To Fold or To Fibrillate? Serendipity in Stability Studies

Szczepankiewicz, Olga

2011

Preface

Proteins are fascinating molecules, complicated to understand by their own and even more so while taking part in various reactions and crowded environments. In order to elucidate some of this complexity, properties of a protein in a simple system are usually investigated. The amino acids in a protein interact non-covalently with each other and the surrounding. Each kind of interaction by itself is well understood but their interplay in a protein system is not easy to predict. Usually, a protein folds into its native and functional structure with optimized interactions. However, under certain conditions, a protein may form different kinds of aggregates. Some of them are just random precipitates and others, such as amyloid fibrils, have a well-ordered structure. Many diseases are associated with amyloid structures of proteins.

In this thesis, I have been looking at different proteins and how these were affected by various mutations. In some cases, the protein stability increased due to improved interactions while in other cases the mutations led to less stable and aggregating proteins. It is all about the balance and interplay of the non-covalent interactions in and around a protein. The title of the thesis, "To Fold or To Fibrillate?" reflects the question of how a protein will be affected by a mutation or by an extrinsic change. Since it is still not easy to predict these effects, it is necessary to perform experiments in order to answer the question. By doing both predictions and experiments, we will hopefully gain even more knowledge about the complexity of proteins.

During my years as a Ph.D. student, I have in part worked with the sweet protein monellin from Serendipity berries. Thanks to my supervisor Sara, I have learned that all unexpected results can be seen as serendipities and that these make our understanding of things wider.

Serendipity is all around; the hard thing is to see it!

Olga Szczepankiewicz
September 2011, Lund, Sweden

Populärvetenskaplig sammanfattning

Det finns tre typer av makromolekyler som är grundläggande för allt liv på jorden: DNA, RNA och proteiner. I varje cell i kroppen finns en kopia av vår arvsmassa, DNA. Denna består av gener som avkodas först till RNA och sedan till tiotusentals proteiner. Alla gener är inte aktiva i varje cell eller i varje tidpunkt. Proteiner kontrollerar nästan alla aspekter av det vi kallar liv. De transporterar syre, bygger upp vävnad, extraherar energi ur den föda vi äter, signalerar mellan och i celler och skyddar oss mot sjukdomar. Blir det fel på ett protein så kan det leda till sjukdomar som t.ex. diabetes, där proteinet insulin produceras i för liten mängd, eller Alzheimers som orsakas av att ett litet protein faller ut i hjärnan. Det är därför av stor vikt att studera olika proteiner och försöka förstå dess komplexa natur.
Man kan tänka sig att ett protein är som ett halsband av pärlor (aminosyror) med olika egenskaper. Det finns 20 olika sorters pärlor; de kan vara stora, små, runda, asymmetriska, blanka och/eller matta. Om du håller ett halsband i din näve så veckar det ihop sig till ett nystan. På samma sätt veckar sig ett protein till en kompakt struktur i vatten. Aminosyror som inte trivs i vatten samlas i mitten av strukturen medan de aminosyrorna som gillar vatten hamnar på ytan av proteinet. Proteinet får gynnsamma interaktioner (entalpivinst) när den veckas men förlorar samtidigt i antal konformationer som det kan anta (entropiförlust). Proteinets entalpi och entropi bestämmer dess stabilitet, det vill säga hur bra strukturen håller ihop. Den unika ordningen av aminosyrorna (pärlorna) i varje protein bestämmer hur proteinet kommer att veckas och vilken biologisk funktion proteinet får. Även om man känner till vilka aminosyror som ingår i ett protein så är det svårt att förutse hur proteinet kommer att veckas. Veckningsprocessen kan också påverkas av yttre faktorer och resultera i att proteinet antar fel struktur som kan vara sjukdomsalstrande.
Ett sätt att öka förståelsen om interaktionerna i ett protein är genom att byta ut (mutera) en eller flera aminosyror i det och se vilken effekt det har på proteinets (mutantens) struktur och stabilitet.

I denna avhandling sammanfattar jag resultaten från fem olika studier. Jag har tillsammans med andra forskare undersökt hur olika mutationer påverkar några proteiners egenskaper. I tre av studierna användes proteinet monellin (artikel I, II och IV). Monellin är ett protein som kommer från ett bär i Ghana. Det binder till samma smakreceptor på

tungan som vanligt socker men binder mycket starkare och ger upphov till längre söthetsupplevelse. Sockermolekylerna lossnar snabbt från receptorn vilket innebär att smakupplevelsen blir kortvarig. Det behövs hundra tusen fler sockermolekyler för att få samma söthetsupplevelse som från en enda monellinmolekyl. Monellin består naturligt av två proteinkedjor, *A* och *B*, som hålls ihop av interaktioner mellan dem.
Vi har tittat på olika mutanter av monellin för att se hur laddningen på proteinet påverkar dess egenskaper så som stabilitet, ihopbindning och söthet. Vi fann att interaktionerna mellan de två fragmenten *A* och *B* kunde korreleras till interaktionerna som håller ihop (stabiliserar) proteinet när *A* och *B* är sammankopplade till en kedja (artikel I). Detta kan utnyttjas när man vill studera proteinets stabilitet. Istället för att använda denatureringsmedel för att mäta när proteinet tappar sin struktur och därmed denatureras kan man titta på hur starkt de två fragmenten binder till varandra och därifrån förutse effekterna på dess stabilitet. När vi undersökte sötheten hos monellinmutanter fann vi att det är viktigt med en positiv laddning hos monellin för att det ska ge en söt smak (artikel II). Söthetsreceptorerna på tungan är negativt laddade och det kan vara anledningen till att en mutant av monellin som är positivt laddad binder till en receptor bättre än en mutant som är negativt laddad.
Fager är ofarliga virus och man kan få dem till att visa upp ett önskat protein. Ett fag-bibliotek är många fager som visar upp var sin mutant av ett protein. Mutanter med specifika egenskaper kan fiskas ut ur biblioteket. Vi framställde ett fag-bibliotek av monellinmutanter i syfte att hitta mutanter med högre stabilitet. De utfiskade mutanterna visade sig istället vara aggregeringsbenägna och vi undersökte korrelationen mellan aggregeringsbenägenheten och stabiliteten. Det visade sig att ju lägre stabilitet en mutant hade desto mer aggregeringsbenägen var den (artikel IV). Eftersom vi i denna studie fiskade ut mutanter på andra grunder än stabilitet bytte vi till en annan metod som heter ”delad GFP” för att försöka hitta stabilare mutanter av ett protein. GFP är ett protein som lyser grönt när det är veckat. Om man delar detta protein i två bitar så lyckas bitarna inte att binda ihop och få sin färg. Om man kopplar två fragment av ett annat protein som kan binda starkt till varandra gör man det också möjligt för fragmenten av GFP att binda ihop och få sin gröna färg. Vi lyckades att stabilisera modellproteinet PGB1 med denna metod (artikel III).
Efter att ha kommit i kontakt med aggregerande proteiner i studie IV ville jag fortsätta med att studera något aggregerande protein från

människa och valde Amyloid beta peptiden (Aβ) som är inblandad i Alzheimers sjukdom. Jag undersökte hur olika förlängningar av denna peptid påverkade dess aggregeringshastighet och fann en korrelation mellan längden av det extra fragmentet och aggregeringshastigheten. Ju längre förlängningen är desto bättre skyddar den mot aggregering. Detta tyder på att en förlängning med icke-aggregerande sekvens försvårar själva aggregeringsprocessen. (artikel V).

Studierna i denna avhandling har ökat vår förståelse av hur proteiner interagerar inom sig själva och att dessa interaktioner ofta är starka nog för att hålla ihop ett protein även om det delas i två separata fragment. Samma slags interaktioner verkar mellan proteiner och vi fann att laddning av olika proteiner har betydelse för hur två proteiner interagerar med varandra. Vi har också studerat förloppet då ett protein bildar aggregat istället för den veckade strukturen och fann att ett protein som innehåller aggregeringsbenägna segment kan skyddas av andra delar av sin sekvens som inte är aggregeringsbenägna.

Det är fortfarande en utmaning att förstå det komplexa samspelet mellan alla interaktioner i och mellan proteiner och mycket spännande forskning återstår!

List of papers and the author´s contribution to the papers

I. Xue WF, **Szczepankiewicz O**, Bauer MC, Thulin E, Linse S, Intra- versus intermolecular interactions in monellin: contribution of surface charges to protein assembly. *J. Mol. Biol*. 2006, 358:1244-55.

I performed the thermal stability studies and analyzed the data. I wrote the corresponding parts of the paper.

II. Xue WF, **Szczepankiewicz O**, Thulin E, Linse S, Carey J, Role of protein surface charge in monellin sweetness. *Biochim. Biophys. Acta*. 2009, 1794:410-20.

I designed and performed the taste study. I performed the urea denaturations by means of CD spectroscopy and analysis of the data. I contributed to the writing of the paper.

III. Lindman S, Hernandez-Garcia A, **Szczepankiewicz O**, Frohm B, Linse S, In vivo protein stabilization based on fragment complementation and a split GFP system. *Proc. Natl. Acad. Sci. USA*. 2010, 107:19826-31.

I and BF performed the fragment complementation studies and analyzed the data. I wrote the corresponding part of the paper.

IV. **Szczepankiewicz O**, Cabaleiro-Lago C, Tartaglia GG, Vendruscolo M, Hunter T, Hunter GJ, Nilsson H, Thulin E, Linse S, Interactions in the native state of monellin, which play a protective role against aggregation. *Mol Biosyst*. 2011, 7:521-32.

SL and I designed the library and initiated the project. I made the gene cloning and the screening. I and SL selected variants which I expressed. I, HN and ET purified the variants. I recorded the CD spectra and performed the thermal denaturation experiments. I analyzed the mutated positions in the crystal structure of the wild-type monellin. I and SL interpreted the results. I wrote the paper and collected improvements from the co-authors.

V. Szczepankiewicz O, Linse B, Frigerio Sala C, Frohm B, Thulin E, Walsh D, Linse S, Terminal extensions retard amyloid formation by an internal dilution mechanism. *Manuscript.*

I, BF and SL performed the kinetic aggregation experiments and analyzed the data. I and SL wrote the paper.

Additional publications not included in the thesis

Berggard T, **Szczepankiewicz O,** Thulin E, Linse S, Myo-inositol monophosphatase is an activated target of calbindin D28k. *J. Biol. Chem.* 2002, 77:41954-9.

Lindman S, Xue WF, **Szczepankiewicz O**, Bauer MC, Nilsson H, Linse S, Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys J.* 2006, 90:2911-21.

Cabaleiro-Lago C, **Szczepankiewicz O,** Linse S, Effect of nano-particles on amyloid aggregation of single chain Monellin mutants. *Manuscript submitted.*

Abbreviations

DNA	Deoxiribonucleic acid
RNA	Ribonucleic acid
MNA	Monellin A-chain
MNB	Monellin B-chain
MN	Monellin
scMN	Single chain monellin
PGB1	B1 domain of protein G
GFP	Green fluorescent protein
Aβ	Amyloid beta
CD	Circular dichroism
NMR	Nuclear magnetic resonance
UV	Ultra violet
ThT	Thioflavin T
N	Native state
U	Unfolded state
J	Joule
K	Kelvin
C	Celsius

One- and three-letter symbols for amino acids:

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

Acknowledgement

Till denna del behövs inga vetenskapliga referenser. Ni som nämns här är själva referenspunkterna i mitt liv! Tack!

Sara, jag slutar aldrig att förundras över din styrka, kunskap, medmänsklighet och den brinnande entusiasm för allt som du tar dig an. Din värme, ditt stöd och din tro på mig har varit avgörande för slutförandet av denna resa. **Eva H, Bodil, Hanna N** & **Björn R,** extraföräldrarna på jobbet. Vilken lyx att börja doktorera på ett ställe som genomsyrades av Er värme och omtanke. Ni fanns där i med- och motgång och fick det mesta att fungera! **Bengt** & **Jan-Erik**, förstår ni vilka gurus ni är i att förmedla kunskap? Er insats och Ert engagemang är ovärderlig, Er kunskap är gränslös! **Kristofer**, tack för ditt stöd och lycka till i undervisningens värld. **Birgitta**, tack för ett fint samarbete och omtanke. **Karin Å**, du är så glad och ditt skratt så ungdomligt, tack för din värme! **Tommy**, du är lugnet i stormen, tack för ditt stöd och uppmuntran. **Ingemar,** du är en cool forskare men det är ändå under din doktorandtid som de roligaste minnena kom till. **Johan**, du är smart, omtänksam och tar dig alltid tid för en pratstund, tack! **Gleb,** thanks for sharing the office with me and for discussions about everything and nothing. **Risto**, tack för ditt vackra leende, din härliga inställning och alla pollipops! **Erik**, mullen, klättermusen och vessikelkungen, tack för din hjälp. **Michalski**, tack för chokladen och den polska andan.
Torbjörn, Hans L, Hans F, Eva T, Bertil, **Mikael A**, **Martin, Celia, Marie G, Carl**, **Ulrich** och **Petra,** tack för ett fantastiskt stöd och för att ni varit så fina arbetskollegor! Good luck to the new people at BPC and in Sara's group: **Tanja, Xiaoting, Karin, Zhiwei, Akram, Shuji, Kalyani, Emelie** and **Filip.** På CPMS vill jag särskilt tacka: **Sinan, Maria A, Thom, Sindra, Hanna H, Kristina, Anders B, Andreas K, Henry, Susanna, Gert, Lo, Salam** och **Folke** för trevliga samtal och stöd. Tack alla andra på **CMPS** för en trivsam arbetsmiljö och att man alltid kan fråga och få hjälp.

I would like to thank my external collaborators: **Björn Linse**, **Janette Carey**, **Armando Hernandez-Garcia**, **Gian Gaetano Tartaglia**, **Michele Vendruscolo**, **Thérèse** & **Garry Hunter**, **Carlo Frigerio Sala** and **Dominic Walsh** for fruitful, valuable and much appreciated contributions to the papers.

Tack till **Jonny** och **Stefan** på vaktmästeriet för många uppiggande skratt! Tack **Paula** på Media-Tryck för fint samarbete kring prinsessboken och andra projekt. Tack **Johan** på tryckeriet i E-huset för att du vågade dig på mina idéer kring omslaget av denna bok! Tack till **Lena L** och **Thor L** för Er professionella hjälp på vägen.

Tack till Er på **Lillskolan** som förmedlat så mycket värme och kunskap till Sixten! Tack till **Ulf, Monica** och **Lars** för den kärlek ni omger Sixten med.

Tack till: **Johan G, Laura, Sanna, Rebecca** & **Sam, Birgitta S, Jenan, Karin A, Kinga, Andy, Anders H, Jörgen, Jonna, Jonathan, Gustav** och **Micke** för stöd och inspiration. Tack **Anders S** för att du får mig att skratta och för att du gjorde ”hemmakontoret” möjligt för mig. **Sara M,** keramikern med känslan för ”det där”. Du förstår och vet. En kopp te med dig är vardagslyx! **Magda H,** göteborgskan med glimten i ögat! Tack för att du gjort min mammatid i Göteborg till en fantastisk epok. Vi med våra små är helt enkelt oslagbara, inte minst enligt Al Gore och Alexander Lervik. **Wejcik,** ditt brinnande intresse för forskning kommer att ta dig långt! Du är en vän som alltid ställer upp, tack! **Aidish,** ditt leende är magiskt och du har introducerat många vackra saker i mitt liv. **Marie B,** omtänksam, klok, stark och alldeles underbar. Så härligt att kunna springa över till Er barfota och mötas av doften av nygräddade scones. **Emcik,** sprakande inspirationskällan! Våra kreativa dagar är som spa för själen! **Mikcik**, jag glömmer inte vår övernattning i stadsparken, leveransen av frukosten på småtimmarna och din flytthjälp på tom mage. Du bär alltid på några vardagsguldkorn i din ficka som du gärna delar med dig av! **Eva H**, få kan få in så mycket på så kort tid som du! Tack för den fina vänskapen och för att du får med mig till gerdahallen. **Åsa** och **Kristina,** vännerna från studietiden. Tio år har gått sen vår resa till Skottland, sen dess har några fått en kotte och en annan en skotte. Nu kan vi snart fira in den sista hatten! **Stincik**, min följeslagare i proteinernas värld och en vän för livet! Du räds inte för galenskaper samtidigt som du är klok, stark och omtänksam när det behövs! Jag vet ingen annan som är så ödmjuk och säker på samma gång! **Sylvek,** min fina ”syster” i nöd och lust. Det strålar om dig och jag beundrar dig för din styrka! Vi kan skratta åt det mesta för det mesta!

Dziękuję całej mojej rodzinie za wsparcie. Szczególne podziękowania dla **babci Helenki**, **babci Bolci**, **cioci Uli**, **cioci Irenki** i **wujka Edzia** za opiekę i miłość! Dziękuję **Rolandowi, Darii, Wiolecie, Grzesiowi** i **Joli** za miłe wspomnienia z dzieciństwa oraz za przyjaźń w dorosłym życiu. (Jag vill tacka hela min familj för all stöd. Särskilt vill jag tacka mormor Helena, farmor Bolcia, faster Ula, moster Irena och mosters man Edziu för att ni alltid funnits där mig och för all Er kärlek! Tack till Roland, Daria, Wioletta, Grzesiu och Jola för oförglömliga minnen från barndomen samt vänskap som varar än idag.)

Kim, tack för att du bodde med oss trots att du åkte på alla barnbaciller. Tack för att du hjälpt mig med hemsidan och att du alltid ställer upp med att ta foton när jag får för mig något galet upptåg.

Filip, Glucik. Hoppas att du hittar något spännande att arbeta med, du har en grym potential! Tack för din input på min avhandling, tur att det inte är du som är min opponent, hehe.

Tato, to chyba dzięki temu, że zabierałeś mnie do akademika no i dzięki temu zdrowemu koziemu mleku, zainteresowałam się nauką o białkach. Dziękuję za wsparcie w trudnych chwilach, miłość oraz za wspaniałe wakcje blisko natury i przemiłych zwierząt. **Agnieszko**, dziekuje Tobie za mile wakacje i wszystkie nasze wyjazdy. (Pappa, det är nog tack vare att jag fick hänga på ditt studentrum samt all den nyttiga getmjölken som jag blev intresserad av vetenskap och proteiner. Tack för ditt stöd i svåra stunder och tack för våra underbara naturnära somrar tillsammans. Agnieszka, tack för fina sommarlov och för alla våra utflykter.)

Mamma, Utan dig hade den här avhandlingen aldrig blivit av. Det är du som är värd en doktorshatt! Tack för din outtömliga kärlek, tålamod, stöd och din underbara inställning till livet! (Mamo, ten doktorat nie zostałby dokończony bez Twojej pomocy. To Ty powinnaś otrzymać kapelusz doktoracki! Dziękuję za Twoją miłość, cierpliwość, wsparcie i Twój fantastyczny pogląd na życie!)

Sixten, min skruttilurk, min älskling, mitt allt! Nu är boken med din teckning på klar! Du är en klok, busig och helt underbar pojke. Jag älskar dig!

Part one: Background to proteins

Introduction to proteins

Proteins are, together with RNA and DNA, the macromolecules fundamental in all life. DNA carries the genetic information which is transcribed to RNA. RNA is translated into amino acids at the ribosome to generate polymers called proteins. Proteins are built up of 20 different amino acids and the sequence of a protein contains the information for the structure that the protein is supposed to obtain. The conformation of a given folded chain is strongly connected to the biological function of that protein. Proteins are involved in numerous processes in living organisms such as enzymatic activities, signaling, transport, control of gene expression and immune response (Anfinsen 1973, Creighton 1993).
Trough evolutionary processes the sequences of natural proteins have emerged so that their unique native states can be formed efficiently even in the complex and crowded environment inside a living cell. Under some conditions, other conformations, such as amyloid fibrils (amyloids), are populated and there is a strict quality control machinery involving molecular chaperones and degradation systems in order to eliminate non-native conformations. A failure in this machinery may result in a wide range of diseases such as Alzheimer´s and Parkinson´s diseases (Dobson 2001, Chiti 2006).
The ability of natural proteins to form amyloids does not violate the hypothesis that a protein sequence codes for a single fold (Anfinsen 1973). Amyloids are formed as a consequence of interactions mainly involving the common polypeptide backbone (Dobson 2001).

Thermodynamics for reactions

According to the first and second law of thermodynamics it is stated that a reaction of reactant A into product B:

$$A \rightleftharpoons B \tag{1}$$

follows, at constant pressure and temperature, the equation:

$$\Delta_r G_{BA} = \mu_B - \mu_A \leq 0 \tag{2}$$

where $\Delta_r G_{AB}$ is the free energy difference of the reaction, i.e. the difference in the chemical potential of the product *B* and reactant *A*. The chemical potentials μ_A and μ_A are described as:

$$\mu_A = \mu_A° + RTln\left(\frac{\gamma_A.[A]}{[A°]}\right) \quad (3)$$

$$\mu_B = \mu_B° + RTln\left(\frac{\gamma_B.[B]}{[B°]}\right) \quad (4)$$

where μ_A^o is the chemical potential at a standard state and γ_A is the activity factor. At low concentrations, the activity factor γ_A can be approximated to 1. *[A]* is the concentration of *A* and $[A^o]$ is the standard state concentration (these parameters are denoted correspondingly for *B*). *R* is the gas constant (8.31451 $JK^{-1}mol^{-1}$) and *T* is the temperature in Kelvin. The standard state is usually chosen to be 1M (ideal solution) in order to be able to compare results. The free energy of the reaction $\Delta_r G_{BA}$ can be written as:

$$\Delta_r G_{BA} = \Delta_r G^o_{BA} + RTlnQ \leq 0 \quad (5)$$

where *Q* is:

$$Q = ([B]/[A]). \quad (6)$$

The term *RTlnQ* represents the entropic contribution upon mixing of components since entropy increases when molecules mix. This term balances the $\Delta_r G^o_{BA}$ until they are equal and the reaction is at equilibrium and *Q* becomes the equilibrium constant *K*:

$$\Delta_r G_{BA} = 0,\ \Delta_r G^o_{BA} = -RTlnK. \quad (7)$$

The term $\Delta_r G^o_{BA}$ in composed of the enthalpy $\Delta_r H^o_{BA}$ and entropy $\Delta_r S^o_{BA}$ and is described as:

$$\Delta_r G^o_{BA} = \Delta_r H^o_{BA} - T\Delta_r S^o_{BA} \quad (8)$$

where $\Delta_r H^o_{BA}$ refers to the internal enthalpy i.e. the energy of the reaction. $\Delta_r S^o_{BA}$ refers to the change in conformational energy of the molecules in water, as if each molecule *A* or *B* was independent of the

rest of the molecules (Atkins 1998).

Intermolecular interactions

All intermolecular interactions are of electrostatic origin. However, it is useful to classify intermolecular interactions into a number of categories. Intermolecular interactions have distance dependence and can be very long-range (Israelachvili 1992). In the following part, intermolecular interactions are described.

van der Waals interactions

The van der Waals interactions are attractive, non-specific and short-range interactions affecting atoms in close contact. They exist between all atoms and molecules and fine-tune the atomic packing in folded proteins. They have the distance dependence of $1/r^6$, where r is the separation distance. These attractive interactions arise between dipoles (Keesom interaction), between dipole and induced dipole (Debye interaction) and between two induced dipoles (dispersion interaction). Dispersion interactions are present between all atoms and molecules and are transient dipolar interactions that arise from fluctuations in the electron densities of nearby atoms. Together with the steric repulsion (atoms cannot be present at the same place at the same time), van der Waals interactions give rise to the Lennard-Jones potential describing the potential energy $U_{AB}(r_{AB})$ of the interaction between atoms A and B at a distance r_{AB}:

$$U_{AB}(r_{AB}) = 4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right]. \quad (9)$$

Here the ε_{AB} is the maximal attractive interaction and σ_{AB} is the point of zero potential. The $1/r^6$ term corresponds to the van der Waals interaction and the $1/r^{12}$ term accounts for the steric repulsion (Israelachvili 1992).

Coulombic interactions

Electrostatic interactions between point charges are called Coulombic interactions and the free energy of such interaction $w(r)$ in a solution is described as:

$$w(r) = \frac{z_1 z_2 e^2}{4\pi\varepsilon_0\varepsilon_r r} \quad (10)$$

where z_1 and z_2 are ionic valences of the charges, ε_0 is the permittivity in vacuum, ε_r is the dielectric constant of the medium, r is the distance between the two charges and e is the elementary charge.
The presence of salt screens Coulombic interactions so that they decay more rapidly than for isolated ions and this is described by the Debye-Hückel theory and the free energy of the interaction $w_{DH}(r)$ becomes:

$$w_{DH} = \frac{1}{\varepsilon_r(T)} \frac{z_1 z_2 e^2}{4\pi\varepsilon_0 r} e^{-\kappa r} \quad (11)$$

where $1/\kappa$ is the Debye length which is proportional to the square root of the ionic strength (Creighton 1993).

Hydrogen bonds

A hydrogen bond is a strong directional dipole-dipole interaction between one electronegative atom (O, N, F and Cl) and a hydrogen atom that is covalently bound to another electronegative atom. The small hydrogen atom gets positively polarized by the covalently bonded electronegative atom which results in a strong interaction with the second electronegative atom.
Water is a very unusual substance due to its high dipole moment and high hydrogen bonding capacity. Each water molecule can participate in four hydrogen bonds forming a tetrahedral coordination which gives rise to a three-dimensional network of water (Israelachvili 1992).

Hydrophobic effect

In the case of simple solutions, there are two opposing forces acting when solutions *A* and *B* are mixed. The driving force for mixing comes from the increase in translational entropy when the system contains both *A* and *B* instead of only *A* or only *B*. Mixing is opposed by loss of enthalpy. The enthalpy and entropy are usually not changing significantly with temperature.
When an apolar molecule, *A*, is transferred from its liquid state into water, it disrupts the strong hydrogen bonds of water molecule network. This gives rise to the hydrophobic effect which is driven by

the attempt of water to satisfy all its hydrogen bonds by expelling the introduced molecules that cannot participate in hydrogen bonds. The free energy of the transfer $\Delta_r G^o_{trans}$ is:

$$\Delta G^o_{trans} = \Delta H^o_{trans} - T\Delta S^o_{trans}. \quad (12)$$

$\Delta C_p{}^o$, the change in heat capacity at constant pressure, is the difference in energy needed to increase the temperature of a solution one degree. A significant increase in $\Delta C_p{}^o$ of the system accompanies the transfer process. $\Delta C_p{}^o$ is described as:

$$\Delta C_p{}^\circ(T) = \frac{d\Delta H^\circ_{trans}(T)}{dT} \quad (13)$$

and is the positive slope of ΔH^o_{trans} versus temperature (Fig. 1). This means that more energy is required when apolar molecules are in water than when they are in themselves. $\Delta C_p{}^o$ gives a temperature dependence of the enthalpic ΔH^o_{trans} and entropic $T\Delta S^o_{trans}$ contributions which are counteracting each other (enthalpy-entropy compensation).

At low temperature, molecules in liquid water participate on average in 3-3.5 hydrogen bonds but around an apolar molecule water molecules may have a coordination of four and thus they are more ordered than the bulk water. This is entropically unfavorable. Therefore, at low temperature, the free energy of transferring an apolar molecule into water results in a decrease of the entropy (Fig. 1).

In contrast, at high temperature, the system behaves more like a simple solution. Water molecules have more motion and the introduction of an apolar molecule does not affect the entropy, instead there is an increase in the enthalpy due to lost hydrogen bonds. At the temperature T_h the enthalpy of transfer is zero and at the temperature T_s the entropy of transfer is zero. The hydrophobicity is the strongest when the entropy of transfer is zero, i.e. at $\Delta_r G^o_{trans}(T_s)$ (Southall 2002) (Fig. 1).

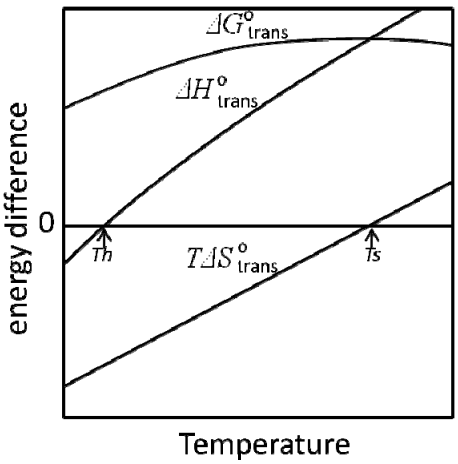

Figure 1. Representative plot of the thermodynamics of transferring apolar molecules into water. The enthalpic ΔH^{o}_{trans} and entropic $T\Delta S^{o}_{trans}$ contributions as well as the free energy ΔG^{o}_{trans} are presented. T_h and T_s are the temperatures at which the entropy and enthalpy of transfer are zero, respectively (Southall 2002).

Intramolecular interactions in proteins

The intramolecular interactions in proteins are of the same origin as the intermolecular interactions described above. The equilibrium between native *N* and unfolded *U* conformation of a protein depends mainly on two large and opposing energy-terms; the hydrophobic effect (governs folding) and the conformational entropy (governs unfolding because the unfolded state represents many more conformations than the native state). Both contributions are temperature dependent which affects their strength at each individual temperature that is studied. This leaves space for other intramolecular interactions to be important for the overall free energy difference of unfolding ΔG^{o}_{UN}. The net stability of a protein is small compared to the individual contribution from different interactions. The structure, function, stability and aggregation propensity of a protein is a result of an interplay of all the interactions and the conformational entropy.

Protein unfolding has many similarities with the hydrophobic effect. If we introduce a protein in its native globular state into water, its surface can participate in the hydrogen bonding network (the entropy of the folded protein is low but the interactions of water are satisfied) while the hydrophobic residues are clustered into a densely packed hydrophobic core excluded from water. If, instead, we introduce an unfolded protein there will be many exposed apolar groups which will

affect the hydrogen bonding of water as described above (the entropy of the protein is high but interactions with water are not favorable). The shift in the $T\Delta S^{o}_{UN}$ term in figure 2 compared to figure 1 is due to a large increase of the chain entropy upon protein unfolding. At the temperature T_h the enthalpy of unfolding is zero and at the temperature T_s the entropy of unfolding is zero. The hydrophobicity is the strongest when the entropy of transfer is zero, i.e. at $\Delta G^{o}_{UN}(T_s)$.At this point the unfolding of the protein is the most unfavorable and therefore this is where the protein is most stable. ΔG^{o}_{UN} is positive under conditions where the native state predominates. At the point T_m, the fraction of the native protein is equal to the fraction of the unfolded protein (Southall 2002).

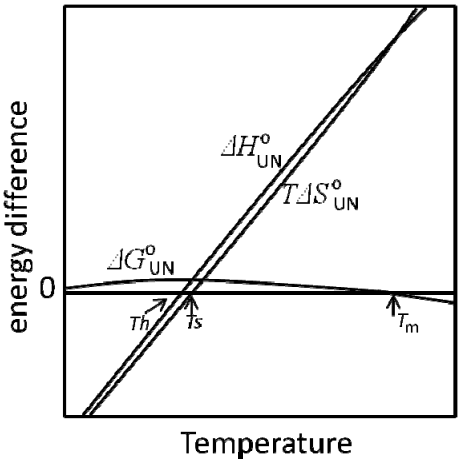

Figure 2. Representative plot of the thermodynamics of protein unfolding. The enthalpic (ΔH^{o}_{UN}) and entropic ($T\Delta S^{o}_{UN}$) contributions as well as the free energy (ΔG^{o}_{UN}) are presented. T_h and T_s are the temperatures at which the entropy and enthalpy of transfer are zero respectively. The fraction of the native protein is equal to the fraction of the unfolded protein at the point T_m (Southall 2002).

The importance of the hydrophobic effect in protein folding has been confirmed in mutational studies where mutations in the hydrophobic core generally have much larger effect on the stability than mutations on the protein surface (Berggård 2001, Dell'Orco 2005).
While the hydrophobic effect and the conformational entropy are the major energy terms in protein folding, the other interactions influence this process and may also be important for other processes such as binding and catalysis. The van der Waals interactions and steric repulsions play an important role in the packing of the hydrophobic

core in the native state.
In the native state the hydrogen bonds of the backbone N-H and C=O are satisfied by organizing into specific secondary structures. In the unfolded state most peptide groups hydrogen bond with water. The contribution of hydrogen bonds to ΔG°_{UN} is small if all hydrogen bonding capabilities are satisfied in both the native and the unfolded state. Hydrogen bonding gives specificity to a small repertoire of secondary structures of proteins since only conformations satisfying all hydrogen bonding capabilities are favored (Dill 1990).
Coulombic interactions can be both stabilizing and destabilizing. It is however believed that proteins contain charges for other reasons than to improve stability. Coulombic interactions are of importance in for example protein function, specificity and kinetics (Tõugu 1996, Linse 2000, Dell'Orco 2005, Xue 2006). Another important aspect of Coulombic interactions is the avoidance of unspecific association (Otzen 2000, Andrè 2004, Sandelin 2007). The Coulombic interactions are dependent on the pH which determines the total charge of the protein. The Coulombic interactions are also dependent on the ionic strength since salt shields charges and thereby determines the extent of interactions between the charges. Most proteins have a net charge which may be unfavorable for the folding. However, proteins need charges to increase solubility and for function (Dill 1990).

Protein folding and stability

Proteins are built up in ribosomes and come out as flexible, disordered polypeptide chains. Protein folding is the search of the free energy minimum and many proteins fold spontaneously into a native conformation as a result of a complex interplay of many different types of interactions between the residues of the polypeptide chain and with the solvent. The major contribution comes from the hydrophobic effect opposed by the loss of conformational freedom in the polypeptide chain. Proteins do not go through all possible conformations when they fold. The general view today is that the energy landscape of folding is shaped as a funnel and that the native stable state can be found by multiple routes down the funnel (Dill 1997). Proteins do not need a specified pathway to fold as long as the native state has the lowest free energy (Linse 2007). The process of protein folding may be very quick (10^{-6} s) or may take some time (10^{-1} s) while the protein goes through different states (Creighton 1993, Dyer 2007).

The native form N usually refers to the functional protein with a specific structure while the unfolded form U refers to a peptide chains with randomly distributed conformations. Protein folding is a cooperative process and essentially all protein molecules are either in the native or unfolded state (Anfinsen 1973). The most common model used for protein stability is the two-state model between these two states:

$$N \rightleftharpoons U \tag{14}$$

with the equilibrium constant K:

$$K = \frac{\gamma_U \cdot [U]}{\gamma_N \cdot [N]}. \tag{15}$$

The activity factors, γ_U and γ_D are assumed to be 1 so that the equilibrium constant K is the ratio between the concentrations of the unfolded and native states:

$$K = \frac{[U]}{[N]}. \tag{16}$$

Protein stability refers only to the difference in standard free energy at equilibrium ΔG^o_{UN} between the unfolded and the native state and do not include the folding route:

$$\Delta G^o_{UN} = G^o_U - G^o_N = -RT\ln K, \tag{17}$$

$$K = e^{-\Delta G^\circ_{UN}/RT}. \tag{18}$$

Since the reaction of unfolding is unimolecular, the standard state concentration does not need to be specified. This also means that the equilibrium distribution between native and unfolded state is concentration independent. This holds as long as the concentration is sufficiently low so that no other reactions are significant.
ΔG^o_{UN} can also be expressed in terms of enthalpy- (ΔH^o_{UN}) and entropy-differences (ΔS^o_{UN}):

$$\Delta G^o_{UN} = \Delta H^o_{UN} - T\Delta S^o_{UN}. \tag{19}$$

ΔH^o_{UN}, enthalpy of unfolding, is the heat of the reaction. ΔS^o_{UN}, entropy of unfolding, includes both the conformational entropy of the polypeptide and the solvation of the protein. ΔG^o_{UN} is positive under conditions where the native state predominates (Fig. 2).
Although the enthalpy and entropy terms are large they counteract each other. Proteins are only marginally stable and the difference in free energy ΔG^o_{UN} between the unfolded and native state is only 10-40 kJ/mol under physiological conditions (Pace 2004).

Protein denaturation

To be able to study the stability, both the native and unfolded states have to be populated at equilibrium. Since, under physiological conditions, the population of the unfolded state is negligible, external means can be used to propagate the unfolded population. Thermal and chemical denaturations are two examples of approaches that are commonly used to increase the population of the unfolded state (Pace 1975). The denaturation transition can be followed by spectroscopic methods that are used to measure the fractions of the native F_N and unfolded F_U protein (Fig. 3).
At each point in the transition the fractions of native and unfolded proteins are:

$$F_N = \frac{[N]}{[N]+[U]} = \frac{1}{1+K},\ F_U = 1 - F_N = \frac{[U]}{[N]+[U]} = \frac{K}{1+K} = \frac{Y_{obs}-Y_N}{Y_U-Y_N}. \quad (20)$$

These equations apply to reversible transitions. The baselines, before Y_N and after Y_U the transition, are assumed to be straight lines and are described as:

$$Y_N = k_N \cdot X + b_N \quad (21)$$

$$Y_U = k_U \cdot X + b_U \quad (22)$$

where k_N and k_U are the slopes, b_N and b_U are the intersections with the Y-axis. X is the denaturating parameter. X_m is the midpoint of the transition and is defined as the point where F_N is equal to F_U.

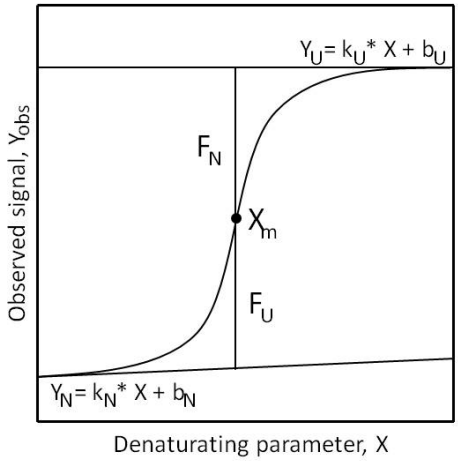

Figure 3. Hypothetical 2-state denaturation process. The baselines are described as straight lines. X_m is the midpoint where F_U=F_N. The transition is steep which means that it is cooperative, the molecules are either in the native or unfolded state.

The observed signal Y_{obs} can be expressed as a function of the baselines and ΔG^o_{UN} and is fitted to the experimental data:

$$Y_{obs} = \frac{Y_N + Y_U * e^{-(\Delta G^\circ UN/RT)}}{1 + e^{-(\Delta G^\circ UN/RT)}}. \quad (23)$$

Y_{obs} can be normalized in order to compare different measurements:

$$Y_{norm} = \frac{Y_{obs} - Y_N}{Y_U - Y_N}. \quad (24)$$

The stability at the condition of interest is obtained by extrapolation. The manner of extrapolation will depend on whether the denaturation is thermal or chemical.
The process of unfolding is usually reversible but this has to be confirmed in each case. Irreversibility may be caused by chemical degradation or precipitation of the protein. Data from irreversible denaturations can be used to compare apparent midpoints of different mutants of a protein and indicate their apparent stability.

Thermal denaturation

Thermal denaturation is obtained by raising the temperature until and beyond the protein changes from native to unfolded state and can be

studied by using calorimetry, circular dichroism (CD) or fluorescence spectroscopy.
The expression of ΔG^{o}_{UN} as a function of temperature is given by the Gibbs-Helmholtz equation (Makhatadze 2004):

$$\Delta G^{\circ}_{UN}(T) = \Delta H^{\circ}_{UN}(T_m)\left[1 - \frac{T}{T_m}\right] + \Delta Cp^{\circ}_{UN}\left[T - T_m - Tln\frac{T}{T_m}\right]. \quad (25)$$

Using the expression for ΔG^{o}_{UN} according to equation (25) in equation (23) gives an expression of the measured signal as a function of the variable temperature. The resulting equation is fitted to the thermal denaturation data. The model is fitted to the measured data using k_N, k_U, b_N, b_U, ΔH^{o}_{UN} and ΔCp^{o}_{UN} as variables. The fit is used to estimate the temperature T_m, which is the temperature where half of the molecules are unfolded, so that F_N is equal to F_U. The stability at any temperature can be estimated by extrapolating from the midpoint temperature to the chosen temperature.

Chemical denaturation

Unfolding of the protein can also be propagated by adding a chemical such as guanidine hydrochloride (GdnHCl) or urea that favors the unfolded state. The process of denaturation can be studied by CD or fluorescence spectroscopy. It is not clear how aforementioned chemicals cause the unfolding but it is believed that they bind to the peptide groups and make them more prone to be solvent exposed. It has been shown experimentally that ΔG^{o}_{UN} depends linearly on the concentration of the denaturant (Pace 1975):

$$\Delta G^{\circ}_{UN}(denaturant) = \Delta G^{\circ}_{UN}(H_2O) - m \cdot [denaturant] \quad (26)$$

where $\Delta G^{o}_{UN}(H_2O)$ is the free energy of unfolding in pure water and m is the linear coefficient. Equation (26) is then used in equation (23) to relate the denaturant concentration to the observed signal in the experiment. The m-value is believed to be related to the change in solvent accessible surface area upon unfolding (Myers 1995). When $\Delta G^{o}_{NU}(\text{denaturant})=0$, the mid-point concentration, C_m, of the unfolding is given by:

$$C_m = \frac{\Delta G^\circ{}_{H_2O}}{m}. \quad (27)$$

Protein binding and assembly

Proteins account for almost all biological functions in living organisms. The intramolecular interactions give rise to the functional protein and the intermolecular interactions with other proteins or ligands allow them to act in different processes.
Protein binding resembles the process of protein folding concerning the enthalpy-entropy compensation. In addition, the gain in energy upon binding must compensate for the loss in entropy when one complex is formed out of two molecules. In the case of ligand binding or protein assembly, there is concentration dependence. Dilution of the whole system governs dissociation. The simplest model of protein assembly is a protein *A* that binds to a protein *B*:

$$A + B \rightleftharpoons AB. \quad (28)$$

Commonly 1M (ideal solution) is chosen as standard state. At equilibrium the activity factors are assumed to be 1 and the equilibrium association constant K_A is:

$$K_A = \frac{\frac{[AB]}{1M}}{\frac{[A]}{1M}\cdot\frac{[B]}{1M}}. \quad (29)$$

The equilibrium dissociation constant K_D is the inverse of K_A and is usually used in biological contexts (Creighton 1993):

$$K_D = \frac{1}{K_A}. \quad (30)$$

Importantly, the standard free energy of complex formation, ΔG^o, will depend strongly on the standard state. However, changes in ΔG^o, the $\Delta\Delta G^o$, due to for example mutations will be independent of the standard state. At a chosen standard state, ΔG^o is:

$$\Delta G^o = -RT\ln(K_A) = \Delta H^o - T\Delta S^o. \quad (31)$$

The equilibrium constant can also be expressed using the total

concentrations $[A]_{tot}$ and $[B]_{tot}$. The concentrations $[A]_{tot}$ and $[B]_{tot}$ must be determined in order to interpret the data:

$$[A]_{tot} = [A] + [AB] \quad (32)$$

$$[B]_{tot} = [B] + [AB] \quad (33)$$

$$K_A = \frac{[AB]}{([A]_{tot} - [AB]) \cdot ([B]_{tot} - [AB])}. \quad (34)$$

The concentration of the complex $[AB]$ is:

$$[AB] = \frac{([A]_{tot} + [B]_{tot} + \frac{1}{K_A})}{2} - \sqrt{\frac{([A]_{tot} + [B]_{tot} + \frac{1}{K_A})^2}{4} - [A]_{tot} \cdot [B]_{tot}}. \quad (35)$$

This expression is often plugged into equations describing binding data (signal Y) to make them dependent on the total concentrations only:

$$Y = (K_D, [A]_{tot}, [B]_{tot}). \quad (36)$$

The degree of saturation is:

$$F_{sat} = \frac{[AB]}{[B]_{tot}} = \frac{[A]}{K_D + [A]}. \quad (37)$$

Protein reconstitution

Protein reconstitution is a case of protein assembly, as described above. Many proteins can be divided into two fragments that, upon mixing, reconstitute to form a functional protein with the same fold as the intact protein although lacking one or more peptide bonds (Finn 1992, Kobayashi 1995, Linse 1997). The loss of the configurational entropy upon assembly of protein fragments into a complex is greater compared to folding of a single chain. As long as this loss can be compensated by the interactions formed in the complex the reconstituted protein forms. The degree of folded structure in the isolated fragments before reconstitution may vary from completely disordered to structure-containing (Carey 2007). A correlation between the stability of the intact protein mutants and the affinity of

reconstituting fragments of the same mutants has been found in several studies (Ruiz-Sanz 1995, Berggård 2001, Xue 2006). This indicates that there is a general correlation between the processes of folding and reconstitution and that protein reconstitution involves the same non-covalent interactions as those involved in folding of the intact protein. Protein reconstitution can be studied under conditions where the protein is maximally stable instead of at denaturating conditions which are far from the physiological conditions (Carey 2006).

Protein aggregation

Proteins may aggregate due to intrinsic as well as extrinsic factors. Aggregated proteins may have amorphous unstructured form or may form amyloid fibrils, which are highly repetitive and ordered structures. Almost all proteins contain segments that can form aggregates. These segments are usually protected within the native conformation but can be exposed upon destabilization of the native state. When hydrophobic regions become exposed to water, the process of aggregation competes with folding (Routledge 2009, Dobson 2001). The propensity of globular proteins to aggregate is often inversely related to the stability of the native state and increased tendency for unfolding seems to promote aggregation (Chiti 2006). Proteins with certain functions were selected trough evolution. This does not mean that the proteins are optimized with respect to stability. Proteins need to be flexible in order to interact with other components and are degraded after carrying out their intended task. The propensity of aggregation can be decreased by binding to different ligands that stabilize the native fold (Masino 2011). A process resulting in aggregated protein may occur due to a defect in the protein sequence or folding, a misbalance in its surrounding or a well-controlled aggregation mechanism. Not every aggregated protein form is dangerous or disease causing. Functional amyloids have been found in many different organisms where they are strictly regulated (Chiti 2006, Otzen 2008).
Amyloid fibrils are characterized by cross-β stacking of peptide segments perpendicular to the long axis of the fiber (Nelson 2005). The formation of amyloid fibrils is preceded by the formation of a wide range of aggregates such as unstructured oligomers and structured proto-fibrils. Recent development makes it possible to distinguish between oligomers and mature fibrils during the process

(Lindgren 2010).
A large number of human diseases are linked to extensive protein amyloid formation. Examples are Alzheimer's and Parkinson's diseases and dialysis-related amyloidosis. The proteins involved in these diseases are amyloid β peptide (Aβ), α-synuclein and β2-microglobulin, respectively. There is often a poor correlation between the aggregated deposits and the symptoms of the disease. Therefore, recent studies have focused on other species of the proteins and it is suggested that the toxic species are small oligomers formed on the pathway towards fibrils (Lambert 1998, Winner 2011).
Most of the aggregation diseases are not associated with genetic mutations but with sporadic events and particularly old age. Proteins have evolved to fold efficiently and to remain correctly folded despite their inherent tendency to aggregate. In order to prevent aggregation, living systems have developed several strategies such as chaperones, folding catalysts and degradation mechanisms. However, in some cases protein aggregation escapes the rescue machinery. Selective pressure during evolution has resulted in proteins that are able to resist aggregation as long as a human is able to reproduce. As we get older, there is not only a greater probability of mutations but aging is also linked to failure in the rescue machinery (Dobson 2001).

To fold or to fibrillate?

In the native state of a protein, the structure is optimized so that most of the hydrophobic residues are situated in the core of the fold while the rest of the residues are located at the surface. Many proteins with substantial differences in amino acid sequences and native structures have been shown to be able to form amyloid fibrils *in vitro* (Chiti 2006). Therefore, the amyloidogenic structure is believed to be generic and many proteins may form amyloids since they contain amyloidogenic segments that can participate in this kind of structure.
The question is, which of the two structures, at a certain condition, is the one with lowest free energy, a monomer in the native state or a monomer in an amyloid structure? A related question is, how is this balance governed by intrinsic and extrinsic factors? Often, the native state is the one that forms fast under native conditions and the transition into the amyloidogenic form takes time due to kinetic barriers. There are reports suggesting a delicate balance governing whether a protein will fold or fibrillate (Tartaglia 2008). In some

conditions, the fibrils may be the most stable form of the protein (Chiti 1999, Baldwin 2011).

Stabilization of proteins

Since stability is the difference in free energy between the native and unfolded state, it can be altered by changes in both states. A protein mutant with improved stability has a larger difference in the free energy between the native and unfolded state compared to the wild type. The increase may come from lowering of the free energy of the native state as well as from increase of the free energy of the unfolded state. Upon a mutation, the different non-covalent interactions are all affected to different degrees. The stability of the new protein mutant is a result of the complex interplay of all these interactions. Changes in electrostatic and van der Waals interactions usually have a larger effect on the native state. Hydrogen bonds are generally equally satisfied in both states. The hydrophobic effect may be increased or decreased depending on how a mutation affects the free energy of both states. Introduction of disulfide bonds and prolines can improve protein stability by restricting the unfolded state and increasing the free energy of this state (Van den Burg 1998).
Most proteins are not optimized for stabilities but rather for functionality. A common problem with proteins used in different applications is their limited stability toward different forms of handling and storage or after administration in the body. Proteins may denaturate, aggregate or undergo proteolytic degradation. The value of many industrially or therapeutically used proteins can therefore be increased if their stability can be improved. It is often useful to combine different approaches in order to be successful. Increased stability may be obtained by repacking of the core as well as by introducing substitutions at the boundary positions and at protein surface (Malakauskas 1998, Martin 2001). Many approaches towards increased stability of different proteins have been performed by rational redesign of the amino acid sequence (Eijsink 2004). Comparison of a protein structure and sequence with its thermostable analogs may give valuable insights into the factors governing stability (Van den Burg 1998). Computational approaches are useful in order to design new variants with potentially improved stability (Malakauskas 1998, Makhatadze 2004). However, it is difficult to predict the effects of a mutation. Combinatorial approaches, such as directed evolution,

allow introduction of several mutations and different combinations of these in a large library of variants (Eijsink 2005). Phage display and split GFP (green fluorescent protein) are two methods for library screening. These methods connect the improved variants to their DNA that can be sequenced or enriched for further rounds of selections. Enzyme activity, protease resistance and reconstitution of fragments are examples of selection strategies applied in order to find variants with improved stabilities (Linse 2000, Robertson 2004, Wunderlich 2006).

Phage display and fragment complementation

Instead of studying the effect of one or a few predesigned mutations on a protein, researchers have increasingly been using combinatorial approaches. Generation and characterization of individual mutant proteins is a labor-intensive process. In contrast, combinatorial approaches, for example those using phage display, offer simultaneous generation of a large amount of protein variants that can be used in a selection experiment to find improved variants. Phage display uses an immobilized target to select the best binding sequences from a peptide or protein library (Scott 1990). Reconstitution in combination with phage display has been used to identify key residues for the interaction between two fragments (Smith 1993, Linse 2000).
Even if it is possible to construct large combinatorial libraries, the sequence space even for a relatively small protein is much larger than any library that can be produced at reasonable laboratory scale. Thus, a rational approach is required for the application of phage display in protein engineering unless incomplete libraries are accepted. Libraries of mutants, generated by using degenerated oligonucleotides, can either be focused to a limited region of the protein so that a small number of amino acids can be randomized or so that restricted mutations can be introduced in a larger number of amino acids. Libraries may also be produced in a completely random fashion by using error-prone PCR (polymerase chain reaction). The design of the phage display experiment and the selection strategy used depend on which questions that are asked. The DNA of selected variants can be sequenced to find patterns and to interpret the favorable mutations that are selected (Hoess 2001).

Split GFP and fragment complementation

In the split GFP (green fluorescent protein) method, two fragments of GFP are used as the reporters for two interacting molecules. These could be two different proteins or fragments of one protein that are able to reconstitute. The concept of reconstitution can thus be used in the split GFP method. The affinity of the GFP fragments is too low for a spontaneous assembly in cells. If interacting partners are fused to the ends of GFP fragments, the local concentrations of the GFP fragments are increased and GFP associates into a native fold which gives rise to chromophore development (Ghosh 2000, Wilson 2004, Magliery 2005). The difference in the fluorescence intensity is coupled to the difference in affinity between variants of the associated fragments. This can be used for optimization of protein stability by investigation of the corresponding reconstitution of the fragments at physiological conditions (Lindman 2009).

Part two: Experimental systems

Monellin

The protein monellin, MN, was originally isolated from the tropical serendipity berry (*Dioscoreophyllum cumminsii*) (Morris 1972). This protein is naturally composed of two fragments, MNA (45aa) and MNB (50aa) (Kohmura 1990) that reconstitute a globular, single domain fold (Bohak 1976). Monellin adopts the β-grasp fold, (one of the most common folds among known protein structures) which is an α-helix packed perpendicularly on a five-stranded anti-parallel β-sheet (Ogata 1987, Day 2003) (Fig. 4). The A-chain forms three strands of the anti-parallel β-sheet whereas the B-chain forms two β-strands separated by an α-helix. Although the biological function of monellin is unknown, the protein is about 100 000 times sweeter than sucrose on a molar basis. A single chain monellin variant, scMN, with retained structure and sweetness has been produced and has increased stability against thermal denaturation (Fig. 4) (Kim 1989). Studies of monellin have focused on identifying the molecular determinants of its sweetness (Morris 1975, Kohmura 1992), stability (Kim 1989, Sung 2001) and aggregation tendency (Morris 1980, Somoza 1993, Konno 1999, Konno 2001).

The consumption of carbohydrates is linked to many diseases like diabetes, obesity and caries. Many attempts have been done in order to design new sweeteners. Sweet tasting proteins like monellin are of great interest from this aspect. Monellin seem to interact with the same sweet-taste receptor, T1R2/T1R3, as the low molecular weight sweeteners. Although the structure of the receptor is not yet resolved, homology studies indicate that the large surface of monellin interacts with an external cavity of the receptor (Temussi 2002).

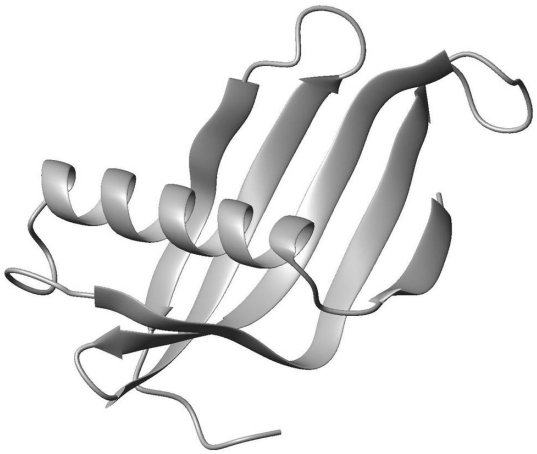

Figure 4. Ribbon structure of single chain monellin which adopts the β-grasp fold with the α-helix packed perpendicularly on a five-stranded anti-parallel β-sheet. The figure was prepared from PDB file 1iv7 using MolMol software (Koradi 1996).

Amyloid beta peptide

Amyloid β peptide (Aβ) is associated with Alzheimer´s disease (AD) which is a neurodegenerative disorder causing senile dementia. The major pathological hallmarks in the brains of AD patients are amyloid plaques, which are extracellular aggregates of insoluble Aβ peptide fibrils, and the intracellular tangles of the protein tau. The most abundant peptides in the extracellular fibrils are Aβ(1–40) and Aβ(1–42), the latter having faster fibrillation kinetics (Fig. 5). Intracellular Aβ is also found and oligomers of Aβ have been shown to interact with a variety of proteins such as tau, which is a protein that supports the microtubular skeleton of axons, which permit passage of nutrients into synapses. Oligomeric Aβ binds to tau protein and this causes a rapid dissociation of tau and the microtubules. The axonal structure

collapses leading to misfunction of synapses and finally to cell death of neurons (Shankar 2009, Rauk 2009).

Aβ is a 40-42 residue peptide cleaved from the C-terminal region of a much larger transmembrane amyloid precursor protein, APP, by the enzymes β-secretase and γ-secretase. The net charge of the Aβ peptide is -3. The long hydrophobic tail (G29–V40 or A42) reduces the solubility and enhances membrane adhesion. The monomeric Aβ in solution is more or less unstructured (Walsh 1999). Another peptide that is proteolytically processed from APP is Aβ(17-40 or 42) by the action of α-secretase and γ-secretase. Variants of Aβ that are both longer and shorter than 42 residues have been found in extracts from brain and cultured cells (Larner 1999). Although Aβ amyloid fibrils show cytotoxic properties, it is believed that Aβ oligomers are the primary cytotoxic species (Lambert 1998, Walsh 2002). The amount of Aβ is controlled by its production, degradation and clearance and it is proposed that a failure in these processes may cause an accumulation of Aβ which triggers the pathogenic cascade (Walsh 2009).

The structure of the Aβ monomer in a fibril consists of two β-strands (for Aβ(1-40) residues 12-24 and 30-40 and for Aβ(1-42) residues 13-21 and 30-39) folded back-to-back across a hairpin turn (Tycko 2003, Sato 2006, Petkova 2006, Chiti 2006). The N-terminal is supposed to be solvent exposed and flexible. Monomers aligning on top of each other form a so called protofilament that twist together to form a fibril (Chiti 2006).

The aggregation kinetics have a sigmoidal appearance which is characteristic of a nucleation dependent polymerization (Chiti 2006). The process starts with a lag phase followed by a rapid exponential growth phase which levels off while reaching equilibrium. The process is stochastic on the microscopic level (few molecules) with a variation in the individual processes. In a sample containing 10^{14} molecules of Aβ the observed kinetic trace, followed by thioflavin T (ThT) fluorescence, is an average over all the individual reactions and is highly reproducible (Hellstrand 2010).

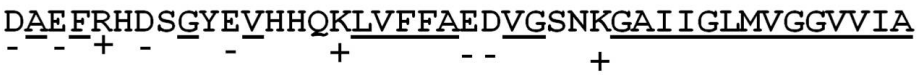

Figure 5. The amino acid sequence of Aβ(1-42). The hydrophobic residues are underlined. The C-terminus G29-A42 is highly hydrophobic while the N-terminal part is more hydrophilic. The positive and negative residues are marked with a plus and minus sign, respectively.

B1 domain of protein G

Streptococcal protein G is a cell surface protein composed of two domains, PGB1 and PGB2. These domains bind to the constant fragment of immunoglobulin G (IgG) with high affinity. The coating with antibodies on the bacterial surface is believed to be a means for the bacteria to escape the host immune system (Goward 1993). B1 domain of protein G (PGB1) is often used as a model protein due to its well-known structure, folding, stability and spectrum. It consists of 56 amino acids with a β-grasp fold that can reconstitute from fragments with a dissociation constant of 9μM (Fig. 6) (Gronenborn 1991, Kobayashi 1995). N-terminal cleavage is observed but is minimized by introduction of the mutation TQ2 (Smith 1994). Asparagine residue followed by a glycine residue is very deamidation prone (Reissner 2003). In order to avoid deamidation, the mutations N8D and N37D were introduced (Lindman 2006). The PGB1 variant with the mutations T2Q, D8N and D37N is called PGB1-QDD and is the variant used in this thesis.

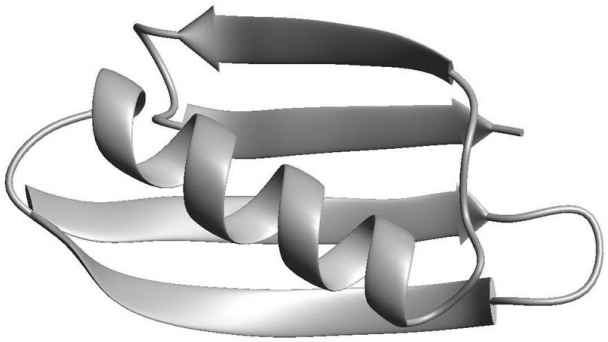

Figure 6. Ribbon structure of PGB1, which adopts the β-grasp fold with the α-helix packed on a four-stranded anti-parallel β-sheet. The figure was prepared from PDB file 3pgb using MolMol software (Koradi 1996).

Green fluorescence protein

Green fluorescent protein (GFP) comes from a jellyfish (*Aequorea Victoria)* and becomes fluorescent upon folding without the requirement of any cofactors (Zimmer 2009). It has a β-barrel structure (eleven β-strands and one α-helix in the interior) and consists of 238 amino acids (Ormö 1996, Tsien 1998). The chromophore of GFP is

composed of S65, Y66 and G67 located in the central α-helix. Upon folding of GFP, these residues undergo an autocatalytic cyclization leading to maturation of the chromophore (Craggs 2009). The wild type protein has several excitation peaks and in order to obtain only one excitation peak at 475 nm and emission at 505 nm, the following mutations were introduced: F64L, S65C, Q80R, Y151L, I167T and K238N (Palm 1997). Upon co-expression of GFP fragments comprising residues 1-157 (GFPN) and 158-238 (GFPC) no fluorescence was observed (Magliery 2005). However, if the two GFP fragments are fused separately to two other protein fragments which are able to reconstitute, the GFP fragments may reconstitute as well and develop fluorescence. This is used in the split GFP method which can be used for *in vivo* screening of protein-protein interactions.

Part three: Methods

Optical spectroscopy

In spectroscopy electromagnetic waves (photons) are used to probe transitions between energy levels of electrons. A transition is possible if a photon matches the energy gap between two energy levels. It is also required that the shape of the electron distribution in the ending energy level is different from the starting level (so that the redistribution of electron density occurs in a non-symmetric fashion; the transition dipole moment is non-zero) (Atkins 1998).
In a protein, the absorption and emission of photons is localized to certain units called chromophores. The backbone peptide unit absorbs light at around 200-220 nm, the aromatic side chains and disulfide bonds between cysteines absorb at 250-300 nm (Creighton 1993).

Absorbance spectroscopy

In absorbance spectroscopy the light that comes through the sample is compared with the incoming light. Trough Lambert-Beer´s law, the absorbance A can be measured:

$$A = \varepsilon \cdot c \cdot l \tag{38}$$

where ε is the extinction coefficient (which has a specific value for

each molecule at each wavelength), c is the concentration and l is the cuvette length (Atkins 1998). The absorbance can be used to obtain the concentration of the sample (keeping in mind the sources of errors from contaminants and light scattering). For proteins, UV (ultra violet) light of 280 nm is typically used to monitor absorbance from aromatic residues (Gill 1989).

Circular dichroism spectroscopy

Natural proteins are made of L-enantiomeric amino acids which, except for glycine, contain at least one chiral atom, C_α. Chiral molecules that are chromophores absorb left- and right-handed circular light to different extent. It is the difference in absorbance between the two components that is measured in circular dichroism spectroscopy.
Far-UV (250-180 nm) circular dichroism (CD) spectroscopy probes the local electronic environment of the peptide backbone and is thus sensitive to protein secondary structure. The secondary structural elements have characteristic spectra and a protein built up of different secondary structures will obtain a spectrum that is a mix of the spectra that each component has by itself.
In Near-UV CD (320-250 nm), the disulfide bonds (broad weak signals throughout the near-UV spectrum) and aromatic signals (F=250-270 nm, Y=270-290 nm and W=280-300 nm) provide information about the tertiary structure.
CD spectroscopy can thus be used in order to study secondary and tertiary structural changes upon introduction of mutations. Also, denaturation processes can be followed by CD spectroscopy.
A sample is illuminated by alternating left- and right-handed circularly polarized light. Since chiral molecules absorb the left- and right-handed circularly polarized light to different extent, there will be a difference in absorbance of the two light components and the outcoming light polarization changes from linear to elliptical form. The ellipticity measured in units of milidegrees (mdeg) can be converted into molar ellipticity $[\Theta]$ with units milidegrees cm^2/decimole according to:

$$[\Theta] = \frac{\Theta}{(c \cdot l)} \tag{39}$$

where c is the concentration and l is the path length.

For far-UV CD measurements on proteins, the mean residue molar ellipticity $[\Theta]_n$ is often used to compare different samples:

$$[\Theta]_n = \frac{\Theta}{(c \cdot l \cdot n)} \quad (40)$$

where n is the number of residues (Kelly 2005).

Fluorescence spectroscopy

Electrons become excited to higher energy levels upon absorption of photons. This event is followed by a vibrational relaxation to the lowest vibrational level in the excited state. For some molecules the return to the ground state is accompanied by a release of a photon. This phenomenon is called fluorescence. The energy is always higher for the excitation than for the emission (emission occurs at longer wavelength than the exciting light). Fluorescence emission spectrum can be used to probe the local electronic environment of aromatic side chains and is thus sensitive to protein tertiary structure.
In proteins, tryptophan (λ_{max}(abs) =280 nm, λ_{max}(em) =348 nm) is the brightest fluorophore followed by tyrosine (λ_{max}(abs) =274 nm, λ_{max}(em) =303 nm) and phenylalanine (λ_{max}(abs) =257 nm, λ_{max}(em) =282 nm). The fluorescence of a tryptophan is sensitive to the polarity of the environment around the fluorescent group which can shift the emission wavelength making it suitable for protein denaturation studies. The shift is towards longer wavelengths in a more polar environment. A buried fluorescent group usually has a higher fluorescence intensity than when it is solvent exposed, where collisions with the solvent quench the fluorescence. Besides tryptophan, tyrosine and phenylalanine, there are other molecules that are fluorescent. GFP has an unusual chromophore formed through a covalent reaction of three of its amino acid residues.
Thioflavin T (ThT) is a fluorescent dye that shows enhanced fluorescence upon binding to amyloid fibrils. It forms micelles in aqueous solution which tend to bind fibrils but the exact mechanism is not well understood (Khurana 2005).

Thioflavin T fluorescence assay-fibrillation kinetics

Aggregation kinetics can be monitored using ThT as a reporter of fibril formation. The excitation wavelength of 440 nm and the emission wavelength of 480 nm are usually used. The aggregation process is followed in a plate reader. The aggregation curves have a sigmoidal appearance with lag phase, elongation phase and finally an equilibrium plateau, characteristic of nucleated polymerization reaction (Chiti 2006). Kinetic parameters are obtained by fitting the following empirical sigmoidal equation to the experimental curves:

$$Y_t = y_0 + \frac{y_{max} - y_0}{1 + e^{-(t - t_{1/2})k}} \tag{41}$$

where Y_t is the fluorescence intensity at time t, y_0 and y_{max} are the initial and maximum fluorescence intensities respectively, $t_{1/2}$ is the time required to reach half of the maximum intensity and k is the apparent first-order aggregation constant (Nielsen 2001). Lag time t_{lag} can be defined as:

$$t_{lag} = t_{1/2} - 2/k. \tag{42}$$

In the case of Aβ and the extended variants in paper V, there is a strong correlation of the lag time and the total monomer concentration at the start of each experiment. The simplest function that fits reasonably well to the lag time versus total Aβ concentration $t_{lag}(c)$ is a power function:

$$t_{lag}(c) = Bc^{\alpha} \tag{43}$$

where c is the Aβ concentration, B is a proportionality constant and α is the exponent (Hellstrand 2010).

Part four: Introduction to papers

I. Intra- versus intermolecular interactions in monellin: contribution of surface charges to protein assembly.

The contribution of Coulombic interactions to protein stability has been investigated in many studies. It seems that nature does not use the Coulombic interactions as the main to optimize stability but rather hydrophobic interactions (Berggård 2001, Dell'Orco 2005, Wunderlich 2005). Electrostatic interactions are long-range and a change in charge will affect all its favorable and unfavorable interactions with the rest of the charges in the protein. Since the charges are closer in space in the folded state, both favorable and unfavorable effects on the free energy are more pronounced in the folded state.
Stability studies require both native and unfolded states to be populated. Different ways of denaturation are used to populate the unfolded state. The same non-covalent interactions that govern folding are also involved in assembly of two fragments of the same protein (Ruiz-Sanz 1995, Berggård 2001). The exceptions are the two extra charges at the cut site. The advantage of studying assembly instead of stability is that the first can be done under physiological conditions.

In this study the role of the Coulombic interactions on monellin assembly as well as stability was investigated. We investigated variants of single chain monellin (scMN), where fragment B is connected to fragment A by a glycine residue, and variants of monellin composed of two fragments A and B (MN) (Kohmura 1990, Spadaccini 2001). Three variants with mutations altering surface charges were designed and the mutations were introduced in both scMN and MN. The mutations were placed in the B fragment and only mutations from aspartic acid to asparagine or glutamic acid to glutamine were allowed in order to not change other properties of the amino acids than the charge. The mutation C41S was introduced in all mutants in order to avoid unwanted dimerization. scMN(+3/-1) and MN(+3/-1) with the net charge of +2 correspond (with the exception of C41S) to wild type scMN and MN respectively. scMN(+3/+5) and MN(+3/+5) and scMN(+3/-5) and MN(+3/-5) are the variants with altered charges on the surfaces with the net charges of +8 and -2, respectively. scMN names denoted with total net charge, scMN+2, scMN-2 and scMN+8,are used below.

CD spectroscopy was used to follow the thermal stability of the scMN mutants at different salt concentrations. Fluorescence spectroscopy and ITC were used to follow the reconstitution of the MN mutants at different salt concentrations. The equilibrium constant K_A (the affinity constant) and the rate constant of association k_{on} were obtained from a multiple-method global fit (Xue 2004).

A charge in a protein interacts favorably or unfavorably with all the other charges present in the rest of the protein. Introduction of a charge in order to match a preexisting charge in a protein is often accompanied by repulsive interactions with other charges. Therefore, the net contribution of Coulombic interactions is the difference between favorable and unfavorable contributions. The thermal stability was lower for scMN+8 compared to scMN+2 and scMN-2, probably due to its high net charge. The stability of this mutant increased with increased salt concentration indicating shielding of repulsive Coulombic interactions by salt (Fig. 6a). The stability of the other two mutants was similar to the wild type scMN (Sung 2001).

The affinity constant between fragment A(+3) and fragment B was not increased upon the change in charge of the B chain from -1 to -5. The affinity constant decreased upon introduction of repulsive electrostatic interactions on the B chain where the charge was changed from -1 to +5 (Fig. 6b). In a study by André *et al.*, it was shown that increased charge difference between interacting partners did not increase the affinity between them. In order to understand this, one has to think of the effects of changed charge on both the free and bound states of the partners. Increased charge difference between interacting partners may increase the affinity. However, this may have an unfavorable effect of the folding process (André 2004).

The rate constant of association was lowest for MN(+3/+5) while it was highest for the MN(+3/-5) complex (Fig. 6c). In the case of MN(+3/+5) both affinity constant and rate constant of association reach a maximum at 150 mM salt (Fig. 6b, c). The free energy of unfolding ΔG^{o}_{NU} of the scMN variants and the association constants of the corresponding MN variants correlate (Fig. 6d). This indicates that the same Coulombic interactions are involved in both folding and reconstitution. There is also a correlation between ΔG^{o}_{NU} of the scMN

variants and the association rate constants of the MN variants while no such correlation is observed between ΔG^{o}_{NU} and the rate cosntants of dissociation k_{off} (Fig. 6e, f).

In a previous study of the protein Calbindin D_{9K}, much larger effects on ΔG^{o}_{NU} and affinity constants were seen upon mutating the hydrophobic residues in the core of the protein than the effects of surface charge mutations. In the case of the hydrophobic mutations, there was a strong correlation between ΔG^{o}_{NU} and affinity constants as well as ΔG^{o}_{NU} and the rate constants of dissociation (Berggård 2001, Dell'Orco 2005).

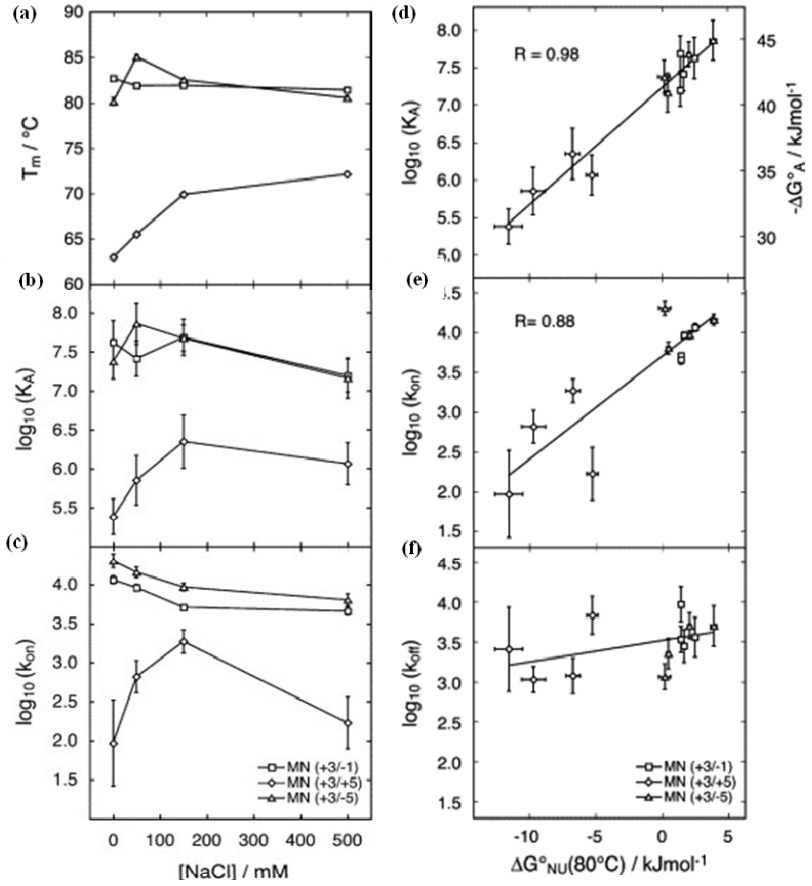

Figure 6. Thermal stability (T_m) of the scMN mutants (a), the affinity constant K_A of MN fragments (b) and the rate constant of association k_{on} of the fragments (c) at different NaCl concentrations. Affinity constants (d), rate constants of association (e) and rate constants dissociation (f) plotted against thermal stability of scMN, free energy of unfolding, ΔG^{o}_{NU} at 80ºC.

Conclusion: While it is difficult to increase affinity and stability by introducing attractive electrostatic interaction, it is easier to decrease affinity by introducing repulsive electrostatic interactions. Upon introducing a change, one has to be aware of the effects on both states in a studied equilibrium. Since charges give rise to long-range electrostatic interactions, all favorable and unfavorable contributions of a new charge have to be taken into consideration. Proteins use a net charge to be soluble but too high charge hinders folding into the native structure. A fine balance is achieved by proteins to obtain both solubility and folding. Compared with hydrophobic interactions the contribution of Coulombic interactions to protein stability and binding is minor. Charge interactions may instead modulate binding by altering the rates of association. The correlation between folding and assembly indicates that the same Coulombic interactions are involved in both folding and assembly. This opens the possibility to use reconstitution as an indirect method of finding variants with potentially increased stability.

II. Role of protein surface charge in monellin sweetness.

The intensely sweet tasting proteins brazzein, mabinlin, monellin, neoculin and thaumatin have been isolated from different tropical or subtropical plants (Morris 1972, van der Wel 1972, Liu 1993, Ming 1994, Shirasuka 2004). Even hen egg white lysozyme elicits a sweet taste even though it is much less sweet than the plant proteins (Masuda 2005). These sweet proteins are of different sizes but a common property is that they have positive net charges at neutral pH. There is no sequence homology among the known sweet proteins and there is little structure similarity between sweet proteins with resolved structures: brazzein, monellin (Fig. 7a, b), neoculin and thaumatin (de Vos 1985, Ogata 1987, Caldwell 1998, Shimizu-Ibuka 2006).
A few studies have shown that the sweet proteins interact with the same receptor, T1R2-T1R3, which is also activated by small sweet molecules, for example sugars and dipeptides (Cagan 1979, Li 2002). The structure of the T1R2-T1R3 receptor is not yet resolved but it has a sequence homology to the mGluR1 receptor for which the structure is resolved. An in silico model for the protein-receptor interaction suggests that monellin, brazzein and thaumatin, although different in structure, contain surfaces that are complementary to a large electronegative cavity in the receptor (Temussi 2002, Morini 2005, Esposito 2006).

We wanted to evaluate the role of surface charges in sweetness of monellin. The single chain mutants of monellin, scMN+2, scMN-2 and scMN+8 were used, previously described in paper I (Xue 2006). We investigated the structure, stability and sweet taste of the mutants. To find general trends of the importance of charges in sweet proteins, we summarized the data for different sweet proteins and charge mutants of these.
2D NMR, far-UV CD, and fluorescence spectroscopy were used in order to investigate if the variants maintain the wild type monellin structure.

The results reveal that the introduced charge changes do not perturb the wild type fold of monellin at the investigated conditions.
The stability of the variants toward denaturation by urea was determined by monitoring the unfolding transition at equilibrium using far-UV CD and fluorescence spectroscopy. Results from these two

methods coincide, which indicates that the proteins unfold in a cooperative manner with simultaneous loss of secondary and tertiary structure. scMN+8 has a lower stability, probably due to increased overall repulsion between the large number of uncompensated positive charges.
The sweetness of the variants was evaluated in a human taste study using a sweetness threshold detection assay. The average threshold concentration of scMN+2 was ca. 70 nM. The scMN+8 variant had slightly reduced sweetness compared to scMN+2 with an average threshold concentration of 460 nM. The sweetness of the variant scMN-2 was not detectable at the highest concentration of 8.2 μM. For comparison, the sweetness threshold for sucrose is 10 mM.

We summarized the threshold data of charged monellin mutants that were known at the point of this work. Two of the variants produced in this work represent the most extreme charge changes introduced to monellin.

Among other single mutation variants of monellin, sweetness is maintained upon removal (with one exception) but not addition of negative charge. No variant with increased positive net charge was found to be sweeter than the wild type monellin which indicates that natural monellin sequence may already be optimized for maximum sweetness. Similarly, André *et al.* found that the affinity between the negative calmodulin and positive peptides is optimized at rather low charge of each partner (André 2004).
The convex surface of monellin, suggested to bind to the receptor, is natively dominated by hydrophobic residues (Fig. 7b) and the strong complex formation with the sweet receptor suggests burial of these groups in the complex. Introduction of charges (Fig. 7c) on this surface of monellin showed significant decrease in the sweetness, which supports the hypothesis that the hydrophobic surface is important for the interaction with the receptor (Temussi 2002, Morini 2005 and Esposito 2006). The polar surface of monellin facing away from the receptor (Fig. 7b) has an excess of positive over negative charges and may modulate the interaction with the receptor in an indirect manner due to the long-range electrostatic effects. The general charge complementarity of monellin with the receptor opens the possibility that electrostatic attraction may keep the partners close to each other promoting reassociation of the partners and a prolonged apparent

association time as a consequence.

Combined mutational data for monellin, brazzein, neoculin, thaumatin and lysozyme, revealed general trends in sweetness and its dependence on the total charge. A rough trend toward reduced sweetness with reduced net positive charge was observed.

Conclusion: Modifications of monellin charges affect the sweetness without significant changes in structure. Among proteins known to be sweet, there is a rough trend towards reduced sweetness with reduced net positive charge which suggests that an overall positive surface electrostatic potential is important for binding to the receptor and the sweet taste.

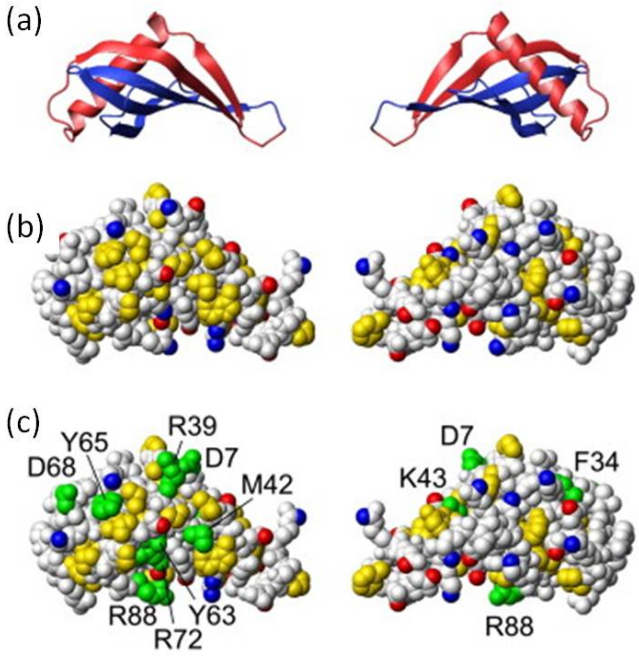

Figure 7. Model structures of scMN. a) Ribbon models of scMN, A chain in blue and B chain in red, viewed from convex (left) and concave (right) sides. b) Space filing models of scMN in the same orientations as in a. c) Picture of charge variants with at least ten-fold reduction in sweet taste (green side chains). Hydrophobic side chains are in yellow, acidic residues have one atom in red and the basic residues have one atom in blue. Figures were prepared from scMN crystal structure 1liv7 from PDB using MolMol software (Koradi 1996).

III. In vivo protein stabilization based on fragment complementation and a split GFP system.

In nature, proteins seem to be optimized for function and not stability. It is often found that there is room for increased stability. Mutations can be introduced in order to lower the free energy of the native state more that for the unfolded state or to increase the free energy of the unfolded state more than for the native state.
Proteins used in different industries need to be stable during processing, storage and final use. Plenty of approaches have been developed in order to increase the stability of different proteins, for example: rational design of mutations, combinatorial libraries of variants (directed evolution) and computational approaches (Malakauskas 1998, Eijsink 2004, Wunderlish 2005). It is often useful to combine different approaches in order to obtain best results.

In this study, we introduce a new approach to stabilize proteins. It is based on the correlation, found previously, between mutational effects on the stability of an intact protein and the effects of the same mutations on the affinity between fragments of the protein (Fig. 8a) (Ruiz-Sanz 1995, Berggård 2001, Xue 2006). This correlation implies that proteins may be stabilized by searching for fragments with enhanced affinity. This can be achieved by using a phage display library in which one fragment of a protein is degenerated and attached to the phage while the other fragment is present in solution. Selection of phages binding to the second fragment may result in complexes with increased affinity (Linse 2000). We initially attempted this approach for monellin, however the aggregation propensity of its sequence led to selection of surface active variants (see paper IV). Here we use split GFP system and fragment complementation. Fragmented GFP has too low affinity to reconstitute in cells but this may be achieved upon fusion of the GFP fragments to fragments of another protein that reconstitutes. The GFP fragments then co-reconstitute and give rise to green fluorescence that can be used as a selection marker (Lindman 2009). The split GFP method promotes reconstituted variants of a protein and eliminates selection of surface active variants. This work is a proof-of-principle study in order to screen for improved affinity between fragments, 1-40 and 41-56, of B1 domain of protein G (PGB1). PGB1 is a small well-studied protein that reconstitutes from fragments (K_D=9μM) and we choose it as a model protein in this study

since there is a lot of data on stabilizing mutations. The variant PGB1-QDD, which we call parent, contains the mutations T2Q, N8D and N37D in order to avoid processing of the sequence. A library of the PGB1 1-40 was designed including mutations which have previously been shown to stabilize and destabilize PGB1. PGB1 mutants with higher affinity between the fragments were obtained by in vivo screening of the library against wild-type 41-56 fragments. Colonies were ranked based on the green fluorescence intensity by human eye.

The DNA was sequenced for the three selected clones with most intense green fluorescence and intact variants of PGB1 1-56 were expressed, purified and evaluated for stability. The intact variant derived from the most intense green fluorescence complex (Top 1) yielded a 12ºC increase in the midpoint of thermal denaturation (Fig. 8b). For the other two variants T_m was increased by 9 (Top 2) and 8ºC (Top 3), respectively and the green fluorescence intensity followed the same trend as in stability. The stabilized mutants all had novel combinations of the mutations in the library, including the additional mutations and a few spontaneous mutations not included in the library design. The mutations affected many different types of interactions in the intact protein as well as between the fragments. For example, mutations L7I, V29F and V39I seem to improve the hydrophobic packing and/or increase the hydrophobic effect, while the mutation A34T may provide better hydrogen bonding in the native state and removal of the charge by the mutation K28T may remove a local electrostatic repulsion in the folded state. This complex combination of all different interactions is difficult to predict in advance, which makes this method very powerful.

Conclusion: We show that the correlation between fragment complementation and protein stability can be used in an in vivo approach based on the split GFP system. A protein that is able to reconstitute can be linked to fragments of GFP so that the GFP fragments become able to reconstitute upon the reconstitution of the protein. The intensity of the green fluorescence emitted by the reconstituted GFP can be used in selection of complexes of the protein with increased affinity.

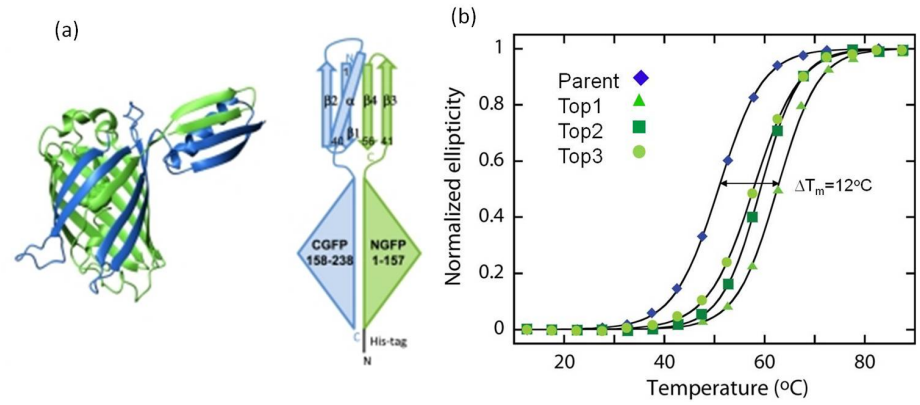

Figure 8. a) Left: A model of the ribbon structure of split GFP with fused PGB1 fragments. The model is produced using MolMol (Koradi 1996) and the PDB files 1ema and 1pgb. Right: Topology of PGB1 fused to GFP fragments. In blue; the N-terminal starting with residues 1-40 of PGB1-QDD, linked to residues 158-238 of GFP. In green, C-terminal starting with residues 1-157 of GFP linked to residues 41-56 of PGB1. b) Normalized thermal unfolding as observed by loss in CD signal at 218 nm. The normalized fit to data is shown as solid lines in all cases.

IV. Interactions in the native state of monellin, which play a protective role against aggregation.

Under native conditions, proteins usually fold into a well-defined globular structure. The hydrophobic effect is the main driving force in folding upon which a hydrophobic core of the protein forms while the hydrophilic amino acids are distributed on the surface.
Proteins are flexible structures participating in different equilibrium systems. Proteins may not only form the native structure but also, under certain conditions, adopt a general amyloidogenic structure. The same kind of non-covalent interactions (Coulombic, van der Waals interactions, hydrophobic effect and hydrogen bonds) are present in both the native and the aggregated state but in different proportions. It is of great importance to understand the underlying intrinsic and extrinsic factors that determine which structure a protein will adopt.
Proteins may contain local regions that are sensitive to aggregation; those regions are usually protected from aggregation since they are buried in the protein structure. The propensity of globular proteins to aggregate is often inversely related to the stability of the native state (Chiti 2006, Tartaglia 2008).

Here we address the role of intrinsic factors in governing native versus aggregated state. We investigated stabilities, fibrillation kinetics and aggregation propensities of single chain monellin and a series of five mutants of this protein. Our parent scMN is composed of the two subdomains of monellin, B and A, connected with a methionine and contains the mutation C41S in order to eliminate disulfide formation. The five mutants, scMN10-14, of parent scMN were retrieved from a phage display library based on their surface activity. All allowed modifications in the mutants are situated in subdomain A and they alter the hydrophobic packing between the subdomains and the surface charge of the protein.

Thermal denaturation was studied using CD spectroscopy. Thioflavin T fluorescence was monitored to follow the fibrillation kinetics. The scMN variant sequences were analyzed by an algorithm to identify the regions of the sequence that are likely to promote aggregation. The algorithm can also take into account the protective role of eventual structured regions against aggregation (Tartaglia 2008, Pechmann 2009).

The variants, scMN10-14, contain in total up to 14 mutations or deletions. Five hydrophobic residues in the hydrophobic core were altered or deleted in all the variants. Each of these five residues is involved in both inter- and intradomain non-covalent interactions. Perturbations of these interactions may lead to destabilization of the protein (by disfavoring the folded state and/or governing of the unfolded states), increased exposure of an aggregation-prone region and thereby enhanced aggregation and formation of amyloid fibrils.

All mutants showed lower stability and faster aggregation kinetics than the parent scMN. There is a correlation between the apparent midpoint of thermal denaturation and the lag phase for fibrillation (Fig. 9a). The least stable mutant displays the shortest lag time for fibrillation, mutants with intermediate stability have intermediate lag times and the most stable mutant has the longest lag time. These results are consistent with previous studies that also reported the presence of a correlation between protein stability and fibrillation time (Chiti 2000, Baden 2008). High net charge of a protein may prevent aggregation due to repulsions between protein molecules. At the same time, a high net charge may interfere with protein stability and increase the exposure of aggregation-prone regions. These two effects need to be balanced in order for a protein to fold correctly. The most stable mutant has a charge of +1 while the rest of the variants have a charge of +3. The parent scMN has a charge of +2 and that seems to be the most optimal charge balance to avoid aggregation.

By analyzing the sequences with the algorithm, two aggregation prone regions in the parent and mutated scMN variants were found. Surprisingly, all mutants had a lower total aggregation propensity based on sequence alone. The aggregation propensities with the correction for structure $\check{Z}_{agg}$ correlate with the fibrillation kinetics (Fig. 9b). It seems that destabilization of the variants of monellin leads to increased exposure of aggregation-prone segments which results in increased tendency for fibril formation. In the most stable parent scMN, the aggregation prone regions are most effectively prevented from aggregation by burial within the folded state.

Conclusion: We suggest that an increase in the aggregation propensity of a protein upon mutations can be explained by the increase in the exposure of the amyloidogenic regions caused by the destabilization of the native state.

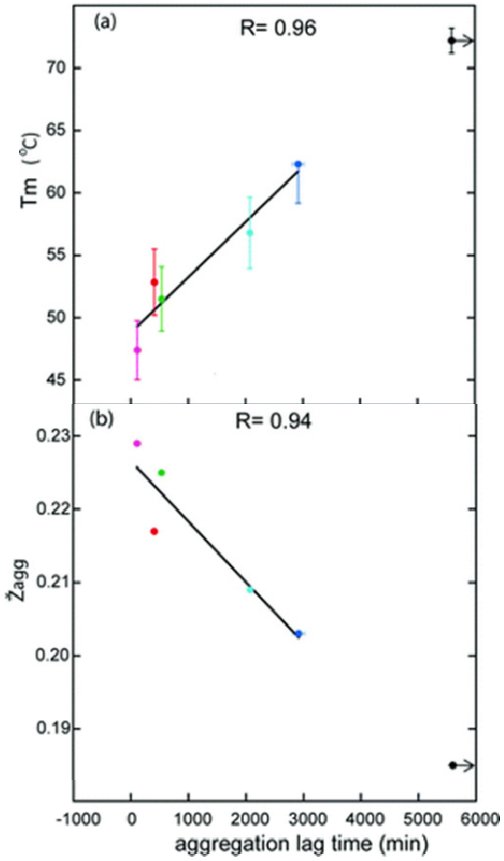

Figure 9. a) Correlation between apparent midpoint of thermal denaturation, T_m, and aggregation lag time for parent scMN and mutants. b) Correlation between predicted aggregation propensity with structural correction, $\check{Z}_{agg}$, and aggregation lag time of parent scMN and mutants. Parent scMN (black), scMN10 (pink), scMN11 (red), scMN12 (green), scMN13 (cyan) and scMN14 (blue).

V. Terminal extensions retard amyloid formation by an internal dilution mechanism

The proteins involved in the amyloidogenic diseases for example Alzheimer's and Parkinson's diseases and dialysis-related amyloidosis are of high interest in order to understand the mechanisms of these diseases.
In Alzheimer's disease, the peptide Aβ is believed to be the main component in the disease-causing mechanism. Aβ is produced by protease cleavage of the large transmembrane protein APP. The biological roles of Aβ and APP are not clear. The most abundant forms of Aβ are 40 or 42 residues long but both shorter and longer variants have been observed. In AD (Alzheimer´s disease) patients, there is a large deposition of aggregated Aβ in the brain but the correlation of the deposition and the symptoms of the disease is poor. Several studies show that small oligomeric entities of Aβ are the toxic species that disturb synapse function and give rise to the symptoms of AD (Lambert 1998, Walsh 2002). Several groups have been studying different lengths of Aβ. It has been found that regions of the hydrophobic C-terminal are crucial for the neurotoxicity and aggregation (Liao 2007). Truncations of the N-terminal result in variants that are more prone to aggregation (Pike 1995, D'Arigo 2009).

In this study, we investigated the effect of N-terminally extended variants on the aggregation kinetics. We produced the wild-type Aβ(M1-42) and six variants of Aβ with the extension of the sequence in steps of 5 residues (Aβ(-5-42), Aβ(-10-42), Aβ(-15-42), Aβ(-20-42), Aβ(-25-42) and Aβ(-30-42)). The added fragments come from the APP sequence. We followed the kinetics of the fibrillation using a continuous thioflavin T fluorescence assay.

We examined a concentration series of each peptide and found that for all peptides the aggregation kinetics are concentration dependent and follow a power function. All the variants have strongly retarded aggregation compared to wild-type. There is a correlation between the logarithm of the half time of the aggregation process ($\log(t_{1/2})$) and the length of the peptide up to the extension of 25 residues (Fig. 10a). We suggest that the retardation is caused by the extended unstructured termini that disturb the aggregation–prone regions from formation of amyloid contacts. As previously shown, truncation of the N-terminus

has an opposite effect (Pike 1995, D'Arrigo 2009). Also for other proteins (α-synuclein, myoglobin and Cks1) retardation is seen upon extensions and acceleration upon truncations of the proteins as long as the aggregation-prone segment stays intact (Kessler 2003, Hoyer 2004, Bader 2006, Corrêa 2011).
In the case of the variant Aβ(-30-42)the trend is broken (Fig 10a). It is slower than the wild-type but faster than the Aβ(-15-42). This could be due to partial folding but the far-UV-CD spectrum of this mutant is highly similar to that of the wild-type reflecting a largely unstructured peptide. Perhaps this extension has an increased affinity to self-associate compared to shorter extensions.

The second part of this work is a Monte Carlo simulation in which a simple model of peptide aggregation was investigated. The model represented a peptide with a C-terminal aggregation-prone region of six residues. The peptide was then extended at the N-terminus with 1, 2, 3 ,4, 5 or 6 residues. The aggregation rate was retarded for all the extended variants and there is a linear correlation between the mid-point of the aggregation process and the extension length over the whole series (Fig. 10b). By lowering the concentration of the peptide in the simulation, the aggregation rate decreases and this effect is similar to the effect of extension.

The correlation between the logarithm of the half time and the length of peptide is also valid throughout all experiments, up to the extension of 25 residues and we suggest that the effect arise from a general physico-chemical effect which we call internal dilution.

Conclusion: Aggregation rate is reduced by terminal extensions outside the aggregation prone region. We suggest that this result arises from a general physico-chemical effect which we call internal dilution.

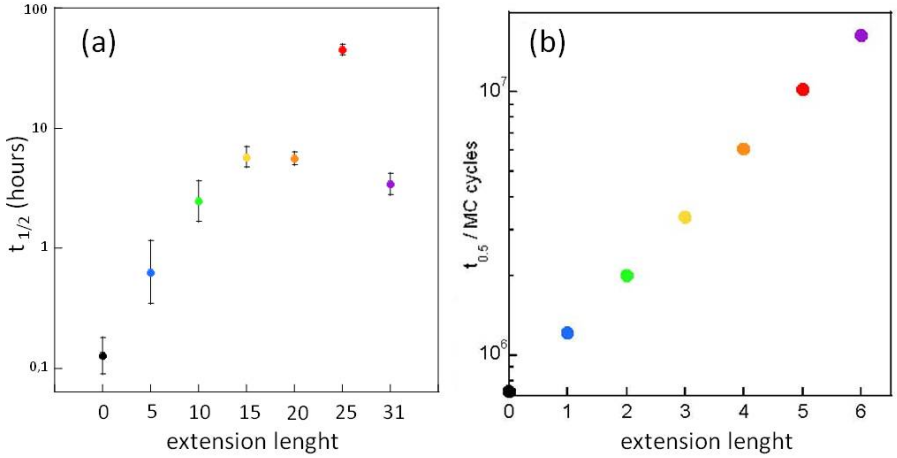

Figure 10. a) Half time of the Aβ-aggregation process at 10 μM as a function of extension length. The average and standard deviation are shown. Aβ(1-42) in black, Aβ(-5-42) in blue, Aβ(-10-42) in green, Aβ(-15-42) in yellow, Aβ(-20-42) in orange, Aβ(-25-42) in red and Aβ(-30-42) in violet. b) The number of simulation steps required until 50% of the fibrillar contacts are formed as a function of the extension length. Extension length of 1, 2, 3 ,4, 5 and 6 residues are shown in black, blue, green, yellow, orange, red and violet, respectively.

Part five: To fold or to fibrillate? Conclusions

- It is difficult to increase stability or affinity by introducing attractive electrostatic interactions. Introduction of repulsive interactions, on the other hand, may lead to decrease in both stability and affinity. The affinity of monellin variants (composed of two fragments) was found to correlate with the stability of single chain monellin, in agreement with previous studies. The association rates also correlated with the stability while there was no such correlation with the dissociation rate.

- An overall positive surface electrostatic potential is important in sweet taste, but overcharging this surface does not increase sweetness.

- When a charge, supposed to compensate an opposite charge, is introduced, intuitively one may think that this charge is attractive but since charges give rise to long-range electrostatic interactions, all favorable and unfavorable contributions of the new charge have to be taken into consideration. One has also to be aware of that a change may influence both states in equilibrium, often in different degrees. Compared with hydrophobic interactions the contribution of Coulombic interactions to protein stability and binding may be minor. Charge interactions may however modulate folding and binding by altering the rates of association and in discrimination of interacting partners.

- The correlation between fragment complementation and protein stability can be used in an *in vivo* approach to stabilize proteins based on the split GFP system. Split GFP system seems to be a successful method for protein stabilization trough fragment assembly.

- It is difficult to predict the impact of a mutation on all the different non-covalent interactions in both the native and the unfolded states. Randomly generated libraries make it possible to evaluate mutations beyond what can be guessed to be important. A combination of several methods like rational design of mutations, combinatorial libraries of variants and computational approaches

may increase the success rate of obtaining a desired effect.

- A correlation between stability and aggregation rate was found for monellin and a series of variants of it. An increase in the aggregation propensity of a protein upon mutations can be explained by the increase in the exposure of the amyloidogenic regions caused by the destabilization of the native state.

- The aggregation rate of an amyloid forming protein may be reduced by terminal extensions outside the aggregation prone sequence. We suggest that this result arises from a general physico-chemical effect which we call internal dilution.

- Protection against aggregation is achieved by stabilizing the structure of the protein or by extending an aggregation-prone sequence with non-aggregating segments.

Part six: References

I. **André**, T. Kesvatera, B. Jonsson, K.S. Akerfeldt and S. Linse, The role of electrostatic interactions in calmodulin–peptide complex formation. Biophys. J. 87 (**2004**), 1929–38.

C.B **Anfinsen**, Principles that govern the folding of protein chains. Science. 181 (**1973**), 223-30.

P.W **Atkins**, Physical chemistry. 6th ed. Oxford University Press **(1998)**.

E.M. **Baden**, E.G. Randles, A.K. Aboagye, J.R. Thompson and M. Ramirez-Alvarado, Structural insights into the role of mutations in amyloidogenesis, J. Biol. Chem. 283 (**2008**), 30950–56.

R. **Bader**, M.A. Seeliger, S.E. Kelly, L.L. Ilag, F. Meersman, A. Limones, B.F. Luisi, C.M. Dobson and L.S. Itzhaki, Folding and fibril formation of the cell cycle protein Cks1. J. Biol. Chem. 281 (**2006**), 18816-24.

A.J. **Baldwin**, T.P. Knowles, G. Tartaglia, A. Fitzpatrick, G. Devlin, S. Shammas, C.A. Waudby, M.F. Mossuto, S.L. Gras, J. Christodoulou, S.J. Anthony-Cahill, P.D. Barker, M. Vendruscolo and C.M. Dobson, Metastability of Native Proteins and the Phenomenon of Amyloid Formation. J. Am. Chem. Soc. 133 (**2011**), 14160-63.

T. **Berggård**, K. Julenius, A. Ogard, T. Drakenberg and S. Linse, Fragment complementation studies of protein stabilization by hydrophobic core residues. Biochemistry 40 (**2001**), 1257–64.

Z. **Bohak** and S.L. Li, The structure of monellin and its relation to the sweetness of the protein. Biochim. Biophys. Acta 427 (**1976**), 153–70.

R.H. **Cagan** and R.W. Morris, Biochemical studies of taste sensation: binding to taste tissue of 3H-labeled monellin, a sweet-tasting protein. Proc. Natl. Acad. Sci. U.S.A. 76 (**1979**), 1692–96.

J.E. **Caldwell**, F. Abildgaard, Z. Dzakula, D. Ming, G. Hellekant and J.L. Markley, Solution structure of the thermostable sweet-tasting protein brazzein. Nat. Struct. Biol. 5 (**1998**), 427–31.

J. **Carey**, S. Lindman, M. Bauer and S. Linse, Protein reconstitution and three-dimensional domain swapping: Benefits and constraints of covalency. Protein Sci. 16 (**2007**), 2317–33.

F. **Chiti** and C.M. Dobson, Review on folding, misfolding and agregation. First atemts to correlate aggregation and phenomenological equation. Annu. Rev. Biochem. 75 (**2006**), 333–66.

F. **Chiti**, N. Taddei, M. Bucciantini, P. White, G. Ramponi and C.M. Dobson, Mutational analysis of the propensity for amyloid formation by a globular protein. EMBO J. 19 (**2000**), 1441–49.

F. **Chiti**, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi and C.M. Dobson, Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. Proc. Natl. Acad. Sci. U.S.A. 96 (**1999**), 3590-94.

D.H.A. **Corrêa** and C.H. Ramos, Amyloid fibril formation by circularly permuted and C-terminally deleted mutants. Int. J. Biol. Macromol. 48 (**2011**), 583-88.

T.D. **Craggs**, Green fluorescent protein: structure, folding and chromophore maturation. Chem. Soc. Rev. 38 (**2009**), 2865-75.

T.E. **Creighton**, Proteins: structures and molecular properties. 2nd ed. W.H. Freeman and Company, New York. (**1993**).

C. **D'Arrigo**, M. Tabaton and A. Perico, N-terminal truncated pyroglutamyl b amyloid peptide Abpy3-42 shows faster aggregation kinetics than the full-length Ab1-42. Biopolymers 91 (**2009**), 861-73.

R. **Day**, D.A. Beck, R.S. Armen and V. Daggett, A consensus view of fold space: combining SCOP, CATH, and the Dali Domain Dictionary. Protein Sci. 12 (**2003**), 2150-60.

A.M. **de Vos**, M. Hatada, H. van der Wel, H. Krabbendam, A.F. Peerdeman and S.H. Kim, Three-dimensional structure of thaumatin I, an intensely sweet protein. Proc. Natl. Acad. Sci. U.S.A. 82 (**1985**), 1406–09.

D. **Dell'Orco**, W.F. Xue, E. Thulin and S. Linse, Electrostatic contributions to the kinetics and thermodynamics of protein assembly. Biophys. J. 88 (**2005**), 1991–2002.

K.A. **Dill**, Dominant forces in protein folding. Biochemistry. 29 (**1990**), 7133-55.

K.A. **Dill** and H.S .Chan, From Levinthal to pathways to funnels. Nat.Struct .Biol. 4 (**1997**),10-19.

C.M. **Dobson**, Structural basis of protein folding and its link with human disease. Philos. Trans. R. Soc. Lond. Biol. Sci. 356 (**2001**), 133–45.

R.B. **Dyer**, Ultrafast and downhill protein folding. Curr. Opin. Struct. Biol. 17 (**2007**), 38-47.

V.G.H**. Eijsink**, A. Bjørk, S. Gåseidnes, R. Sirevåg, B. Synstad, B. van den Burg and G. Vriend, Rational engineering of enzyme stability. J. Biotechnol. 113 (**2004**), 105–120.

V.G.H. **Eijsink**, S. Gåseidnes, T.V. Borchert and B. van den Burg, Directed evolution of enzyme stability. Biomol. Eng. 22 (**2005**), 21–30.

V. **Esposito**, R. Gallucci, D. Picone, G. Saviano, T. Tancredi and P.A. Temussi, The importance of electrostatic potential in the interaction of sweet proteins with the sweet taste receptor. J. Mol. Biol. 360 (**2006**), 448–56.

B.E. **Finn**, J. Kordel, E. Thulin, P. Sellers and S. Forsen, Dissection of calbindin D9k into two Ca(2+)-binding subdomains by a combination of mutagenesis and chemical cleavage. FEBS Letters 298 (**1992**), 211–14.

I. **Ghosh**, A.D Hamilton and L. Regan, Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. J. Am. Chem. Soc. 122 (**2000**), 5658-59.

S.C. **Gill** and P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182 (**1989**), 319-26.

C.R. **Goward**, M.D. Scawen, J.P. Murphy and T. Atkinson. Molecular evolution of bacterial cell-surface proteins. Trends Biochem. Sci.18 (**1993**), 136-40.

A.M. **Gronenborn**, D.R Filpula, N.Z Essig, A. Achari, M. Whilton, P.T. Wingfield and G.M Clore, A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. Science. 253 (**1991**), 657–61.

E. **Hellstrand**, B. Boland, D.M. Walsh and S. Linse, Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. ACS Chem. Neurosci, 1 (**2010**), 13–18.

RH **Hoess**, Protein design and phage display. Chem Rev. 101 (**2001**), 3205-18.

W. **Hoyer**, D. Cherny, V. Subramaniam and T.M. Jovin, Impact of the acidic C-terminal region comprising amino acids 109-140 on a-synuclein aggregation in vitro. Biochemistry 43 (**2004**) 16233-42.

J.N. **Israelachvili**, Intermolecular and surface forces. 2nd ed. Academic Press, London (**1992**).

S.M. **Kelly**, T.J. Jess and N.C. Price, How to study proteins by circular dichroism. Biochim. Biophys. Acta. 1751 (**2005**), 119-39.

J.C. **Kessler**, J.C. Rochet and P.T. Lansbury, The N-terminal repeat domain of a-synuclein inhibits β-sheet and amyloid fibril formation. Biochemistry 42 (**2003**), 672-78.

R. **Khurana**, C. Coleman, C. Ionescu-Zanetti, S.A. Carter, V. Krishna, R.K. Grover, R. Roy and S. Singh, Mechanism of thioflavin T binding to amyloid fibrils. J Struct.

Biol. 151(**2005**), 229-38.

S.H. **Kim**, C.H. Kang, R.Kim, J.M. Cho, Y.B. Lee and T.K. Lee, Redesigning a sweet protein: increased stability and renaturability. Protein Eng. 2 (**1989**), 571–75.

N. **Kobayashi**, S. Honda, H. Yoshii, H. Uedaira and E. Munekata, Complement assembly of two fragments of the streptococcal protein G B1 domain in aqueous solution. FEBS Lett. 366 (**1995**), 99–103.

M. **Kohmura**, N. Nio and Y. Ariyoshi, Complete amino acid sequence of the sweet protein monellin. Agric. Biol. Chem. 54 (**1990**), 2219–24.

M. **Kohmura**, N. Nio and Y. Ariyoshi, Solid-phase synthesis of [AsnA16]-, [AsnA22]-, [GlnA25]-, and [AsnA26]monellin, analogues of the sweet protein monellin. Biosci. Biotechnol. Biochem. 56 (**1992**), 472-76.

T. **Konno**, K. Murata and K. Nagayama, Amyloid-like aggregates of a plant protein: a case of a sweet-tasting protein, monellin. FEBS Lett. 454 **(1999)**, 122–26.

T. **Konno**, Multistep nucleus formation and a separate subunit contribution of the amyloidgenesis of heat-denatured monellin. Protein Sci. 10 **(2001)**, 2093–2101.

R. **Koradi**, M. Billeter and K. Wuthrich, MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. 14 (**1996**), 51–55, 29–32.

M.P. **Lambert**, A.K. Barlow, B.A. Chromy, C. Edwards, R. Freed, M. Liosatos, T.E. Morgan, I. Rozovsky, B. Trommer, K.L. Viola, P. Wals, C. Zhang, C.E. Finch, G.A. Krafft and W.L. Klein, Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. Proc. Natl. Acad. Sci. U.S.A. 95 (**1998**), 6448-53.

A.J. **Larner**, Hypothesis: amyloid beta-peptides truncated at the N-terminus contribute to the pathogenesis of Alzheimer's disease. Neurobiol. Aging. 20 (**1999**), 65-69.

X. **Li**, L. Staszewski, H. Xu, K. Durick, M. Zoller and E. Adler, Human receptors for sweet and umami taste. Proc. Natl. Acad. Sci. U.S.A. 99 (**2002**), 4692–96.

M.Q. **Liao**, Y.J. Tzeng, L.Y. Chang, H.B. Huang, T.H. Lin, C.L. Chyan and Y.C. Chen, The correlation between neurotoxicity, aggregative ability and secondary structure studied by sequence truncated Abeta peptides. FEBS Lett. (**2007**), 1161-65.

M. **Lindgren** and P. Hammarström, Amyloid oligomers: spectroscopic character-rization of amyloidogenic protein states. FEBS J. 277 (**2010**), 1380-88.

S. **Lindman**, I. Johansson, E. Thulin and S. Linse, Green fluorescence induced by

EF-hand assembly in a split GFP system. Protein Sci. 18 (**2009**), 1221–29.

S. **Lindman**, W.F. Xue, O. Szczepankiewicz, M.C. Bauer, H. Nilsson and S. Linse, Salting the charged surface: pH and salt dependence of protein G B1 stability. Biophys. J. 90 (**2006**), 2911-21.

S. **Linse** and B. Linse, Protein folding through kinetic discrimination. J. Am. Chem. Soc. 129 (**2007**), 8481-86.

S. **Linse**, E. Thulin, L.K. Gifford, D. Radzewsky, J. Hagan, R.R. Wilk and K.S Akerfelt, Domain organization of calbindin D(28k) as determined from the association of six synthetic EF-hand fragments. Protein Sci. 6 (**1997**), 2385–96.

S. **Linse**, M. Voorhies, E. Norstrom and D.A. Schultz, An EF-hand phage display study of calmodulin subdomain pairing. J. Mol. Biol. 296 (**2000**), 473–86.

X. **Liu**, S. Maeda, Z. Hu, T. Aiuchi, K. Nakaya and Y. Kurihara, Purification, complete amino acid sequence and structural characterization of the heat-stable sweet protein, mabinlin II. Eur. J. Biochem. 211 (**1993**), 281–87.

T.J. **Magliery**, C.G. Wilson, W. Pan, D. Mishler, I. Ghosh, A.D. Hamilton and L. Regan, Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. J. Am. Chem. Soc. 127 (**2005**), 146-57.

G.I. **Makhatadze**, V.V. Loladze, A.V. Gribenko and M.M. Lopez, Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions. J. Mol. Biol. 336 (**2004**), 929–42.

S.M. **Malakauskas** and S.L. Mayo, Design, structure and stability of a hyperthermophilic protein variant. Nat. Struct. Mol. Biol. 5 (**1998**), 470–75.

A. **Martin**, V. Sieber and F.X. Schmid, *In-vitro* selection of highly stabilized protein variants with optimized surface. J. Mol. Biol. 309 (**2001**), 717-26.

L. **Masino**, G. Nicastro, L. Calder, M. Vendruscolo and A. Pastore, Functional interactions as a survival strategy against abnormal aggregation. FASEB J. 25 (**2011**), 45-54.

T. **Masuda**, N. Ide and N. Kitabatake, Structure–sweetness relationship in egg white lysozyme: role of lysine and arginine residues on the elicitation of lysozyme sweetness. Chem. Senses 30 (**2005**), 667–81.

D. **Ming** and G. Hellekant, Brazzein, a new high-potency thermostable sweet protein from *Pentadiplandra brazzeana B*. FEBS. Lett. 355 (**1994**), 106–08.

G. **Morini**, A. Bassoli and P.A. Temussi, From small sweeteners to sweet proteins: anatomy of the binding sites of the human T1R2_T1R3 receptor. J. Med. Chem. 48

(**2005**), 5520–29.

J.A. **Morris** and R.H. Cagan, Purification of monellin, the sweet principle of *Dioscoreophyllum cumminsii*. Biochim. Biophys. Acta 261 (**1972**), 114–22.

J.A. **Morris** and R.H. Cagan, Effects of denaturants on the sweet-tasting protein monellin. Proc. Soc. Exp. Biol. Med. 150 (**1975**), pp. 265-70.
J.A. **Morris** and R.H. Cagan, Formation of oligomeric monellin in protein denaturants. Proc Soc Exp Biol Med. 64 (**1980**), 351-54.

J.K. **Myers**, C.N. Pace, and J.M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci. 4 (**1995**), 2138–48.

R. **Nelson**, M.R. Sawaya, M. Balbirnie, A.Ø. Madsen, C. Riekel, R. Grothe and D. Eisenberg, Structure of the cross-β spine of amyliod-like fibrils. Nature 435 (**2005**), 773–78.

L. **Nielsen,** R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V.N. Uversky and A.L. Fink, Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. Biochemistry. 40 (**2001**), 6036-46.

C. **Ogata**, M. Hatada, G. Tomlinson, W.C. Shin and S.H. Kim, Crystal structure of the intensely sweet protein monellin. Nature 328 (**1987**), 739–42.

M. **Ormö**, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien and S.J. Remington, Crystal structure of the Aequorea victoria green fluorescent protein. Science. 273 (**1996**), 1392-95.

D.E. **Otzen**, O. Kristensen and M. Oliveberg, Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. Proc. Natl. Acad. Sci. U.S.A. 97 (**2000**), 9907-12.

D. **Otzen** and P.H. Nielsen, We find them here, we find them there: functional bacterial amyloid. Cell Mol. Life Sci. 65 (**2008**) 910-27.

C.N. **Pace**, The stability of globular proteins. CRC Crit. Rev. Biochem. 3 (**1975**), 1-43.

C.N. **Pace**, S. Treviño, E. Prabhakaran and J.M. Scholtz, Protein structure, stability and solubility in water and other solvents. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359 (**2004**), 1225-34.

G.J. **Palm**, A. Zdanov, G.A. Gaitanaris, R. Stauber, G.N. Pavlakis and A. Wlodawer, The structural basis for spectral variations in green fluorescent protein. Nat. Struct. Biol. 4 (**1997**), 361-65.

S. **Pechmann**, E.D. Levy, G.G. Tartaglia and M. Vendruscolo, Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc. Natl. Acad. Sci. U.S.A. 106 (**2009**), 10159–64.

A.T. **Petkova**, W.M. Yau and R. Tycko, Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. Biochemistry 45 (**2006**), 498-512.

C.J. **Pike**, M.J. Overman and C.W, Cotman. Amino-terminal deletions enhance aggregation of beta-amyloid peptides in vitro. J Biol Chem. 270 (**1995**), 23895-98.

A, **Rauk,** The chemistry of Alzheimer's disease. Chem. Soc. Rev. 38 (**2009**), 2698-15.

K.J. **Reissner** and D.W. Aswad, Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals? Cell Mol. Life Sci. 60 (**2003**), 1281-95.

D.E. **Robertson** and B.A. Steer, Recent progress in biocatalyst discovery and optimization. Curr. Opin. Chem. Biol. 8 (**2004**), 141–49.

K.E. **Routledge**, G.G. Tartaglia, G.W. Platt, M. Vendruscolo and S.E. Radford, Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. J. Mol. Biol. 389 (**2009**), 776-86.

J. **Ruiz-Sanz**, G.d. Prat-Gay, D.E. Otzen and A.R. Fersht, 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 (**1995**), 1695–1701.

E. **Sandelin**, A. Nordlund, P.M. Andersen, S.S. Marklund and M. Oliveberg, Amyotrophic lateral sclerosis-associated copper/zinc superoxide dismutase mutations preferentially reduce the repulsive charge of the proteins. J. Biol. Chem. 282 (**2007**), 21230-36.

T. **Sato**, P. Kienlen-Campard, M. Ahmed, W. Liu, H. Li, J.I. Elliott, S. Aimoto, S.N. Constantinescu, J.N. Octave and S.O. Smith, Inhibitors of amyloid toxicity based on beta-sheet packing of Abeta40 and Abeta42. Biochemistry. 45 (**2006**), 5503-16.

J.K. **Scott** and G.P. Smith, Searching for peptide ligands with an epitope library. Science 249 (**1990**), 386-90.

G.M. **Shankar** and D.M. Walsh, Alzheimer's disease: synaptic dysfunction and Abeta. Mol. Neurodegener. 4 (**2009**), 48.

A. **Shimizu-Ibuka**, Y. Morita, T. Terada, T. Asakura, K. Nakajima, S. Iwata, T. Misaka, H. Sorimachi, S. Arai and K. Abe, Crystal structure of neoculin: insights into its sweetness and taste-modifying activity. J. Mol. Biol. 359 (**2006**), 148–58.

Y. **Shirasuka**, K. Nakajima, T. Asakura, H. Yamashita, A. Yamamoto, S. Hata, S. Nagata, M. Abo, H. Sorimachi and K. Abe, Neoculin as a new taste-modifying protein occurring in the fruit of Curculigo latifolia. Biosci. Biotechnol. Biochem. 68 (**2004**), 1403–07.

C.K. **Smith**, J.M. Withka and L. Regan, A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. Biochemistry 33 (**1994**), 5510-17.

G.P. **Smith**, D.A. Schultz and J.E. Ladbury, A ribonuclease S-peptide antagonist discovered with a bacteriophage display library. Gene 128 (**1993**), 37-42.

J.R. **Somoza**, F. Jiang, L. Tong, C.H. Kang, J.M. Cho and S.H. Kim, Two crystal structures of a potently sweet protein. Natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution. J. Mol. Biol. 234 (**1993**), 390–404.

N.T. **Southall**, K.A. Dill and A.D. J. Haymet, A view of the hydrophobic Effect. J. Phys. Chem. 106 (**2002**), 521-33.

R. **Spadaccini**, O. Crescenzi, T. Tancredi, N. de Casamassimi, G. Saviano and R. Scognamiglio, A. Di Donato and P.A. Temussi, Solution structure of a sweet protein: NMR study of MNEI, a single chain monellin. J. Mol. Biol. 305 (**2001**), 505–14.

Y.H. **Sung**, J. Shin, H.J. Chang, J.M. Cho and W. Lee, Solution structure, backbone dynamics, and stability of a double mutant single-chain monellin. structural origin of sweetness. J. Biol. Chem. 276 (**2001**), 19624–30.

G.G. **Tartaglia**, A.P. Pawar, S. Campioni, C.M. Dobson, F. Chiti and M. Vendruscolo, Prediction of aggregation-prone regions in structured proteins. J. Mol. Biol. 380 (**2008**), 425–36.

P.A. **Temussi**, Why are sweet proteins sweet? Interaction of brazzein, monellin and thaumatin with the T1R2–T1R3 receptor. FEBS Lett. 526 (**2002**), 1–4.

R.Y. **Tsien**, The green fluorescent protein. Annu. Rev. Biochem. 67 (**1998**), 509-44.

R**. Tycko**, Insights into the amyloid folding problem from solid-state NMR. Biochemistry 42 (**2003**), 3151-59.

V. **Tõugu** and T. Kesvatera, Role of ionic interactions in cholinesterase catalysis. Biochim. Biophys. Acta. 1298 (**1996**) 12-30.

B. **Van den Burg**, G. Vriend, O.R. Veltman, G. Venema and V.G. Eijsink, Engineering an enzyme to resist boiling. Proc. Natl. Acad. Sci. U.S.A. 95 (**1998**), 2056-60.

H. **van der Wel** and K. Loeve, Isolation and characterization of thaumatin I and II,

the sweet-tasting proteins from *Thaumatococcus daniellii Benth*. Eur. J. Biochem. 31 (**1972**), 221–25.

D.M. **Walsh**, D.M. Hartley, Y. Kusumoto, Y. Fezoui, M.M. Condron, A. Lomakin, G.B. Benedek, D.J. Selkoe and D.B. Teplow, Amyloid beta-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. J. Biol. Chem. 274 (**1999**), 25945-52.

D.M. **Walsh**, I. Klyubin, J.V. Fadeeva, W.K. Cullen, R. Anwyl, M.S. Wolfe, M.J. Rowan and D.J. Selkoe, Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation *in vivo*. Nature 416 (**2002**), 535-39.

D.M. **Walsh**, E. Thulin, A. Minuogue, T. Gustafsson, E. Pang, D.B. Teplow and S. Linse, A facile method for expression and purification of the Alzheimer disease-associated amyloid β-peptide. FEBS J. 276 (**2009**), 1266-81.

C.G. **Wilson**, T.J. Magliery and L. Regan, Detecting protein-protein interactions with GFP-fragment reassembly. Nat. Methods 1 (**2004**), 255–262.

B. **Winner**, R. Jappelli, S.K. Maji, P.A. Desplats, L. Boyer, S. Aigner, C. Hetzer, T. Loher, M. Vilar , S. Campioni, C. Tzitzilonis, A. Soragni, S. Jessberger, H. Mira, A. Consiglio, E. Pham, E. Masliah, F.H. Gage and R. Riek, *In vivo* demonstration that alpha-synuclein oligomers are toxic. Proc. Natl. Acad. Sci. U.S.A. 108 (**2011**), 4194-99.

M. **Wunderlich**, A. Martin and F.X. Schmid, Stabilization of the cold shock protein CspB from Bacillus subtilis by evolutionary optimization of Coulombic interactions. J. Mol. Biol. 347 (**2005**), 1063–76.

M. **Wunderlich** and F.X. Schmid, In vitro evolution of a hyperstable GB1 variant. J. Mol. Biol. 363 (**2006**), 545–57.

W.F. **Xue**, J. Carey and S. Linse, Multi-method global analysis of thermodynamics and kinetics in reconstitution of monellin. Proteins: Struct. Funct. Genet. 57 (**2004**), 586–95.

W.F. **Xue**, O. Szczepankiewicz, M.C. Bauer, E. Thulin and S. Linse, Intra- versus intermolecular interactions in monellin: contribution of surface charges to protein assembly. J. Mol. Biol. 358 (**2006**), 1244–55.

M. **Zimmer**, GFP: from jellyfish to the Nobel prize and beyond. Chem. Soc. Rev. 38 (**2009**), 2823-32.