diehl *et al.*, 2010). A survey of 28 protein–ligand complexes shows that 10, of which 3 are Gal3C-studies, have a more rigid bound state (Caro *et al.*, 2017).

Two examples of protein–protein interaction with increased backbone dynamics upon complexation are presented in a review (Jarymowycz and Stone, 2006). i) Protein kinase A bound to the anchoring protein HT31pep (Fayos *et al.*, 2003) and ii) N-TIMP-1 bound to MMP-3 (Arumugam *et al.*, 2003). Both have an open, water-exposed binding site. A rationale behind a flexible bound state could be that the dynamics provide a means of transient sampling of states in the free energy landscape and by this improving affinity and allosteric signalling.

ii) Another difference between MD and NMR is the larger dispersion in the NMR-data for the radial distribution of entropy (Paper III: Fig. 9).

iii) The summed NMR-data (backbone and side-chain conformational entropy) shows an increased flexibility in order O → P → M, while the MD-simulations gives the opposite M→P→O (Paper III: Fig. 11). The ensemble refinement method supports, to some extent, the NMR-view in this aspect. In this comparison of the methods it is important to keep in mind that the MD simulation analyse the entropy based on fluctuations in all dihedral angles in the complex. In the protein as well as in the ligand. The NMR data also study complexes, but only subsets of bond fluctuations in the protein.

This was a challenging project/paper. A large amount of NMR-measurements are performed, about 50 2D-NMR relaxation experiments (on 500, 600 and 800 MHz, for each complex) accompanied with assignment experiments. Also, as mentioned in the section 'Spectral density modelling and generalized order parameters', the interpretation of the model-free data is not trivial. Nuclear spin relaxation studies are demanding if the goal, as in this case, is to achieve *quantitative* measures of the protein dynamics.

Paper IV: halogen bonds in drug design

This study resembles the paper III in the sense that it investigates ligand-binding on the model Gal3C based on minor changes of the ligand. Here the structure–thermodynamic relationship for halogen bonds between the ligand X=F, Cl, Br, and I substituents and the backbone carbonyl group of Gly182 in the protein was studied, based on recent findings (Zetterberg *et al.*, 2018).[40] As for paper III a multi-method approach is applied involving synthesis of ligands, X-ray crystallography, computer simulations, ITC, and NMR. The NMR-studies do not involve relaxation methods, and the entropy cannot be quantified. Relaxation experiments that will investigate dynamics are proposed.

The NMR-samples were prepared by pooling two of the three ITC-measurements into an NMR-tube, the low protein concentration (~0.06 mM) required considerably long ^{15}N-HSQC experiments of 22h.

There was large chemical shift perturbation, $\Delta\delta_{\mathrm{eff}} = \{\,[\Delta\delta(\,^{1}\mathrm{H})]^2 + [\Delta\delta(^{15}\mathrm{N})]^2\,\}^{1/2}$ around Gly182 (paper IV: Fig. 9A and paper IV: Fig. S3), one dimensional shift perturbations $\Delta\delta(^{15}\mathrm{N})$ was followed specifically. Figure 23 shows the sequence of the protein where the large perturbations are found, and it is based on data seen in paper IV: Fig. 9C and Fig. S3.

The explanation for the orange, upward arrows, representing a deshielding trend (increase in δ) of larger X (i.e. in order F → Cl → Br → I), goes as follows: The larger σ-hole along the halogen series, the more will the electrons surrounding side chain Asn174, backbone amide Glu184 and Gly182 become depleted/deshielded. The opposite trend of shielding (green downward arrow), is seen for the side chain amide of Trp181. This might be explained by the electronegative middle part of halogen bond, as predicted form theory (Clark *et al.*, 2007).

The causes of the shielding trend for backbone amide Arg183 and Trp181 are not as clear. Tentative mechanisms are presented in the manuscript section 'NMR chemical shift perturbations correlate with binding enthalpy'.

The trend for $\Delta\delta(\,^{1}\mathrm{H})$ is not as apparent. The ^{1}H-shifts are in general more difficult to interpret, as described in section 'Local diamagnetic and paramagnetic term'. Due to the pronounced effects on the chemical shifts, the model might be suitable for theoretical calculations of chemical shifts.

[40] Five α-D-thio galactopyranoside derivates which vary in one meta-substituent of a phenyl group (H, F, Cl, Br and I).

A plot of the ITC-measured enthalpy vs $\Delta\delta(^{15}\mathrm{N})$ (paper IV: Fig. 10) shows correlation for the perturbed residues around Gly182. The ITC-data also reveal enthalpy-entropy compensation. The MD simulations in this study reveals many interesting findings, for example that the solvation entropy of the water molecules around the binding site reproduces trends in the binding entropies from ITC.

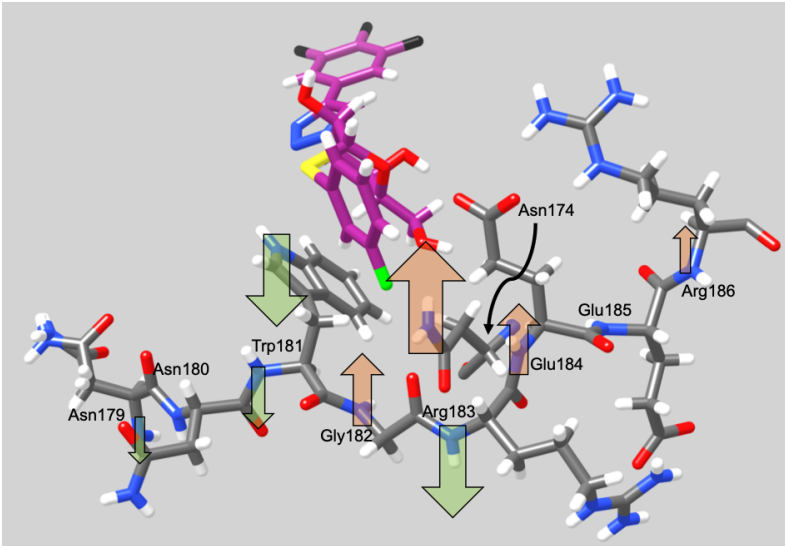


Figure 23. Image of residues 174, and 179–186 showing qualitative trends in chemical shift perturbation when changing X (green atom) in C–X bond where X=F, Cl, Br and I. Grey (protein) and magenta (ligand) are C, blue – N, red – O, white – H, yellow – S, black – F, and green – X. ^{15}N-chemical shifts give information of electron density at the site of amide nitrogen atoms. The trend in the halogen series is represented as a green downward arrow for a shielding trend – i.e. a decrease in δ, an electron surplus at that site – and as an orange upward arrow for a deshielding trend – i.e. an increase in δ, an electron depletion at that site. The arrows start at the NH and its size shows the magnitude of the trend. NH–Asn180 and NH–Glu185 did not show significant effect, due to side chains directed outwards from the binding site. Other amides without arrow lack NMR-data. Data used to generate the arrows are found in paper IV: Fig. 9C and Fig. S3.

Paper V: pH-titrations of histidines in lactose–galectin-3C complex

Here the NMR-experiments aided the neutron crystallography experiments in search for hydrogen bonding patterns in the Gal3C–ligand interaction. Three crystal structures were studied and NMR pH-titration experiments, directed on the four histidine side chains (His), were performed on the Gal3C–lactose complex. The aim was to determine the degree of protonation, and the *tautomeric* form of the histidines. This knowledge is useful for location of hydrogens in the protein. Computational studies depend on the degree of protonation, and the result from this study is used for the MD-simulations in paper III for example. The histidine imidazole ring has three tautomeric forms since two N can be protonated, Nδ1 and Nϵ2: i) The common and stable form H–Nϵ, ii) the unusual H–Nδ1, and iii) the positive form H+ with both Nδ1 and Nϵ2 protonated.

Lactose binds weakly to Gal3C with K_d = 231 μM (Diehl *et al.*, 2010) resulting in a low protein concentration due to dilution with lactose solution (Gal3C is saturated, see Figure 2.). ^{13}C-HSQC experiments were used, and two samples for the titrations, one for going down in pH (7.4 → 4.9) and one for going up (7.4 → 8.3). Unfolding of Gal3C starts at pH 5.3–5.2 and is almost complete at pH 4.9.

In general, the ^{1}H-, ^{13}C- or ^{15}N- chemical shift of any titratable residue (electrically charged side-chains) changes with pH. The largest change occurs normally closest to the site of protonation, and the change in electric environment influences the chemical shifts of all ring nuclei in a proportional manner. This means for example that the shift perturbation vs pH of any nuclei in the ring should give the same pK_a. All histidines in Gal3C (158, 208, 217 and 223) were visible in the NMR-spectra (Figure 24). Ideally all four should be traced when pH is varied, for the Cδ2-shift as well as for the Cϵ1-shift. Also, we should be able to follow the shifts in the ^{1}H- and ^{13}C-dimension and get 4×4 titration curves (^{1}Hϵ1, ^{13}Cϵ1, ^{1}Hδ2 and ^{13}Cδ2 vs pH) from one titration series of HSQC spectra. However, only His158 show δ2-shifts. His223 show them partly. We decided to rely on the ^{13}Cϵ1-shift only.

Vila et al. comment on the ^{13}Cϵ1-shift: "the ^{13}Cϵ1-nucleus is not a very sensitive indicator of the changes of forms of the imidazole ring of His" (Vila *et al.*, 2011)[41], hence information regarding which tautomer histidine possesses has to be found in combination with the ^{13}Cϵ1-shift.

His158 binds via a hydrogen bond to lactose and it is the only histidine in the binding pocket. The fact it does not titrate means it is consistently in H–Nδ1 or H–Nϵ and not in H+. The existence of the Cδ2-peak at ~127 ppm in the ^{13}C-HSQC provides evidence for the Nδ1 tautomer (Vila *et al.*, 2011). A peak at (6.5, 119) ppm appears at pH 5.88, not seen at next higher pH. This could be, what Vila et al refers to as, the '119-ppm peak', a mix of the H+ and the Nϵ2-tautomer. The peak disappears around the same pH. Quote from Paper V: "The presence of this unusual H-ND1 tautomer is consistent with the fact that NE2 of His158 is a critical H-bond acceptor for HO4 of lactose."

His208 has a well-defined pK_a = 6.2. The major tautomer at pH 7.5 (the pH for the neutron crystallography experiment) is Nϵ2, since Nϵ2 in general is favoured over Nδ1 and we see no peak ~127 ppm. At this pH there is ~5% H+. The neutron data shows two conformations of this residue.

His217 has pK_a = 5.15 ± 0.20. The major tautomer of His217 at pH 7.5 is Nϵ2, based on same arguments as for His208. At pH 7.5 there is ~1% H+.

His223 has tautomer Nϵ2 which is confirmed by the ^{15}N HSQC spectra where Nϵ2-peaks are seen. They would not appear if it was a hydrogen bond.

The comparison with neutron structures broadly agrees with the NMR-findings and it is summarized in paper V: table 3. Important findings from the neutron data is the central

[41] Besides being a good paper this article from 2011 fascinates since one of the authors, Harold Scheraga, is cited in the section 'Titration of Histidine side chains' for his NMR research on histidine side chain in 1966. Impressing 45 years between the publications on the same topic.

role of Arg162 in keeping the conserved H-bond network, since Arg162 participates in three highly directional H-bonds to lactose.

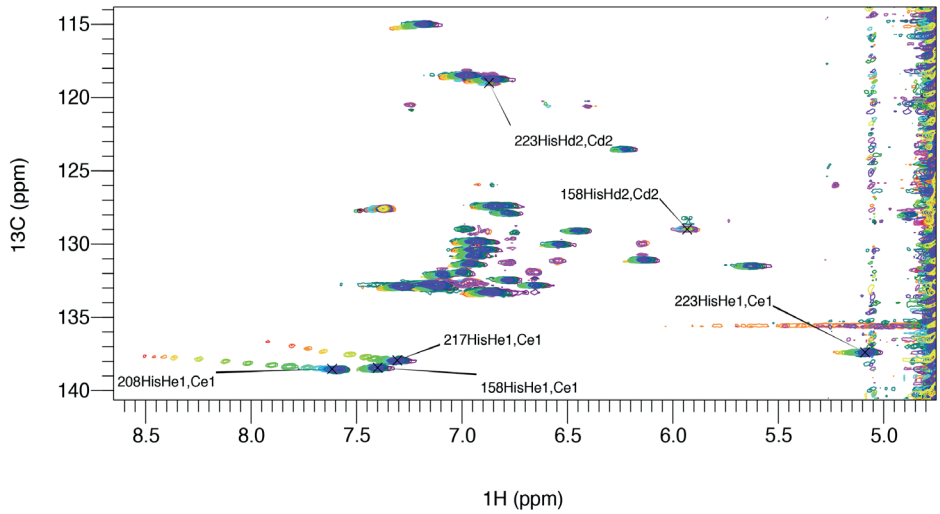


Figure 24. Lactose bound Gal3C, ct-^{13}C-HSQC in the aromatic region. Overlaid spectra for 17 pH ranging 4.9 (red) – 8.3 (purple). All four Cϵ are seen, and two Cδ. The spectra show Cϵ moves for His208 and His217, but not for His158 and His223. Both negative and positive peaks in shown the spectra. The unassigned peaks are the other aromatic residues. Ulrich Weininger did the assignment.

The last graph

Often things happen which the researcher cannot explain. When measuring the pH in the NMR-tube *after* the NMR-experiments, it appears that the lower pH-values are up-regulated when compared with the 'before'-measurement. The pH-measurement *before* was used in the publication. The last graph of the thesis (Figure 25) show this. It might be related to the unfolding process at low pH, or to something else.

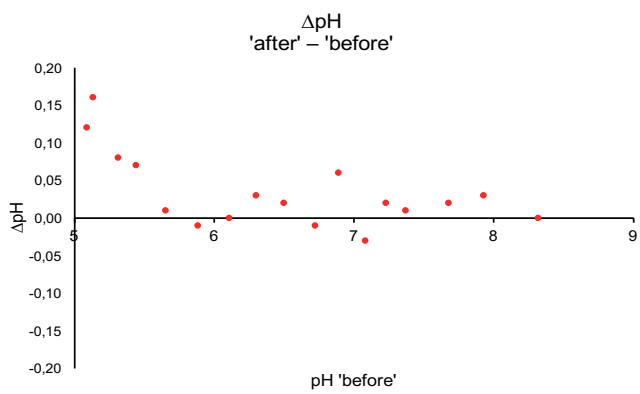


Figure 25. Comparison of pH in NMR tubes with Gal3C + lactose before and after NMR experiment. A hysteresis phenonmena is seen. pH is upregulated when measured the second time, order of days after the first.


KEMICENTRUM

Acknowledgement

First and foremost, I thank my supervisor, **Mikael Akke** for all support throughout this adventure known as doctoral studies. Thanks for accepting me as a Ph D student. Working with you, and with your group has been a rewarding experience that I will keep with me throughout my career and life. I thank my co-supervisor **Kristofer Modig** for all much-appreciated help and insights that you willingly share.

Dear colleagues and friends expected to be mentioned here since the convention propose so. Due to my age and character it would be many names. Every person I talk to (living or dead) means much to me. I decided however, since I believe in institutions – in the broadest definition – to list institutions instead of people. Strong institutions, of different kind and format, with traditions, informal and formal rules, are more important than individuals for the persistence of our society. I *prefer to believe* that people in strong institutions work for the institution more than for themselves. **Lund University** is a very strong institution founded 350 years ago. I'm a product of ideas and love delivered from the University, and I consider it *the* most important institution for this thesis. I specify other institutions:

To **Department of Biophysical Chemistry**, t h a n k s! To all highly valued colleagues at **Department of Biochemistry** at the **Center for Molecular Protein Science**, *thanks*! Before every visit to the CMPS lunchroom I've known I will meet friendly and fascinating people. I'm thankful to **Kemicentrum** as a scenic building, as the site for my undergraduate studies at **LU/LTH** 1998–2003, and as the site for my doctoral studies. I'm thankful to *all* **teachers** and **coaches** I've had. Thanks to the **DECREC** collaboration, to the rewarding **European Molecular Biology Organization** (EMBO)-courses in Basel 2017 and Berlin 2015. Thanks to **Swedish NMR Centre** for great generosity, and to **Lund Protein Production Platform** and to **Malvern Panalytical**. **Varian Associates** (deceased) and **Bruker corporation** were/are important institutions within the amazing **NMR-community**. Finally, thanks to the **Wallenberg family**, and to *all* **Swedish taxpayers** – what would I do without you?

The marriage – including all its modern varieties – is perhaps the strongest institution ever created? My wife, however, is not an institution, she is **Charlotta**, an extraordinary woman, *thank YOU* for always standing strong and beautiful by my side.

References

Ahlner, A. *et al.* (2013) 'PINT: a software for integration of peak volumes and extraction of relaxation rates', *J Biomol NMR*, 56(3), pp. 191–202.

Akke, M., Brüschweiler, R. and Palmer, A. G. (1993) 'NMR order parameters and free energy - an analytical approach and its application to cooperative calcium(2+) binding by calbindin D9k', *Journal of the American Chemical Society*. American Chemical Society, 115(21), pp. 9832–9833.

Alexander, P. *et al.* (1992) 'Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures', *Biochemistry*. American Chemical Society, 31(14), pp. 3597–3603.

Arumugam, S. *et al.* (2003) 'Increased Backbone Mobility in β-Barrel Enhances Entropy Gain Driving Binding of N-TIMP-1 to MMP-3', *Journal of Molecular Biology*, 327(3), pp. 719–734.

Atkins, P. W. and De Paula, J. (2006) *Atkins' physical chemistry / Peter Atkins, Julio de Paula*. Oxford : Oxford Univ. Press, 20068. ed.

Atkins, P. W. and De Paula, J. (2010) 'Physical Chemistry for the Life Sciences', in. Oxford University Press.

Aue, W. P., Bartholdi, E. and Ernst, R. (1976) 'Two-dimensional spectroscopy. Application to nuclear magnetic resonance', *The Journal of Chemical Physics*. American Institute of Physics, 64(5), pp. 2229–2246.

Auffinger, P. *et al.* (2004) *Halogen bonds in biological molecules*.

Bain, A. D. (2003) 'Chemical exchange in NMR', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 43(3–4), pp. 63–103.

Ban, D. *et al.* (2012) 'Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead', *Journal of Magnetic Resonance*, 221, pp. 1–4.

Bax, A. (2011) 'Triple resonance three-dimensional protein NMR: Before it became a black box', *Journal of Magnetic Resonance*, 213(2), pp. 442–445.

Becker, E. D. (2000) 'High Resolution NMR Theory and Chemical Applications, 3rd ed.', in. Academic Press, p. 424.

Bellelli, A. and Carey, J. (2017) 'Proteins with Multiple Binding Sites', in *Reversible Ligand Binding*. Chichester, UK: John Wiley & Sons, Ltd, pp. 75–115.

Ben-Naim, A. (2011) 'Entropy: Order or Information', *Journal of Chemical Education*, 88(5), pp. 594–596.

Bille, A. *et al.* (2019) 'Stability and Local Unfolding of SOD1 in the Presence of Protein Crowders', *The Journal of Physical Chemistry B*. American Chemical Society, 123(9), pp. 1920–1930.

Björck, L. and Kronvall, G. (1984) 'Purification and some properties of streptococcal protein G, a novel IgG-binding reagent.', *Journal of immunology (Baltimore, Md. : 1950)*, 133(2), pp. 969–974.

Bloch, F. (1946) 'Nuclear Induction', *Physical Review*. American Physical Society, 70(7–8), pp. 460–474.

Bloch, F., Hansen, W. W. and Packard, M. (1946) 'The Nuclear Induction Experiment', *Physical Review*. American Physical Society, 70(7–8), pp. 474–485.

Bloembergen, N., Purcell, E. M. and Pound, R. V (1948) 'Relaxation effects in nuclear magnetic resonance absorption', *Physical Review*, 73(7), pp. 679–712.

Bodenhausen, G., Kogler, H. and Ernst, R. (1984) 'Selection of coherence-transfer pathways in NMR

pulse experiments', *Journal of Magnetic Resonance (1969)*, 58(3), pp. 370–388.

Bodenhausen, G. and Ruben, D. J. (1980) 'Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy', *Chemical Physics Letters*, 69(1), pp. 185–189.

Bondar, A.-N. and White, S. H. (2012) 'Hydrogen bond dynamics in membrane protein function.', *Biochimica et biophysica acta*. NIH Public Access, 1818(4), pp. 942–50.

Bondi, A. (1964) 'van der Waals Volumes and Radii', *The Journal of Physical Chemistry*. American Chemical Society, 68(3), pp. 441–451.

Boza-Serrano, A. *et al.* (2019) 'Galectin-3, a novel endogenous TREM2 ligand, detrimentally regulates inflammatory response in Alzheimer's disease', *Acta Neuropathologica*.

Bradbury, J. H. and Scheraga, H. A. (1966) 'Structural Studies of Ribonuclease .24. Application of Nuclear Magnetic Resonance Spectroscopy to Distinguish between Histidine Residues of Ribonuclease', *Journal of the American Chemical Society*, 88(18), pp. 4240-.

Brautigam, C. A. *et al.* (2016) 'Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions', *Nat. Protocols*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 11(5), pp. 882–894.

Caro, J. A. *et al.* (2017) 'Entropy in molecular recognition by proteins', *Proceedings of the National Academy of Sciences*, 114(25), pp. 6563–6568.

Carr, H. Y. and Purcell, E. M. (1954) 'Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments', *Physical Review*, 94(3), pp. 630–638.

Cavanagh, J. *et al.* (2007) *Protein NMR spectroscopy [electronic resource] : principles and practice*. Amsterdam ; Boston : Academic Press, c2007 2nd ed.

Cavanagh, J. and Rance, M. (1993) 'Sensitivity-Enhanced NMR Techniques for the Study of Biomolecules', in Webb, G. A. (ed.) *Annual Reports on NMR Spectroscopy*. Academic Press, pp. 1–58.

Chaires, J. B. (2008) 'Calorimetry and thermodynamics in drug design', in *Annual review of biophysics*, pp. 135–151.

Chandler, D. (1987) *Introduction to Modern Statistical Mechanics*. Oxford University Press.

Chandler, D. (2005) 'Interfaces and the driving force of hydrophobic assembly', *Nature*, 437(7059), pp. 640–647.

Charlier, C., Cousin, S. F. and Ferrage, F. (2016) 'Protein dynamics from nuclear magnetic relaxation', *Chemical Society Reviews*. The Royal Society of Chemistry.

Chary, K. V. R. and Govil, G. (2008) *NMR in biological systems : from molecules to humans*. Springer.

Chodera, J. D. and Mobley, D. L. (2013) 'Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design', *Annual review of biophysics*, 42, pp. 121–142.

Clark, T. *et al.* (2007) 'Halogen bonding: the σ-hole', *Journal of Molecular Modeling*, 13, pp. 291–296.

Clarkson, M. W. (2007) *Webpage: The model-free dynamics formalism of Lipari and Szabo*.

Clore, G. M. *et al.* (1990) 'Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins', *Journal of the American Chemical Society*. American Chemical Society, 112(12), pp. 4989–4991.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) 'Protein backbone angle restraints from searching a database for chemical shift and sequence homology', *Journal of Biomolecular NMR*. Kluwer Academic Publishers, 13(3), pp. 289–302.

Cowie, J. M. and Toporowski, P. M. (1961) 'Association in Binary Liquid System Dimethyl Sulphoxide-Water', *Canadian Journal of Chemistry-Revue Canadienne De Chimie*, 39(11), pp. 2240-.

CRC Handbook of Chemistry and Physics (2018). CRC Press/Taylor & Francis, Boca Raton, FL.

d'Auvergne, E. J. and Gooley, P. R. (2008a) 'Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces', *J Biomol NMR*. Springer Netherlands, 40(2), pp. 107–119.

d'Auvergne, E. J. and Gooley, P. R. (2008b) 'Optimisation of NMR dynamic models II. A new

methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor', *J Biomol NMR*. Springer Netherlands, 40(2), pp. 121–133.

Delaglio, F. *et al.* (1995) 'NMRPipe: A multidimensional spectral processing system based on UNIX pipes', *J Biomol NMR*, 6(3), pp. 277–293.

Demirel, Y. (2014) *Nonequilibrium Thermodynamics Transport and Rate Processes in Physical, Chemical and Biological Systems*. 3rd edn. Elsevier.

Diehl, C. *et al.* (2010) 'Protein Flexibility and Conformational Entropy in Ligand Design Targeting the Carbohydrate Recognition Domain of Galectin-3', *Journal of the American Chemical Society*, 132(41), pp. 14577–14589.

Douty, K. (2016) 'NMR pioneers reflect on Silicon Valley: a conversation with Martin Packard and Weston Anderson', *Magnetic Resonance in Chemistry*, 54(10), pp. 782–786.

Dumic, J., Dabelic, S. and Flögel, M. (2006) 'Galectin-3: An open-ended story', *Biochimica et Biophysica Acta - General Subjects*, pp. 616–635.

Dyson, H. J. and Wright, P. E. (2001) 'Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states', *Methods in Enzymology*, pp. 258–270.

Ernst, R. and Anderson, W. A. (1966) 'Application of Fourier Transform Spectroscopy to Magnetic Resonance', *Review of Scientific Instruments*. American Institute of Physics, 37(1), pp. 93–102.

Eu, B. C. (2004) *Generalized Thermodynamics The Thermodynamics of Irreversible Processes and Generalized Hydrodynamicse*. Edited by A. van der Merve. Dordrecht: Kluwer Academic Publishers.

Farrow, N. A. *et al.* (1994) 'Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by N-15 NMR Relaxation', *Biochemistry*, 33(19), pp. 5984–6003.

Favro, L. D. (1960) 'Theory of the Rotational Brownian Motion of a Free Rigid Body', *Physical Review*. American Physical Society, 119(1), pp. 53–62.

Fayos, R. *et al.* (2003) 'Induction of Flexibility through Protein-Protein Interactions', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 278(20), pp. 18581–18587.

Ferrage, F. (2012) 'Protein Dynamics by 15N Nuclear Magnetic Relaxation', in Shekhtman, A. and Burz, D. S. (eds) *Protein NMR Techniques*. Totowa, NJ: Humana Press, pp. 141–163.

Ferreira de Freitas, R. and Schapira, M. (2017) 'A systematic analysis of atomic protein-ligand interactions in the PDB.', *MedChemComm*, 8(10), pp. 1970–1981.

Fersht, A. R. (1987) 'The hydrogen bond in molecular recognition', *Trends in Biochemical Sciences*. Elsevier, 12, pp. 301–304.

Fleming, P. J. and Rose, G. D. (2005) 'Do all backbone polar groups in proteins form hydrogen bonds?', *Protein science : a publication of the Protein Society*. Wiley-Blackwell, 14(7), pp. 1911–7.

Forsén, S. and Hoffman, R. A. (1963) 'Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance', *The Journal of Chemical Physics*, 39(11), pp. 2892–2901.

Fox, J. M. *et al.* (2018) 'The Molecular Origin of Enthalpy/Entropy Compensation in Biomolecular Recognition', *Annu. Rev. Biophys*, 47, pp. 223–250.

Frueh, D. P. (2014) 'Practical aspects of NMR signal assignment in larger and challenging proteins', *Progress in Nuclear Magnetic Resonance Spectroscopy*. Pergamon, 78, pp. 47–75.

Fushman, D. and Cowburn, D. (2001) 'Nuclear magnetic resonance relaxation in determination of residue-specific 15N chemical shift tensors in proteins in solution: Protein dynamics, structure, and applications of transverse relaxation optimized spectroscopy', in Thomas L. James, V. D. and Uli, S. (eds) *Methods in Enzymology*. Academic Press, pp. 109-IN1.

Geschwindner, S., Ulander, J. and Johansson, P. (2015) 'Ligand Binding Thermodynamics in Drug Discovery: Still a Hot Tip?', *Journal of Medicinal Chemistry*. American Chemical Society, 58(16), pp. 6321–6335.

Glasstone, S. (1947) *Thermodynamics For Chemists*. Van Nostrand Publisher.

Gronenborn, A. M. *et al.* (1991) 'A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G', *Science*, 253(5020), pp. 657–661.

Gronenborn, A. M. and Clore, G. M. (1993) 'Identification of the Contact Surface of a Streptococcal Protein G Domain Complexed with a Human Fc Fragment', *Journal of Molecular Biology*, 233(3), pp. 331–335.

Grzesiek, S. (2017) *Lecture: Basics of NMR relaxation and dynamics – EMBO Practical Course on NMR, Basel, August 7–12, 2017*. Basel.

Grzesiek, S. and Bax, A. (1992a) 'An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins', *Journal of Magnetic Resonance (1969)*, 99(1), pp. 201–207.

Grzesiek, S. and Bax, A. (1992b) 'Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR', *Journal of the American Chemical Society*. American Chemical Society, 114(16), pp. 6291–6293.

Gutowsky, H. S., McCall, D. W. and Slichter, C. P. (1953) 'Nuclear Magnetic Resonance Multiplets in Liquids', *The Journal of Chemical Physics*. American Institute of Physics, 21(2), pp. 279–292.

Habchi, J. *et al.* (2014) 'Introducing Protein Intrinsic Disorder', *Chemical Reviews*. American Chemical Society, 114(13), pp. 6561–6588.

Halle, B. (2009) 'The physical basis of model-free analysis of NMR relaxation data from proteins and complex fluids', *Journal of Chemical Physics*, 131(22), p. 22.

Halle, B. (ed.) (2016) *Biophysical Chemistry, course material KFK032, Lund University*. Media-Tryck, Lund University.

Halle, B. and Wennerström, H. (1981) 'Interpretation of magnetic resonance data from water nuclei in heterogeneous systems', *The Journal of Chemical Physics*. AIP, 75(4), pp. 1928–1943.

Hansen, A. L. and Kay, L. E. (2011) 'Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups', *J Biomol NMR*, 50(4), pp. 347–355.

Hansen, A. L. and Kay, L. E. (2014) 'Measurement of histidine pKa values and tautomer populations in invisible protein states', *Proceedings of the National Academy of Sciences*, 111(17), pp. E1705–E1712.

Harris, R. K. (1983) *Nuclear Magnetic Resonance Spectroscopy - A Physicochemical View*. 2nd edn. Pitman Publishing Inc.

Hassel, O. (1970) *Nobel Prize Lecture: Structural aspects of interatomic charge-transfer bonding*.

Helgstrand, M., Härd, T. and Allard, P. (2000) 'Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange', *J Biomol NMR*, 18(1), pp. 49–63.

Hemminger, W. and Sarge, S. M. (1998) 'DEFINITIONS, NOMENCLATURE, TERMS AND LITERATURE', in Brown, M. E. (ed.) *Handbook of Thermal Analysis and Calorimetry*. Elsevier Science B.V.

Hill, A. V. (1913) 'The Combinations of Haemoglobin with Oxygen and with Carbon Monoxide. I', *Biochemical Journal*. Portland Press Ltd, 7(5), p. 471.

van Holde, K. E., Johnson, W. C. and Ho, P. S. (2006) *Principles of Physical Biochemistry*. Pearson Prentice HaU.

Homans, S. W. (2007) 'Dynamics and Thermodynamics of Ligand--Protein Interactions', *Top Curr Chem*. Edited by T. Peters. Berlin, Heidelberg: Springer Berlin Heidelberg, 273, pp. 51–82.

Hore, P. J. (1995) *Nuclear Magnetic Resonance. Oxford Chemistry Primers*. Oxford University Press.

Hubbard, P. S. (1972) 'Rotational Brownian Motion', *Physical Review A*. American Physical Society, 6(6), pp. 2421–2433.

Ignjatovic, M. M. (2019) *Theoretical studies of protein-ligand binding*. Lund University.

J. Hajduk, P. *et al.* (2000) 'NMR-Based Screening of Proteins Containing 13C-Labeled Methyl Groups', *Journal of the American Chemical Society*, 122(33), pp. 7898–7904.

Jameson, C. J. and Gutowsky, H. S. (1964) 'Calculation of Chemical Shifts. I. General Formulation and

the Z Dependence', *Journal of chemical physics*, 40, p. 2310.

Jarymowycz, V. A. and Stone, M. J. (2006) 'Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences', *Chemical Reviews*. American Chemical Society, 106(5), pp. 1624–1671.

Jaynes, E. T. (1957) 'Information Theory and Statistical Mechanics', *Physical Review*, 106(4), pp. 620–630.

Jeener, J. (1971) 'Ampere International Summer School, Basko Polje, Jugoslavia, 1971.' Basko Polje: unpublished.

Jeffrey, G. A. (1997) *An Introduction to Hydrogen Bonding*. Osford University Press.

van Kampen, N. G. (2007) *Stochastic Processes in Physics and Chemistry*. 3rd edn. North Holland.

Kanelis, V., Forman-Kay, J. D. and Kay, L. E. (2001) 'Multidimensional NMR Methods for Protein Structure Determination', *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)*. John Wiley & Sons, Ltd, 52(6), pp. 291–302.

Karplus, M. (1959) 'Contact Electron-Spin Coupling of Nuclear Magnetic Moments', *The Journal of Chemical Physics*. American Institute of Physics, 30(1), pp. 11–15.

Kaupp, M. (2004) 'Interpretation of NMR Chemical Shifts', in *Calculation of NMR and EPR Parameters*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 293–306.

Kauzmann, W. (1959) 'Some Factors in the Interpretation of Protein Denaturation', *Advances in Protein Chemistry*. Academic Press, 14, pp. 1–63.

Kay, L. E., Keifer, P. A. and Saarinen, T. (1992) 'Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy With Improved Sensitivity', *Journal of the American Chemical Society*, 114(26), pp. 10663–10665.

Kay, L. E., Torchia, D. A. and Bax, A. (1989) 'Backbone dynamics of proteins as studied by n-15 inverse detected heteronuclear NMR-spectroscopy - application to staphylococcal nuclease', *Biochemistry*, 28(23), pp. 8972–8979.

Keeler, J. (2002) *Course material: Understanding NMR spectroscopy*.

Keller, S. *et al.* (2012) 'High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis', *Analytical Chemistry*. American Chemical Society, 84(11), pp. 5066–5073.

Keller, T. W. (2019) 'NMR's pivotal years: 1969–1972', *Journal of Magnetic Resonance*. Academic Press, 306, pp. 6–11.

Knight, J. (2002) 'Heat', *Science of everyday things, Vol 1 Real-Life Chemistry*. Gale Group – Thomson Learning.

Korzhnev, D. M. *et al.* (2001) 'NMR studies of Brownian tumbling and internal motions in proteins', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 38(3), pp. 197–266.

Kovermann, M., Rogne, P. and Wolf-Watz, M. (2016) 'Protein dynamics and function from solution state NMR spectroscopy', *Quarterly Reviews of Biophysics*. Cambridge University Press, 49, p. e6.

Kowalewski, J. (1990) 'Nuclear Spin Relaxation in Diamagnetic Fluids: Part 1. General Aspects and Inorganic Applications', in Webb, G. A. (ed.) *Annual Reports on NMR Spectroscopy*. Academic Press, pp. 307–414.

Kroenke, C. D., Rance, M. and Palmer, A. G. (1999) 'Variability of the 15N Chemical Shift Anisotropy in Escherichia coli Ribonuclease H in Solution', *Journal of the American Chemical Society*. American Chemical Society, 121, pp. 10119–10125.

Kronvall, G. (1973) 'A surface component in group A, C, and G streptococci with non-immune reactivity for immunoglobulin G.', *Journal of immunology (Baltimore, Md. : 1950)*, 111(5), pp. 1401–6.

Kronvall, G. (ed.) (2013) 'Perspective on Receptins and Resistance', in *Symposium at Nobel Forum Karolinska Institutet, March 5 2013*. Stockholm.

Kumar, R., Misini Ignjatović, M., *et al.* (2019) 'Structure and energetics of ligand–fluorine interactions with galectin-3 backbone and side-chain amides – insight into solvation effects and multipolar

interactions', *Chemical Science*, p. submitted.

Kumar, R., Ignjatović, M. M., *et al.* (2019) 'Structure and Energetics of Ligand–Fluorine Interactions with Galectin-3 Backbone and Side-Chain Amides: Insight into Solvation Effects and Multipolar Interactions', *ChemMedChem*. John Wiley & Sons, Ltd, 14(16), pp. 1528–1536.

Ladbury, J. E., Klebe, G. and Freire, E. (2010) 'Adding calorimetric data to decision making in lead discovery: a hot tip', *Nat Rev Drug Discov*. Nature Publishing Group, 9(1), pp. 23–27.

Lamb, W. E. (1941) 'Internal Diamagnetic Fields', *Physical Review*. American Physical Society, 60(11), pp. 817–819.

Leffler, H. *et al.* (2002) 'Introduction to galectins', *Glycoconjugate Journal*, 19(7–9), pp. 433–440.

Leiter, D. J. and Leiter, S. (2003) *A to Z of physicists*. Facts On File.

Lepur, A. *et al.* (2012) 'Ligand induced galectin-3 protein self-association.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 287(26), pp. 21751–6.

Levine, I. N. (2009) *Physical chemistry*. 6th edn. McGraw-Hill.

Levitt, M. H. (2008) *Spin Dynamics: Basics of Nuclear Magnetic Resonance Second Edition*. John Wiley & Sons Ltd.

Li, D. W. and Brüschweiler, R. (2009) 'A Dictionary for Protein Side-Chain Entropies from NMR Order Parameters', *Journal of the American Chemical Society*, 131(21), pp. 7226–7227.

Lindman, S. *et al.* (2006) 'Electrostatic Contributions to Residue-Specific Protonation Equilibria and Proton Binding Capacitance for a Small Protein†', *Biochemistry*. American Chemical Society, 45(47), pp. 13993–14002.

Lipari, G. and Szabo, A. (1982a) 'Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules . 1. Theory and range of validity', *Journal of the American Chemical Society*, 104(17), pp. 4546–4559.

Lipari, G. and Szabo, A. (1982b) 'Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results', *Journal of the American Chemical Society*, 104(17), pp. 4559–4570.

Löw, C. *et al.* (2008) 'Structural insights into an equilibrium folding intermediate of an archaeal ankyrin repeat protein.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(10), pp. 3779–84.

Luginbuhl, P. and Wüthrich, K. (2002) 'Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 40(3), pp. 199–247.

Ma, B. *et al.* (2002) 'Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations.', *Protein science : a publication of the Protein Society*. Wiley-Blackwell, 11(2), pp. 184–97.

Markley, J. L. (1975) 'Observation of histidine residues in proteins by nuclear magnetic resonance spectroscopy', *Accounts of Chemical Research*. American Chemical Society, 8(2), pp. 70–80.

Mayer, K. L. *et al.* (2003) 'Covariation of backbone motion throughout a small protein domain', *Nature Structural & Molecular Biology*. Nature Publishing Group, 10(11), pp. 962–965.

McCammon, A. and Harvey, S. (1987) 'Dynamics in proteins and nucleic acids', in. Cambridge University Press.

McConnell, H. M. (1957) 'Theory of Nuclear Magnetic Shielding in Molecules. I. Long-Range Dipolar Shielding of Protons', *The Journal of Chemical Physics*. American Institute of Physics, 27(1), pp. 226–229.

Meiboom, S. and Gill, D. (1958) 'Modified Spin-Echo Method for Measuring Nuclear Relaxation Times', *Review of Scientific Instruments*, 29(8), pp. 688–691.

Millet, O. *et al.* (2000) 'The Static Magnetic Field Dependence of Chemical Exchange Linebroadening Defines the NMR Chemical Shift Time Scale', *Journal of the American Chemical Society*, 122(12), pp. 2867–2877.

Millet, O. *et al.* (2002) 'Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in C-13-labeled and fractionally H-2-enriched proteins in solution', *Journal of the American Chemical Society*, 124(22), pp. 6439–6448.

Mizukami, T., Sakuma, Y. and Maki, K. (2016) 'Statistical Mechanical Model for pH-Induced Protein Folding: Application to Apomyoglobin', *The Journal of Physical Chemistry B*. American Chemical Society, 120(34), pp. 8970–8986.

Moncrief, J. W. (2000) 'Antoine Laurent Lavoisier', *Science and Its Times*. Gale Group – Thomson Learning.

Morowitz, H. (1978) *Foundations of Bioenergetics*. Academic Press.

Morris, G. A. and Freeman, R. (1979) 'Enhancement of nuclear magnetic-resonance signals by polarization transfer', *Journal of the American Chemical Society*, 101(3), pp. 760–762.

Muhandiram, D. R. *et al.* (1995) 'Measurement of 2H T1 and T1.rho. Relaxation Times in Uniformly 13C-Labeled and Fractionally 2H-Labeled Proteins in Solution', *Journal of the American Chemical Society*. American Chemical Society, 117(46), pp. 11536–11544.

Müller, I. (2007) *A history of thermodynamics : the doctrine of energy and entropy*. Springer.

Müller, K., Faeh, C. and Diederich, F. (2007) 'Fluorine in pharmaceuticals: Looking beyond intuition', *Science*, pp. 1881–1886.

Myszka, D. G. *et al.* (2003) 'The ABRF-MIRG'02 Study: Assembly State, Thermodynamic, and Kinetic Analysis of an Enzyme/Inhibitor Interaction', *Journal of Biomolecular Techniques : JBT*. The Association of Biomolecular Resource Facilities, 14(4), pp. 247–269.

Nash, L. K. (1966) *Elements of Statistical Thermodynamics*. Addison-Wesley Publishing Company.

Newberry, R. W. and Raines, R. T. (2016) 'A prevalent intraresidue hydrogen bond stabilizes proteins', *Nature Chemical Biology*, 12(12), pp. 1084–1088.

Newlaczyl, A. U. and Yu, L.-G. (2011) 'Galectin-3 – A jack-of-all-trades in cancer', *Cancer Letters*. Elsevier, 313(2), pp. 123–128.

Nicholas, M. P. *et al.* (2010) 'Nuclear spin relaxation in isotropic and anisotropic media', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 57(2), pp. 111–158.

Nielsen, J. E. (2008) 'Analyzing Protein NMR pH-Titration Curves', in Ralph, A. W. and David, C. S. (eds) *Annual Reports in Computational Chemistry*. Elsevier, pp. 89–106.

Nodet, G. and Abergel, D. (2007) 'An overview of recent developments in the interpretation and prediction of fast internal protein dynamics', *European Biophysics Journal*. Springer-Verlag, 36(8), pp. 985–993.

Norrsell, U. (2000) 'Magnus Gustaf Blix (1849-1904); Neurophysiological, Physiological, and Engineering Virtuoso', *Journal of the History of the Neurosciences*, 9(3), pp. 238–249.

Olejniczak, M. *et al.* (2012) 'A simple scheme for magnetic balance in four-component relativistic Kohn–Sham calculations of nuclear magnetic resonance shielding constants in a Gaussian basis', *The Journal of Chemical Physics*. American Institute of Physics, 136(1), p. 14108.

Olsson, T. S. G. *et al.* (2008) 'The Thermodynamics of Protein–Ligand Interaction and Solvation: Insights for Ligand Design', *Journal of Molecular Biology*. Academic Press, 384(4), pp. 1002–1017.

Onsager, L. (1931) 'Reciprocal Relations in Irreversible Processes. II.', *Physical Review*. American Physical Society, 38(12), pp. 2265–2279.

Overhauser, A. W. (1953) 'Polarization of Nuclei in Metals', *Physical Review*. American Physical Society, 92(2), pp. 411–415.

Pace, C. N. (1990) 'Conformational stability of globular proteins', *Trends in Biochemical Sciences*, 15(1), pp. 14–17.

Palmer, A. G. *et al.* (1991) 'Sensitivity Improvement In Proton-Detected 2-Dimensional Heteronuclear Correlation NMR-Spectroscopy', *Journal of Magnetic Resonance*, 93(1), pp. 151–170.

Palmer, A. G. (2004) 'NMR characterization of the dynamics of biomacromolecules', *Chemical Reviews*, 104(8), pp. 3623–3640.

Palmer, A. G. (2014) ‘Chemical exchange in biomacromolecules: Past, present, and future’, *Journal of Magnetic Resonance*, 241, pp. 3–17.

Pauling, L., Corey, R. B. and Branson, H. R. (1951) ‘The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 37(4), pp. 205–11.

Paulini, R., Müller, K. and Diederich, F. (2005) ‘Orthogonal Multipolar Interactions in Structural Chemistry and Biology’, *Angewandte Chemie International Edition*. John Wiley & Sons, Ltd, 44(12), pp. 1788–1805.

Pelton, J. G. *et al.* (1993) ‘Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques’, *Protein Sci*, 2(4), pp. 543–558.

Perozzo, R., Folkers, G. and Scapozza, L. (2004) ‘Thermodynamics of protein-ligand interactions: History, presence, and future aspects’, *Journal of Receptors and Signal Transduction*, 24(1–2), pp. 1–52.

Persson, E. and Halle, B. (2008) ‘Nanosecond to Microsecond Protein Dynamics Probed by Magnetic Relaxation Dispersion of Buried Water Molecules’, *Journal of the American Chemical Society*. American Chemical Society, 130, pp. 1774–1787.

Pervushin, K. *et al.* (1997) ‘Attenuated T2 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.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 94(23), pp. 12366–71.

Peterson, K. (2018) *Molecular basis for galectin-ligand interactions : Design, synthesis and analysis of ligands*. Lund University.

Peterson, K. *et al.* (2018) ‘Systematic Tuning of Fluoro-galectin-3 Interactions Provides Thiodigalactoside Derivatives with Single-Digit nM Affinity and High Selectivity’, *Journal of Medicinal Chemistry*. American Chemical Society, 61(3), pp. 1164–1175.

Pettersen, E. F. *et al.* (2004) ‘UCSF chimera - A visualization system for exploratory research and analysis’, *Journal of Computational Chemistry*, 25(13), pp. 1605–1612.

Price, W. S. (2009) *NMR studies of translational motion*. Cambridge University Press.

Proctor, W. G. and Yu, F. C. (1950) ‘The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound’, *Physical Review*. American Physical Society, 77(5), pp. 717–717.

Purcell, E. M., Torrey, H. C. and Pound, R. V (1946) ‘Resonance Absorption by Nuclear Magnetic Moments in a Solid’, *Physical Review*. American Physical Society, 69(1–2), pp. 37–38.

Pustovalova, Y., Mayzel, M. and Orekhov, V. Y. (2018) ‘XLSY: Extra-Large NMR Spectroscopy’, *Angewandte Chemie International Edition*. John Wiley & Sons, Ltd, 57(43), pp. 14043–14045.

Rabi, I. I. *et al.* (1938) ‘A New Method of Measuring Nuclear Magnetic Moment’, *Physical Review*. American Physical Society, 53(4), p. 318.

Ramsey, N. F. (1950a) ‘Magnetic shielding of nuclei in molecules’, *Physical Review*, 78(6), pp. 699–703.

Ramsey, N. F. (1950b) ‘The Internal Diamagnetic Field Correction in Measurements of the Proton Magnetic Moment’, *Physical Review*. American Physical Society, 77(4), p. 567.

Redfield, A. G. (1957) ‘On the Theory of Relaxation Processes’, *Ibm Journal of Research and Development*, 1(1), pp. 19–31.

Richarz, R. and Wüthrich, K. (1978) ‘Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly- X- L - Ala-OH’, *Biopolymers*. John Wiley & Sons, Ltd, 17(9), pp. 2133–2141.

Rule, G. S. and Hitchens, K. T. (2006) *Fundamentals of Protein NMR Spectroscopy [electronic resource] / by Gordon S. Rule, T. Kevin Hitchens*, *Focus on Structural Biology: 5*. Dordrecht : Springer, 2006.

Sandström, J. (1982) *Dynamic NMR spectroscopy*. Academic Press: New York.

Sattler, M., Schleucher, J. and Griesinger, C. (1999) 'Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34(2), pp. 93–158.

Sauer-Eriksson, A. E. *et al.* (1995) 'Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG.', *Structure (London, England : 1993)*, 3(3), pp. 265–78.

Sauerwein, A. C. and Hansen, D. F. (2015) 'Relaxation Dispersion NMR Spectroscopy', in Berliner, L. (ed.) *Protein NMR: Modern Techniques and Biomedical Applications*. Boston, MA: Springer US, pp. 75–132.

Scholfield, M. R. *et al.* (2017) 'Structure–Energy Relationships of Halogen Bonds in Proteins', *Biochemistry*, 56(22), pp. 2794–2802.

Schulz, G. E. and Schirmer, R. H. (1979) *Principles of protein structure*. Springer-Verlag.

Seetharaman, J. *et al.* (1998) 'X-ray Crystal Structure of the Human Galectin-3 Carbohydrate Recognition Domain at 2.1-Å Resolution', *Journal of Biological Chemistry*, 273(21), pp. 13047–13052.

Sharp, K. (2019) 'Companion Simulations and Modeling to NMR-Based Dynamical Studies of Proteins', *Methods in Enzymology*. Academic Press, 615, pp. 1–41.

Sharp, K. and Matschinsky, F. (2015) 'Translation of Ludwig Boltzmann's Paper "On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium"', *Entropy*, 17, pp. 1981–2009.

Shen, Y. *et al.* (2008) 'Consistent blind protein structure generation from NMR chemical shift data.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(12), pp. 4685–90.

Sklenář, V., Torchia, D. and Bax, A. (1987) 'Measurement of carbon-13 longitudinal relaxation using 1H detection', *Journal of Magnetic Resonance (1969)*. Academic Press, 73(2), pp. 375–379.

Skrynnikov, N. R., Millet, O. and Kay, L. E. (2002) 'Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions', *Journal of the American Chemical Society*, 124(22), pp. 6449–6460.

Slater, J. C. (1939) *Introduction to Chemical Physics*. 1st edn. McGraw-Hill Book Company.

Sørensen, O. W. *et al.* (1983) 'Product operator formalism for the description of NMR pulse experiments', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 16(0), pp. 163–192.

Spera, S. and Bax, A. (1991) 'Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts', *Journal of the American Chemical Society*. American Chemical Society, 113(14), pp. 5490–5492.

Stone, M. J. *et al.* (1992) 'Backbone dynamics of the Bacillus subtilis glucose permease IIA domain determined from nitrogen-15 NMR relaxation measurements', *Biochemistry*, 31(18), pp. 4394–4406.

Takeuchi, K., Baskaran, K. and Arthanari, H. (2019) 'Structure determination using solution NMR: Is it worth the effort?', *Journal of Magnetic Resonance*. Academic Press, 306, pp. 195–201.

Taylor, J. C. (2001) *Hidden unity in nature's laws*. Cambridge University Press.

Tellinghuisen, J. and Chodera, J. D. (2011) 'Systematic errors in isothermal titration calorimetry: Concentrations and baselines', *Analytical Biochemistry*, 414(2), pp. 297–299.

Thijssen, V. L. *et al.* (2015) 'Galectin expression in cancer diagnosis and prognosis: A systematic review', *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. Elsevier, 1855(2), pp. 235–247.

Tjandra, N. *et al.* (1995) 'Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation', *Journal of the American Chemical Society*, 117(50), pp. 12562–12566.

Tomlinson, J. H. *et al.* (2010) 'Structural origins of pH-dependent chemical shifts in the B1 domain of protein G', *Proteins: Structure, Function, and Bioinformatics*. Wiley Subscription Services, Inc., A Wiley Company, 78(14), pp. 3000–3016.

Trent Franks, W. *et al.* (2005) 'Magic-Angle Spinning Solid-State NMR Spectroscopy of the β1 Immunoglobulin Binding Domain of Protein G (GB1): 15N and 13C Chemical Shift Assignments and Conformational Analysis', *Journal of the American Chemical Society*, 127(35), pp. 12291–12305.

Vallurupalli, P., Bouvignies, G. and Kay, L. E. (2012) 'Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation', *Journal of the American Chemical Society*. American Chemical Society, 134(19), pp. 8148–8161.

van de Ven, F. J. M. (1995) *Multidimensional NMR in Liquids Basic Principles and Experimental Methods*. WILEY-VCH Verlag.

Verteramo, M. L. (2019) *Design, Synthesis and Thermodynamic Studies of Galectin Ligands*. Lund University.

Verteramo, M. L. *et al.* (2019) 'Interplay between Conformational Entropy and Solvation Entropy in Protein-Ligand Binding', *J. Am. Chem. Soc*. UTC, 141, pp. 2012–2026.

Vila, J. A. *et al.* (2011) 'Assessing the fractions of tautomeric forms of the imidazole ring of histidine in proteins as a function of pH', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(14), pp. 5602–5607.

Vold, R. L. *et al.* (1968) 'Measurement of Spin Relaxation in Complex Systems', *The Journal of Chemical Physics*. American Institute of Physics, 48(8), pp. 3831–3832.

Vranken, W. F. *et al.* (2005) 'The CCPN data model for NMR spectroscopy: development of a software pipeline', *Proteins-Structure Function and Bioinformatics*, 59, pp. 687–696.

Walker, O., Varadan, R. and Fushman, D. (2004) 'Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from N-15 relaxation data using computer program ROTDIF', *Journal of Magnetic Resonance*, 168(2), pp. 336–345.

Wand, A. J. and Sharp, K. (2018) 'Measuring Entropy in Molecular Recognition by Proteins', *Annu. Rev. Biophys.*, 47.

Wangsness, R. K. and Bloch, F. (1953) 'The Dynamical Theory of Nuclear Induction', *Physical Review*, 89(4), pp. 728–739.

Wertz, J. (1955) 'Nuclear And Electronic Spin Magnetic Resonance', *Chemical Reviews*, 55(5), pp. 829–955.

Wetlaufer, D. B. (1973) 'Nucleation, rapid folding, and globular intrachain regions in proteins.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 70(3), pp. 697–701.

Wikström, M. (1995) *Biophysical Studies on Immunoglobulin-binding Bacterial Surface Proteins*. Lund University.

Williamson, M. P. (2013) 'Using chemical shift perturbation to characterise ligand binding', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73, pp. 1–16.

Wishart, D. S. *et al.* (1995) '1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects', *J Biomol NMR*, 5(1), pp. 67–81.

Wishart, D. S. and Case, D. A. (2002) 'Use of chemical shifts in macromolecular structure determination', in Thomas L. James, V. D. U. S. (ed.) *Methods in Enzymology*. Academic Press, 33–34.

Woessner, D. E. (1962) 'Nuclear Spin Relaxation in Ellipsoids Undergoing Rotational Brownian Motion', *The Journal of Chemical Physics*, 37(3), pp. 647–654.

Woessner, D. E. (2007) 'Relaxation Effects of Chemical Exchange', in *eMagRes*. John Wiley & Sons.

Yang, D. W. and Kay, L. E. (1996) 'Contributions to Conformational Entropy Arising from Bond Vector Fluctuations Measured from NMR-Derived Order Parameters: Application to Protein Folding', *Journal of Molecular Biology*, 263(2), pp. 369–382.

Yang, R. Y., Hsu, D. K. and Liu, F. T. (1996) 'Expression of galectin-3 modulates T-cell growth and apoptosis.', *Proceedings of the National Academy of Sciences*, 93(13), pp. 6737–6742.

Zetterberg, F. R. *et al.* (2018) 'Monosaccharide Derivatives with Low-Nanomolar Lectin Affinity and High Selectivity Based on Combined Fluorine-Amide, Phenyl-Arginine, Sulfur-π, and Halogen Bond Interactions', *ChemMedChem*, 13(2), pp. 133–137.

Zhao, H., Piszczek, G. and Schuck, P. (2015) 'SEDPHAT – A platform for global ITC analysis and global multi-method analysis of molecular interactions', *Methods*, 76, pp. 137–148.

The thesis investigates protein dynamics using nuclear magnetic resonance. It is a continuation of ideas proposed by Empedocles of Akragas (492–432 BC), who suggested universe was built from four elements, air, earth, fire and water, as the painting shows. Aristotle (384–322 BC) added the fifth, *the quintessence*, or the aether – of divine nature – coloured white.

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

ISBN 978-91-7422-694-2

Biophysical Chemistry
Department of Chemistry
Faculty of Engineering (LTH)
Lund University

9 789174 226942