



# LUND UNIVERSITY

## Perplexing Protein Puzzles

Lindman, Stina

2010

[Link to publication](#)

*Citation for published version (APA):*

Lindman, S. (2010). *Perplexing Protein Puzzles*. [Doctoral Thesis (compilation), Biophysical Chemistry]. Department of Biophysical Chemistry, Lund University.

*Total number of authors:*

1

### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

# Perplexing Protein Puzzles

Stina Lindman

Doctoral thesis  
Department of Biophysical Chemistry  
Lund University, 2010



**LUND**  
UNIVERSITY

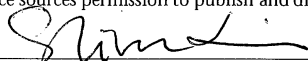
Akademisk avhandling för avläggande av filosofie doktorsexamen vid tekniska fakulteten vid Lunds universitet, att offentligen försvaras i hörsal B, Kemacentrum, fredagen den 26 februari 2010, klockan 10.00. Fakultetsopponent är Professor Torleif Härd, Sveriges lantbruksuniversitet, Uppsala, Sverige.

Organization LUND UNIVERSITY  Department of Biophysical Chemistry P.O. Box 124 SE-22100 Lund, Sweden		Document name DOCTORAL DISSERTATION	
		Date of issue 2010-01-18	
Author(s) Stina Lindman		Sponsoring organization The Research School in Pharmaceutical Sciences (FLÄK)	
Title and subtitle Perplexing Protein Puzzles			
Abstract <p>Protein structure and stability are inherent in the amino acid sequence and governed by non-covalent interactions. The cooperation between forces is, however, perplexing and not well understood. In order to elucidate and predict protein folding and stability, detailed studies of non-covalent interactions are required.</p> <p>Presented in this thesis are detailed studies on electrostatic interactions in proteins and their contribution to protein stability for PGB1, the protein G B1 domain from Streptococcus sp. Electrostatic interactions were investigated both by means of thermal and chemical denaturation at various pH and salt concentrations using CD spectroscopy and differential scanning calorimetry, and by site-specific pKa-value determination using heteronuclear NMR spectroscopy. Investigations of electrostatic interactions both in the folded and unfolded states of the protein were conducted. The results show strong electrostatic coupling between charges in the native protein that was not efficiently screened by salt. In the unfolded state most charge-charge interactions were reduced. Novel methods in analyzing NMR titration data are introduced and used to reveal electrostatic coupling. Moreover, the relative contribution of pKa-shift arising from direct Coulomb interactions, hydrogen bonding and desolvation was elucidated. Using experimentally determined pKa-values of the folded and unfolded states the pH dependent stability could be accurately calculated and compared to denaturation studies.</p> <p>This thesis moreover contains results from research aiming at stabilizing proteins. The same non-covalent interactions govern the stability of an intact protein chain and the affinity between fragments of the chain, which is made use of in the novel method. Reconstituting fragments of a protein were cloned into a split-GFP system where assembly of fused fragments was visualized by green fluorescence. It was shown that the fluorescence intensity correlated to the affinity between fused fragments. In a subsequent study the fluorescence intensity was used to screen for stabilized variants from a small, focused library of PGB1. Increased green fluorescence revealed enhanced stability of intact chain as observed from denaturation studies. Based on thermodynamic principles and in vivo screening the method shows promising results in screening of larger libraries for optimization of protein stability.</p>			
Key words: Electrostatic interactions; protein stability; protein reconstitution; pKa-values; protein G B1; split-GFP; calbindin D9k; protein denaturation; heteronuclear NMR; pH titrations			
Classification system and/or index terms (if any):			
Supplementary bibliographical information:		Language English	
ISSN and key title:		ISBN 978-91-628-8020-0	
Recipient's notes		Number of pages 197	Price
		Security classification	

Distribution by (name and address)

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature



Date

2010-01-18

# Perplexing Protein Puzzles

Stina Lindman

Doctoral thesis  
Department of Biophysical Chemistry  
Lund University, 2010



**LUND**  
UNIVERSITY

Front cover: PGB1 illustrated as a perplexing jigsaw puzzle

Supervisor: Sara Linse

Examination committee: Professor Bo Jönsson  
Dept. of Theoretical Chemistry, Lund University

Professor Cecilia Emanuelsson  
Dept. of Biochemistry, Lund University

Assistant Professor Birthe Kragelund  
Dept. of Biology, University of Copenhagen

**Perplexing Protein Puzzles**

Copyright © 2010 Stina Lindman

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

ISBN 978-91-628-8020-0

Printed by Media-Tryck, Lund 2010

## PREFACE

Ever since childhood I loved jigsaw puzzles; putting a piece in the right place is very rewarding. The same feeling applies to finding pieces in the protein puzzle. Unlike a jigsaw puzzle the number of pieces to complete the protein puzzle is undefined and there is no template to follow. Rather, the more we understand about proteins the more there is to study. The title of this thesis is broad and hence it is impossible to provide an answer to all questions it imposes. Therefore, this thesis will be far from completing the protein puzzle; nevertheless I feel that interesting results have been obtained in the fields of protein electrostatics and stabilization.

The past five years I have devoted to the fascinating world of proteins. I have studied protein interactions within different proteins, between different proteins, between proteins and ligands using a wide range of methods. The introductory part of this thesis provides the background to my work at biophysical chemistry and should aid in reading the appended papers. The first part contains a general introduction to protein structure, folding and stability followed by the protein systems and methods used. In the last part I will briefly introduce you to the seven papers forming the basis of this thesis.

I hope you will enjoy reading about some pieces in the perplexing protein puzzle!

Stina Lindman  
January 2010, Lund, Sweden

# POPULÄRVETENSKAPLIG SAMMANFATTNING PÅ SVENSKA

Proteiner är uppbyggda av aminosyror sammankopplade enligt en förutbestämd mall till långa kedjor. Mallen är den genetiska koden där all information om hur proteinet skall se ut finns. När den långa kedjan av aminosyror bildats "veckar" proteinet ihop sig till en specifik tre-dimensionell struktur. Proteinveckning kan liknas med att lägga pussel fast där bitarna sitter som på ett pärlband. Pusselbitarna, aminosyrorna, finns i 20 olika varianter med olika form, storlek och laddning. Innan proteinet "veckat" sig finns det många sätt att lägga ut bitarna på men när pusslet är lagt och bitarna pusslats ihop för att passa varandra perfekt finns det bara ett sätt att lägga pusslet. Till skillnad från ett pussel är proteiner flexibla och är inte fast i det "veckade" pusslet utan kan till olika grad veckla upp och ihop sig hela tiden. Alla proteiner byggs upp av samma 20 sorters pusselbitar men varje protein får i det färdiglagda pusslet en unik tre-dimensionell struktur för att kunna utföra en specifik funktion.

Ibland har det blivit fel i den genetiska koden så att pusselbitar bytts ut mot andra, det har skett en mutation. När detta inträffar kan strukturen drastiskt förändras; om viktiga mittenbitar bytts ut finns det risk för att hela pusslet förstörs men om kantbitarna ändras blir konsekvenserna sällan lika förödande. I många fall är det just mutationer som orsakar sjukdomar eftersom proteinet har tappat eller fått en annan funktion. Proteinforskare introducerar ofta mutationer i proteiner för att se vilken effekt de får i det färdiglagda pusslet och ibland förbättrar faktiskt utbyta pusselbitar proteinet. Det är inte bara proteinets funktion som påverkas av hur pusslet är lagt utan även dess stabilitet. Ett proteins stabilitet beror inte bara av hur väl pusselbitarna passar ihop i det lagda pusslet utan det är skillnaden mellan det olagda och lagda pusslet som ger stabiliteten. För att stabilisera ett protein kan man således antingen gynna det lagda pusslet eller missgynna det olagda pusslet. En mutation som ger samma effekt i lagt och olagt pussel stabiliserar därmed inte proteinet. I den här avhandlingen har i främsta hand proteinstabilitet studerats, dels för att förstå vilka krafter som bidrar till stabilitet dels för att utveckla en metod för att stabilisera proteiner.

Hur väl pusselbitarna passar och hur hårt de sitter ihop jämfört med att vara utspridda, protein stabilitet, kan förändras med olika tillsatser. Till exempel kan vissa pusselbitars laddning förändras om man ändrar pH (vätejonkoncentrationen) i provet. När pusselbitarna ändrar laddning kan vissa aminosyror attrahera eller repellera varandra mer eller mindre än tidigare och stabiliteten av pusslet förändras. Jag har i artiklarna I-IV studerat hur stabiliteten av ett visst protein förändras när specifika pusselbitar ändrar laddning. Vidare har jag med hjälp av kärnmagnetisk resonans (NMR) spektroskopi bestämt hur varje negativt laddad pusselbit ändrar sin laddning som funktion av pH. Utifrån detta har växelverkan mellan laddade pusselbitar kunnat utredas. Det vi såg var att proteinets stabilitet påverkas mycket av förändring i pH och att samma typ av pusselbit men med olika position i proteinet

har olika laddning och olika benägenhet att laddas upp. Vi tittade även på pusselbitarnas laddning i det utvecklade proteinet, det olagda pusslet, och såg att här var det mycket mindre skillnad i laddning och benägenhet att bli uppladdade. Utifrån laddningarna på varje pusselbit i lagt och olagt pussel kunde även stabiliteten av det lagda pusslet förklaras.

I artiklarna VI och VII har pusslet utvecklats genom att pusselbitar från två olika proteiner fogades ihop. Här delades aminosyrasekvenserna av grönt fluorescerande protein (GFP) och ett kalciumbindande protein (calbindin) upp i två delar. En del från GFP sammanlänkades sedan med en del av calbindin och den andra delen av calbindin länkades ihop med den andra delen av GFP. Nu var pusslet inte bara en enda kedja av pusselbitar från ett protein utan två kedjor av pusselbitar från två olika proteiner. Pusslandet av dessa proteinsekvenser studerades och kunde verifieras med hjälp av att GFP blir grönt vid korrekt pusslande. Resultaten visade att intensiteten av det gröna ljuset från GFP kunde sammankopplas med hur väl pusslet var lagt vid olika förhållanden. Detta faktum möjliggör att använda metoden för att stabilisera proteiner och utnyttjades i studien som presenteras i artikel VII. Här visade vi att utifrån en stor samling mutanter med olika stabilitet kunde de stabilaste mutanterna plockas ut baserat på starkast grönt fluorescerande ljus.

Trots att man vet mycket om proteiner idag så kan man fortfarande inte förutsäga från en utvecklad pusselsekvens hur det färdiglagda pusslet kommer att se ut. Sådan kunskap skulle kunna hjälpa till att förstå varför vissa mutationer är sjukdomsalstrande eller hitta botemedel mot sjukdomar. Mina resultat har ökat förståelsen för vilka krafter, främst mellan laddningar, som håller samman ett protein i det pusslade tillståndet och nya sätt att analysera interaktioner mellan laddningar. Vidare har vi skapat grunden för en lovande metod att optimera proteinpussel och stabilisera proteiner.



## LIST OF PAPERS

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals. The papers are appended at the end of the thesis.

- I. Stina Lindman, Wei-Feng Xue, Olga Szczepankiewicz, Mikael C. Bauer, Hanna Nilsson, Sara Linse  
**Salting the charged surface: pH and salt dependence of protein G B1 stability**  
*Biophysical Journal*, 2006, **90**, 2911-2921. © Biophysical Society
- II. Stina Lindman, Sara Linse, Frans A.A. Mulder, Ingemar André  
**Electrostatic contributions to residue-specific protonation equilibria and proton binding capacitance for a small protein**  
*Biochemistry*, 2006, **45**, 13993-14002. © American Chemical Society
- III. Stina Lindman, Sara Linse, Frans A.A. Mulder, Ingemar André  
**pK<sub>a</sub> values for side-chain carboxyl groups of a PGB1 variant explain salt and pH-dependent stability**  
*Biophysical Journal*, 2007, **92**, 257-266. © Biophysical Society
- IV. Stina Lindman, Mikael C. Bauer, Mikael Lund, Carl Diehl, Frans A.A. Mulder, Mikael Akke, Sara Linse  
**Electrostatic interactions in unfolded states under native conditions through fragment pK<sub>a</sub> values**  
*Manuscript*
- V. Jannette Carey, Stina Lindman, Mikael C. Bauer, Sara Linse  
**Protein reconstitution and three-dimensional domain swapping: Benefits and constraints of covalency**  
*Protein Science*, 2007, **16**, 2317-2333. © The Protein Society
- VI. Stina Lindman, Ida Johansson, Eva Thulin, Sara Linse  
**Green fluorescence induced by EF-hand assembly in a split GFP system**  
*Protein Science*, 2009, **18**, 1221-1229. © The Protein Society
- VII. Stina Lindman, Armando Hernandez-Garcia, Olga Szczepankiewicz, Birgitta Frohm, Sara Linse  
**In vivo protein stabilization based on thermodynamic principles and a split GFP system**  
*Manuscript*

Other papers by the author not included in the thesis.

Stina Lindman, Iseult Lynch, Eva Thulin, Hanna Nilsson, Kenneth A. Dawson, Sara Linse

**Systematic investigation of the thermodynamics of HSA adsorption to N-isopropylacrylamide/N-tert-butylacrylamide copolymer nanoparticles. Effects of particle size and hydrophobicity**

*Nano Letters*, 2007, 7, 914-920

Tommy Cedervall, Iseult Lynch, Stina Lindman, Tord Berggård, Eva Thulin, Hanna Nilsson, Kenneth A. Dawson, Sara Linse

**Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles**

*Proceedings of the National Academy of Science USA*, 2007, **104**, 2050-2055

Sara Linse, Celia Cabaleiro-Lago, Wei-Feng Xue, Iseult Lynch, Stina Lindman, Eva Thulin, Sheena E. Radford, Kenneth A. Dawson

**Nucleation of protein fibrillation by nanoparticles**

*Proceedings of the National Academy of Science USA*, 2007, **104**, 8691-8696

Celia Cabaleiro-Lago, Fiona Quinlan-Pluck, Iseult Lynch, Stina Lindman, Aedin M. Minogue, Eva Thulin, Dominique M. Walsh, Kenneth A. Dawson, Sara Linse

**Inhibition of amyloid beta protein fibrillation by polymeric nanoparticles**

*Journal of American Chemical Society*, 2008, **130**, 15437-15443.

Tommy Cedervall, Stina Lindman, Tord Berggård, Eva Thulin, Hanna Nilsson, Sara Linse

**Nanopartiklar i biologiska system**

*Kemivärlden Biotech*, 2008, **5**, 30-32

Erika Gustafsson, Cecilia Forsberg, Karin Haraldsson, Stina Lindman, Lill Ljung, Christina Furebring.

**Purification of truncated and mutated Chemotaxis Inhibitory Protein of Staphylococcus aureus-an anti-inflammatory protein**

*Protein Expression and Purification*, 2009, **63**, 95-101

Erika Gustafsson, Anna Rosén, Karin Barchan, Kok P. M. van Kessel, Karin Haraldsson, Stina Lindman, Cecilia Forsberg, Lill Ljung, Karin Bryder, Björn Walse, Pieter-Jan Haas, Jos A.G. van Strijp, Christina Furebring

**Directed evolution of Chemotaxis Inhibitory Protein of Staphylococcus aureus generates biologically functional variants with reduced interaction with human antibodies**

*Protein Engineering, Design and Selections*, 2010, **23**, 91-101

## MY CONTRIBUTIONS TO THE PAPERS

- I. I performed thermal denaturations, isoelectric focusing and SPR measurements. I expressed and purified the  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled protein. I performed and analyzed NMR spectroscopy. I wrote the first version of the paper.
- II. I expressed and purified the  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled protein. IA and I made peak assignments. I performed pH titrations. IA and I performed data analysis. IA and I wrote the paper and collected improvements from S Linse and FM.
- III. IA and I performed data analysis. IA and I wrote the paper and collected improvements from S Linse and FM.
- IV. MB, MA, S Linse and I initiated the project. MB and I expressed and purified the  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled fragments. MB, CD and I performed NMR spectroscopy and peak assignments. MB and I performed pH titrations. MB and I performed data analysis. I wrote the first version of the paper and collected improvements from co-authors.
- V. JC, MB, S Linse and I performed literature search and wrote the paper.
- VI. IJ and I performed CD and fluorescence spectroscopy. I performed co-expression and kinetics *in vivo*. S Linse and I wrote the paper.
- VII. S Linse and I designed the library and initiated the project. BF and I expressed and purified the selected variants. I performed and analyzed thermal denaturations. I wrote parts of the paper.

## ABBREVIATIONS AND SYMBOLS

$B_0$	applied magnetic field
BiFC	bimolecular fluorescence complementation
CD	circular dichroism
$C_p$	heat capacity
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
$e$	elementary charge, $1.602 \cdot 10^{-19}$ C
$\epsilon_0$	permittivity of vacuum, $8.854 \cdot 10^{-12}$ F/m
$\epsilon_r$	dielectric constant
G	Gibb's free energy
$\gamma$	gyromagnetic ratio
GdnHCl	guanidine hydrochloride
GFP	green fluorescent protein
H	enthalpy
$h$	Planck's constant, $6.626 \cdot 10^{-34}$ J·s
$\hbar$	Planck's constant/ $2\pi$
IgG	immunoglobulin G
$n_H$	Hill parameter
NMR	nuclear magnetic resonance
PGB1	B1 domain of protein G
PGB1-QDD	PGB1 with T2Q, N8D and N37D
pI	isoelectric point
R	gas constant, $8.314$ J/(K·mol)
RNA	ribonucleic acid
S	entropy
$\sigma$	shielding constant
SPR	surface plasmon resonance
UV	ultra violet
wt	wild type

### One- and three-letter symbols for amino acids:

Alanine=Ala=A	Leucine=Leu=L
Arginine=Arg=R	Lysine=Lys=K
Asparagine=Asn=N	Methionine=Met=M
Aspartic acid=Asp=D	Phenylalanine=Phe=F
Cysteine=Cys=C	Proline=Pro=P
Glutamic acid=Glu=E	Serine=Ser=S
Glutamine =Gln=Q	Threonine=Thr=T
Glycine=Gly=G	Tryptophan=Trp=W
Histidine=His=H	Tyrosine=Tyr=Y
Isoleucine=Ile=I	Valine=Val=V

Mutations are written as: N8D, where asparagine 8 is mutated to aspartate

# CONTENTS

Preface .....	v
Populärvetenskaplig sammanfattning på svenska.....	vi
List of papers.....	viii
My contributions to the papers.....	x
Abbreviations and symbols .....	xi
Contents .....	xii
<b>1. Background: protein structure and interactions .....</b>	<b>1</b>
1.1 Introduction to proteins.....	1
1.2 Protein folding and stability .....	2
1.3 Protein denaturation .....	5
1.3.1 Thermal denaturation .....	6
1.3.2 Chemical denaturation and salt effects .....	7
1.4 Stabilization of proteins.....	8
1.5 Intramolecular forces.....	9
1.5.1 Hydrogen bonding .....	9
1.5.2 Hydrophobic effect.....	10
1.5.3 Van der Waals interactions .....	10
1.5.4 Conformational entropy .....	11
1.5.5 Electrostatic interactions .....	11
1.6 The denatured state.....	17
1.6.1 The Gaussian-chain model .....	18
1.7 Intermolecular interactions.....	19
1.8 Ligand binding.....	19
1.9 Protein reconstitution .....	21
<b>2. Protein systems selected .....</b>	<b>23</b>
2.1 Protein G B1.....	23
2.2 Calbindin D9k.....	24
2.3 Green fluorescent protein.....	25
<b>3. Methods.....</b>	<b>27</b>
3.1 Circular dichroism spectroscopy.....	27
3.2 Fluorescence spectroscopy .....	28
3.3 Bimolecular fluorescence complementation.....	30
3.3.1 Split GFP system .....	30
3.4 Differential scanning calorimetry .....	31
3.5 Surface plasmon resonance technology .....	32
3.6 Nuclear magnetic resonance spectroscopy .....	33

3.6.1 Chemical shift .....	34
3.6.2 Isotope labeling.....	35
3.6.3 Two and multidimensional NMR spectroscopy .....	35
<b>4. Introduction to papers .....</b>	<b>37</b>
4.1 Electrostatic interactions in PGB1; papers I, II, III and IV .....	37
4.1.1 Paper I.....	37
4.1.2 Paper II.....	38
4.1.3 Paper III .....	39
4.1.4 Paper IV .....	40
4.1.5 Conclusions papers I-IV .....	40
4.2 Reconstitution, domain swapping and general protein energetics; paper V .....	41
4.2.1 Paper V.....	41
4.3 Split GFP in protein stabilization; Papers VI and VII.....	41
4.3.1 Paper VI.....	41
4.3.1 Paper VII.....	42
<b>5. Conclusions.....</b>	<b>44</b>
<b>6. Acknowledgement.....</b>	<b>45</b>
<b>7. References .....</b>	<b>48</b>

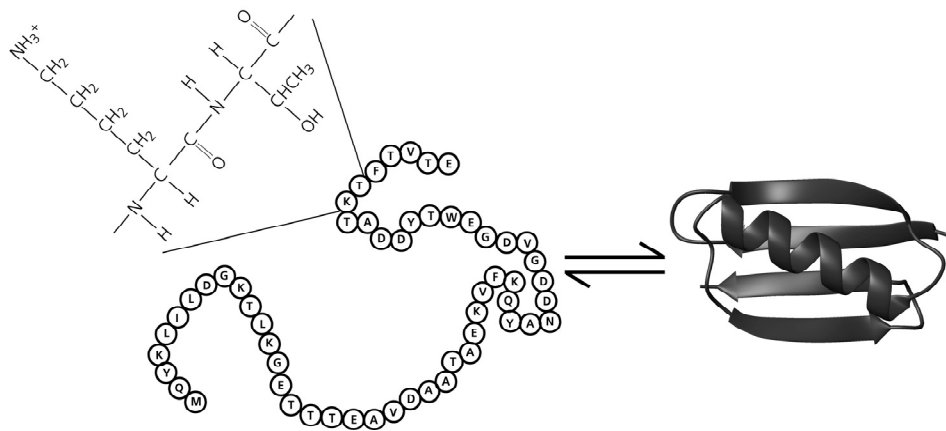


# 1. BACKGROUND: PROTEIN STRUCTURE AND INTERACTIONS

## 1.1 Introduction to proteins

Proteins are one of the macromolecules of life. They are responsible for memory, learning and other higher functions such as enzymatic activities, signaling, transport, control of gene expression and immune response (1). Largely diverse yet so similar, all proteins are built from the same 20 building blocks: the amino acids. The delicate and fine choice between different amino acids creates structures that are functional in very different environments such as oily cell membranes or salty environments where halophilic bacteria live.

The amino acid sequence for a given protein is encoded in our genes. The three letter code from DNA, transcribed into RNA, is translated at the ribosome into amino acids to generate polymers of very specific length and sequence. This specific sequence folds into a well ordered three-dimensional structure with biological activity. In Figure 1 the primary (amino acid sequence) and tertiary structures of a 56-amino acids protein are shown.



**Figure 1.** Proteins are made of amino acids that are linked through peptide bonds to form polypeptides. This polypeptide chain folds into a well-defined three-dimensional structure. A 56-amino acid protein is shown where the structure of a dipeptide unit is explicitly shown. The denatured state to the left (with unknown structure) is in equilibrium with the native state to the right (with known structure).



The word protein has been around since 1838 when the Swedish scientist Jöns Jacob Berzelius coined the word (2). In 1819 the first amino acid (leucine) was discovered and by 1900 it was a general knowledge that proteins were built from amino acids (2). Since then significant knowledge has been obtained about proteins but still these polymers are mysterious in several respects. For example, it is not yet possible to predict the structure of a given amino acid sequence. Therefore, further knowledge about proteins is still required and crucial in many areas. For example, we know that there are many diseases such as Alzheimer's, Parkinson's and Huntington's diseases that are associated with misfolding of proteins but we do not know the underlying mechanisms (3). Furthermore, the number of therapeutical proteins is growing. For the development of cures for diseases or finding new drug targets detailed understanding of the forces involved in proteins is essential.

## 1.2 Protein folding and stability

If folding of a polypeptide chain was the search through all possible conformations it would take longer than the age of the universe to find the native state. Since proteins can fold on a microsecond time scale (4), this is not the case. These contradictory statements were declared 40 years ago is the Levinthal paradox and it was realized that proteins do not go through all conformations in the search of the native state (5, 6). Protein folding is the search for the free energy minimum and all the information needed to fold into the native state is inherent in the amino acid sequence (7, 8). Today the protein folding free energy landscape is regarded as more rugged than in the early studies, however, many proteins are still considered to fold fast and without intermediates while we know that other proteins fold slowly with many intermediates, getting stuck in kinetic traps or needing chaperones to fold (9-12). After folding into the native state the protein is still highly dynamic with flexibility to be functional in biological processes (13). Protein folding mechanisms, kinetics, pathways and dynamics are enormous fields that are beyond the scope of this thesis, which more or less exclusively deals with equilibrium thermodynamics.

Proteins appear to be in either of two states, the native or denatured (Figure 1), while partially folded structures are rare. The residues are acting together to favor either the native or the denatured state so that protein folding is a highly cooperative process. The stability of a protein is defined in terms of the difference in free energy between the denatured (D) and native (N) states,  $\Delta G_{DN}^{\circ}$ . The denatured state is in equilibrium with the native state but has a very low population under native conditions. Little is known about the structure of the denatured protein, but has a crucial role in stability since the free energy of the native state is always compared to the free energy of the denatured state. The native state has many stabilizing non-covalent interactions but they are offset by large conformational entropy in the denatured state (see section 1.5). Hence, there are large opposing entropy and enthalpy terms called entropy-enthalpy compensation which is due to a large  $\Delta C_{p,DN}^{\circ}$ . When summed together the free energy difference is normally small, only

10-60 kJ/mol, meaning that most proteins are only marginally stable under physiological conditions (14, 15). This is, however, enough to ensure that almost all protein molecules are in the native state under biologically relevant conditions. Subtracting two large terms to give one small net result also introduces a large uncertainty that is one reason why it is hard to calculate protein stability accurately.

Establishing the framework of protein stability the easiest and most widely used model is to assume that the protein is in equilibrium between only two states (12):



where the free energies of the two states are

$$G_N = G_N^\circ + RT \ln([N]/1M), \quad G_D = G_D^\circ + RT \ln([D]/1M) \quad (2)$$

and the free energy difference is

$$G_D - G_N = \Delta G_{DN} = \Delta G_{DN}^\circ + RT \ln([D]/[N]) = \Delta G_{DN}^\circ + RT \ln(K) \quad (3)$$

At equilibrium, the free energy difference ( $\Delta G_{DN}$ ) is zero, and then the standard free energy difference is:

$$\Delta G_{DN}^\circ = -RT \ln(K) \quad (4)$$

Expressed in this way,  $\Delta G_{DN}^\circ$  is a positive number under conditions where the protein is in the native state. The free energy minimum is always a balance between the lowest enthalpy (H) and highest entropy (S) (Equation 5). To be more correct, any process always seeks to find the highest entropy but this can be achieved by increasing the entropy in the system or in the surroundings. When minimizing the enthalpy in the system, heat is given off that will increase the entropy in the surroundings and hence the entropy is increased in the overall process.  $\Delta G^\circ$  is introduced as a concept to take care of both the system and the surroundings at the same time and is smaller than 0 for a spontaneous process.

$$\Delta G_{DN}^\circ = \Delta H_{DN}^\circ - T \Delta S_{DN}^\circ \quad (5)$$

For proteins, the enthalpy lies in the non-covalent intramolecular interactions while the entropy lies in the number of conformations the polypeptide chain can adopt and is much larger for the unfolded peptide due to the flexibility of the chain. As one can see already from this expression,  $\Delta G_{DN}^\circ$  depends on temperature. If  $\Delta H_{DN}^\circ$  and  $\Delta S_{DN}^\circ$  are positive numbers and  $\Delta H_{DN}^\circ \geq T \Delta S_{DN}^\circ$  (as in the case for the comparison between denatured and native protein) one can see that on increasing the

temperature, the  $T\Delta S_{DN}^{\circ}$  term will eventually become larger than  $\Delta H_{DN}^{\circ}$ . At this temperature  $\Delta G_{DN}^{\circ}$  will change sign and the protein unfolds. To make things more complicated  $\Delta H_{DN}^{\circ}$  and  $\Delta S_{DN}^{\circ}$  are not constant but change with temperature according to:

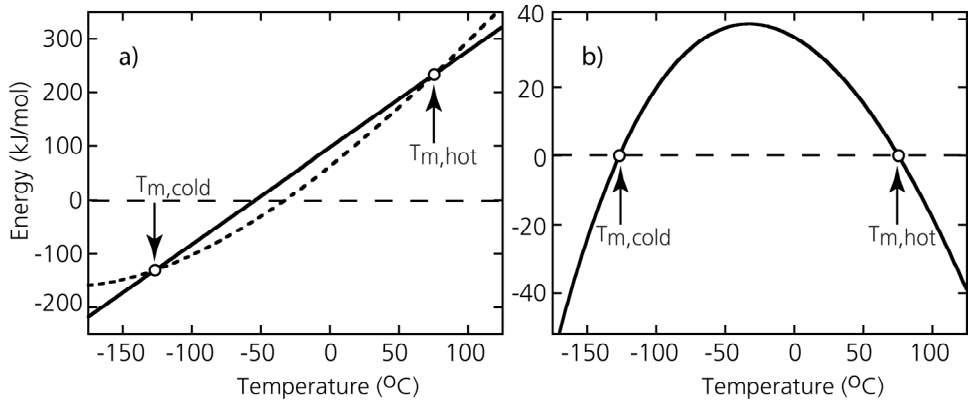
$$\Delta H_{DN}^{\circ}(T) = \Delta H_{DN,ref}^{\circ} + \Delta C_{p,DN}^{\circ}(T - T_{ref}) \quad (6)$$

and

$$\Delta S_{DN}^{\circ}(T) = \Delta S_{DN,ref}^{\circ} + \Delta C_{p,DN}^{\circ} \ln(T/T_{ref}) \quad (7)$$

where  $\Delta C_{p,DN}^{\circ}$  is the heat capacity at constant pressure. The heat capacities for the native and denatured states are not constant but the difference between them varies much less so  $\Delta C_{p,DN}^{\circ}$  is usually assumed to be constant with temperature.  $\Delta H_{DN,ref}^{\circ}$  and  $\Delta S_{DN,ref}^{\circ}$  are the enthalpy and entropy at a given reference temperature. If choosing the unfolding temperature,  $T_m$ , as the reference temperature and recognizing that  $\Delta G_{DN}^{\circ} = 0$  at this temperature,  $\Delta G_{DN}^{\circ}$  can be written as:

$$\Delta G_{DN}^{\circ}(T) = \Delta H_{DN}^{\circ}(T_m) \left[ 1 - \frac{T}{T_m} \right] + \Delta C_{p,DN}^{\circ} \left[ T - T_m - T \cdot \ln \left( \frac{T}{T_m} \right) \right] \quad (8)$$



**Figure 2.** Temperature dependence of a)  $\Delta H_{DN}^{\circ}$  (solid line),  $T\Delta S_{DN}^{\circ}$  (dashed line) and b)  $\Delta G_{DN}^{\circ}$  for PGB1-QDD at pH 5. The figure is generated using values determined in paper I ( $\Delta H_{DN}^{\circ} = 233$  kJ/mol,  $\Delta C_{p,DN}^{\circ} = 1.8$  kJ/(mol·K),  $T_m = 75.5^{\circ}\text{C}$ ). The cold and heat denaturation points are indicated but for this protein the cold denaturation occurs at an extremely low temperature and is just a hypothetical point. Note the different scales of the y-axis in a and b and that the curve in b arises from the small difference between the curves in a.

From Figure 2 one can see that both  $\Delta H^{\circ}_{DN}$  and  $T\Delta S^{\circ}_{DN}$  are large terms and show significant temperature dependence. However, the terms give opposite effects so that  $\Delta G^{\circ}_{DN}$  has less temperature dependence and lower magnitude. Moreover, there are two temperatures where  $\Delta H^{\circ}_{DN} = T\Delta S^{\circ}_{DN}$  and  $\Delta G^{\circ}_{DN} = 0$  corresponding to cold and heat denaturation. Above the heat denaturation temperature and below the cold denaturation temperature, respectively the entropy term becomes larger than the enthalpy term so that the denatured state is favored.

### 1.3 Protein denaturation

Richard B. Anfinsen was one of the pioneers in showing that many proteins can be reversibly denatured (reviewed in (7)) and denaturation studies are still very common ways to study protein stability. In protein denaturation one starts with the native protein and then changes the properties to favor the denatured state. The fractions of native ( $F_N$ ) and denatured ( $F_D$ ) protein are

$$F_N = \frac{[N]}{[N] + [D]} = \frac{1}{1 + K}, \quad F_D = \frac{[D]}{[N] + [D]} = \frac{K}{1 + K} \quad (9)$$

where  $K = [D]/[N]$ .

Common methods for denaturation studies are circular dichroism (CD) spectroscopy, fluorescence spectroscopy and differential scanning calorimetry (DSC). For denaturation studies with these methods, the baselines before ( $Y_N$ ) and after ( $Y_U$ ) the actual unfolding are commonly assumed to be straight lines.

$$Y_N = k_N \cdot X + b_N \quad (10)$$

$$Y_D = k_D \cdot X + b_D \quad (11)$$

$k_N$  and  $k_D$  are the slopes,  $b_N$  and  $b_D$  are the intercepts and  $X$  is the changing variable.  $Y$  is the measured signal, ellipticity (for CD), fluorescence or  $C_p$  (for DSC). Thermal and solvent denaturations are two common ways to denature a protein and are discussed in sections 1.3.1-2.

All the equations and concepts described above and in the following sections apply only to reversible unfolding. In some cases proteins are irreversibly denatured, which usually consists of unfolding followed by aggregation and many times is due to a local, rather than a global unfolding (16). Even though no thermodynamic parameters can be extracted, irreversible denaturations can still be useful and give

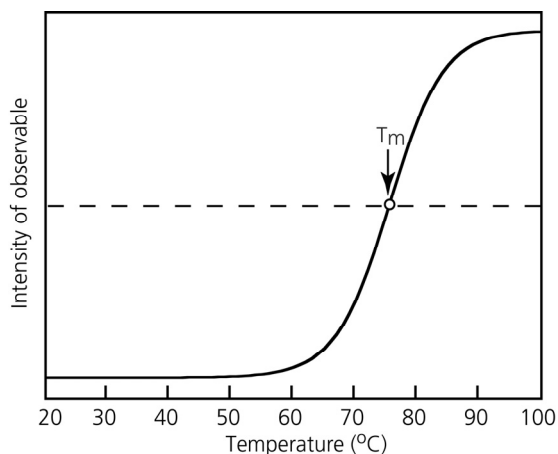
stability indications when for example comparing the apparent melting temperature of different mutants of the same protein.

### 1.3.1 Thermal denaturation

Changing the temperature is an easy and common way to denature a protein (Figure 3) where increased temperature favors the denatured state as a consequence of the larger conformational space the protein can explore. From Figure 2 one can see that there are two temperatures where the entropy term ( $T\Delta S_{DN}^{\circ}$ ) becomes larger than the enthalpy term and the protein is half way unfolded. Using Equations 4, 9, 10 and 11 the following equation is fitted to thermal denaturation data measured with CD or fluorescence spectroscopy:

$$Y_{\text{obs}} = \frac{(k_N \cdot X + b_N) + (k_D \cdot X + b_D) \cdot e^{-(\Delta G^{\circ}_{DN}(T)/RT)}}{1 + e^{-(\Delta G^{\circ}_{DN}(T)/RT)}} \quad (12)$$

The temperature dependence of  $\Delta G^{\circ}_{DN}$  is described in Equation 8. From fitting these equations to experimental data, the midpoint temperature,  $T_m$ , can be derived. At this temperature half of the molecules are denatured so  $F_N = F_D$ .  $k_N$ ,  $k_D$ ,  $b_N$ ,  $b_D$  and  $\Delta H^{\circ}_{DN}$  are variables in the fit. The heat capacity,  $\Delta C^{\circ}_{p,DN}$ , is either left as a variable in the fit or determined by, for example, calorimetry.



**Figure 3.** Hypothetical 2-state thermal denaturation using the same parameters for  $T_m$ ,  $\Delta H^{\circ}_{DN}$  and  $\Delta C^{\circ}_{p,DN}$  as used in Figure 2 and fictive baselines. The cooperative transition is apparent where the molecules are either native or denatured; this is due to the collection of non-covalent interactions favoring the native state that are overcome by the large conformational entropy at temperatures above  $T_m$ .

From the midpoint temperature one can extrapolate back to physiological conditions and from that derive the stability ( $\Delta G^{\circ}_{DN}$ ) under these conditions. Thermal denaturations at several pH values and salt concentrations are presented in paper I. For thermal denaturation with DSC the unfolding is monitored through the heat capacity of the system (see section 3.4).

### 1.3.2 Chemical denaturation and salt effects

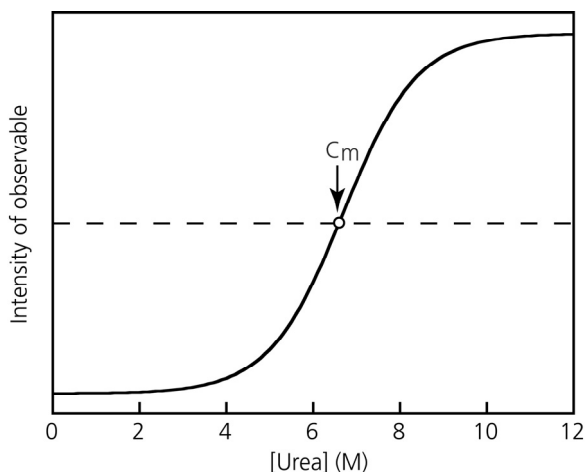
Another straight-forward way to denature a protein is to, in a stepwise manner, add a chemical to the protein solution that favors the denatured state. Examples of such denaturants are urea or guanidine hydrochloride (GdnHCl). These chemicals are preferred solvents for the peptide groups and hence favor the solvent exposed denatured state (17). Urea has the disadvantage that a high concentration (several molar) is needed to unfold a protein while GdnHCl is a salt so that electrostatics cannot be studied at the same time. When studying chemical denaturation almost the same approach and equations are used as for thermal denaturation. CD, fluorescence and UV spectroscopy are common methods to study the denaturation process. If there is a reversible two-state unfolding and assuming that the free energy of unfolding by denaturant is linear, then:

$$\Delta G^{\circ}_{DN}(\text{denaturant}) = \Delta G^{\circ}_{DN}(\text{H}_2\text{O}) - m \cdot [\text{denaturant}] \quad (13)$$

Using Equations 4, 9, 10, 11 and 13 the following equation can be fitted to data:

$$Y_{\text{obs}} = \frac{(k_N \cdot X + b_N) + (k_D \cdot X + b_D) \cdot e^{-((\Delta G^{\circ}_{DN}(\text{H}_2\text{O}) - m \cdot [\text{denaturant}]) / RT)}}{1 + e^{-((\Delta G^{\circ}_{DN}(\text{H}_2\text{O}) - m \cdot [\text{denaturant}]) / RT)}} \quad (14)$$

$\Delta G^{\circ}_{DN}(\text{H}_2\text{O})$  is the free energy of unfolding in pure water, the m-value is related to the change in solvent accessible surface area upon unfolding where a large m-value indicates a large increase upon unfolding (18).  $C_m$  is the denaturant concentration at the transition midpoint and is obtained when  $\Delta G^{\circ}_{DN}(\text{denaturant})=0$  so that  $C_m = \Delta G^{\circ}_{DN}(\text{H}_2\text{O})/m$ . A hypothetical urea denaturation is shown in Figure 4.



**Figure 4.** Hypothetical 2-state urea denaturation where  $\Delta G_{\text{DN}}^{\circ}(\text{H}_2\text{O})=19.4$  kJ/mol,  $m=2.9$  kJ/(mol·M) and  $T=298.15$  K.

Changes in pH are also used effectively to modulate the stability of proteins where protonation/deprotonation changes the free energy differentially in the native and denatured state. pH-dependent stability is discussed further in section 1.5.5.

Salt is not usually considered as a denaturant, but modulates protein stability. Addition of salt to a protein solution usually increases the solubility at low concentration, but at higher concentrations the salt normally causes the protein molecules to aggregate (19, 20). The first phenomenon is called salting-in and the second salting-out where salting-out depends on the nature of the salt. The Hofmeister series (reviewed by Baldwin (21)) describes the effect of various salts.

## 1.4 Stabilization of proteins

Denaturation, aggregation and proteolytic degradation are problems associated with many proteins used as therapeutics or in other biotechnical applications. The therapeutic and technical importance of these proteins may therefore increase if their stability can be improved. Most proteins are not optimized for stability, but rather for functionality. Spontaneous mutations are mostly destabilizing and hence proteins are only sufficiently stable to maintain function. This opens up the possibility to optimize protein stability by protein engineering. Methods to stabilize proteins are important tools in biotechnology but can also uncover the principles important for protein folding and stability

Small changes in the amino acid sequence are usually tolerated without disrupting the fold of the protein, but different mutants can have very different stability. Popular methods to identify stabilizing mutations are structure-guided design and comparison

to thermostable analogs (22), computational design (23-25) or directed evolution (26-29). These studies have shown that it is often difficult and complex to predict the stability effect of a certain mutation and that random libraries commonly give the best results (28). The problem to face is, however, to limit the sequence space and a combination of the three methods mentioned above generally gives stability improvement (29). General results from stability optimizations are that the core is usually well packed and not easily improved (30, 31) unless several substitutions are made or buried charges are replaced by hydrophobic residues (32). Mutations at boundary positions (23) and at the protein surface have in several cases given fruitful stabilization effects (26, 33, 34). There are also studies showing that introduction of proline residues and disulphide bonds can improve protein stability (22, 35-37) .

## 1.5 Intramolecular forces

The largest free energy terms involved in protein stability are the hydrophobic effect and conformational entropy. These terms are much greater than  $\Delta G_{\text{DN,max}}^{\circ}$ , but are opposing and similar in magnitude and hence leave importance to other non-covalent interactions. Therefore, the three-dimensional structures of proteins are governed by non-covalent interactions. These interactions are much weaker than covalent bonds and the contributions from each residue are small and go in different directions. However, by accounting for all the residual components for a whole protein the interactions add to a significant number and are responsible for the cooperative nature of protein folding.

### 1.5.1 Hydrogen bonding

Hydrogen bonds are attractive interactions between two electronegative atoms that share a hydrogen atom. It is a dipole-dipole interaction that becomes exceptionally strong due to the small size of the proton whose electron density is easily depleted by the donor and acceptor atoms (38). The hydrogen still remains closer to the donor atom ( $\sim 1 \text{ \AA}$ ) than the acceptor atom ( $> 1.5 \text{ \AA}$ ) and the interaction is primarily of electrostatic origin. The strength of the hydrogen bond depends on the electronegativity and the orientation of the atoms but in general the energy of a hydrogen bond lies in the order of 10-40 kJ/mol and is much stronger than van der Waals interactions ( $\sim 1 \text{ kJ/mol}$ ) but weaker than Coulombic interactions ( $\sim 500 \text{ kJ/mol}$ ) (38).

Almost all N-H and C=O groups of the protein backbone are involved in hydrogen bonds and are responsible for the regular secondary structural elements such as  $\alpha$ -helices and  $\beta$ -sheets. The nitrogen is the donor atom, the oxygen is the acceptor atom and the hydrogen is shared between the two. There are also hydrogen bonds between side-chains and between side-chains and back-bone in the protein interior; some of which are very strong (39, 40) and highly conserved (41).



Hydrogen bonding is not the driving force for protein folding but rather a means for polar groups in the protein interior to deal with the unfavorable hydrophobic environment. The protein has hydrogen bonding capacity both in the native state (between peptide groups) and denatured state (between peptide groups and water). However, it seems that all hydrogen bonds have to be satisfied in the protein because unsatisfied hydrogen bonds in the protein interior give a large unfavorable contribution to the stability (42).

### 1.5.2 Hydrophobic effect

The hydrophobic effect is the tendency for nonpolar molecules to form aggregates of like molecules in aqueous solution. In a regular solution the mixing of two components is driven by the increase in translational entropy upon mixing, while usually the enthalpy increases. For this system, the entropy not included in the translational entropy is usually very small. Both the entropy and enthalpy are more or less temperature independent and the free energy is governed by the magnitude of the enthalpy change. In contrast to this simple solution, the signature of the hydrophobic effect is that both the enthalpy and excess entropy are largely temperature dependent. The reason for the large temperature dependence is a large  $\Delta C_{p, DN}^{\circ}$ . At low temperatures the free energy of transferring a non-polar molecule into water is signified by a decrease in entropy. One contribution to this decrease is due to the fact that the water molecules have to order themselves around the non-polar molecule in order not to lose hydrogen bond possibilities. At high temperature, an increase in enthalpy gives a large contribution to the free energy of transfer. At this temperature the molecules are not ordered so the entropy loss is small but rather, water loses hydrogen bonding possibilities.

Protein folding and stability show many similarities to the hydrophobic effect where in the native state the hydrophobic residues are clustered into a densely packed hydrophobic core excluded from water. Moreover, proteins show unfolding enthalpies and entropies that are temperature dependent due to a large  $\Delta C_{p, DN}^{\circ}$ . This also results in that proteins exhibit two temperatures where  $\Delta G_{DN}^{\circ}$  is zero; one high and one low temperature corresponding to cold and heat denaturation temperatures, respectively (Figure 2). Moreover, most proteins exhibit a maximum stability around room temperature where the solubility of non-polar molecules is expected to be the lowest. Due to these similarities with the hydrophobic effect, it is hypothesized that hydrophobic interactions are the major driving force for protein folding, first suggested by Kauzmann (43) and further discussed by Dill (44). For more discussions in the subject, see the following references (43-46).

### 1.5.3 Van der Waals interactions

Van der Waals interactions exist between all atoms and molecules and are therefore non-specific. These close-range interactions are a collection of forces that arise from

dipole-dipole interactions, dipole-induced dipole interactions and interactions between two induced dipoles. Commonly van der Waals interactions in combination with exchange repulsion are modeled as a Lennard-Jones potential energy function,

$$U(r) = 4\epsilon_{LJ} \left[ \left( \frac{\sigma_{LJ}}{r} \right)^{12} - \left( \frac{\sigma_{LJ}}{r} \right)^6 \right] \quad (15)$$

where  $\epsilon_{LJ}$  is the depth of the potential well and  $\sigma_{LJ}$  is the point of zero potential. The  $1/r^6$  term is the collection of attractive van der Waals forces while the  $1/r^{12}$  term accounts for the exchange repulsion.

Even though van der Waals interactions are not dominant forces in protein folding they play an important role in the packed and optimized hydrophobic core. These forces can be calculated for model systems but their contributions to protein stability are not easily predicted (45).

#### 1.5.4 Conformational entropy

Upon folding a protein into the restricted native state there is an enormous loss in conformational entropy. The loss in conformational entropy is the major force opposing protein folding and has a large impact on the free energy difference between the native and denatured states. The magnitude of the conformational entropy seems to be residue dependent and is not easily determined (45). However, there are studies trying to calculate the conformational entropy for different residue types (47).

Disulphide bonds are introduced by nature to restrict the number of conformations of the denatured state. By lowering the conformational entropy of the denatured state the relative stability of the native state is increased (1). However, there are studies indicating that disulphide bonds also lower the enthalpy of the native state (48, 49).

#### 1.5.5 Electrostatic interactions

Electrostatic interactions are long-range forces that play a key role in proteins. They are responsible for many of the functions that proteins display and tune protein folding and stability (50-54). Charges direct proteins and substrates to their correct location, regulate enzyme activity and are important for ion binding (55-63). Moreover they prevent aggregation by improving protein solubility (52, 64, 65).

##### *Electrostatic framework*

All non-covalent interactions except for exchange repulsion and dispersion attraction are more or less of electrostatic origin and can be described as the interaction of point

charges. Coulomb's law describes the interaction between point charges which is very large in vacuum.

$$U_{ij} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} = \frac{e^2 z_i z_j}{4\pi\epsilon_0 r_{ij}} \quad (16)$$

where  $q_i$  and  $q_j$  are the charges of residue  $i$  and  $j$  respectively,  $r$  the distance between them,  $e=1.602\cdot 10^{-19}$  C is the elementary charge,  $\epsilon_0=8.854\cdot 10^{-12}$  F/m the permittivity of vacuum and  $z$  is the charge number. As obvious from Equation 16 the interaction decays as  $1/r$  and is therefore long-range compared to, for example, the van der Waals interactions described above.

Coulomb's law describes the interaction of charges in vacuum, but in solution the description is complicated due to the solvent-charge interactions. In a polar or polarizable solvent the solvent molecules will orient themselves around the charges, which will result in a decreased effective charge, depending on the polarity or polarizability of the solvent. Polarization means that the charge distribution changes in an external electric field and polarizability refers to how easily the charge distribution of the system changes. It is very hard to calculate all interactions for the individual solvent molecules and instead the solvent molecules have been averaged out to one macroscopic constant. This constant is called the dielectric constant ( $\epsilon_r$ ) and gives a measure of the polarizability of the solvent compared to vacuum. A solvent described with a dielectric constant is referred to a dielectric continuum where the interaction between two charges is written as:

$$G_{ij} = \frac{1}{\epsilon_r(T)} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} = \frac{1}{\epsilon_r(T)} \frac{e^2 z_i z_j}{4\pi\epsilon_0 r_{ij}} \quad (17)$$

Due to the temperature dependence of  $\epsilon_r$  the energy in Equation 17 is a free energy because it also accounts for the entropy. For the interaction between two charges of same sign in a solvent with a low dielectric constant it is enthalpically unfavorable to bring the charges together while in a solvent with a high dielectric constant it is entropically unfavorable to bring the charges together (38). As apparent from Equation 17 the effective interaction energy is reduced by  $1/\epsilon_r$ . For water  $\epsilon_r$  is 78.5 at 25°C (38) and hence charge-charge interactions in water have only 1.3% of the free energy of interaction in vacuum. Even though the dielectric constant is a macroscopic property there are many attempts to define  $\epsilon_r$  of the protein interior without a general consensus (66, 67).

Introducing salt into the system screens charge-charge interactions and in dilute salt solutions the Debye- Hückel theory (68) can describe the interaction energy between two charges:

$$G_{ij} = \frac{1}{\epsilon_r(T)} \frac{e^2 z_i z_j}{4\pi\epsilon_0 r_{ij}} e^{-\kappa r_{ij}}, \quad (18)$$

where  $\kappa=1/D$  and  $D=3.04/1^{1/2}$  Å in water at 25°C.

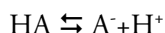
### *Protein charges and titrations*

In proteins there are seven residues that are potentially charged depending on the pH, the ionic strength and the local electrostatic environment. As many as 29% of the amino acids in proteins are titratable (69).  $pK_a$  values are used to describe the charge state of a given residue. When  $pH=pK_a$  it is equally probable for this particular residue to be protonated as deprotonated. The  $pK_a$  values for individual amino acid types have been measured several times and are referred to as model  $pK_a$  values since there are no other charges in the vicinity affecting the values. The model values for charged residues occurring in proteins are reported in Table 1. As seen from the table several residues carry a charge at neutral conditions where most proteins function.

**Table 1.** Model  $pK_a$ -values of the charged groups in proteins. The structure of the side-chain under physiological conditions is shown. Values from Nozaki and Tanford (70).

Residue type	$pK_{a,model}$	Structure at neutral pH
Asp	4.0	$-CH_2-COO^-$
Glu	4.4	$-CH_2-CH_2-COO^-$
Tyr	9.6	$-CH_2$ -phenol
Cys	10.8	$-CH_2-SH$
C-terminus	3.8	$-COO^-$
Lys	10.4	$-CH_2-CH_2-CH_2-CH_2-NH_3^+$
Arg	12.0	$-CH_2-CH_2-CH_2-NH=C=(NH_2)_2^+$
His	6.3	$-CH_2$ -imidazole- $(NH)_2^+$
N-terminus	7.5	$-NH_3^+$

The acid/base equilibrium is described as:



with the acid dissociation constant,  $K_a$

$$K_a = \frac{a_{A^-} a_{H^+}}{a_{AH}} = \frac{\gamma_{A^-} [A^-] a_{H^+}}{\gamma_{AH} [AH]} \quad (19)$$

with the assumption  $\gamma_{A^-} = \gamma_{HA} = 1$  the acid dissociation constant can be written as:

$$K_a = \frac{[A^-] a_{H^+}}{[AH]} \quad (20)$$

where  $\text{pH} = -\log_{10}(a_{H^+})$  and  $\text{p}K_a = -\log_{10}K_a$

For a negatively charged group the fractional deprotonation as a function of pH is:

$$F_{A^-} = \frac{[A^-]}{[HA] + [A^-]} = \frac{10^{-\text{p}K_a}}{10^{-\text{pH}} + 10^{-\text{p}K_a}} = \frac{1}{1 + 10^{\text{p}K_a - \text{pH}}} \quad (21)$$

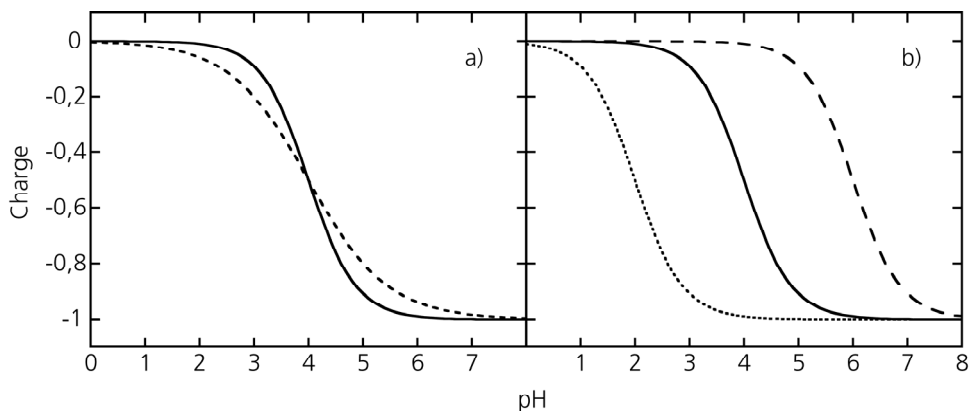
This equation is called the Henderson-Hasselbalch equation and describes many pH titration events that are not affected by other charges. In proteins there are multiple charges affecting each other (Figures 5 and 6). Many times a modified version of Equation 21 is used to improve the fit to data (71):

$$F_{A^-} = \frac{1}{1 + 10^{n_H(\text{p}K_a - \text{pH})}} \quad (22)$$

The Hill parameter ( $n_H$ ) also has a physical meaning (see paper II) and tells about the electrostatic coupling in the system. Most charges in a protein are surface exposed and facilitate interactions with charged molecules of opposite sign and prevent aggregation between similarly charged proteins.

#### *Determination of electrostatic interactions experimentally*

Compared to other non-covalent interactions in proteins it is rather easy to study electrostatic interactions. The importance of a specific charged residue can be investigated by mutagenesis. Removal or introduction of a charge can give very insightful information (15, 24, 26, 53, 72-77). Moreover, electrostatics can be investigated by means of pH titrations, which can give information of the charge state of the protein as a whole (78-80) or for specific residues in the protein (50). Salt efficiently screens electrostatics in a protein and can be used as an additional tool to mutagenesis and pH titrations (81-84).



**Figure 5.** Hypothetical titration of aspartate. Residue specific pH-titrations can give manifestations of different interactions. a) Electrostatic interactions in proteins are commonly observed as elongated titration curves (dashed line) compared to the ideal case following Equation 21 (solid line). b) Titration processes with highly shifted  $pK_a$  values can stem from desolvation of the charge (upshift, dashed line) or strong hydrogen bonds involving the carboxylate (downshift, dotted line).

By the use of NMR spectroscopy one can get residue specific information about the electrostatics in the protein (see papers II-IV). By monitoring the chemical shift as a function of pH the  $pK_a$  values of each titrating residue in the protein can be obtained (for a review about  $pK_a$  values obtained by NMR see (85)). Comparing the  $pK_a$  values obtained in this way with the values in Table 1 can give insight into the electrostatic coupling and other interactions in the protein. A downshifted  $pK_a$  value usually reports on a favorable electrostatic interaction (86) or strong hydrogen bond (39) while an upshifted  $pK_a$  value reports on either unfavorable electrostatic interactions (74) or a buried charge (71, 87, 88) (Figure 5). Electrostatic interactions between surface exposed residues normally give  $pK_a$  values with relatively small shifts.  $pK_a$  values of buried residues, on the other hand, can generate  $pK_a$  values that are shifted by several units (87). From the  $pK_a$  values the pH dependent stability of the protein can be calculated.

$$\partial \Delta G_{DN}^{\circ}(\text{pH}) = \sum_i \int_{\text{pH}} 2.3RT(q_D^i - q_N^i) \partial \text{pH} \quad (23)$$

where  $q_D$  and  $q_N$  are the charges of residue  $i$  in the denatured and native states respectively and are calculated from the  $pK_a$  values in the denatured and native states:

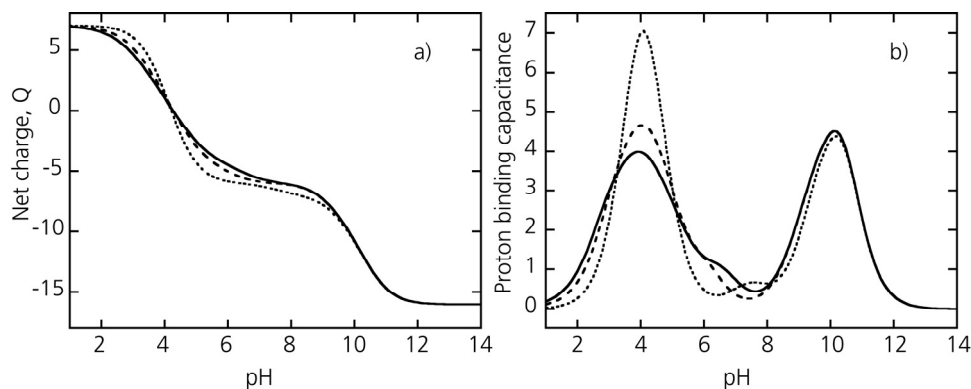
$$q^i = \frac{z}{1 + 10^{n_H z (\text{pH} - pK_{a,i})}} \quad (24)$$

where  $z$  is  $+1/-1$  for basic and acidic residues respectively. The main obstacle here is to obtain  $pK_a$  values in the denatured state, but this can be circumvented by measuring  $pK_a$  values of protein fragments (see paper IV).

The net charge of the protein as a function of pH can be obtained from pH titrations of the whole protein or by summing up the charges in the protein. The pH where the net charge is zero is called the isoelectric point (pI). At the pI the protein still contains charges but the negative and positive charges are balanced. Due to the zero net charge the protein is normally least soluble and most aggregation prone at this pH (52, 89, 90) but for some proteins the highest thermodynamic stability is observed at this pH (91). The net charge of the protein can be calculated according to:

$$Q(\text{pH}) = \sum_i \frac{-1}{1 + 10^{n_H(\text{p}K_{a,i} - \text{pH})}} + \sum_j \frac{1}{1 + 10^{n_H(\text{pH} - \text{p}K_{a,j})}} \quad (25)$$

The first term gives the net charge of all the negative residues in the protein and the second term gives the charge of the positive residues in the protein. An example of calculated net charge for a small protein is shown in Figure 6a.



**Figure 6.** a) The net charge in PGB1-QDD as a function of pH. The net charge determined from residue specific  $pK_a$  values in low salt (solid line) and 0.5 M NaCl (dashed line) calculated using Equation 25. The dotted line represents the net charge calculated according to the model values presented in Table 1. The  $pK_a$  values of the lysines could not be determined and hence the model values were used. b) The proton binding capacitance of PGB1-QDD as a function of pH determined from the negative derivative of the curves in a). For a real protein at low salt the proton binding capacitance curve is extended due to electrostatic coupling and  $pK_a$  shifts.

## *Contribution of electrostatic interactions to stability*

The contribution of electrostatic interactions to protein stability is not understood. There are several studies trying to solve this problem but the general conclusion is that electrostatic interactions are very context dependent and hence can be stabilizing in some cases and sometimes destabilizing. However, it is generally believed that proteins contain charges for other reasons than to improve the stability.

Surface exposed charges are not as large in magnitude due to the efficient screening from the surrounding water and generally give quite low effect on protein stability as exemplified in a study where 18 surface charges were removed without affecting the stability (73). However, other studies have shown that surface salt bridges can stabilize the protein (76) or that surface charge removal can give fruitful stability enhancement (92). As a fact it has been proposed that protein stability can indeed be increased by optimization of surface electrostatics (24, 72).

When only considering charge-charge interactions of opposite sign, in the protein interior with a low dielectric constant, these are stabilizing. However, there is a large desolvation penalty accompanied with bringing a charge to a medium with a low  $\epsilon_r$ . It is argued that to have a single charge completely buried in the protein would be so unfavorable, as to unfold the protein and hence all charges inside proteins are surrounded by local protein dipoles and internal water molecules (66). However, it was recently shown that a stable protein has high tolerance for introduction of ionizable residues in the protein interior but with major shifts in  $pK_a$ -values (93). In the same protein a buried valine was replaced by aspartate and glutamate to illustrate the  $pK_a$  shifts accompanied with desolvating charges. The Asp had a  $pK_a$  of 8.9 and the Glu had a  $pK_a$  value of 8.8 which are +4.9 and +4.4 pH units higher than the intrinsic  $pK_a$  values of these residues (78, 87). These large shifts are rarely seen in proteins which indicates that single charges seldom are buried to that extent and that buried residues can have normal  $pK_a$  values due to local flexibility and water penetration (94). In contrast, it is believed that a protein can have two oppositely charged residues buried. This ion couple is sometimes called a salt-bridge and can have large functional importance (86). However, the contribution of salt bridges to protein stability is sometimes stabilizing and other times destabilizing (53).

## **1.6 The denatured state**

As pointed out in section 1.2 the denatured state is of great importance for protein stability and is thus briefly discussed below. The denatured state is in equilibrium with the native state and has a very low population under native conditions which makes it hard to study. Numerous studies have tried to picture the denatured state and there are indications that for many proteins this state is not completely unfolded (95-97). Moreover, there are several studies suggesting that the denatured state adopts conformations with dominant contributions from polyproline II structures (reviewed in (96)). Interactions persisting in the denatured state can have profound effects and



drastically stabilize or destabilize the native state. Studies have showed that electrostatic interactions in the denatured state destabilize the native state (98, 99). Introducing disulphide bonds, prolines or changing glycine to alanine stabilizes the native state by increasing the free energy of the denatured state (22).

It is of large interest to model the denatured state accurately in order to calculate protein stability. There are different indirect ways to study the denatured state such as mutational studies (98, 100) and more direct ways where the residual structure is investigated by hydrogen exchange of acid denatured protein (101). Moreover, large fragments of the protein have been used to model the denatured state ((102-104) and paper IV) as well as unfolded monomers of folded heterodimers (105).

### 1.6.1 The Gaussian-chain model

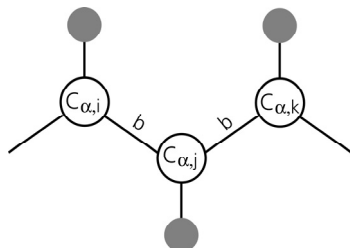
One way of treating the denatured state is to model it as a flexible polymer where the distance between any two residues ( $r$ ) is not fixed but varies according to (106):

$$p(r) = 4\pi^{-2} (3/2\pi d^2)^{3/2} \exp(-3r^2/2d^2) \quad (26)$$

where the root mean square distance between residues ( $d$ ) is dependent on the number of peptide bonds separating the two residues ( $n$ ), the effective bond length and a shift ( $s$ ) to account for that side-chains are stretching out from the back-bone:

$$d = bn^{1/2} + s \quad (27)$$

The mean electrostatic interaction energy between two residues can be calculated according to the Debye-Hückel theory using numerical integration. This model has been very useful to account for local and non specific electrostatic interactions in unfolded proteins. In a number of studies the pH-dependent stability calculated using the Gaussian chain model was shown to better reproduce stability data compared to using model values or a native-like model (106, 107).



**Figure 7.** Visualization of the Gaussian chain model where  $b=7.5 \text{ \AA}$  and  $n=1$  between  $i$  and  $j$  and  $n=2$  between  $i$  and  $k$ . To account for that the distance of interest is between two side-chains the shift  $s$  ( $5 \text{ \AA}$ ) is added according to Equation 27.

## 1.7 Intermolecular interactions

Interactions between molecules are governed by the same noncovalent molecular forces as in the intramolecular case. The understanding of inter molecular interactions is crucial in deciphering many biological aspects. For example it is crucial when optimizing a therapeutic protein that binds to a receptor (108) or when trying to characterize the interactions between proteins and nanoparticles (109-112). One difference between intramolecular and intermolecular interactions is configurational entropy. Configurational entropy has to do with the number of possible configurations of the molecular system and is the opposing force for assembly of molecules. This is the reason for the concentration dependence of association where a high concentration drives assembly.

There are several techniques to identify protein intermolecular interactions such as isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR) and surface plasmon resonance (SPR). One technique used and described in more detail in section 3.6 is Bimolecular fluorescence complementation (BiFC) and in particular a split GFP method.

## 1.8 Ligand binding

There is no difference between the models describing protein-ligand binding and protein-protein association. The ligand has a broad definition and can be a small ion (113), another protein (114) or a large nanoparticle (111). Affinity is a concept to describe how strongly the ligand is bound and is shown in the equilibrium constant. Usually the inverse of the association constant (dissociation constant) is reported and gives a measure of the free ligand concentration needed to reach half saturation (Figure 8a).

The simplest binding model is where the protein (P) binds only one ligand (L). Consider the following equilibrium:



If choosing the standard state to be 1 M the association constant is

$$K_A = \frac{[PL](1 \text{ M})}{[P][L]} \quad (29)$$

and the standard free energy is

$$\Delta G^\circ = -RT \ln(K_A) \quad (30)$$

Note here that the equilibrium constant and the standard free energy are dependent on the standard state. The complex is formed according to:

$$\frac{d[\text{PL}]}{dt} = k^{\text{on}}[\text{P}][\text{L}] - k^{\text{off}}[\text{PL}] \quad (31)$$

At equilibrium the complex concentration is constant so that:

$$k^{\text{on}}[\text{P}][\text{L}] = k^{\text{off}}[\text{PL}] \quad (32)$$

where

$$\frac{k^{\text{on}}}{k^{\text{off}}} = K_A \quad \text{and} \quad K_A = \frac{1}{K_D} \quad (33,34)$$

The degree of saturation is:

$$F_{\text{sat}} = \frac{[\text{PL}]}{[\text{P}]_{\text{tot}}} = \frac{[\text{L}]_{\text{free}}}{K_D + [\text{L}]_{\text{free}}} \quad (35)$$

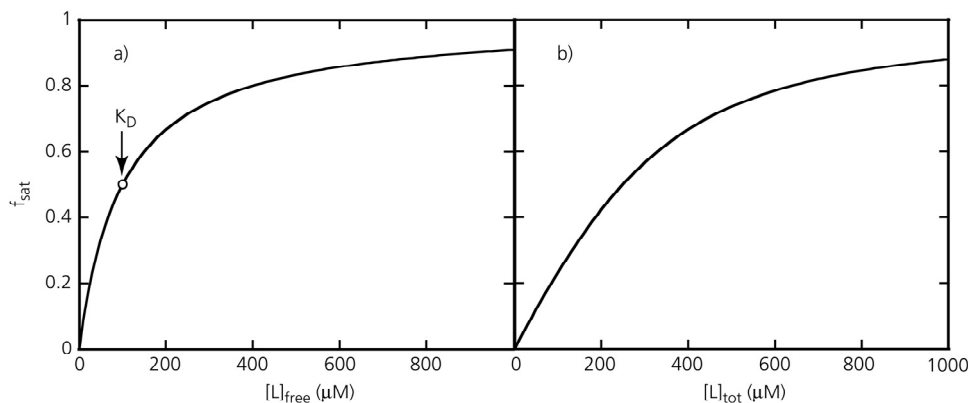
Usually one only knows the total ligand concentration,  $[\text{L}]_{\text{tot}}$ , and if stating the degree of saturation in terms of the  $[\text{L}]_{\text{tot}}$  instead the equation becomes more complicated:

$$F_{\text{sat}} = \frac{\frac{[\text{L}]_{\text{tot}} - [\text{P}]_{\text{tot}} - K_D}{2} + \sqrt{\frac{([\text{P}]_{\text{tot}} - [\text{L}]_{\text{tot}} + K_D)^2}{4} + [\text{L}]_{\text{tot}} \cdot K_D}}{K_D + \frac{[\text{L}]_{\text{tot}} - [\text{P}]_{\text{tot}} - K_D}{2} + \sqrt{\frac{([\text{P}]_{\text{tot}} - [\text{L}]_{\text{tot}} + K_D)^2}{4} + [\text{L}]_{\text{tot}} \cdot K_D}} \quad (36)$$

By following a binding reaction (e.g. using spectroscopy) with known total protein and ligand concentrations and fitting Equation 36 to data the dissociation constant can be determined. An example of a hypothetical binding curve using Equation 36 is shown in Figure 8b. Using the definition of the association and dissociation constants defined in Equation 29,  $K_A$  is dimensionless and is the more rigorous way of writing. However, using Equation 33 and 34 or ignoring the standard state in Equation 29,  $K_A$  and  $K_D$  have units in terms of  $M^{-1}$  and  $M$  respectively. This way of writing aid in knowing the free ligand concentration required to reach half saturation and is used frequently in the biological literature.

Just as for protein folding, the free energy of association is a balance between the lowest enthalpy and the highest entropy and association reactions display the same enthalpy-entropy compensation. Upon association the molecules lose configurational entropy but might gain in enthalpy and solvent entropy by the release of water from the binding site (reviewed in (13)). There are also studies showing that the conformational entropy of a protein upon binding a ligand can actually increase (115).

In many cases the protein has many binding sites for ligands and the binding of one ligand affects the binding of subsequent ligands. In these instances it is of interest to determine microscopic and macroscopic binding constants and the cooperativity of binding (116, 117). In the case of proton binding to ionizable groups there are usually too many ligands to obtain microscopic binding constants but the cooperativity can be observed in the shapes of the binding curves (see paper II).



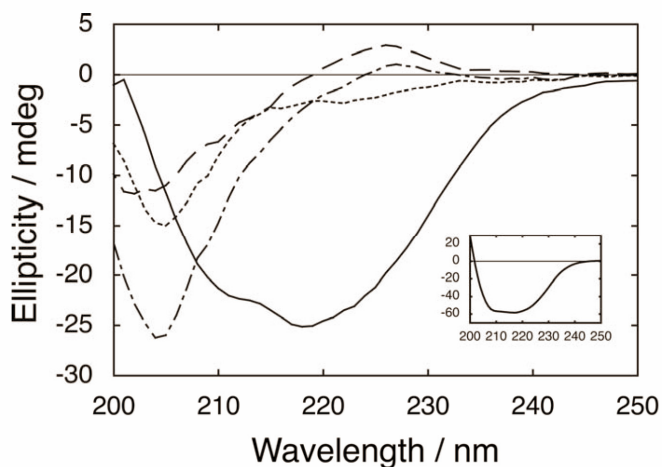
**Figure 8.** Hypothetical ligand binding obtained by using Equation 35 (a) and 36 (b) with  $K_D=100 \mu\text{M}$  and  $[P]_{\text{tot}}=300 \mu\text{M}$ .

## 1.9 Protein reconstitution

The phenomenon of protein reconstitution, or fragment complementation, was observed over 50 years ago for ribonuclease by Richards (118) and refers to the ability of dissected proteins to reassemble a functional protein with the same fold as the intact protein. It is now observed for many proteins (119, 120) and is thought to be an intrinsic property of proteins (119). The same molecular forces, described above, are involved in protein reconstitution as in folding of the intact chain. The main difference is the configurational entropy associated with assembly of molecules. Due to this close correlation between folding and reconstitution it is believed that there is a correlation of the affinity between fragments and the stability of the intact chain. This has been shown in a few cases (54, 121, 122) where the affinity between

reconstituting fragments of mutants is correlated with the stability of the intact chain of the same mutants. The correlation suggests that protein reconstitution can be a model to study the relative strength of different kinds of interactions contributing to protein stability under native conditions. Moreover, it proposes that by optimizing the affinity, the stability is also enhanced. The main advantage of protein reconstitution is that the contributions to the affinity between fragments can be measured under conditions where the protein is maximally stable. This is in contrast to denaturation studies that are investigated far from physiological conditions.

Fragment complementation is observed for all proteins used in this thesis (Figure 11) even though the affinity between the fragments span over several orders of magnitude. Experimentally, reconstitution can be studied using a wide range of biophysical methods such as NMR spectroscopy, CD spectroscopy, fluorescence spectroscopy, absorbance spectroscopy, SPR technology, ITC and gel electrophoresis. A detailed review about protein reconstitution is found in paper V.



**Figure 9.** Visualization of reconstitution for PGB1-QDD using CD spectroscopy. The spectrum for the N41 fragment (dashed line), C16 fragment (dotted line), the sum (dash-dotted line) and the mixture of equimolar of each fragment (solid line). The inset shows the spectrum of intact PGB1-QDD. The figure was adapted from (123).

## 2. PROTEIN SYSTEMS SELECTED

The results presented in this thesis were obtained mainly using three proteins that will be described briefly in the following sections. These proteins were selected for their characteristics to fulfill the needs of each project.

### 2.1 Protein G B1

Protein G is an immunoglobulin G (IgG) binding, multidomain cell surface protein, found in certain strains of *Streptococcus* and belongs to the same class of bacterial IgG binding proteins (124) as proteins A and L<sup>1</sup>. Protein A and G bind competitively to the same part of IgG even though there is little sequence or structural homology between the IgG binding domains (114). The coating with antibodies on the bacterial surface is thought to be a means for the bacteria to evade the host immune system (125).

Protein G has at least two domains, called B1 and B2 (abbreviated as PGB1 and PGB2), that bind the constant fragment of IgG with high affinity. The binding site on PGB1 was investigated and is thought to involve strand 3 and the helix (114, 126, 127). PGB1 is a 6.2 kDa protein consisting of 56 amino acids with a  $\beta$ -grasp fold<sup>2</sup> as determined both by NMR and crystallography (128, 129) (Figure 11). The protein has two  $\beta$ -hairpins (two antiparallel  $\beta$ -strands connected by a turn) and a parallel sheet formed by the N- and C-terminal strands. Altogether, the strands are grasping the helix. The protein has a buried tryptophan but contains no prolines or cysteines and is easy to express and purify. Due to its characteristics it has been one of the most widely used model proteins in biochemistry and biophysics.

The protein is thought to fold according to a two-state model, is very stable to high temperatures and denaturant concentration and is folded over a wide pH range (129, 130). The thermodynamic properties of the protein are well-characterized revealing a low  $\Delta C_{p, DN}^{\circ}$ . This is believed to be the reason for the high melting temperature despite a low  $\Delta G_{DN}^{\circ}$  at room temperature (130). The folding of PGB1 is initiated by the second  $\beta$ -hairpin (131) but the order of folding can be changed if the first hairpin is stabilized (132). The pK<sub>a</sub> values of the acidic groups in the protein were determined by proton NMR spectroscopy (133) and their possible ion-pairing to lysines (134).

---

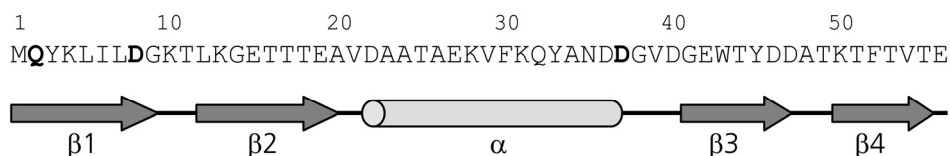
<sup>1</sup> Protein A and L: IgG binding cell surface proteins found in *Staphylococcus aureus* and *Peptostreptococcus magnus* respectively.

<sup>2</sup> Two-layer alpha+beta architecture in which four beta-strands form a mixed beta-sheet. Strands 1 and 4 are parallel and an alpha-helix connects strands 2 and 3 to form a right-handed beta-alpha-beta unit.

PGB1 can reconstitute between fragments comprising residues 1-40 and 41-56 even though with low affinity (135-137). Moreover, it is shown that the C-terminal  $\beta$ -hairpin folds at low temperatures and high salt concentration (123, 135, 138-140).

In the wild type (wt) protein N-terminal cleavage is observed but is minimized in the T2Q mutant (141). Asparagines followed by a glycine are very deamidation prone, especially at high temperatures and pH (142). In order to avoid mixtures of several species due to deamidation, asparagines 8 and 37 were mutated to aspartate. The new variant of PGB1 hence has the following mutations T2Q, N8D and N37D and is called PGB1-QDD. This variant is used in all PGB1 studies in this thesis.

The sequence and characteristics of PGB1-QDD are given in Figure 10 and Table 2.



**Figure 10.** Sequence and topology of PGB1-QDD. The mutated residues are in bold.

**Table 2. Characteristics of PGB1-QDD.** Thermodynamics and affinity to IgG of intact protein determined at pH 7.5 and low salt (150 mM for IgG binding studies).  $\Delta G^\circ$  was extrapolated back to 25°C or 0 M urea for thermal and urea denaturations respectively.  $\Delta H^\circ$  was determined at  $T_m$  or  $C_m$  and  $\Delta C_p^\circ$  was obtained from global fit (91). The isoelectric point was determined from residue specific  $pK_a$  values (143). The affinity between fragments was determined at pH 5.5, low salt and 25°C (123).

$T_m^*$	$\Delta G^{\circ\dagger}$	$\Delta H^{\circ\dagger}$	$\Delta C_p^\circ \ddagger$	$C_m^{\text{I}}$	$\Delta G^{\circ\dagger}$	$m^\S$	pI	log $K_A$ QDD•IgG	log $K_A$ N41•C16
56.3	16.2	198.9	1.8	4.1	12.1	2.9	4.3	7.74	4.3

Units expressed as  $^* \text{ }^\circ\text{C}$ ,  $^\dagger \text{ kJ/mol}$ ,  $^\ddagger \text{ kJ/(K}\cdot\text{mol)}$ ,  $^{\text{I}} \text{ M}$ ,  $^\S \text{ kJ/(mol}\cdot\text{M)}$

## 2.2 Calbindin $D_{9k}$

Calbindin  $D_{9k}$  is a mammalian calcium binding protein that is expressed in many organs such as the small intestine and kidney where calcium is taken up. The protein is believed to be involved in calcium uptake, buffering and transport. Bovine Calbindin  $D_{9k}$  (8.5 kDa) consists of 75 amino acids structured into two EF-hands which bind one calcium ion each (144) (Figure 11). Many of the residues are charged at neutral pH; 13 Glu, 4 Asp and 10 Lys; which give the protein a net negative charge. The protonation behaviors for most charged residues were investigated and are presented in references (145-147). The calcium binding occurs without any large

conformational change and is very strong and cooperative (148). Calbindin D<sub>9k</sub> is highly stable with a melting temperature of 85°C for the calcium free form (149).

The proline at position 43 has a significant cis-trans isomerization (150, 151). To avoid a mixture of protein structures a mutant is commonly used, where the P43M variant is a frequent choice. This variant also facilitates cyanogen bromide cleavage to obtain two fragments. These fragments reconstitute with high affinity and in a calcium dependent manner. The effect of hydrophobic core mutations was investigated and shown to be important for the affinity (121) while electrostatic interactions showed smaller effects on the affinity (152).

## 2.3 Green fluorescent protein

Green fluorescent protein (GFP) comes from the jellyfish *Aequorea victoria* and as the name implies fluoresces in green. Three scientists, Osamu Shimomura, Martin Chalfie and Roger Tsien were awarded the Nobel Prize in 2008 for the discovery and development of GFP. The discovery was made in the 1960's and today the use of GFP is widespread and applied in many areas of science and medicine (153).

The 27 kDa protein consists of 238 amino acids that fold into a  $\beta$ -barrel structure<sup>3</sup> made of 11  $\beta$ -strand with a single  $\alpha$ -helix crossing the interior (154, 155) (Figure 11). The astonishing aspect of this protein is that the green fluorescence is intrinsic in the amino acid sequence. Residues 65, 66 and 67 build up the chromophore that is located in the central helix. When the protein has folded into a near native structure these residues undergo an autocatalytic cyclization, with the use of molecular oxygen, to form a 4-(p-hydroxybenzylidene)-5-imidazolinone (154-156). The chromophore of wt GFP from *A. victoria*, comprising Ser65, Tyr66 and Gly 67, alternates between two ionization states with different excitation peaks at 395 and 475 nm. The different states, however, have similar emission peaks at 508 and 503 nm. By mutating serine 65 to threonine or cysteine the anionic species is favored and absorption occurs solely at 475 nm (157, 158). The GFP studies in this thesis, paper VI and VII, use a S65C variant with excitation at 475 nm and emission at 505 nm. The folding and stability of GFP has been investigated in depth. GFP can be reversibly denatured and the protein folds quickly but chromophore formation is slow (156, 157). However, GFP is found to misfold and aggregate and it is found that the protein folds better below 37°C (157).

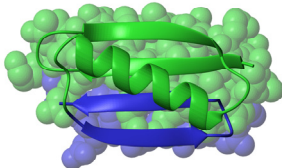
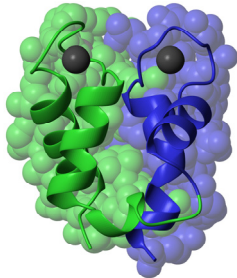
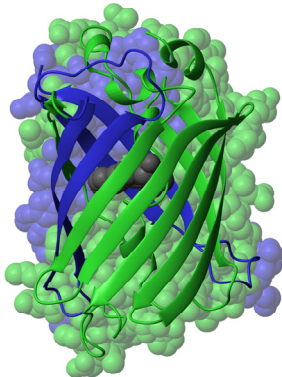
There are several GFP variants with different properties. Mutating, mainly the chromophore residues, the protein could fluoresce in different colors from blue to yellow (155, 159). By optimizing another fluorescent protein, DsRed from the coral

---

<sup>3</sup> several  $\beta$ -strands forming a closed structure in which the first strand is hydrogen bonded to the last



*Discosoma striata*, the red part of the spectrum was incorporated (159, 160). There are also mutants of GFP that have changed the properties of the protein rather than the color. There are variants with increased solubility, stability and improved fluorescence intensity (64, 65, 155, 157, 161)).

Protein:	Reconstituting fragments:	3-D structure:
<b>Protein G B1</b>	res 1-40 + res 41-56  $K_D=9 \mu\text{M}$	
<b>Calbindin D9k</b>	res 1-43 + res 44-75  $K_D=3 \text{ pM}$	
<b>Green fluorescent protein</b>	res 1-157 + res 158-238	

**Figure 11.** All proteins used in this thesis can reconstitute from fragments (GFP needs a fusion partner to reassociate) and the N- and C-terminal fragments are shown in green and blue respectively. The affinities between PGB1 and calbindin fragments were determined in (121, 137) respectively, but not determined for GFP. The figure was prepared using the pdb files 1pgb (PGB1), 4icb (calbindin D9k) and 1ema (GFP) and MOLMOL (162).

## 3. METHODS

### 3.1 Circular dichroism spectroscopy

Circular dichroism (CD) arises from the fact that chiral molecules (with chromophores) absorb left- and right-circular polarized light to different extents. It is the difference in absorbance between the two components that is measured and hence the signal can be both positive and negative. A drawback is that it has much lower sensitivity compared to regular absorbance since it measures a difference.

According to Beer-Lamberts law, the absorbance is:

$$A = \epsilon \cdot c \cdot l \quad (37)$$

$\epsilon$  is the absorption coefficient ( $M^{-1}cm^{-1}$ ),  $c$  is the concentration (M) and  $l$  is the sample pathlength (cm). CD is the difference in  $\epsilon$  between left (L) and right (R) polarized light according to:

$$\Delta\epsilon = \epsilon_L - \epsilon_R \quad (38)$$

The molar ellipticity,  $[\theta]$  (given in  $deg \cdot cm^2/dmol$ ), is related to  $\Delta\epsilon$ :

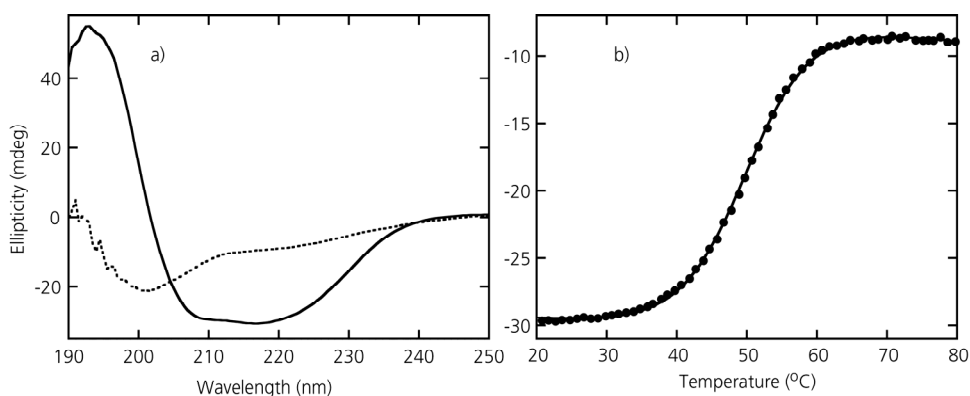
$$[\theta] = 3298 \cdot \Delta\epsilon \quad (39)$$

Briefly, in the CD spectrometer a light-source first produces high energy light that is passed through a monochromator to produce linearly polarized monochromatic light. The light goes through a photoelastic modulator to produce alternating left- and right- circular polarized light. The detector measures the light after the sample and if the sample contains chiral molecules the left- and right circular-polarized light will be absorbed to different extents. The voltage difference for the two components is measured and transformed into millidegrees of ellipticity.

Proteins are made by L-enantiomeric amino acids with a chiral  $C_\alpha$  (except for Gly) and will absorb left and right polarized light to different extents. In the far-UV CD region (250-170 nm) the peptide backbone is the chiral chromophore that gives rise to the CD signal. The far-UV CD signal depends on the secondary structural elements where the peptide bonds are located, but cannot except in rare cases, give information about the amino acids building up the structures. The secondary structural elements have characteristic spectra, but in a protein with a mixture of elements this will be reflected in the total signal (Figure 12a). Hence the CD spectrum can be used to determine the secondary structure of a protein, but with uncertainty. In  $\alpha$ -helices there will be a synergy effect giving rise to strong signals. The far-UV signal is sensitive since all peptide bonds contribute.

In the near UV CD region (350-250 nm) information about the tertiary structure can be obtained. The chromophores are the aromatic residues (Phe 250-270 nm, Tyr 270-290 nm and Trp 280-300 nm) and disulphide bonds (broad signals throughout the near UV-spectrum). Moreover, this region can, in combination with the far-UV region, give information about whether the protein has a well folded tertiary structure or simply secondary structural elements. The near-UV signal is very insensitive and a high protein concentration is usually needed.

CD is commonly used to observe structural differences between mutants and structural differences upon binding. Moreover, CD is frequently used to study protein denaturation (see section 1.3.1 and Figure 12b). For a recent review about CD see Kelly et al. (163).



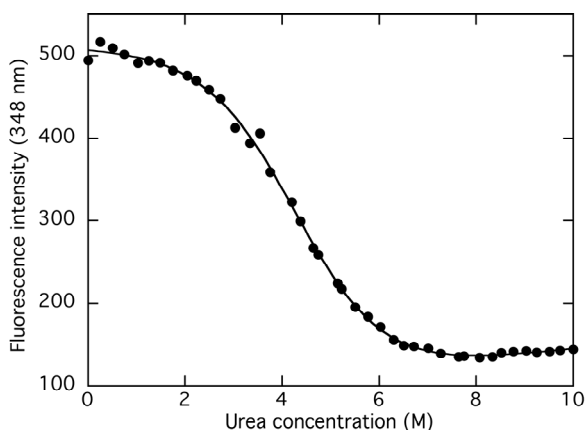
**Figure 12.** a) Comparison of far-UV CD spectra for PGB1-QDD at 20°C (solid line) and 95°C (dotted line). b) Thermal denaturation as observed from the loss in signal at 218 nm (solid circles) and fitting Equation 12 to data (solid line).

### 3.2 Fluorescence spectroscopy

Normally at room temperature molecules are in the electronic ground state. Upon absorption of photons, the electrons become excited to higher energy levels. In general the molecules are excited to higher vibrational levels as well. The excitation event is very fast but is followed by a slow vibrational relaxation to the lowest vibrational level in the excited state. The energy lost in this process is radiationless and the energy is lost as heat. For some molecules the following return to the ground state is accompanied by release of a photon, where this phenomenon is called fluorescence. The emitted photon always has lower energy than the incoming one.

Proteins contain three amino acids that fluoresce, tryptophan, tyrosine and phenylalanine<sup>4</sup>. Of these intrinsic fluorophores Trp dominates the fluorescence in most proteins. Moreover, the Trp fluorescence can be vastly different when located in different environments. A buried tryptophan usually has higher fluorescence than when solvent exposed, where collisions with the solvent quench the fluorescence. The emission wavelength can also shift significantly since the fluorescence is sensitive to the polarity of the environment with a shift towards longer wavelengths in a more polar environment. This makes Trp fluorescence suitable for protein denaturation studies (Figure 13).

There are also proteins with intrinsic fluorescence that does not stem from Trp, Tyr or Phe. The most widely known example is the green fluorescent protein (GFP) that shows green fluorescence (see section 2.3). The fluorophore is encoded in the protein sequence and is composed of three amino acids that make a cyclization reaction in the protein core (155, 159, 160). More examples of proteins with intrinsic fluorescence are DS red and derivatives thereof (160, 164).



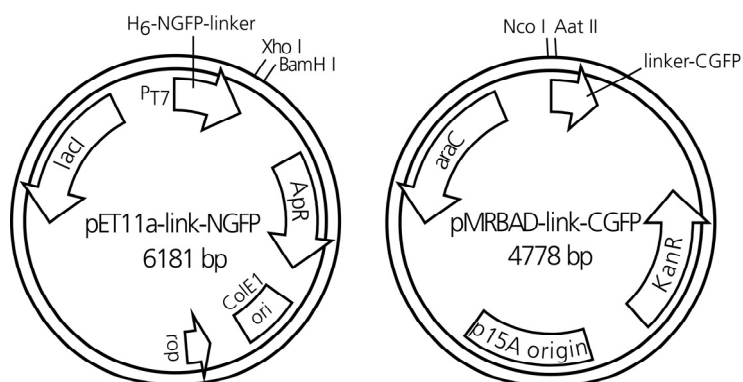
**Figure 13.** Urea denaturation of PGB1-QDD following the tryptophan fluorescence. Trp 43 is buried in the protein core in the native state but changes fluorescence drastically when the tryptophan becomes solvent exposed in the denatured protein.

---

<sup>4</sup>tryptophan  $\lambda_{\max}$  (abs)= 280,  $\lambda_{\max}$  (em)= 348, tyrosine  $\lambda_{\max}$  (abs)= 274,  $\lambda_{\max}$  (em)= 303 and phenylalanine  $\lambda_{\max}$  (abs)= 257,  $\lambda_{\max}$  (em)= 282

### 3.3 Bimolecular fluorescence complementation

Bimolecular fluorescence complementation (BiFC) is an assay based on the reconstitution of fragments of a fluorescent protein with the aid of interacting partners fused to the fragments. The method is very useful to visualize interactions *in vivo* and no exogenous components perturbing the cells are required. Many of the methods rely on complementation of GFP or derivatives thereof. BiFC has been proven to be useful to show interactions in various cells as bacteria, yeast, plants and mammalian cells (165-171). One major drawback of the method is the slow rate of the reactions required for fluorophore maturation (172). Therefore the complex formation cannot be studied in real-time and the complex formation is usually irreversible. These aspects of the methods, however, also contribute with some advantages in that transient and low-affinity interactions can be observed that are out of reach with other methods. For a review about BiFC, see for example Kerppola 2008 and 2009 (172, 173).



**Figure 14.** Vectors used in the split GFP system. The fusion partners are ligated into the vectors using the XhoI/BamHI and NcoI/AatII restriction sites for pET11a-link-NGFP and pMRBAD-link-CGFP, respectively. The figure was prepared using the vector sequences found at <http://www.yale.edu/reganlab/publications.html> and the program pDRAW32 found at (<http://www.acaclone.com>).

#### 3.3.1 Split GFP system

One variant of BiFC is the split GFP method where GFP is the fluorescent protein. In one system developed by the group of Lynne Regan, GFP is dissected between residues 157 and 158 to yield two fragments. These fragments have too low affinity to be able to reconstitute, but if interacting partners are fused to the ends of the fragments the local concentration of the GFP fragments is increased and GFP

associates to form the native fold with chromophore development (168, 169, 171, 174).

The cloning into this system makes use of two different vectors (Figure 14) in order to control the expression levels of the different fragments (169). Moreover, the vectors have different cloning sites with separate restriction pairs. The two vectors have distinct promoters where one vector is induced by Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and one with arabinose. The antibacterial resistance is also different with ampicillin and kanamycin resistance for pET11a-link-NGFP and pMRBAD-link-CGFP respectively. Hence, bacterial growth on a medium containing both antibiotics ensures incorporation of both plasmids.

### 3.4 Differential scanning calorimetry

In DSC the heat capacity,  $C_p$ , of the solution of interest is compared to the  $C_p$  of a reference solution and monitored as a function of temperature. The cells containing the solutions are kept at constant pressure in an adiabatic shield. The difference in applied power required to maintain the sample and reference cells at the same temperature is accurately monitored as a function of temperature.

DSC is a common way of studying heat denaturation of proteins (see section 1.3.1). The  $C_p$  of a protein undergoing an unfolding event changes as in Figure 15. Before the unfolding the heat capacity of the native state,  $C_{p,N}$ , is displayed and after the unfolding the heat capacity of the denatured state,  $C_{p,D}$  is shown. The difference between the two is  $\Delta C_{p,DN}$ .

$$\Delta C_{p,DN} = C_{p,D} - C_{p,N} \quad (40)$$

where the observed heat capacity is

$$C_{p,obs} = F_N C_{p,N} + C_{p,exc} + F_D C_{p,D} \quad (41)$$

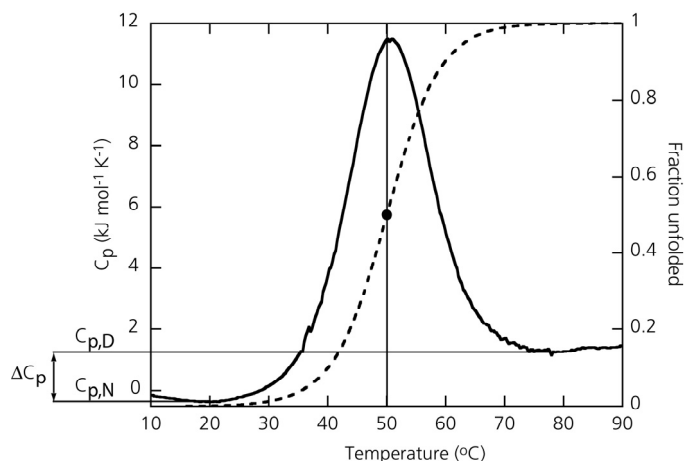
The excess heat capacity,  $C_{p,exc}$ , observed during the denaturation event is due to that the supplied energy is used to shift the protein molecule to structures with higher energy. Hence, during the transition more heat is required in order to increase the temperature of the solution. The area under the excess  $C_p$  curve is the enthalpy change of the denaturation,  $\Delta H_{cal}$ .

$$\Delta H_{cal} = \int C_p^{exc} dT \quad (42)$$

This enthalpy is model independent. If assuming a two-state model the enthalpy of denaturation can also be calculated using van't Hoff's equation

$$\Delta H_{\text{vHoff}} = RT^2 \frac{d \ln K}{dT} \quad (43)$$

Comparison of the two enthalpies ( $\Delta H_{\text{vHoff}}/ \Delta H_{\text{cal}}$ ) gives a measure of how well the denaturation follows a two state process. A ratio below one could be due to that the protein has intermediates or forms oligomers (175).



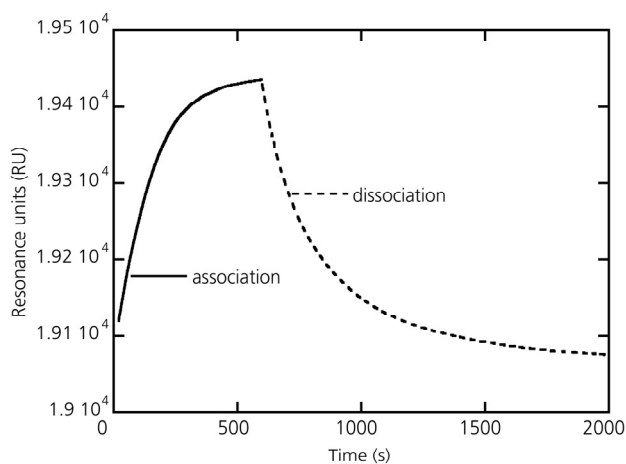
**Figure 15.** DSC of PGB1-QDD at pH 7.5 and low salt. The raw DSC trace (after buffer subtraction) is shown as a solid line. The fraction denatured is shown in the dashed line and the melting temperature is shown with a filled circle and solid vertical line.  $C_{p,N}$ ,  $C_{p,D}$  and  $\Delta C_{p,DN}$  are indicated to the left.

### 3.5 Surface plasmon resonance technology

SPR technology is a useful method to study protein interactions because it is possible to obtain both affinities and kinetics. Moreover, it is a general method that does not rely on specific chromophores and it has a broad affinity range. A drawback is that interactions are studied at a surface, instead of in solution, and surface effects may perturb the binding parameters.

The optics of an SPR instrument consists of a metal coated prism. Electromagnetic radiation from a light source is used to excite surface plasmons in the metal film. The plasmons are only excited by light at a certain incidence angle. At this resonance angle the reflected light has a sharp minimum and it is this angle dependent intensity that is measured. The resonance angle is dependent on the refractive index of the sample above the metal film and association and dissociation of macromolecules on the surface changes the refractive index. This makes the resonance signal proportional to

the mass bound at the surface. In a typical experiment one component is immobilized at or close to the metal film and the other component, dissolved in the buffer of choice is continuously flown over the surface. The amount bound is monitored as a function of time and gives the association phase (Figure 16). After saturation of the surface, one switches to flowing buffer over the surface and then dissociation is measured (Figure 16). Dissociation should be concentration independent and usually a single exponential can be fitted to these data. Association is on the other hand concentration dependent so in order to extract the association rate constant from the equation fitted to association data one needs to know the concentration.



**Figure 16.** Binding of PGB1-QDD to IgG as determined from SPR measurements where the association and dissociation phases are indicated. IgG was immobilized by amine coupling to a C5 chip and PGB1-QDD flown over the chip. The pH was 7.5 and the ionic strength 150 mM.

### 3.6 Nuclear magnetic resonance spectroscopy

After the development of Fourier transform NMR spectroscopy (176) and multidimensional NMR spectroscopy (177) the method experienced an enormous growth in studying proteins. NMR spectroscopy is one of the few methods used to determine protein structures, but it is also a very useful method to provide a dynamic picture at atomic resolution of events going on in biological processes. For example, NMR spectroscopy can be used to measure ligand binding of potential drug candidates to a specific receptor, both in terms of affinities and which residues are involved (178). Moreover, it is very useful in determining kinetics and dynamics (179, 180). Following, will be a short section about NMR theory. For a more elaborate description the reader is referred to Cavanagh et al. (181).



All nuclei have an intrinsic angular momentum called spin which is expressed by the nuclear spin quantum number  $I$ . An NMR active nucleus has a nonzero  $I$  that gives rise to angular momentum of the nuclei. For proteins one mostly studies the nuclei  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  that have  $I=1/2$ . The spin angular momentum,  $\mathbf{I}$ , is a vector quantity, but due to the quantum mechanical nature of spin only one Cartesian component can be specified simultaneously. Normally the z-component is expressed,

$$I_z = \hbar m \quad (44)$$

where  $m$  is the magnetic quantum number that has  $2I+1$  values in integral steps ( $-I, -I+1, \dots, I-1, I$ ). So for  $I=1/2$  and  $m=-1/2, 1/2$  the value of  $I_z = -1/2\hbar, 1/2\hbar$ . Nuclei with  $I \neq 0$  also have a nuclear magnetic moment,  $\mu$ , which is defined as

$$\mu = \gamma \cdot \mathbf{I} \quad (45)$$

where the constant,  $\gamma$ , is the gyromagnetic ratio and has distinct values for different nuclei, for example,  $\gamma_{^1\text{H}} = 26.75 \cdot 10^7 \text{ rad} \cdot \text{T}^{-1} \cdot \text{s}^{-1}$ ,  $\gamma_{^{13}\text{C}} = 6.73 \cdot 10^7 \text{ rad} \cdot \text{T}^{-1} \cdot \text{s}^{-1}$  and  $\gamma_{^{15}\text{N}} = -2.71 \cdot 10^7 \text{ rad} \cdot \text{T}^{-1} \cdot \text{s}^{-1}$ . Hence the magnetic moment is different for separate nuclei, even though they have the same spin angular momentum.

In the absence of an external magnetic field the spin angular momentum vectors do not have a preferred orientation since all values of  $m$  have the same energy. In the presence of a magnetic field the different spins will have different energies. If just considering the z-components the energy is:

$$E = -\mu_z B_0 = -\hbar m \gamma B_0 \quad (46)$$

where  $\hbar$  is Planck's constant ( $h$ ) divided by  $2\pi$  and  $B_0$  the applied magnetic field. The spin oriented in the same direction as the magnetic field has the lowest energy. Thus, in a magnetic field the  $2I+1$  different energy levels will have a slightly unequal population (according to the Boltzmann distribution) and will depend on the nuclei and magnetic field strength. It is this small population difference that is measured and is the reason why NMR spectroscopy is a very insensitive method compared to other spectroscopic methods.

### 3.6.1 Chemical shift

The NMR frequency of a nucleus in a molecule depends primarily on the field strength and the gyromagnetic ratio,  $\gamma$ , of the nucleus. For example  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  will be observed at very distinct frequencies since they have different  $\gamma$ . However, not all protons or carbon nuclei will be observed at exactly the same frequency but rather this will depend on the local environment. This slight shift in frequency is called the

chemical shift and is measured in parts per million (ppm) compared to a reference frequency of a standard molecule. The reason for chemical shift is that the actual magnetic field experienced by a nucleus located in a molecule is somewhat different compared to the magnetic field the nucleus would experience if the electrons were removed. The applied magnetic field induces motions of the electrons generating secondary magnetic fields decreasing or enhancing the applied magnetic field. The magnetic field observed by the nucleus is hence smaller or larger than the applied magnetic field and is referred to as the nucleus being shielded or deshielded with the shielding constant  $\sigma$  and the frequency of the nucleus is:

$$\nu = \frac{\gamma B_0 (1 - \sigma)}{2\pi} \quad (47)$$

However, since  $\sigma$  depends on the applied magnetic field and is difficult to determine the chemical shift is measured relative to a reference compound:

$$\delta = 10^6 \frac{(\nu - \nu_{\text{ref}})}{\nu_{\text{ref}}} \quad (48)$$

The chemical shift differences (in ppm) measured at different field strengths can conveniently be compared. The reference compound is usually one with shielded nuclei so that this signal is observed at the low frequency end of the spectrum and most other nuclei of the same kind in other local environments get positive  $\delta$ .

### 3.6.2 Isotope labeling

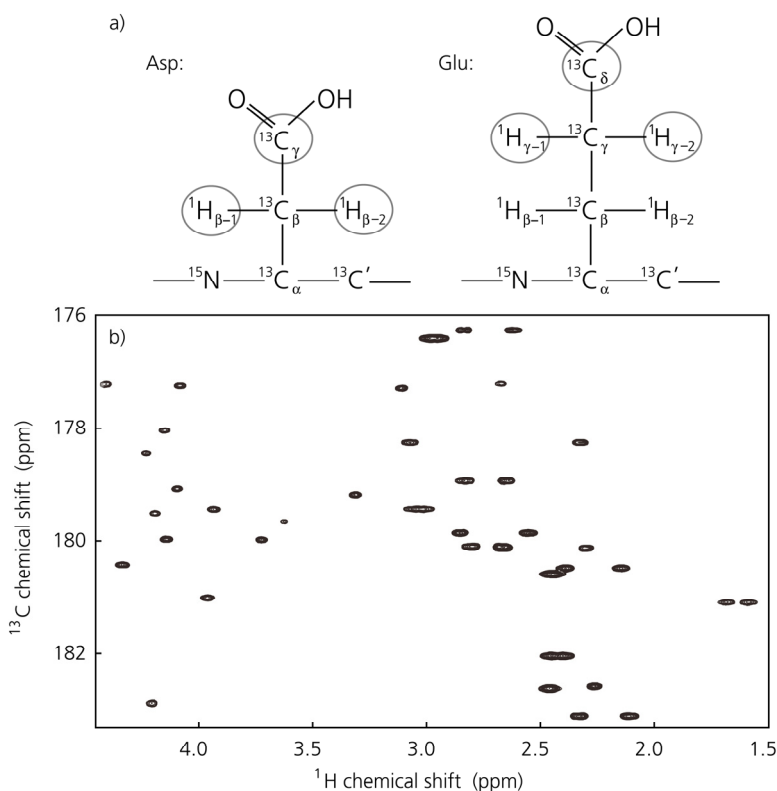
Proteins are mostly composed of hydrogen, carbon, nitrogen and oxygen. The natural isotopes of these residues are  $^1\text{H}$ ,  $^{12}\text{C}$ ,  $^{14}\text{N}$  and  $^{16}\text{O}$ . Of these it is only  $^1\text{H}$  that has a nuclear spin quantum number of  $\frac{1}{2}$  that is the most convenient spin to study. Therefore, when using NMR, one normally expresses proteins in media enriched with isotopes more suitable for NMR, where the most common isotopes are  $^{13}\text{C}$  and  $^{15}\text{N}$ .

### 3.6.3 Two and multidimensional NMR spectroscopy

When dealing with protein NMR spectroscopy the traditional one-dimensional NMR experiment is not enough to resolve all the atoms in the molecule. More dimensions have to be invoked to obtain a reasonable resolution. Therefore, protein NMR spectroscopists today mostly deal with two- or multidimensional NMR spectroscopy. In these experiments the chemical shift of one nucleus is correlated with the chemical shift of at least one other nucleus.

One of the most common two dimensional experiments for proteins is the  $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) experiment (182) where the

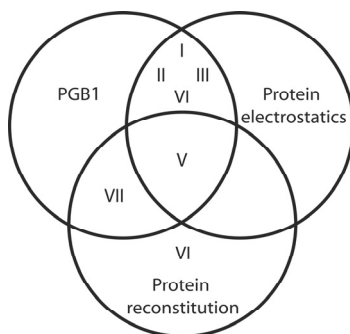
backbone amide nitrogens and protons are correlated in a two dimensional experiment and gives a fingerprint of the protein. An example of a spectrum from such an experiment is shown in paper I, Figure 1a. In this thesis NMR spectroscopy has mostly been used to determine the protonation/deprotonation events occurring at negatively charged residues in proteins. By doing so, information about electrostatic interactions in proteins can be obtained at atomic resolution. Here one utilizes that the chemical shift is sensitive to changes in electrostatic potential at nearby atoms. One experiment used in these studies correlates the side-chain carboxyl carbon chemical shift with the  $\beta$ - or  $\gamma$ -protons for Asp and Glu respectively (183). In Figure 17 such a spectrum is shown for PGB1-QDD. As the carbonyl group becomes deprotonated the  $^{13}\text{C}$  chemical shift increases and can easily be determined as a function of pH. See paper IV, Figure 3 for an illustration of how the chemical shift changes with pH.



**Figure 17.** a) The nuclei that are correlated in the spectrum (circled) for Asp and Glu. Since there are two  $\beta$ - and  $\gamma$ - protons for Asp and Glu respectively there can be two proton signals for every carbon signal stemming from that the protons can have slightly different local environments. b) H(C)CO spectrum for PGB1-QDD at pH 5. Signals to the right show side-chain carboxyl groups. To the left back-bone carbonyl groups are also displayed.

## 4. INTRODUCTION TO PAPERS

During my thesis work I have had the opportunity to study different aspects of various protein puzzles which has resulted in many papers where seven of these papers form the basis of this thesis. In Figure 18 a Venn diagram illustrates the connections between the presented papers. Following will be a short description of the different projects and papers.



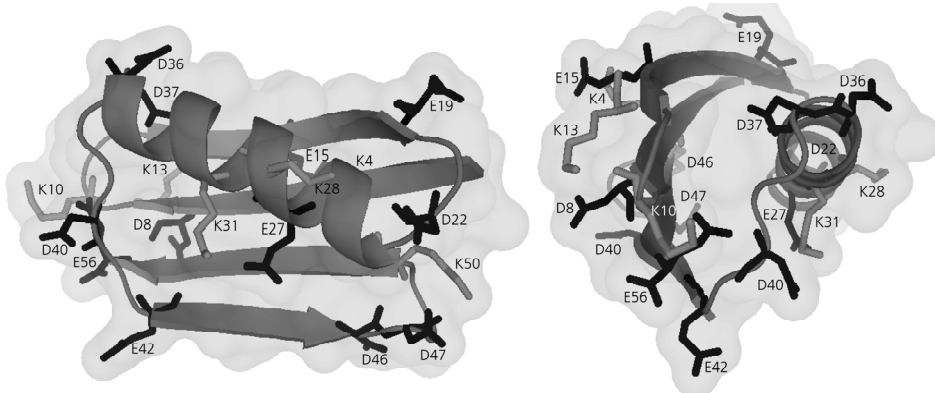
**Figure 18.** The papers in this thesis are put into categories in a Venn diagram (184) where connections between them are shown.

### 4.1 Electrostatic interactions in PGB1; papers I, II, III and IV

Proteins have at neutral pH a high number of charges, which affect protein stability even though they are not introduced to improve the stability. Elucidation of electrostatic interactions in proteins is a highly complex matter since there are many charges interacting in a coupled manner. The high charge density of PGB1-QDD makes this an excellent protein for studies of electrostatics in proteins. The protein is small, stable and has many surface exposed charges; 7 Asp, 5 Glu and 6 Lys (Figure 19).

#### 4.1.1 Paper I

pH modulates the charge distribution in a protein and is hence a means to study the effect of electrostatic interactions in proteins. Introduction of salt is anticipated to screen most electrostatic interactions expecting the stability to be more or less pH independent in a salt solution. However, the net stability might be affected by salt because attraction and repulsion might be of different magnitudes.



**Figure 19.** PGB1-QDD is a highly charged system with 20 surface exposed charges at pH 7. The protein is shown in two orientations, related by a 90 degree rotation out of the paper where negatively and positively charged side-chains are shown in black and grey respectively. The figure was prepared using the pdb file 1pgb and PyMOL (DeLano Scientific) where T2Q, N8D and N37D were mutated into the structure.

In paper I the pH-dependent stability of PGB1-QDD studied at different salt concentrations is presented. The stability was measured both by thermal and chemical denaturation using several methods; CD and fluorescence spectroscopy and DSC. At low salt concentration, the protein showed a clear pH-dependent stability with an optimum around pH 4.5, which is close to the isoelectric point. At high and low pH the stability decreased but with a larger loss at high pH indicating that the protein is more tolerant to acidic than basic pH.

Due to the high charge density in the protein salt was anticipated to diminish the pH-dependence in the stability. At physiological salt concentration, however, the pH dependence was surprisingly only mildly reduced. Still at 2 M salt the stability was pH-dependent and the screening of charges was different at low and high pH. It was found that salt was able to screen electrostatic interactions at low, but not at high net charge. From the drastic drop in stability between pH 6 and 10 one could reason that there must be residues that have unequal  $pK_a$  values in the native and denatured states persisting even at a high salt concentration. At the low pH values salt screened efficiently suggesting that residues titrating in this region have the same  $pK_a$  values in the native and denatured states at a high salt concentration. In all, the results from the study put an interest in  $pK_a$  determinations at different salt concentrations.

#### 4.1.2 Paper II

$pK_a$  values of individual residues can give useful insight into electrostatic interactions and contributions of electrostatics to protein stability. By determining the  $pK_a$  values

at different salt concentrations information about direct electrostatic interactions can be obtained. In order to get a better understanding of the pH-dependent stability of PGB1-QDD heteronuclear NMR spectroscopy was used to get residue specific  $pK_a$ -values at two salt conditions; no added salt and 0.5 M NaCl. From the  $^{13}\text{C}$  chemical shift of the carbonyl carbon of Asp and Glu residues as a function of pH the  $pK_a$  values were determined. The results showed that there was a large spread in the  $pK_a$  values where some residues had highly upshifted values whereas some residues had highly downshifted values. The titration curves were investigated in more detail and new tools to investigate electrostatic coupling were derived. For some residues the modified Henderson-Hasselbalch Equation (22) did not fit to data due to asymmetric titration curves so instead the data was fitted with a model taking into account electrostatic interactions with other residues and gave excellent agreement.

The derivatives of the titration data were calculated to show the proton binding capacitance as a function of pH and this gave information of electrostatic coupling in the protein. An asymmetry in proton binding capacitance as a function of pH could also be captured. Moreover, it was observed that residues titrating in the same interval had extended capacitance curves while residues with highly shifted  $pK_a$  values had capacitance curves similar to ideal Asp and Glu titrating in the absence of other charges. Electrostatic coupling was further investigated from the pH dependence of the  $pK_a$  values. It was shown that for an ideal titration event, following Equation 21, there is no pH dependence in the  $pK_a$  values while for a titration event following Equation 22 there is a linear dependence.

The  $pK_a$  values shifted surprisingly little with the addition of 0.5 M salt. On the other hand, the capacitance curves changed and were more similar to ideal. It was reasoned that the electrostatic coupling was largely screened by salt even though not visualized in shifted  $pK_a$  values. Moreover, it was suggested that the highly shifted  $pK_a$  values at low salt concentration were shifted for other reasons than direct charge-charge interactions. Titration data from  $^1\text{H}$  chemical shifts were also analyzed and showed that proton data are unreliable due to the sensitivity to other events occurring in the protein.

### 4.1.3 Paper III

The pH-dependent stability of a protein can be directly calculated from  $pK_a$  values in the native and denatured states (Equation 23) and be compared to experimental denaturation data such as obtained in paper I. In paper III this comparison was performed; the  $pK_a$ -values presented in paper II were used to calculate the pH-dependent stability and compared to the denaturation data derived in paper I. Since no  $pK_a$  values of the denatured state were obtained these had to be calculated and a Gaussian chain model (section 1.6.1) was used. Using this model for the denatured state and experimental values in the native state there was a discrepancy between calculated and measured stability. Better comparison was obtained when shifts in the

denatured state were introduced. The results indicated either evenly shifted  $pK_a$  values in the denatured state or residual structure that introduced electrostatic interactions yielding shifted  $pK_a$  values for some residues.

From the pH dependence of  $^{13}\text{C}$  chemical shifts of the backbone carbonyl groups specific interactions could be revealed. Carbonyl groups displaying pH dependent chemical shifts with an apparent  $pK_a$  value traced hydrogen bonding patterns. Even though the calculated pH-dependent stability explained major parts of the experimental stability curve there were still unanswered questions. The question marks arose mainly from the denatured protein with an urge to study this state.

#### 4.1.4 Paper IV

Since protein stability is defined as the difference between the native and denatured states it is of great importance to study the denatured state. However, it is hard to monitor this state under native conditions due to the low population. Usually one has to rely on mutational studies, calculations or use denaturants to investigate interactions in the denatured state. In the study presented in paper IV we used unfolded fragments of PGB1-QDD as models for the denatured state of the protein. Using heteronuclear NMR spectroscopy  $pK_a$  values in the unfolded fragments were determined and showed that most electrostatic interactions were removed. Calculations were performed to obtain shifts in  $pK_a$  values due to cleavage and introduction of new termini. Based on the  $pK_a$  values in the fragments (corrected for the cleavage) and  $pK_a$  values in the native state presented in papers II and III, the pH dependent stability was reexamined and was found to give a good comparison to stability data. The result indicated that there is no residual structure in the unfolded state and using fragments as a representation seems to be an accurate model.

#### 4.1.5 Conclusions papers I-IV

The protein under investigation, PGB1, is a bacterial protein functioning under physiological conditions; i.e. 150 mM salt and pH 7. However, the pH of optimal stability, 4.5 and low salt, is far from this pH and salt concentration and hence it can be concluded that the charges are not introduced to increase protein stability. Rather, the high charge density of the protein presumably prevents the protein from aggregation with other protein molecules of the same and different kinds.

Other general findings from the study of electrostatics in the model protein PGB1-QDD are that in a protein without specific ion-pairs it is hard to resolve all pair interactions in the protein. Rather, the protein should be seen as a network of charges with a wide range of  $pK_a$  values. Moreover, the titration events of charges titrating in the same pH interval are affecting each other and give broad charge capacitance curves. Residues titrating in ranges where few other residues titrate have shifted  $pK_a$  values but the titration processes are close to ideal.

High concentration of sodium chloride screens electrostatic interactions in the low pH range but only marginally in the high pH range. A possible explanation is that charges titrating in this pH interval (D37 and the N-terminus) have shifted  $pK_a$  values of non-electrostatic character such as desolvation and hydrogen-bonding. The downshifted  $pK_a$  values, on the other hand, arise due to both electrostatic interactions and hydrogen bonding where the former is efficiently screened by salt. The pH dependent stability can accurately be calculated from the  $pK_a$  values in the native and denatured states, but better models for the denatured state are needed. Moreover, stability data obtained without lengthy extrapolation are necessary to move the accuracy forward. Possibly affinity studies under native conditions using reconstituting fragments can play a role here.

## **4.2 Reconstitution, domain swapping and general protein energetics; paper V**

### *4.2.1 Paper V*

Paper V is a review article and is included in the thesis because it connects the projects (Figure 18) and adds some thoughtful general insights into protein energetics that aid in reading the other papers.

The close connection between protein reconstitution, domain swapping and folding of a single chain blurs the distinction between folding and binding. The concepts are actually manifestations of the same non-covalent interactions. Reconstitution studies can give valuable insight into protein folding and stability and possibly help to understand even higher assemblies such as amyloids.

## **4.3 Split GFP in protein stabilization; Papers VI and VII**

### *4.3.1 Paper VI*

In reconstitution, fragments of a protein reassemble and fold into the same three-dimensional structure as the intact chain in its native state. This has been utilized in split GFP methods where two fragments of GFP are used as reporters for interacting partners. The fragments of GFP have too low affinity for one another to allow for spontaneous reassembly, but if two interacting partners are fused to the two GFP fragments GFP can be reconstituted and display green fluorescence. Hence, a green sample indicates interactions between fused partners.

This method has been used many times before to detect interactions on a qualitative level. Here it is used on a more quantitative level to see if the fluorescence intensity reflects affinity differences. To investigate this, the two EF-hands of calbindin D<sub>9k</sub> were fused to the split GFP fragments. One fragment contained the N-terminal parts of GFP and calbindin, GFPN-EF1, and one fragment the C-terminal parts of



calbindin and GFP, EF2-GFPC. When the two fragments were expressed in *E. coli* on agar plates the resulting colonies turned green. This proved an interaction between the fragments. Moreover, the expression in *E. coli* was monitored over time in 96-well plates and showed stronger fluorescence intensity compared to a reference construct with lower affinity indicating that the high affinity between the EF-hands was reflected in the bright GFP fluorescence.

The complex was also purified and studied *in vitro* and this showed that the complex was associated and folded up to higher temperature in the presence compared to absence of calcium as visualized from the loss in GFP fluorescence. This suggests that the affinity between the fragments in the split GFP system follows the same calcium dependence as observed for the EF-hands themselves. To test this SPR was used to determine the affinity between the fragments and showed that the affinity between the EF-hands in the split GFP system was increased drastically compared to the affinity between the EF-hands themselves. Moreover, the affinity showed the same calcium dependence as observed for the fragments themselves.

From the results it was concluded that affinity differences resulted in fluorescence intensity differences. Based on this finding it is tempting to use the method to screen a library for variants of increased fluorescence and investigate if this in turn would be correlated to higher affinity between fragments and higher stability of the corresponding intact chains without the fusion to GFP. If this was the case the method would be suited for stability optimization.

#### 4.3.1 Paper VII

The results presented in paper VI suggests that reconstitution in the split GFP method is suitable for optimizing protein stability. An excellent test case is PGB1-QDD as this protein has moderate affinity between reconstituting fragments. Enhanced affinity would, based on the previous study, be reflected in increased fluorescence compared to the parent protein. Moreover, PGB1 has been stabilized before so that mutations with known stability enhancement could be introduced into the library to show proof of principle.

As a starting point the fragments consisting of residues 1-40 and 41-56 of parent PGB1-QDD were cloned into the vectors containing the split GFP fragments. Co-expression of the plasmids yielded faint green colonies and indicated successful reassembly of PGB1-QDD and GFP. In the next step a small and condensed library of the N-terminal fragment of PGB1-QDD was cloned into one of the split GFP vectors. The library was designed to contain mutations with stability increasing and decreasing effects. Due to changes in the genetic code several other mutations arose to generate a library of 3456 possible mutations.

From co-expression of the plasmid containing the C-terminal part of GFP plus the library of residues 1-40 of PGB1-QDD with the plasmid containing the N-terminal part of GFP plus residues 41-56 of PGB1-QDD green colonies arose. After several

rounds of re-streaking of colonies and visual inspection in comparison to parent PGB1-QDD construct, colonies with brighter fluorescence could be identified. The colonies were ranked according to green fluorescence and the DNA of the split GFP plasmids containing PGB1-QDD 1-40 was sequenced. The selected sequences contained similar mutations compared to the previously reported stabilizing mutations.

To test if the enhanced green fluorescence also reflected increased stability of the full-length protein the DNA for residues 1-40 of the three top clones and the wt DNA of residues 41-56 was produced in expression vectors. The top three variants were expressed, purified and tested in mass spectrometry and gel filtration to ensure purity and verify that the protein was monomeric. Spectral analysis and thermal denaturation using CD spectroscopy at physiological conditions was performed for the top three variants and compared to PGB1-QDD. The spectra of the clones were almost identical to the parent protein. Results and analysis of temperature denaturation studies showed that the top three candidates all were stabilized compared to the parent protein. The Top1, 2 and 3 candidates showed 12, 9 and 8°C increase in  $T_m$  compared to the parent protein. Altogether the results showed that reconstitution in the split GFP method was useful to identify variants with increased stability. The increased GFP fluorescence reflected increased stability of the protein fused to the GFP fragments.

In continuing studies it would be of interest to investigate the affinity between fragments constituting the top three clones and perform denaturation studies at different pH to see if the introduced mutations only shifted the pH optimum or are stabilized throughout the pH range. Based on the promising results general guidelines for using the method to stabilize a protein were set up.

## 5. CONCLUSIONS

The work presented in this thesis contains detailed investigations about non-covalent interactions in proteins. A detailed description about the non-covalent interactions governing protein folding and stability are required for predicting the structure of an amino acid sequence. Charge interactions are crucial for proteins, but their contributions are complicated to calculate. The results from paper I-IV can aid in refining models treating electrostatic interactions in proteins.

Protein optimization is normally used to improve the functionality of a protein. However, many times proteins lack stability and also need to be improved in this aspect. Techniques that can find stabilized variants without the need for clever design are useful in wide areas of research and development such as in the pharmaceutical industry. The results presented in papers VI and VII suggest that reconstitution in split GFP is a promising method to optimize protein stability.

## 6. ACKNOWLEDGEMENT

Och så det roligaste, att tacka alla underbara människor som funnits där för mig på ett eller annat sätt...

Först och främst ett stort tack till min fantastiska handledare **Sara**, som möjliggjort denna avhandling. Du har lotsat mig fram på den krokiga doktorandstigen utan att tappa tron på mig. Jag beundrar din makalösa förmåga och kompetens i allt från vetenskap till att skriva barnböcker, springa snabbt och vara en livsnjutare. Tack för din aldrig trytande ork, din enorma omtänksamhet och för att du inspirerar alla i din omgivning!

Övriga seniorer: **Mikael A**, biträdande handledare, med en aldrig sinande optimism och en förmåga att styra upp ett manuskript. **Bertil**, tack för dina djupgående diskussioner och förklaringar vid seminarier och i kurskompendier. Du har en förmåga att sätta fingret på det som annars är otydligt och svårt att förstå. **Torbjörn**, tack för att du alltid tar dig tid till stora som små NMR-problem. **Kristofer**, med din pedagogiska förmåga blir allt lättförståeligt. Tack för att du tar dig an PGB1. **Bengt**, din aldrig trytande entusiasm att förklara och diskutera gör det fantastiskt lärorikt, stimulerande och roligt att undervisa tillsammans med dig. Tack för svaren på alla mina termodynamiska problem och dina tänkvärda reflektioner över det viktiga i livet! **Jan-Erik**, den atmosfär du sprider har gjort det utvecklande men även väldigt trevligt att undervisa. Alltid lika kul att höra om dina resor och äventyr!

I Saras grupp: **Olga**, du är en fantastisk vän och kollega med en förmåga att sprida värme till alla i din omgivning. Genom åren har det blivit en rad galna fester (en del riktigt B), äpplepallande och oförglömliga resor. Tack för att du varit sardinen vid min sida på soffan i Montreal, promenad polaren på FLÄK-möten, och den söta sömerskan av sidenslipsen till Albin. Doktorerande utan dig hade inte innefattat Maltamys, Kanada cykling, klädkod och hade inte varit hälften så roligt och färgstarkt. **Mikael**, din omtänksamhet och genorisitet sprider sig till all i din omgivning. Tack för trevligt resällskap i Kalifornien, fantastiska middagar, inspiration i löpspåret och för att du alltid tagit dig tid till mina frågor och problem. **Erik**, en skämtsam prick med huvudet på skaft, tack för trevligt sällskap på Vandarhem Linné, för din förmåga att förklara vetenskap och för dina spexiga bidrag. **Tommy**, livskonstnären och politikern alltid med ett varmt hjärta och ett skämt till hands. **Martin**, den stressfria norrlänningen med skön stil. Tack för dyket i Spanien! **Celia**, the Spanish roommate sending out happiness to all in the vicinity.

Övriga doktorander och post docs: **Johan**, man slutar aldrig förvånas över din både intellektuella och sociala kompetens. Tack för den sköna stämningen du sprider! **Carl**, tack för all ovärderlig NMR hjälp, sällskap i löpspåret på FLÄK-möten och utmaningar på alpingympan. Doktor **Erik**, tack för att jag fått följa dig i fotspåren i avhandlingsarbetet och för trevligt sällskap ända sedan forskarskoletiden. **Gleb**, the superman and joker of the room. Thanks for never being serious, helping me with

computer problems and offering croissants. Your proofreading was very appreciated along with your puzzling title. **Michal**, the Polish Postdoc longing to become a part of the SOG room. Thanks for offering many laughs!

I labbet: **Hanna**, alltid med tid för omtanke och tänkvärda konversationer. Tack för gott samarbete, många rena proteiner och alla gånger du räddat mig i labbet. **Eva**, tack för att du bryr dig, för all kunskap du förmedlar och för välbehövliga hejarrop i och utanför vasaloppsspåret. **Birgitta**, superroligt att ha dig på bpc och tack för gott samarbete. Tack för din öppenhet, omtänksamhet, humor och värme! **Hans**, tack för alla gånger du har räddat min dator ur knipa och alla instrument du klarat livhanken på. Stort tack för alla trevliga konversationer och det lugn du sprider kring dig!

Tidigare medarbetare: **Eva**, alltid mellan en hundpromenad och ett gerdapass med tusen järn i elden och en blick som få att se hur andra mår. Tack för fin vänskap och för att du hållt reda på oss förvirrade forskare. **Bodil**, lugnet i stormen på sekretariatet med ett otroligt varmt hjärta. Tack för allt stöd när experimenten misslyckats eller kurserna varit övermäktiga. **Ingemar**, jag vet att detta är jobbigt för dig men jag har så mycket att tacka dig för. Tack för måndagskatten, komplimang 2005 och allt du faktiskt har lärt mig! Kul att ha dig tillbaka och sist men inte minst, tack för tian. **Wei-Feng**, reskompanjonen i Grekland och Kalifornien. Tack för all kunskap du förmedlat, godiset du fyllde resväskan med i Grekland och den nattliga expeditionen till Hollywoodskylten. **Ulrika**, den ultimata rumskompisen med exemplarisk ordning och reda. Tack för trevligt sällskap och all NMR hjälp. **Frans**, din kompetens och distans till jobbet är oslagbar. Tack för alla trevliga samarbeten! **Mel**, the year with you in the lab was awesome! Thanks for nice runs, great parties, tasty dinners and for hosting me in SF. Tack till **Patrik, Houman, Jonas, Magnus, Kaare, Monica, Cynthia, Jenny** som alla varit till stor hjälp och förgyllt min doktorandtid. **Jannette**, with a never ending enthusiasm for science. Thanks for everything you have taught me about ligand binding and protein reconstitution. **Armando**, your hard work really improved this thesis. Thanks for taking care of my desk while I was away. **Ida**, mycket trevligt att ha dig i gruppen och tack för vetenskapliga bidrag.

Grabbarna på Teokem, **Mikael, Bosse, Gunnar, Björn, Martin Trul, Martin Tur**, m.fl.; tack för trevligt sällskap, tjuvnyp, surströmming, galna fester och vetenskapligt utbyte. Saknar er!

CMPS-folket; tack för det varma välkommandet hos er.

Tack till alla på Alligator som gjorde min tid hos er så trevlig! Särskilt tack till **Erika** för trevligt samarbete och allt du lärt mig från husbesiktning till protein optimering och projektplanering. **Mats** och **Tina**, tack för att jag fick möjlighet att göra projekt hos er med inblick i företagsvärlden och ovärderliga erfarenheter till följd.

Killarna på vaktmästeriet **Stefan, Christer** och **Johnny**, tack för hjälp med med gastuber, lån av bord oh stolar och för ert trevliga humor.

Alla vänner utanför jobbet: **Sussie**, med få förunnad styrka, omtänksamhet och känsla för design. Tack för alla upplevelser vi delat från helvetesvadet i Sarek till bebisbad och

delad seger på vasaloppet! Även tack för allt stöd från dig och Jonas under slutspurten av skrivandet. **Jonas**, den bästa grannen man kan tänka sig med inneboende noggrannhet både på labbet och gällande städningen där hemma. Tack för att du alltid ställer upp, för att du korrekturläst avhandlingen och för lärdomar i livsnjutning. **Axel**, 9-månaders charmtrollet alltid med ett leende bakom nappen. **Sus**, **Ilka**, **Linn**, **Isabel**, **Ida**, **Emma** och **Jessica**, simmarkompisarna som distanserat mig från jobbet. Alltid lika trevligt när vi ses. **Eken** och **Tessan** tack för fantastiska resor till Sägbäcken, goda middagar, pepparkakshustävlingar och allt annat som fått mig att koppla bort jobbet.

Familjen: **Mamma**, så beundransvärd med din stabilitet, drivkraft och förmåga att njuta! Tack för att du alltid finns till hands för att hjälpa eller bara snacka lite. **Pappa**, tack för att du aldrig slutar tro på mig och för hjälp med vetenskapliga problem. Syskonen **Åsa**, **Gudrun** och **Rune** med familjer. Tack för att ni finns! **Tone**, jag beundrar din kreativitet, optimism, uthållighet och omtänksamhet ända in i det sista. **Niklas**, det är så mycket jag skulle vilja tacka dig för att det skulle kunna bli en egen bok. Tack för att du stöttat och trott på mig och är min fasta punkt när annat är skakigt. Kärleken till dig blir bara starkare för varje dag! **Albin**, du är söt, snäll, smart, rolig, mysig, busig och har verkligen fått mig på fall! Tack för att du hjälpt mig att inte gräva ner mig för mycket i denna avhandling. Du är nog dessutom (förutom möjligen Axel) den enda som kommer att läsa avhandlingen med alla sinnen.

## 7. REFERENCES

1. Creighton, T. E. (1993) *Proteins: structures and molecular properties*, 2nd ed., W.H. Freeman and Company, New York.
2. Tanford, C., and Reynolds, J. A. (2001) *Nature's robots: A history of proteins* Oxford University Press, Oxford.
3. Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem.* 75, 333-366.
4. Dyer, R. B. (2007) Ultrafast and downhill protein folding, *Curr. Opin. Struct. Biol.* 17, 38-47.
5. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol.* 4, 10-19.
6. Levinthal, C. (1968) Are there pathways for protein folding, *J. Chim. Phys. Chim. Biol.* 65, 44-45.
7. Anfinsen, C. B. (1973) Principles that govern folding of protein chains, *Science.* 181, 223-230.
8. Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961) Kinetics of formation of native ribonuclease during oxidation of reduced polypeptide chain, *Proc. Natl. Acad. Sci. USA.* 47, 1309-1314.
9. Oliveberg, M., and Wolynes, P. G. (2005) The experimental survey of protein-folding energy landscapes, *Q. Rev. Biophys.* 38, 245-288.
10. Feldman, D. E., and Frydman, J. (2000) Protein folding in vivo: the importance of molecular chaperones, *Curr. Opin. Struct. Biol.* 10, 26-33.
11. Ferguson, N., and Fersht, A. R. (2003) Early events in protein folding, *Curr. Opin. Struct. Biol.* 13, 75-81.
12. Jackson, S. E. (1998) How do small single-domain proteins fold?, *Fold. Des.* 3, R81-R91.
13. Teilum, K., Olsen, J. G., and Kragelund, B. B. (2009) Functional aspects of protein flexibility, *Cell. Mol. Life. Sci.* 66, 2231-2247.
14. Pace, C. N. (1990) Conformational stability of globular-proteins, *Trends Biochem. Sci.* 15, 14-17.
15. Pace, C. N., Trevino, S., Prabhakaran, E., and Scholtz, J. M. (2004) Protein structure, stability and solubility in water and other solvents, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1225-1234.
16. Eijsink, V. G. H., Bjørk, A., Gåseidnes, S., Sirevåg, R., Synstad, B., van den Burg, B., and Vriend, G. (2004) Rational engineering of enzyme stability, *J. Biotechnol.* 113, 105-120.
17. Shaw, K. L., Scholtz, J. M., Pace, C. N., and Grimsley, G. R. (2009) Determining the conformational stability of a protein using urea denaturation curves, in *Methods Mol. Biol.* pp 41-55, Humana Press Inc.
18. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m-values and heat-capacity changes - relation to changes in accessible surface-areas of protein unfolding, *Protein Sci.* 4, 2138-2148.

19. Green, A. A. (1931) Studies in the physical chemistry of the proteins VIII. The solubility of hemoglobin in concentrated salt solutions. A study of the salting out of proteins, *J. Biol. Chem.* *93*, 495-516.
20. Green, A. A. (1932) Studies in the physical chemistry of the proteins X. The solubility of hemoglobin in solutions of chlorides and sulfates of varying concentration, *J. Biol. Chem.* *95*, 47-66.
21. Baldwin, R. (1996) How Hofmeister ion interactions affect protein stability, *Biophys. J.* *71*, 2056-2063.
22. Van den Burg, B., Vriend, G., Veltman, O. R., and Eijsink, V. G. H. (1998) Engineering an enzyme to resist boiling, *Proc. Natl. Acad. Sci. USA.* *95*, 2056-2060.
23. Malakauskas, S. M., and Mayo, S. L. (1998) Design, structure and stability of a hyperthermophilic protein variant, *Nat. Struct. Mol. Biol.* *5*, 470-475.
24. Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T., Loladze, V. V., and Makhataдзе, G. I. (2006) Protein stability and surface electrostatics: A charged relationship, *Biochemistry.* *45*, 2761-2766.
25. Korkegian, A., Black, M. E., Baker, D., and Stoddard, B. L. (2005) Computational thermostabilization of an enzyme, *Science.* *308*, 857-860.
26. Wunderlich, M., Martin, A., and Schmid, F. X. (2005) Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of coulombic interactions, *J. Mol. Biol.* *347*, 1063-1076.
27. Wunderlich, M., Martin, A., Staab, C. A., and Schmid, F. X. (2005) Evolutionary protein stabilization in comparison with computational design, *J. Mol. Biol.* *351*, 1160-1168.
28. Wunderlich, M., and Schmid, F. X. (2006) In vitro evolution of a hyperstable GB1 variant, *J. Mol. Biol.* *363*, 545-557.
29. Eijsink, V. G. H., Gåseidnes, S., Borchert, T. V., and van den Burg, B. (2005) Directed evolution of enzyme stability, *Biomol. Eng.* *22*, 21-30.
30. Finucane, M. D., and Woolfson, D. N. (1999) Core-directed protein design. II. Rescue of a multiply mutated and destabilized variant of ubiquitin, *Biochemistry.* *38*, 11613-11623.
31. Mooers, B. H. M., Datta, D., Baase, W. A., Zollars, E. S., Mayo, S. L., and Matthews, B. W. (2003) Repacking the core of T4 lysozyme by automated design, *J. Mol. Biol.* *332*, 741-756.
32. Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. (1995) Are buried salt bridges important for protein stability and conformational specificity, *Nat. Struct. Biol.* *2*, 122-128.
33. Martin, A., Sieber, V., and Schmid, F. X. (2001) In-vitro selection of highly stabilized protein variants with optimized surface, *J. Mol. Biol.* *309*, 717-726.
34. Sanchez-Ruiz, J. M., and Makhataдзе, G. I. (2001) To charge or not to charge?, *Trends Biotechnol.* *19*, 132-135.
35. Prajapati, R. S., Das, M., Sreeramulu, S., Sirajuddin, M., Srinivasan, S., Krishnamurthy, V., Ranjani, R., Ramakrishnan, C., and Varadarajan, R.



- (2007) Thermodynamic effects of proline introduction on protein stability, *Proteins*. *66*, 480-491.
36. Takano, K., Higashi, R., Okada, J., Mukaiyama, A., Tadokoro, T., Koga, Y., and Kanaya, S. (2009) Proline effect on the thermostability and slow unfolding of a hyperthermophilic protein, *J. Biochem.* *145*, 79-85.
  37. Wahlberg, E., and Härd, T. (2006) Conformational stabilization of an engineered binding protein, *J. Am. Chem. Soc.* *128*, 7651-7660.
  38. Israelachvili, J. (1991) *Intermolecular and Surface Forces*, 2 ed., Elsevier.
  39. Thurlkill, R. L., Grimsley, G. R., Scholtz, M., and Pace, C. N. (2006) Hydrogen bonding markedly reduces the pK of buried carboxyl groups in proteins, *J. Mol. Biol.* *362*, 594-604.
  40. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pK(a) values, *Proteins*. *61*, 704-721.
  41. Worth, C. L., and Blundell, T. L. (2009) Satisfaction of hydrogen-bonding potential influences the conservation of polar sidechains, *Proteins*. *75*, 413-429.
  42. Fleming, P. J., and Rose, G. D. (2005) Do all backbone polar groups in proteins form hydrogen bonds?, *Protein Sci.* *14*, 1911-1917.
  43. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation, *Adv. Protein Chem.* *14*, 1-63.
  44. Dill, K. A. (1990) Dominant forces in protein folding, *Biochemistry*. *29*, 7133-7155.
  45. Baldwin, R. L. (2007) Energetics of protein folding, *J. Mol. Biol.* *371*, 283-301.
  46. Southall, N. T., Dill, K. A., and Haymet, A. D. J. (2002) A view of the hydrophobic effect, *J. Phys. Chem. B*. *106*, 521-533.
  47. Daquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding, *Proteins*. *25*, 143-156.
  48. Kuroki, R., Inaka, K., Taniyama, Y., Kidokoro, S., Matsushima, M., Kikuchi, M., and Yutani, K. (1992) Enthalpic destabilization of a mutant human lysozyme lacking a disulfide bridge between Cysteine-77 and Cysteine-95, *Biochemistry*. *31*, 8323-8328.
  49. Vaz, D. C., Rodrigues, J. R., Sebald, W., Dobson, C. M., and Brito, R. M. M. (2006) Enthalpic and entropic contributions mediate the role of disulfide bonds on the conformational stability of interleukin-4, *Protein Sci.* *15*, 33-44.
  50. Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2009) A summary of the measured pK values of the ionizable groups in folded proteins, *Protein Sci.* *18*, 247-251.
  51. Pace, C. N., Grimsley, G. R., and Scholtz, J. M. (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility, *J. Biol. Chem.* *284*, 13285-13289.

52. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A., and Pace, C. N. (2001) The effect of net charge on the solubility, activity, and stability of ribonuclease Sa, *Protein Sci.* 10, 1206-1215.
53. Bosshard, H. R., Marti, D. N., and Jelesarov, I. (2004) Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings, *J. Mol. Recognit.* 17, 1-16.
54. Xue, W. F., Szczepankiewicz, O., Bauer, M. C., Thulin, E., and Linse, S. (2006) Intra- versus intermolecular interactions in monellin: Contribution of surface charges to protein assembly, *J. Mol. Biol.* 358, 1244-1255.
55. Linse, S., Brodin, P., Johansson, C., Thulin, E., Grundström, T., and Forsén, S. (1988) The role of protein surface charges in ion binding, *Nature.* 335, 651-652.
56. Linse, S., Johansson, C., Brodin, P., Grundström, T., Drakenberg, T., and Forsén, S. (1991) Electrostatic contributions to the binding of Ca<sup>2+</sup> in calbindin D9k, *Biochemistry.* 30, 154-162.
57. Harris, T. K., and Turner, G. J. (2002) Structural basis of perturbed pK(a) values of catalytic groups in enzyme active sites, *IUBMB Life.* 53, 85-98.
58. Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites, *J. Biol. Chem.* 273, 27035-27038.
59. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance, *Nature.* 358, 347-351.
60. Ciriolo, M. R., Battistoni, A., Falconi, M., Filomeni, G., and Rotilio, G. (2001) Role of the electrostatic loop of Cu,Zn superoxide dismutase in the copper uptake process, *Eur. J. Biochem.* 268, 737-742.
61. Kangas, E., and Tidor, B. (2001) Electrostatic complementarity at ligand binding sites: application to chorismate mutase, *J. Phys. Chem. B* 105, 880-888.
62. Otzen, D. E., Kristensen, O., and Oliveberg, M. (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: A structural clue to amyloid assembly, *Proc. Natl. Acad. Sci. USA.* 97, 9907-9912.
63. Tougu, V., and Kesvatera, T. (1996) Role of ionic interactions in cholinesterase catalysis., *Biochim. Biophys. Acta.* 1298, 12-30.
64. Lawrence, M. S., Phillips, K. J., and Liu, D. R. (2007) Supercharging proteins can impart unusual resilience, *J. Am. Chem. Soc* 129, 10110-10112.
65. Vendruscolo, M., and Dobson, C. M. (2007) Chemical biology - More charges against aggregation, *Nature.* 449, 555-555.
66. Warshel, A., and Åqvist, J. (1991) Electrostatic energy and macromolecular function, *Annu. Rev. Biophys. Biophys. Chem.* 20, 267-298.
67. Warshel, A., Sharma, P. K., Kato, M., and Parson, W. W. (2006) Modeling electrostatic effects in proteins, *Biochim. Biophys. Acta.* 1764, 1647-1676.

68. Debye, P., and Hückel, E. (1923) The theory of electrolytes I. The lowering of the freezing point and related occurrences, *Physikalische Zeitschrift*. 24, 185-206.
69. Jordan, I. K., Kondrashov, F. A., Adzhubei, I. A., Wolf, Y. I., Koonin, E. V., Kondrashov, A. S., and Sunyaev, S. (2005) A universal trend of amino acid gain and loss in protein evolution, *Nature*. 433, 633-638.
70. Nozaki, Y., and Tanford, C. (1967) Examination of titration behavior, *Methods Enzymol*. 11, 715-734.
71. Chivers, P. T., Prehoda, K. E., Volkman, B. F., Kim, B. M., Markley, J. L., and Raines, R. T. (1997) Microscopic pK(a) values of Escherichia coli thioredoxin, *Biochemistry*. 36, 14985-14991.
72. Grimsley, G., Shaw, K., Fee, L., Alston, R., Huyghues-Despointes, B., Thurlkill, R., Scholtz, J., and Pace, C. (1999) Increasing protein stability by altering long-range coulombic interactions, *Protein Sci*. 8, 1843-1849.
73. Gudiksen, K. L., Gitlin, I., Yang, J., Urbach, A. R., Moustakas, D. T., and Whitesides, G. M. (2005) Eliminating positively charged lysine  $\epsilon$ -NH<sub>3</sub><sup>+</sup> groups on the surface of carbonic anhydrase has no significant influence on its folding from sodium dodecyl sulfate, *J. Am. Chem. Soc*. 127, 4707-4714.
74. Laurents, D. V., Huyghues-Despointes, B. M. P., Bruix, M., Thurlkill, R. L., Schell, D., Newsom, S., Grimsley, G. R., Shaw, K. L., Trevino, S., Rico, M., Briggs, J. M., Antosiewicz, J. M., Scholtz, J. M., and Pace, C. N. (2003) Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2), *J. Mol. Biol*. 325, 1077-1092.
75. Loladze, V. V., and Makhatadze, G. I. (2002) Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable, *Protein Sci*. 11, 174-177.
76. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: Guidelines for protein engineering, *J. Mol. Biol*. 327, 1135-1148.
77. Makhatadze, G. I., Loladze, V. V., Gribenko, A. V., and Lopez, M. M. (2004) Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions, *J. Mol. Biol*. 336, 929-942.
78. Dwyer, J. J., Gittis, A. G., Karp, D. A., Lattman, E. E., Spencer, D. S., Stites, W. E., and Garcia-Moreno, B. (2000) High apparent dielectric constants in the interior of a protein reflect water penetration, *Biophys. J*. 79, 1610-1620.
79. Garcia-Moreno, B., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S., and Stites, W. E. (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein, *Biophys. Chem*. 64, 211-224.
80. Whitten, S. T., and Garcia-Moreno, B. (2000) pH dependence of stability of staphylococcal nuclease: Evidence of substantial electrostatic interactions in the denatured state, *Biochemistry*. 39, 14292-14304.

81. Lee, K. K., Fitch, C. A., and Garcia-Moreno E., B. (2002) Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein, *Protein Sci.* *11*, 1004-1016.
82. Lee, K. K., Fitch, C. A., Lecomte, J. T., and Garcia-Moreno, E. B. (2002) Electrostatic effects in highly charged proteins: salt sensitivity of pKa values of histidines in staphylococcal nuclease., *Biochemistry.* *41*, 6556-6567.
83. Perez-Jimenez, R., Godoy-Ruiz, R., Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (2004) The efficiency of different salts to screen charge interactions in proteins: A Hofmeister effect?, *Biophys. J.* *86*, 2414-2429.
84. Spencer, D. S., Xu, K., Logan, T. M., and Zhou, H.-X. (2005) Effects of pH, salt, and macromolecular crowding on the stability of FK506-binding protein: An integrated experimental and theoretical study, *J. Mol. Biol.* *351*, 219-232.
85. Szakács, Z., Kraszni, M., and Noszál, B. (2004) Determination of microscopic acid-base parameters from NMR-pH titrations, *Anal. Bioanal. Chem.* *378*, 1428-1448.
86. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme., *Biochemistry.* *29*, 2403-2408.
87. Karp, D. A., Gittis, A. G., Stahley, M. R., Fitch, C. A., Stites, W. E., and Garcia-Moreno, B. (2007) High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp, *Biophys. J.* *92*, 2041-2053.
88. Dyson, H. J., Jeng, M. F., Tennant, L. L., Slaby, I., Lindell, M., Cui, D. S., Kuprin, S., and Holmgren, A. (1997) Effects of buried charged groups on cysteine thiol ionization and reactivity in Escherichia coli thioredoxin: Structural and functional characterization of mutants of Asp 26 and Lys 57, *Biochemistry.* *36*, 2622-2636.
89. Schmittschmitt, J. P., and Scholtz, J. M. (2003) The role of protein stability, solubility, and net charge in amyloid fibril formation, *Protein Sci.* *12*, 2374-2378.
90. Cohn, E. J. (1922) Studies in the physical chemistry of the proteins. I. The solubility of certain proteins at their isoelectric points, *J. Gen. Physiol.* *4*, 697-722.
91. Lindman, S., Xue, W. F., Szczepankiewicz, O., Bauer, M. C., Nilsson, H., and Linse, S. (2006) Salting the charged surface: pH and salt dependence of protein G B1 stability, *Biophys. J.* *90*, 2911-2921.
92. Akke, M., and Forsén, S. (1990) Protein stability and electrostatic interactions between solvent exposed charged side chains., *Proteins.* *8*, 23-9.
93. Isom, D. G., Cannon, B. R., Castañeda, C. A., Robinson, A., and García-Moreno, E. (2008) High tolerance for ionizable residues in the hydrophobic interior of proteins, *Proc. Natl. Acad. Sci. USA.* *105*, 17784-17788.
94. Harms, M. J., Schlessman, J. L., Chimenti, M. S., Sue, G. R., Damjanovic, A., and Garcia-Moreno, B. (2008) A buried lysine that titrates with a normal

- pK(a): Role of conformational flexibility at the protein-water interface as a determinant of pK(a) values, *Protein Sci.* 17, 833-845.
95. Bowler, B. E. (2007) Thermodynamics of protein denatured states, *Mol. Biosyst.* 3, 88-99.
  96. Shi, Z. S., Chen, K., Liu, Z. G., and Kallenbach, N. R. (2006) Conformation of the backbone in unfolded proteins, *Chem. Rev.* 106, 1877-1897.
  97. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science.* 295, 1719-1722.
  98. Cho, J.-H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and Kinetics of Non-native Interactions in Protein Folding: A Single Point Mutant Significantly Stabilizes the N-terminal Domain of L9 by Modulating Non-native Interactions in the Denatured State, *J. Mol. Biol.* 338, 827-837.
  99. Pace, C. N., Alston, R. W., and Shaw, K. L. (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, *Protein Sci.* 9, 1395-1398.
  100. Shortle, D. (1995) Staphylococcal nuclease: A showcase of m-value effects, *Adv. Protein Chem.* 46, 217-247.
  101. Jeng, M. F., and Englander, S. W. (1991) Stable submolecular folding units in a noncompact form of cytochrome c, *J. Mol. Biol.* 221, 1045-1061.
  102. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry.* 38, 4896-903.
  103. Pujato, M., Bracken, C., Mancusso, R., Cataldi, M., and Tasayco, M. L. (2005) pH dependence of amide chemical shifts in natively disordered polypeptides detects medium-range interactions with ionizable residues, *Biophys. J.* 89, 3293-3302.
  104. Pujato, M., Navarro, A., Versace, R., Mancusso, R., Ghose, R., and Tasayco, M. L. (2006) The pH-dependence of amide chemical shift of Asp/Glu reflects its pK(a) in intrinsically disordered proteins with only local interactions, *Biochim. Biophys. Acta.* 1764, 1227-1233.
  105. Marti, D. N., and Bosshard, H. R. (2004) Inverse electrostatic effect: Electrostatic repulsion in the unfolded state stabilizes a leucine zipper, *Biochemistry.* 43, 12436-12447.
  106. Zhou, H. X. (2002) A Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins, *Proc. Natl. Acad. Sci. USA.* 99, 3569-3574.
  107. Zhou, H. X. (2004) Polymer models of protein stability, folding, and interactions, *Biochemistry.* 43, 2141-2154.
  108. Gustafsson, E., Forsberg, C., Haraldsson, K., Lindman, S., Ljung, L., and Furebring, C. (2009) Purification of truncated and mutated Chemotaxis

- Inhibitory Protein of *Staphylococcus aureus*-an anti-inflammatory protein, *Protein Expr. Purif.* *63*, 95-101.
109. Cabaleiro-Lago, C., Quinlan-Pluck, F., Lynch, I., Lindman, S., Minogue, A. M., Thulin, E., Walsh, D. M., Dawson, K. A., and Linse, S. (2008) Inhibition of amyloid beta protein fibrillation by polymeric nanoparticles, *J. Am. Chem. Soc.* *130*, 15437-15443.
  110. Cedervall, T., Lynch, I., Lindman, S., Berggård, T., Thulin, E., Nilsson, H., Dawson, K. A., and Linse, S. (2007) Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles, *Proc. Natl. Acad. Sci. USA.* *104*, 2050-2055.
  111. Lindman, S., Lynch, I., Thulin, E., Nilsson, H., Dawson, K. A., and Linse, S. (2007) Systematic investigation of the thermodynamics of HSA adsorption to N-iso-propylacrylamide/N-tert-butylacrylamide copolymer nanoparticles. Effects of particle size and hydrophobicity, *Nano Lett.* *7*, 914-920.
  112. Linse, S., Cabaleiro-Lago, C., Xue, W. F., Lynch, I., Lindman, S., Thulin, E., Radford, S. E., and Dawson, K. A. (2007) Nucleation of protein fibrillation by nanoparticles, *Proc. Natl. Acad. Sci. USA.* *104*, 8691-8696.
  113. Linse, S., Helmersson, A., and Forsén, S. (1991) Calcium binding to calmodulin and its globular domains, *J. Biol. Chem.* *266*, 8050-8054.
  114. Sauer-Eriksson, A. E., Kleywegt, G. J., Uhl, M., and Jones, T. A. (1995) Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG, *Structure.* *3*, 265-278.
  115. Diehl, C., Genheden, S., Modig, K., Ryde, U., and Akke, M. (2009) Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3, *J. Biomol. NMR.* *45*, 157-169.
  116. Forsén, S., and Linse, S. (1995) Cooperativity: over the Hill, *Trends Biochem. Sci.* *20*, 495-497.
  117. Acerenza, L., and Mizraji, E. (1997) Cooperativity: A unified view, *Biochim. Biophys. Acta.* *1339*, 155-166.
  118. Richards, F. M. (1958) On the enzymatic activity of subtilisin-modified ribonuclease, *Proc. Natl. Acad. Sci. USA.* *44*, 162-166.
  119. Carey, J., Lindman, S., Bauer, M., and Linse, S. (2007) Protein reconstitution and three-dimensional domain swapping: Benefits and constraints of covalency, *Protein Sci.* *16*, 2317-2333.
  120. Håkansson, M., and Linse, S. (2002) Protein reconstitution and 3D domain swapping, *Curr. Protein. Pept. Sci.* *3*, 629-642.
  121. Berggård, T., Julenius, K., Ogard, A., Drakenberg, T., and Linse, S. (2001) Fragment complementation studies of protein stabilization by hydrophobic core residues, *Biochemistry.* *40*, 1257-1264.
  122. Ruiz-Sanz, J., Prat-Gay, G. d., Otzen, D. E., and Fersht, A. R. (1995) Protein fragments as models for events in protein folding pathways - protein engineering analysis of the association of 2 complementary fragments of the barley chymotrypsin inhibitor-2 (Ci-2), *Biochemistry.* *34*, 1695-1701.

123. Bauer, M. C., Xue, W. F., and Linse, S. (2009) Protein GB1 folding and assembly from structural elements, *Int. J. Mol. Sci.* *10*, 1552-1566.
124. Housden, N. G., Harrison, S., Roberts, S. E., Beckingham, J. A., Graille, M., Stura, E., and Gore, M. G. (2003) Immunoglobulin-binding domains: Protein L from *Peptostreptococcus magnus*, *Biochem. Soc. Trans.* *31*, 716-718.
125. Goward, C. R., Scawen, M. D., Murphy, J. P., and Atkinson, T. (1993) Molecular evolution of bacterial cell-surface proteins, *Trends Biochem. Sci.* *18*, 136-140.
126. Frick, I. M., Wikström, M., Forsén, S., Drakenberg, T., Gomi, H., Sjöbring, U., and Björck, L. (1992) Convergent evolution among immunoglobulin-G-binding bacterial proteins, *Proc. Natl. Acad. Sci. USA.* *89*, 8532-8536.
127. 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, *J. Mol. Biol.* *233*, 331-335.
128. Gallagher, T., Alexander, P., Bryan, P., and Gilliland, G. L. (1994) Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR., *Biochemistry.* *33*, 4721-9.
129. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G, *Science.* *253*, 657-661.
130. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (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.* *31*, 3597-3603.
131. McCallister, E. L., Alm, E., and Baker, D. (2000) Critical role of beta-hairpin formation in protein G folding, *Nat. Struct. Biol.* *7*, 669-673.
132. Nauli, S., Kuhlman, B., and Baker, D. (2001) Computer-based redesign of a protein folding pathway, *Nat. Struct. Biol.* *8*, 602-605.
133. Khare, D., Alexander, P., Antosiewicz, J., Bryan, P., Gilson, M., and Orban, J. (1997) pKa measurements from nuclear magnetic resonance for the B1 and B2 immunoglobulin G-binding domains of protein G: comparison with calculated values for nuclear magnetic resonance and X-ray structures., *Biochemistry.* *36*, 3580-3589.
134. Tomlinson, J. H., Ullah, S., Hansen, P. E., and Williamson, M. P. (2009) Characterization of salt bridges to lysines in the protein G B1 domain, *J. Am. Chem. Soc.* *131*, 4674-4684.
135. Honda, S., Kobayashi, N., Munekata, E., and Uedaira, H. (1999) Fragment reconstitution of a small protein: Folding energetics of the reconstituted immunoglobulin binding domain B1 of streptococcal protein G, *Biochemistry.* *38*, 1203-1213.

136. Kobayashi, N., Honda, S., and Munekata, E. (1999) Fragment reconstitution of a small protein: Disulfide mutant of a short C-terminal fragment derived from streptococcal protein G, *Biochemistry*. *38*, 3228-3234.
137. Kobayashi, N., Honda, S., Yoshii, H., Uedaira, H., and Munekata, E. (1995) Complement assembly of two fragments of the streptococcal protein GB1 domain in aqueous solution, *FEBS Lett.* *366*, 99-103.
138. Blanco, F. J., Rivas, G., and Serrano, L. (1994) A short linear peptide that folds into a native stable beta hairpin in aqueous solution, *Nat. Struct. Biol.* *1*, 584-590.
139. Blanco, F. J., and Serrano, L. (1995) Folding of protein G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements, *Eur. J. Biochem.* *230*, 634-649.
140. Honda, S., Kobayashi, N., and Munekata, E. (2000) Thermodynamics of a beta-hairpin structure: Evidence for cooperative formation of folding nucleus, *J. Mol. Biol.* *295*, 269-278.
141. Smith, C. K., Withka, J. M., and Regan, L. (1994) A thermodynamic scale for the beta-sheet forming tendencies of the amino acids., *Biochemistry*. *33*, 5510-5517.
142. Reissner, K. J., and Aswad, D. W. (2003) Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals?, *Cell. Mol. Life Sci.* *60*, 1281-1295.
143. Lindman, S., Linse, S., Mulder, F. A. A., and Andre, I. (2007) pK(a) values for side-chain carboxyl groups of a PGB1 variant explain salt and pH-dependent stability, *Biophys. J.* *92*, 257-266.
144. Szebenyi, D. M. E., and Moffat, K. (1986) The refined structure of vitamin-D-dependent calcium-binding protein from bovine intestine - Molecular details, ion binding, and implications for the structure of other calcium-binding proteins, *J. Biol. Chem.* *261*, 8761-8777.
145. Kesvatera, T., Jönsson, B., Thulin, E., and Linse, S. (1996) Measurement and modelling of sequence-specific pK(a) values of lysine residues in calbindin D-9k, *J. Mol. Biol.* *259*, 828-839.
146. Kesvatera, T., Jönsson, B., Thulin, E., and Linse, S. (1999) Ionization behavior of acidic residues in calbindin D-9k, *Proteins*. *37*, 106-115.
147. Kesvatera, T., Jönsson, B., Thulin, E., and Linse, S. (2001) Focusing of the electrostatic potential at EF-hands of calbindin D-9k: Titration of acidic residues, *Proteins*. *45*, 129-135.
148. Linse, S., Brodin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmnden, K., Grundström, T., and Forsén, S. (1987) Structure-function-relationships in EF-Hand Ca<sup>2+</sup>-binding proteins - Protein engineering and biophysical studies of calbindin D9k, *Biochemistry*. *26*, 6723-6735.
149. Linse, S., Thulin, E., and Sellers, P. (1993) Disulfide bonds in homodimers and heterodimers of EF-hand subdomains of calbindin-D(9k) - Stability, calcium-binding, and NMR-studies, *Protein Sci.* *2*, 985-1000.



150. Chazin, W. J., Kordel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T., and Forsén, S. (1989) Proline isomerism leads to multiple folded conformations of calbindin D9k - Direct evidence from two-dimensional H1-NMR spectroscopy, *Proc. Natl. Acad. Sci. USA.* *86*, 2195-2198.
151. Kordel, J., Forsén, S., Drakenberg, T., and Chazin, W. J. (1990) The rate and structural consequences of proline cis-trans isomerization in calbindin-D9k - NMR-studies of the minor (cis-Pro43) isoform and the Pro43Gly mutant, *Biochemistry.* *29*, 4400-4409.
152. Dell'Orco, D., Xue, W. F., Thulin, E., and Linse, S. (2005) Electrostatic contributions to the kinetics and thermodynamics of protein assembly, *Biophys. J.* *88*, 1991-2002.
153. Zimmer, M. (2009) GFP: from jellyfish to the Nobel prize and beyond, *Chem. Soc. Rev.* *38*, 2823-2832.
154. Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein, *Science.* *273*, 1392-1395.
155. Tsien, R. Y. (1998) The green fluorescent protein, *Annu. Rev. Biochem.* *67*, 509-544.
156. Craggs, T. D. (2009) Green fluorescent protein: structure, folding and chromophore maturation, *Chem. Soc. Rev.* *38*, 2865-2875.
157. Zimmer, M. (2002) Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior, *Chem. Rev.* *102*, 759-781.
158. Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N., and Wlodawer, A. (1997) The structural basis for spectral variations in green fluorescent protein, *Nat. Struct. Biol.* *4*, 361-365.
159. Sample, V., Newman, R. H., and Zhang, J. (2009) The structure and function of fluorescent proteins, *Chem. Soc. Rev.* *38*, 2852-2864.
160. Verkhusha, V. V., and Lukyanov, K. A. (2004) The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins, *Nat. Biotechnol.* *22*, 289-296.
161. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nat. Biotechnol.* *24*, 79-88.
162. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures, *J. Mol. Graph.* *14*, 51-55.
163. Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism, *Biochim. Biophys. Acta.* *1751*, 119-139.
164. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp red fluorescent protein, *Nat. Biotechnol.* *22*, 1567-1572.

165. Hu, C. D., Chinenov, Y., and Kerppola, T. K. (2002) Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation, *Mol. Cell.* *9*, 789-798.
166. Barnard, E., McFerran, N. V., Trudgett, A., Nelson, J., and Timson, D. J. (2008) Development and implementation of split-GFP-based bimolecular fluorescence complementation (BiFC) assays in yeast, *Biochem. Soc. Trans.* *36*, 479-482.
167. Barnard, E., McFerran, N. V., Trudgett, A., Nelson, J., and Timson, D. J. (2008) Detection and localisation of protein-protein interactions in *Saccharomyces cerevisiae* using a split-GFP method, *Fungal Genet. Biol.* *45*, 597-604.
168. Ghosh, I., Hamilton, A. D., and Regan, L. (2000) Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein, *J. Am. Chem. Soc.* *122*, 5658-5659.
169. Magliery, T. J., Wilson, C. G. M., Pan, W. L., Mishler, D., Ghosh, I., Hamilton, A. D., and Regan, L. (2005) Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: Scope and mechanism, *J. Am. Chem. Soc.* *127*, 146-157.
170. Ohad, N., Shichrur, K., and Yalovsky, S. (2007) The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation, *Plant Physiol.* *145*, 1090-1099.
171. Wilson, C. G., Magliery, T. J., and Regan, L. (2004) Detecting protein-protein interactions with GFP-fragment reassembly, *Nat. Methods.* *1*, 255-262.
172. Kerppola, T. K. (2008) Biomolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells, *Annu. Rev. Biophys.* *37*, 465-487.
173. Kerppola, T. K. (2009) Visualization of molecular interactions using bimolecular fluorescence complementation analysis: Characteristics of protein fragment complementation, *Chem. Soc. Rev.* *38*, 2876-2886.
174. Lindman, S., Johansson, I., Thulin, E., and Linse, S. (2009) Green fluorescence induced by EF-hand assembly in a split GFP system, *Protein Sci.* *18*, 1221-1229.
175. Makhatadze, G. I. (2004) in *Protein Folding Handbook* pp 70–98.
176. Barthold, E., and Ernst, R. R. (1973) Fourier Spectroscopy and Causality Principle, *J. Magn. Reson.* *11*, 9-19.
177. Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) 2-Dimensional spectroscopy - Application to Nuclear Magnetic-Resonance, *J. Chem. Phys.* *64*, 2229-2246.
178. Meyer, B., and Peters, T. (2003) NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors, *Angew. Chem. Int. Ed. Engl.* *42*, 864-890.
179. Ishima, R., and Torchia, D. A. (2000) Protein dynamics from NMR, *Nat. Struct. Mol. Biol.* *7*, 740-743.

180. Mittermaier, A., and Kay, L. E. (2006) New tools provide new insights in NMR studies of protein dynamics, *Science*. 312, 224.
181. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., and Skelton, N. J. (2007) *Protein NMR spectroscopy: principles and practice*, 2nd edn. ed., Elsevier, San Diego.
182. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, *Chem. Phys. Lett.* 69, 185-189.
183. Yamazaki, T., Yoshida, M., and Nagayama, K. (1993) Complete assignments of magnetic resonances of ribonuclease H from *Escherichia coli* by double- and triple-resonance 2D and 3D NMR spectroscopies, *Biochemistry*. 32, 5656-5669.
184. Venn, J. (1881) *Symbolic Logic*, Macmillan and Company, London.