Structure determination and thermodynamic stabilization of an engineered protein-protein complex

Elisabet Wahlberg

School of Biotechnology
Royal Institute of Technology

Stockholm 2006
Cover illustration: $^{15}$N-HSQC spectra of the Z$_{SPA-1}$ affibody, with or without different amount of Z domain added (0%, 50% and 100%). Spectra recorded 30 °C in 20 mM phosphate buffer pH 6.8.
The interaction between two 6 kDa proteins has been investigated. The studied complex of micromolar affinity (Kd) consists of the Z domain derived from staphylococcal protein A and the related protein ZSPA-1, belonging to a group of binding proteins denoted affibody molecules generated via combinatorial engineering of the Z domain. Affibody-target protein complexes are good model systems for structural and thermodynamic studies of protein-protein interactions. With the Z:ZSPA-1 pair as a starting point, we determined the solution structure of the complex and carried out a preliminary characterization of ZSPA-1. We found that the complex contains a rather large (ca. 1600 Å²) interaction interface with tight steric and polar/nonpolar complementarity. The structure of ZSPA-1 in the complex is well-ordered in a conformation that is very similar to that of the Z domain. However, the conformation of the free ZSPA-1 is best characterized by comparisons with protein molten globules. It shows a reduced secondary structure content, aggregation propensity, poor thermal stability, and binds the hydrophobic dye ANS. This molten globule state of ZSPA-1 is the native state in the absence of the Z domain, and the ordered state is only adopted following a stabilization that occurs upon binding. A more extensive characterization of ZSPA-1 suggested that the average topology of the Z domain is retained in the molten globule state but that it is represented by a multitude of conformations. Furthermore, the molten globule state is only marginally stable, and a significant fraction of ZSPA-1 exists in a completely unfolded state at room temperature. A complete thermodynamic characterization of the Z:ZSPA-1 pair suggests that the stabilization of the molten globule state to an ordered three helix structure in the complex is associated with a significant conformational entropy penalty that might influence the binding affinity negatively and result in an intermediate-affinity (µM) binding protein. This can be compared to a dissociation constant of 20-70 nM for the complex Z:Fc of IgG where Z uses the same binding surface as in Z:ZSPA-1. Structure analyses of Z in the free and bound state reveal an induced fit response upon complex formation with ZSPA-1 where a conformational change of several side chains in the binding surface increases the accessible surface area with almost 400 Å² i.e. almost half of the total interaction surface in the complex. Two cysteine residues were introduced at specific positions in ZSPA-1 for five mutants in order to stabilize the conformation of ZSPA-1 by disulfide bridge formation. The mutants were thermodynamically characterized and the binding affinity of one mutant showed an improvement by more than a factor of ten. The improvement of the introduced cysteine bridge correlates with an increase in binding enthalpy rather than with entropy. Further analysis of the binding entropy suggests that the conformational entropy change in fact is reduced but its favorable contribution is opposed by a less favorable desolvation enthalpy change. These studies illustrate the structural and thermodynamic complexity of protein-protein interactions, but also that this complexity can be dissected and understood. In this study, a comprehensive characterization of the ZSPA-1 affibody has gained insight into the intricate mechanisms involved in complex formation. These theories were supported by the design of a ZSPA-1 mutant with improved binding affinity.
Till Johan
Main References

This thesis is based on the following publications, referred to in the text by their Roman numerals.


*Joint first authors contributing equally to the work
Content

Introduction ........................................................................................................................................ 1
1 Proteins ........................................................................................................................................ 1
2 Protein Structure .......................................................................................................................... 3
3 Protein Stability ............................................................................................................................ 7
3.1 Non-covalent interactions .......................................................................................................... 9
  3.1.1 Electrostatic interactions ..................................................................................................... 9
  3.1.2 van der Waals interactions .................................................................................................. 10
  3.1.3 Hydrogen bonds ................................................................................................................ 11
  3.1.4 Hydrophobic interactions ................................................................................................... 11
3.2 Total interactions in the native state .......................................................................................... 13
3.3 The denatured state .................................................................................................................. 14
3.4 Molten globules ....................................................................................................................... 14
4 Protein-protein interactions ......................................................................................................... 16
  4.1 Conformational changes on complex formation .................................................................... 16
  4.2 Structural properties of protein interfaces ............................................................................. 17
  4.3 The size of the interface .......................................................................................................... 19
  4.4 Characteristics of hot spots ................................................................................................... 19
  4.5 Consensus binding site and adaptivity in hot spots ............................................................ 20
  4.6 Affinity and specificity .......................................................................................................... 21
5 Protein engineering ....................................................................................................................... 23
  5.1 Protein engineering methods ................................................................................................ 23
  5.2 Engineered binding proteins from alternative protein scaffolds ........................................ 26
6 Affibody molecules ....................................................................................................................... 28
  6.1 Protein A ................................................................................................................................ 28
  6.2 The Z domain ........................................................................................................................ 28
  6.3 Affibody molecules ................................................................................................................ 29
Present investigation .......................................................................................................................... 32
7 Structural analysis of an affibody-target pair (I, II) .................................................................... 33
  7.1 The Z:ZSPA-1 system ............................................................................................................. 33
  7.2 Structure analysis .................................................................................................................... 34
8 Biophysical characterization of the affibody ZSPA-1 (II, III) .................................................... 39
  8.1 Molten globule characteristics ............................................................................................. 39
  8.2 Structural and dynamic studies ............................................................................................. 43
  8.3 Characterisation of four ZSPA-1 mutants .............................................................................. 44
9 Thermodynamic stabilization of the Z:ZSPA-1 complex .............................................................. 46
  9.1 The Z:ZSPA-1 binding thermodynamics .............................................................................. 47
  9.2 Disulfide engineering (IV) ..................................................................................................... 48
  9.3 Biophysical properties of the cysteine-mutant 3:F5C/S39C ................................................ 49
  9.4 Results of all cysteine mutants .............................................................................................. 51
10 Conclusion and outlook .............................................................................................................. 55
11 Abbreviations ............................................................................................................................. 57
12 Acknowledgements .................................................................................................................... 58
13 References ................................................................................................................................... 59
Introduction

1 Proteins

Proteins are among the most abundant organic molecules in almost all living systems. In fact, apart from water, proteins constitute most of the cell mass. (Alberts et al. 2002). Looking at the processes going on in a cell, proteins execute many of these. Over time, evolution has optimized the characteristics of proteins, including their structures, their stabilities and abilities to interact, to tailor this class of molecules for the properties needed for biological function. From a chemical perspective, their structural complexity and functional diversity is overwhelming. Proteins are found to have many different functions, for example to provide structural rigidity as keratin in hair, to regulate metabolic processes such as insulin, to transport other molecules as seen for haemoglobin, to protect a host against intruding substances as the antibodies and to speed up chemical reactions as seen for enzymes.

Despite their functional diversity proteins have many properties in common. They are all built up of a long string of amino acids, typically between 50 and 2000, connected to each other through peptide bonds, hence the name polypeptide. The amino acid sequence contains all the information needed for a three-dimensional structure to form even though help from other molecules in the cell may be needed on the way. The folded protein is stabilized by many weak noncovalent interactions and covalent disulfide bridges can contribute as well. The native state is a critical factor for the biological function of a protein.

Through the human genome project we now know that a human cell can synthesize about 20 000 to 25 000 different proteins (Uhlen and Ponten 2005). For some of these proteins we know their function and what other molecules they interact with but for many we don’t know more than their potential amino acid sequence. There is a lot of effort spent to annotate the function of these unknown proteins. Because structures have been evolutionary more conserved than DNA and amino acid sequences, knowing the three-dimensional structure of a protein might help to find their function, when sequence homology studies fail. Many proteins need to interact with other molecules in order to function. It can be with an ion, a protein, DNA or a membrane for example. In the cell, large networks of proteins interacting with each other in complicated schemes are found.
Another aspect of studying protein structures and interactions is to gain a deeper understanding of the molecular mechanisms of protein stability and molecular recognition. This is important for the attempts to predict protein properties such as structure, effect of mutations on protein stability or why some proteins misfold under certain circumstances. Misfolded proteins are thought to be involved in about 20 diseases such as Alzheimer’s and Parkinson’s disease, type II diabetes, bovine spongiform encephalitis (BSE) for example (Horwich 2002; Dobson 2003; Selkoe 2003). Prediction algorithms will also aid in the improvement and design of new binders and inhibitors to disease related targets.
2 Protein Structure

The Danish biochemist Kai Linderstrøm-Lang introduced the terms primary, secondary and tertiary structure to stress that protein structures can be divided into different levels of organisation (Branden and Tooze 1999). One of the biggest challenges in bioscience has for a long time been the ability to predict the three-dimensional protein structure, directly from the amino acid sequence. This is commonly referred to as the folding problem. Although large advances have been made in recent years (Baker and Eaton 2004) structures of proteins still need to be determined by experimental methods using NMR or X-ray crystallography. Looking in the Brookhaven protein data bank (www.pdb.org) by June 2006 the atomic coordinates for about 35 000 protein structures are deposited.

Proteins are built of 20 different amino acids, each with different chemical properties. They all have the same fundamental structure, which consists of a central alpha-carbon (C\(_{\alpha}\)) attached to; a hydrogen (H\(_{\alpha}\)), an amino group and a carboxyl group. What gives the amino acids their unique properties is the side chain connected to C\(_{\alpha}\), which is different for each amino acid. Amino acids are usually divided into three families based on the chemical character of their side chain; they may be hydrophobic, polar or charged. An amino acid important for this thesis is cysteine. Cysteine contains a thiol-group (-SH) in its side chain that can oxidize to form a covalent disulfide bond with another cysteine that is near in the structure but might be far away in sequence. Disulfide bridges commonly act to stabilize the three-dimensional fold of proteins or link different polypeptide chains together. Because the interior of a cell has a reductive environment, disulfide bonds are usually found in secreted proteins.

Already in 1902, Emil Fischer and Franz Hofmeister independently established that proteins were made of chains of covalently linked amino acids (Dill 1990). In 1953, Frederick Sanger presented the first complete sequence of a protein, the hormone insulin, and it was not until then it was established that a given protein had its own unique amino acid sequence (Sanger and Tuppy 1951; Sanger and Thompson 1953). Sanger also showed that insulin was built of only L-amino acids connected by peptide bonds. The amino acid sequence is often referred to as primary structure.

The peptide bond is planar due to a considerable double-bond character and as a consequence rotation around the C\(^{-}\) - N bond is prevented (Branden and Tooze 1999; Berg et al. 2002).
Therefore the peptide bond is rigid and the bond distances and angles are known. This reduces the degree of freedom of the backbone of the polypeptide chain. Still rotation around two bonds on either side of the C\(\alpha\) atom, N–C\(\alpha\) and C\(\alpha\)–C’ called phi (\(\phi\)) and psi (\(\psi\)) respectively, can rotate freely. Consequently, if the \(\phi\) and \(\psi\) angles of all amino acids are determined with high accuracy then the conformation of the polypeptide backbone is determined. Another important feature of proteins is that each peptide bond has both a hydrogen donor (NH) and a hydrogen acceptor (CO) that can be engaged in hydrogen bond formation.

As the protein is synthesized, interactions begin to take place among various amino acids along the growing chain. There are two regular periodic structures formed by proteins, the \(\alpha\)-helix and the \(\beta\)-pleated sheet, predicted by Linus Pauling and Robert Corey in 1951 (Pauling and Corey 1951 (a-c); Pauling et al. 1951). They considered which conformations that were sterically allowed and could form backbone hydrogen bonds between different amino acids. Five years later the \(\alpha\)-helix was experimentally verified by John Kendrew’s structure of myoglobin (Kendrew et al. 1958).
Due to steric collisions between atoms in the backbone or side chain, most combinations of $\phi$ and $\psi$ angles are not allowed. The Indian biophysicist G.N. Ramachandran (Ramachandran and Sasisekharan 1968) first calculated which angles that were allowed in a polypeptide and plotted these against each other in a so-called Ramachandran plot.

When plotting the $\phi$ and $\psi$ angles of the polypeptide backbone from known structures into a Ramachandran diagram they will cluster into three different regions for all amino acids except glycine. Since the side chain of glycine only consists of a hydrogen it can adopt many more conformations than the other residues. Therefore glycine has a special role in proteins allowing more freedom in the backbone conformation.

In the $\alpha$-helix the polar carbonyl and amide groups of the polypeptide backbone form hydrogen bonds between amino acid $i$ and $i+4$, which is energetically favourable. The backbone C=O and N-H in the first and the last turn of a helix are missing an intra-chain hydrogen binding partner and instead usually form hydrogen bonds to side chains, called N-cap or C-cap residues, or to water (Presta and Rose 1988; Richardson and Richardson 1988; Serrano et al. 1992). In central positions of an $\alpha$-helix, straight-chain amino acids are in general favourable, because they do not suffer as much from conformational restriction in the helix, as bulky residues. Glycine and proline are least favourable but also residues with side chains that can form hydrogen bonds to
the main chain of the denatured state are destabilizing (Lyu et al. 1990; O’Neil and DeGrado 1990; Horovitz et al. 1992; Blaber et al. 1993; Myers et al. 1996). At the ends of a helix, glycine might be favourable though.

The β-structure is formed when two or more extended strands come together connected by hydrogen bonds. They can be arranged in parallel or anti parallel direction. The propensity of amino acids for β-sheet formation is more complex than for α-helices because they depend both on secondary and tertiary structure. However, experimental results suggest that β-branched amino acids such as threonine and valine usually are favoured in β-strands, because they are allowed a larger degree of freedom (Fersht 1999). Glycine is disfavoured due to a higher energetic cost of being restricted.

The organisation in space of the polypeptide chain or covalently connected chains is referred to as tertiary structure. The tertiary structure of a protein is stabilized by many weak non-covalent interactions between the secondary structure elements and occasional by disulfide bridges. In water solution, like inside a cell, nonpolar groups are in general placed in the interior and polar groups mainly on the surface (Branden and Tooze 1999). However in an apolar environment such as in a membrane, the amino acid distribution might be reversed. Tertiary structures can be divided into three groups based on their main secondary structure contents (Branden and Tooze 1999); the alpha-domain structures, beta structures and alpha/beta structures.

Some proteins are formed by association of two or more separately synthesised polypeptide chains forming a quaternary structure. The oxygen carrying protein hemoglobin for example is composed of four myoglobin-like polypeptide chains, called α and β chains. The formation of such an oligomeric complex often has a functional role, but it might also provide stabilization toward unfolding and degradation (Stites 1997).
3 Protein Stability

From a thermodynamic perspective, proteins are only marginally stable. A small change in temperature, buffer composition or pH can convert the folded protein sample into a denatured state. But from a biological perspective it is important that the stability is “just right”. The protein needs to be stable enough to maintain its native form and function at physiological conditions but it also needs to exhibit some flexibility depending on its functional mechanism. Further, it needs to be rapidly turned over in response to the cells changing needs. This is mainly accomplished by many weak and diverse noncovalent interactions that together stabilize the fold of the protein. Even though protein stability has been studied over several decades, all factors involved are still not completely understood (Weiss et al. 2001).

The stability of a protein is determined by the difference in Gibbs free energy between the native state and a denatured state. The conformational stability of a protein ($\Delta G^\circ_{\text{unfold}}$) usually falls in the range 5-10 kcal/mol (Dill 1990), not much more than the energy from a single hydrogen bond. The equilibrium constant ($K_{\text{unfold}}$) between the two states can be related to Gibbs energy as

$$\Delta G^\circ_{\text{unfold}} = -RT \ln(K_{\text{unfold}})$$

where $R$ is the gas constant and $T$ is the temperature given in Kelvin.

There are two important contributors to the Gibbs energy, the enthalpy and the entropy. The enthalpy contribution to Gibbs energy derives from the noncovalent interactions within the polypeptide chain relative to the solvent, such as hydrogen bonds, ion pairing, and van der Waals interactions. The folding of a protein decreases the conformational freedom of the polypeptide chain at the same time as it releases ordered water molecules. The overall contribution of the entropy to Gibbs energy is mainly the difference between these two processes. The relation of enthalpy and entropy to Gibbs energy of unfolding is given by the thermodynamic expression

$$\Delta G^\circ_{\text{unfold}} = \Delta H^\circ_{\text{unfold}} - T\Delta S^\circ_{\text{unfold}}$$
where \( \Delta H^\circ_{\text{unfold}} \) and \( \Delta S^\circ_{\text{unfold}} \) is the enthalpy and entropy differences between the native and the denatured state respectively. Both \( \Delta H^\circ_{\text{unfold}} \) and \( T\Delta S^\circ_{\text{unfold}} \) can reach several hundred kcal/mol but it turns out that they almost cancel each other. This makes the prediction of \( \Delta G^\circ_{\text{unfold}} \) very sensitive even to small errors.

The heat capacity change, \( \Delta C^\circ_p \), is defined as the first derivative of the function of enthalpy with respect to temperature and is a measure of the capacity of a substance to store energy. Both the enthalpy and entropy are variables of temperature (Brandts 1964; Lumry et al. 1966; Freire 2001) and their temperature dependence can be written as a function of \( \Delta C^\circ_p \). If \( \Delta C^\circ_p \) can be considered to be temperature independent in the temperature interval \( \Delta T \), i.e. \( \Delta C^\circ_p = (\Delta H^\circ/\Delta T) \) and \( \Delta C^\circ_p = T(\Delta S^\circ/\Delta T) \), Gibbs energy can be written as follows,

\[
\Delta G^\circ (T) = \Delta H^\circ (T_R) + \Delta C^\circ_p (T - T_R) - T [\Delta S^\circ (T_R) + \Delta C^\circ_p \ln(T/ T_R)]
\]

where \( T_R \) is some reference temperature. If \( T_R \) is equal to \( T_m \) (\( T_m \) is the melting temperature), by definition \( \Delta H^\circ (T_m) = T_m \Delta S^\circ (T_m) \) and consequently \( \Delta G^\circ (T_m) = 0 \), which gives

\[
\Delta G^\circ (T) = \Delta H^\circ (T_m) \left(1 - T/ T_m\right) + \Delta C^\circ_p \left[(T - T_m) - T \ln(T/ T_m)\right]
\]

The temperature stability of a protein is completely defined if the three thermodynamic parameters, \( \Delta H^\circ \), \( \Delta S^\circ \) (at some reference temperature) and \( \Delta C^\circ_p(T) \) are known.

The temperature dependence of \( \Delta H^\circ \) and \( \Delta S^\circ \) converts Gibbs energy from a linear to a curved function. This means that the function will cross zero twice, one at high temperature (heat denaturation) and one at low temperature (cold denaturation) (Becktel and Schellman 1987; Franks 1995). For most proteins however the temperature for cold denaturation is far below the freezing point for water so it will not be detected (Fersht 1999).
A consequence of the non-linear function of $\Delta G^\circ$ with respect to temperature is that a protein with a lower melting temperature ($T_m$) can be more stable at room temperature than a protein with higher $T_m$. Because of that it is not possible to predict a protein's temperature stability from the stability measured in room temperature with denaturants.

3.1 Non-covalent interactions

A balance of opposing forces determines the stability of the native state. The forces, important in the folded state includes electrostatic and van der Waals interactions, hydrogen bonds and the hydrophobic effect (Freire 2001) while the conformational entropy favours the unfolded state. (Razvi and Scholtz 2006)

3.1.1 Electrostatic interactions

The strength of electrostatic interactions between charges is inversely proportional to the distance between the charges and the dielectric constant of the medium. In some occasions electrostatic interactions can be very strong even challenging the strength of covalent bonds. However, the effect of the dielectric constant is large in a polar medium such as water ($\varepsilon = 78 \text{ J}^{-1}\text{C}^2\text{m}^{-1}$) (Atkins and de Paula 2006) and the potential energy decreases by almost two orders of
Protein Stability

magnitude compared to vacuum and gives a typically value a few kcal/mol (Dill 1990). This, together with the relatively low frequency of ion-pairs in proteins, suggests that their contribution to protein stability is probably small (Barlow and Thornton 1983; Sali et al. 1991). However, due to the fine balance of stabilizing forces in the native state, even small contributions are important. Charged amino acids are most commonly located at the surface of the protein, exposed to solvent (Barlow and Thornton 1983). It is energetically unfavourable to bury charges in the core of a protein but if a salt-bridge is buried, it is usually important for functional reasons and it will be stronger due to the often low dielectric environment in the interior of a protein. A problem associated with calculations of electrostatic interactions is to estimate the dielectric constant within the protein. This value can vary greatly because the interior may be composed of both biopolymer and water. A value of $\varepsilon = 2 - 4$ is often used (Fersht 1999).

3.1.2 van der Waals interactions
Interactions involving polar and nonpolar groups can occur through van der Waals interactions. They originate from permanent or induced dipole moments and have a distance dependence of the inverse power of six. Polar groups attain a permanent electric dipole moment due to partial charges on their atoms and they will try to orientate their dipole moments relative each other to minimize their potential energy. A nonpolar group can acquire a temporary dipole moment in response to a nearby electric field, resulting from an ion or a dipole that cause a distortion of the electron distribution of the group. The only interaction between nonpolar groups is through dispersion or London interactions; commonly they are grouped together with van der Waals forces. Here a transient dipole arises from fluctuations of the position of electrons, which results in a partially positive and negative charge. While the dipole exists it can polarize another group generating a second dipole. Even as the first dipole will go on continuously changing the direction and magnitude, perhaps within $10^{-16}$ seconds, the other dipole will follow it (Atkins and de Paula 2006). The magnitude of the interaction depends on how polarizable the two participating groups are (Fersht 1999). If the group contains large atoms with electrons at some distance from the nucleus the electrons will be less controlled by the nucleus and more polarizable. The distance between the atoms are also important, the closer the atoms, the stronger the attractive force, until the atoms are actually in contact (at the van der Waals radius) at which point the forces become repulsive and very strong. The strength of van der Waals interactions is small, approximately 0.1-0.2 kcal/mol each (Fersht 1999; Atkins and de Paula 2006) but they are many and together they will make a significant contribution to the overall stability of the molecule (Fersht 1999).
3.1.3 Hydrogen bonds
The hydrogen bond is a particularly important force in biological systems. The classical view of hydrogen bonds is that they are formed between a hydrogen covalently attached to an electronegative atom and another electronegative atom. In proteins the most common electronegative atoms are oxygen and nitrogen. This means that each peptide bond has both a hydrogen donor and a hydrogen acceptor. These donor/acceptor groups are mainly used to form hydrogen bonds that stabilize the secondary structure building blocks. Except from ion-ion interactions that can be very strong the hydrogen bond is usually the strongest noncovalent interaction within the protein. The strength depends on the electronegativity, distance between and orientation of the atoms involved and is usually in the range 2-10 kcal/mol (Dill 1990). The optimal orientation between the three atoms involved is linear and the typical distance between an amide proton (=NH) and a carbonyl oxygen (-CO) is 2.85 to 3.00 Å and for O···H a typical distance is 1.85 to 2.00 Å. Apart from the classical view, it is now known that hydrogen bonds are influenced to different extent by coulombic, covalent and van der Waals interactions. The relative contribution of the different forces depends on the donor-acceptor pair involved and their contact geometry. One example of very weak hydrogen bonds, with a large van der Waals influence, is observed when the hydrogen bond donor/acceptor is weak, as when carbon represents the electronegative atom in the C-H···O bond. One example of these more exotic hydrogen bonds have been described by Taylor and Kennard, who discovered this phenomenon when analysing a number of protein structures in 1982 (Taylor and Kennard 1982 JACS. Involvement of different C-H donor groups and aromatic pi-system acceptors of amino acid side chains have also been described (Brandl 2001).

3.1.4 Hydrophobic interactions
The hydrophobic interactions play an important role in many biological processes such as membrane and micelle formation, protein interactions, in addition to protein stability. Together with the determination of the first crystallographic structure of a protein in 1958 (Kendrew et al. 1958) came proof that the architecture of proteins consists of a core with predominantly buried hydrophobic amino acids. This showed that the structure of proteins to some extent is driven by the avoidance of hydrophobic groups for water. A question that followed was if the avoidance of oil (i.e. hydrophobic for proteins) for water possibly were the result of an affinity of oil for oil or an affinity of water for water. The opinion of today is that the avoidance of oil for water mainly is caused by the water-water hydrogen bonding, because the water interactions are stronger than the dispersion interactions attracting oil to oil (Southall et al. 2002). The effect of hydrophobic
interactions is best described by the exclusion of nonpolar atoms from contact with water, as seen when mixing oil and water.

The hydrophobic interaction can be divided into two components; one component results from dispersion interactions as discussed above. The other component called the hydrophobic effect is entropic at room temperature and accompanied with a large change in heat capacity. Several models have been proposed to explain the hydrophobic effect and they are still debated. One reason is that different opinions exist of how to treat the ordering of water molecules. According to the so called “iceberg model” presented by Frank and Evans in 1945 (Frank and Evans 1945;), the hydrophobic effect is a result of Natures strive to preserve the number and quality of hydrogen bonds, worth about 6 kcal/mol each (Fersht 1999), such that water molecules lines up around an exposed nonpolar body. However, this arrangement decreases the number of possibilities each lined water molecule have to form hydrogen bonds and results in a decrease in entropy of the solvent. Just as when mixing oil with water, the unfolding of a protein is accompanied by a large increase in heat capacity. This is mainly thought to be the result of breaking the well-ordered hydrogen bonds around the protein as the thermal energy increases, similar to melting ice (Dill 1990). The large heat capacity change is the reason for the strong temperature dependence of the entropy and enthalpy.

Figure 3.2 A schematic drawing illustrating the dependence of $\Delta G^{\text{unfold}}$, $\Delta H^{\text{unfold}}$ and $\Delta S^{\text{unfold}}$ on temperature, adapted from (Southall et al. 2002). $T_h$ and $T_s$ are the temperatures where the unfolding enthalpy and entropy, respectively, are zero.
3.2 Total interactions in the native state

The native state is represented by the conformation having the lowest free energy. The general opinion of today is that the hydrophobic interactions are the dominant force for protein folding and stability and responsible for the compact conformation of the nonpolar core. This opinion is supported by the unusual temperature dependence of protein stability. Proteins show both heat and cold denaturation, similar to dissolving a nonpolar solute in water (Privalov and Gill 1988). Further, nonpolar solvents cause proteins to unfold (Dill 1990). Additional evidence supporting the idea of hydrophobic interactions as the dominant force comes from the large increase in heat capacity upon unfolding (Privalov et al. 1979; Pace et al. 1988; Privalov and Gill 1988). It is also known from structural analyses that mainly nonpolar amino acids form the core of proteins. Hydrogen bonds were for a long time believed to be the driving force for protein folding and stability, as first suggested by Mirsky and Pauling in 1936 (Mirsky and Pauling 1936) but as for reasons discussed above this view has changed. However, when the polar groups are buried upon hydrophobic collapse, the formation of hydrogen bonds becomes highly favourable and is responsible for the inner architecture of the protein such as formation of $\alpha$-helices and $\beta$-sheet. This is consistent with the fact that solvents, capable of forming strong hydrogen bonds, destabilize $\alpha$-helices and $\beta$-sheets (Dill 1990).

A denatured state forms many interactions to water molecules, which bind directly to the backbone, the polar side chains and also cluster around hydrophobic side chains. Consequently, in a denatured state, the polypeptide chain experiences a large flexibility but water molecules are restrained. When the protein folds it exchanges the interactions to water for interactions with itself and the water molecules are released. At the same time, the polypeptide chain becomes restrained. As a consequence of this delicate balance, no type of interaction is unimportant, and even weak interactions may contribute significantly to stability, both in a negative and positive way (Dill 1990).
3.3 The denatured state

The denatured state actually consists of an ensemble of states with similar energy. It has a high conformational freedom but might not be completely unfolded; instead having some residual structure left (Freund et al. 1996).

To study protein stability, proteins often need to be studied in their denatured state. This state can be achieved by changing the temperature or pH of the sample, by addition of a chemical denaturant such as urea or guanidinium hydrochloride (GuHCl) or by increasing the pressure (Fersht 1999). In general, small proteins denature reversibly, however they have a higher tendency to aggregate and precipitate in the denatured state when the denaturant is removed. Temperature denaturation of proteins occurs as a consequence of an increase in entropy ($T\Delta S_{\text{unfold}}$) with temperature, and when it exceeds $\Delta H_{\text{unfold}}$ the protein will unfold. During cold denaturation the enthalpy term decreases, and again the protein will unfold when entropy becomes the dominating term (Fersht 1999). Urea and GuHCl act by solubilising both side chains and polypeptide backbone of the protein (Schellman 1994; Zou et al. 1998). Because the denatured state is more exposed than the native state the denatured state will be more stabilized than the native state by such denaturant (Fersht 1999).

3.4 Molten globules

The molten globule state was first described as a state that could be induced for certain proteins by different conditions, including low pH, addition of chemical denaturants or removal of cofactor. It has been frequently studied as a representative for intermediates in the folding process. A molten globule state is defined as a partly organized globular state, which has formed most of its secondary structure content, but is less compact than the native state, and lacking the proper interactions within the core. The tendency to expose hydrophobic regions can be visualized by binding to the fluorescent dye ANS, which has an affinity for hydrophobic regions. The molten globule state should not be seen as a single state; instead it is represented by an ensemble of rapidly interconverting conformations (Semisotnov et al. 1991; Dobson 1994; Ptitsyn 1995; Schulman et al. 1997; Regan 2003).
Figure 3.3 Chemical structure of a) guanidine hydrochloride (GuHCL), b) urea, c) trimethyl-N-amineoxide (TMAO), and d) 8-anilino-1-naphtalenesulfonic acid (ANS). Figure kindly provided by C. Lendel.
4 Protein-protein interactions

If observing a living cell, there is an enormous amount of activities going on at any given moment. To do the right thing at the right time, molecules need a way to communicate and between proteins this communication is primarily mediated by protein-protein interactions. Because of their importance for biological processes a detailed understanding of these interactions is a primary goal for protein science of today.

Not surprisingly many human diseases originate from abnormal protein-protein interactions. Malfunctioning endogenous proteins can result in either loss of an essential interaction or formation of a complex at an inappropriate time or location. This is valid for some of our most feared diseases of today such as cancer and the group of protein-misfolding diseases (Dobson 2002; Ryan and Matthews 2005). Obviously, a deep understanding of the molecular mechanisms of protein-protein interactions will be very important for the development of new binding molecules capable of binding, inhibiting or activating certain proteins or protein-protein complexes.

4.1 Conformational changes on complex formation

There are various levels of conformational changes that occur upon complex formation and different models are used to describe them. Some proteins form complexes with only small side chain rotations and no change in backbone conformation, commonly referred to as rigid body or lock and key interactions, originally introduced by Emil Fischer in 1894 (Goh et al. 2004). Other proteins experience changes also in their backbone conformation, often also including hinge, loop and/or domain movements. The mechanism can be illustrated by a ligand which induces a conformational change of the protein interface to fit the ligand as a “hand in a glove”. The mechanism is especially relevant for enzyme complexes which often undergo domain shifts when binding substrates (Gerstein et al. 1993; Jones and Thornton 1996). This model is often referred to as induced fit, introduced by Daniel Koshland in 1958 (Koshland 1958). In the pre-existing equilibrium hypothesis, (Tsai et al. 1999; Goh et al. 2004) the interface in the native state of the protein is represented by an ensemble of conformations. These conformations can be visualized as occupants of different energy minima formed by the ruggedness at the bottom of the folding funnel (Tsai et al. 1999). The ligand is supposed to bind to only the active conformation and
consequently shifting the equilibrium toward the active conformation. This hypothesis is supported by the crystal structures of an antibody caught in two different conformations in its free state and each state is shown to bind different ligands (James et al. 2003). Flexibility in certain regions allows the protein to adopt different conformations and provides a possibility to bind structurally different ligands. This gives an opportunity to functional diversity without being dependent of evolution (James and Tawfik 2003; Goh et al. 2004) Flexible regions in proteins also provides a mechanism to control the formation of multi protein complexes, for example; a binds first b and fold to be able to bind c (Caflisch 2003). An extreme form of induced fit is seen for the group of proteins whose native state is intrinsically unstructured (Wright and Dyson 1999; Tompa 2002; Uversky 2002). Many of the intrinsically unstructured proteins (IUPs) are involved in key cellular processes, such as transcription, translation, signal transduction and in the cell cycle, they are believed to be quite common (Tompa 2002). As first recognized in site specific DNA binding (Spolar and Record 1994) folding upon binding is accompanied by a large decrease in conformational entropy. This might be a mechanism for protein to uncouple affinity from specificity, which is of importance for many transient forming complexes. Another interesting function of IUPs is that they often preserve an extended conformation upon complex formation. They are therefore are able to bind with a large interface area, forming multiple contact points to the target relative to its size (Tompa 2002).

4.2 Structural properties of protein interfaces

Protein-protein interactions show a remarkable variation both with respect to the size of their binding interface and to the nature of their interaction. Jones and Thornton introduced a classification where they divided protein-protein complexes into four groups (Jones and Thornton 1996). One group consists of homodimeric or oligomeric proteins, which often are obligate complexes of proteins which are often unstable in their monomeric forms. Other classes are enzyme-inhibitor and antibody-protein complexes. Last is the class of heterodimeric proteins, a class that is steadily growing with more structures becoming available.

Because structures of complexes consisting of protease-inhibitors and antibody-antigens have been over-represented in the protein data bank, earlier results from structural analyses have been based primarily on them. In a compiled study, Lo Conte, Chothia and Janin have analyzed the atomic structure of binding interfaces for 75 protein-protein complexes. Of these, 32 are heterodimeric protein complexes, obligate complexes are excluded (Lo Conte et al. 1999). Some
general principles can be drawn from this and other studies in this area (Jansson et al. 1996; Stites 1997; Bogan and Thorn 1998; Chakrabarti and Janin 2002).

First, protein recognition sites generally show high chemical and topological complementarity, with alignment of polar and nonpolar residues. The packing density is comparable to the interior of proteins and ordered water molecules within the interface contribute to the close packing and as well as providing polar contacts between the proteins.

Second, the average interface area is large. The distribution about the average is wide, but half of the complexes fit within the “standard size” of $1600 \pm 400 \, \text{Å}^2$. The surface is relatively flat compared to typical enzyme-ligand complexes. Protein interfaces larger than standard size usually divide into multiple patches, of which at least one is of standard size (Chakrabarti and Janin 2002).

Third, many protein interfaces contain a solvent excluded core of amino acids lined with a rim of amino acids that are solvent exposed to some extent and might function to seal the central area from solvent. An interface of standard size contains about 22 amino acids from each protein, about half forming the core.

Fourth, while interfaces of obligate protein complexes often have a hydrophobic character, resembling the interior of proteins, interfaces of nonobligate protein-protein complexes are very similar to the solvent exposed surfaces of proteins in general. The binding surface is generally not more hydrophobic than the rest of the protein surface, even though large variations occur (Jones and Thornton 1996; Lo Conte et al. 1999). One exception are the core residues at the interface of the proteins studied by Lo Conte and colleagues that has a somewhat higher percentage of nonpolar atoms 63% compared to 56% of the surface on average. The amino acid composition shows only minor differences from the rest of the surface. They are richer in aromatic residues: tyrosine, phenylalanine, tryptophane, and also in histidine and methionine. They are also depleted in charged residues: aspartate, glutamate and lysine, but not in arginine. However, the difference in amino acid composition compared to other protein surfaces is small.
4.3 The size of the interface

Lo Conte and colleagues found that the overall range of interface areas is wide, ranging from 1150 to 4660 Å². The interface area is defined as the area of both proteins that become inaccessible to solvent upon binding. Most complexes were found to cluster in the lower end of this range with 55 out of 75 (70%) within so called standard size, 1250-2000 Å². 27% of the proteins complexes though, had a larger interface in the range 2000-4660 Å², which is similar in size to the interaction surface of oligomeric proteins. However, they were found to be less hydrophobic and contain more charged groups than oligomeric assemblies. This is not surprising since the non-obligate complexes need to be stable and soluble in water while obligate protein complexes might not (Lo Conte et al. 1999). A similar feature of oligomeric proteins and protein-protein complexes with large interfaces is that binding often is followed by large conformational changes.

There is an average number of ten hydrogen bonds in a recognition area of which almost one third involves a charged side-chain. Salt bridges are also common. The carbonyl oxygen is the most common hydrogen bond acceptor within the interface. (Jones and Thornton 1996; Lo Conte et al. 1999).

4.4 Characteristics of hot spots

Alanine scanning mutagenesis has shown that residues in so-called hot spots are important for binding affinity (Clackson and Wells 1995; Bogan and Thorn 1998; Hu et al. 2000). This is further supported by crystal structures showing high complementarity between the hotspots of both molecules with buried charged residues forming ion-pairs and hydrophobic residues of one protein fitting snugly into a cavity of the other. By definition a hot spot mutation acts by destabilizing the bound state ensemble relative to the unbound (DeLano 2002). In a pioneering study, Wells and colleagues studied the binding interface of the human growth hormone (hGH) and the extracellular domain of the receptor (hGHbp) (Wells 1991; Clackson and Wells 1995). They used an alanine scanning approach to substitute individual amino acids in the binding interface to alanine in order to identify side chains important to the binding affinity of the complex. They found that substitution of some amino acids within the core region of the interface had larger effect on affinity than others, forming a hotspot for interaction. This effect
could not be correlated to the amount of buried surface area. Especially the substitution of either
one of two tryptophan residues of hGHbp destabilized the complex more than 4.5 kcal/mol. They also noticed that substitution of residues at the rim, which correspond to half of the interface area, had little effect on binding affinity.

4.5 Consensus binding site and adaptivity in hot spots

Bogan and Thorn have compiled and analyzed 2325 alanine mutants within the binding interfaces of 22 protein-protein complexes, where the change in free energy of binding is known (Bogan and Thorn 1998). They found that three amino acids; tryptophan, tyrosine and arginine occurred more frequently than other residues in hot spots, while leucine, methionine, serine, threonine and valine were rarely found. They also suggested that the main function of the area surrounding the core is functioning as an o-ring, forming a tight seal excluding solvent from the core.

A few proteins have more than one binding partner and it has been observed that they have a tendency to use the same hotspot for binding different proteins. Structural analyses of hot spot regions show that they can adopt different conformations to different target molecules. An interesting example is the Fc domain of IgG, which binds to four targets (SPA, streptococcal protein G, rheumatoid factor and neonatal Fc-receptor) using the same consensus region involving CH$_2$ and CH$_3$ (Bogan and Thorn 1998; DeLano et al. 2000). In one study peptides binding to Fc were selected using phage-display technology (described in section 5.1) (De Lano Wells Science 2000). All three isolated peptide binders competed for binding to the consensus region of Fc, which excluded functional reasons for targeting the same surface. The crystal structures of Fc, bound to different proteins, show that the consensus region undergoes a series of conformational changes to form a complementary surface to each binding protein. Theses changes include side chain rotations and the relative orientation of CH$_2$ and CH$_3$ changes.

Similarly, when phage-display technology has been used to select binders to hormones these binders often seem to target the hot spot for interaction (Pillutla et al. 2002; Sidhu 2003). Development of small ligands that bind to protein-protein interfaces has challenged researchers for a long time. It has been considered difficult due to the large and flat binding site and also because of absence of natural small molecule binders to use as a starting point. In a recent published article Thanos and colleagues (Thanos et al. 2006) describe a small molecule, designed
independently of the receptor, which binds IL-2 using the same hot spot as the IL-2α receptor but trap IL-2 in a different conformation. Despite that the small molecule only uses about one third of the contacts compared to IL-2Rα, they have almost the same affinity. Optimized electrostatic interactions and by trapping IL-2 in a conformation that provides deeper cavities and better packing with fewer buried water in the interface is most likely the reasons for the more efficient binding.

It seems that hot spots possess an inherit property which makes them attractive for binding by other proteins and peptides. Possibly, this is influenced by the higher frequency of certain amino acids in hot spots together with an increased conformational adaptivity in the area.

4.6 Affinity and specificity

The same interactions that are involved in maintaining protein stability are also mediating protein-protein interactions. A distinct difference is that protein-protein interactions can be favoured by both enthalpy and entropy. The association constant is determined by the free energy difference between the bound and the free states of the interacting proteins. In addition to providing the association constant a thermodynamic characterization of the interaction also gives insight into the mechanism of binding, for example entropic effects origin from hydration/dehydration or conformational changes or enthalpic contributions from bond formation (Doyle 1997). If the binding enthalpies are determined for different temperatures, then also the change in heat capacity $\Delta C^\circ_{\text{p,bind}}$ can be determined. The affinity of a protein to another protein refers to the strengths of binding while its specificity refers to its restriction of binding to one preferred ligand over another. The association equilibrium between compound M1 and M2 can be written:

$$M_1 + M_2 \leftrightarrow M_1M_2$$

The strength of the interaction is described by the association constant $K_a$, or dissociation constant $K_d$ and are related to the rate constants, $k_{\text{on}}$ and $k_{\text{off}}$ for association and dissociation respectively.

$$K_a = [M_1[M_2]/[M_1][M_2] = k_{\text{on}}/k_{\text{off}} = 1/K_d$$
These constants are related to Gibbs free energy for binding as:

$$\Delta G^\circ_{\text{bind}} = -RT \ln K_a = \Delta H^\circ_{\text{bind}} - T \Delta S^\circ_{\text{bind}}$$

Where $R$ is the gas constant (1.9872 cal/K mol) and $T$ is the temperature in Kelvin and $\Delta H_{\text{bind}}$ and $\Delta S_{\text{bind}}$ is the enthalpy and entropy of binding respectively.
5 Protein engineering

In the late 1970s the technical advancements had made the time ready for molecular biology to step into a new area, the ability to specifically modify a protein by changing the coding gene. The first to describe this approach was Michael Smith and colleagues in 1978 when they successfully inserted a point mutation to change the phenotype of the bacteriophage φX174, using oligonucleotides on single stranded DNA (Hutchison et al. 1978). This was soon followed by others and in a study Wilkinson and colleagues mutated residues in the enzyme tyrosyl-transfer RNA synthetase to evaluate their contribution. They successfully anticipated an amino acid substitution with improved catalytic activity and they thereby coined the expression; protein engineering (Wilkinson et al. 1984). Protein engineering has had a tremendous impact on protein science during the last decades. It has made it possible to improve properties of proteins and to re-design proteins with completely new functions. The field can typically involve anything from a single amino acid substitution to the deletion or addition of entire domains, where unwanted or desired functionalities can be removed or added. It has also given us insight into areas of protein folding pathways (Fersht and Daggett 2002) and stability through the quantification of the energetic contribution of noncovalent interactions (Matthews 1987; Shortle 1989; Pakula and Sauer 1990; van den Burg and Eijsink 2002), and also in specificity and catalysis as for example in the attempt to convert the specificity of trypsin into chymotrypsin (Graf et al. 1987; Perona et al. 1995; Venekei et al. 1996).

5.1 Protein engineering methods

Site-directed mutagenesis has proven to be a useful method for introducing specific changes into DNA. It can involve a single point mutation or a combination of point mutations, deletions or insertions. The methods to perform site-directed mutagenesis were first based on the use of single-stranded DNA, where M13 phages or phagemid vectors were used as template for the mutation (Zoller and Smith 1982; Zoller and Smith 1983). The mutation is introduced by synthetic oligonucleotides containing one or more mismatched nucleotides, which are hybridized to the template and used as a primer for a DNA polymerase-assisted synthesis of the complementary strand. The hetero-duplex DNA is transformed into E. coli for amplification and. Mutants are obtained by various screening methods and subsequently sequenced. Since then many improvements have been made to increase the yield of mutants relative to wild type, see for
example (Kunkel 1985; Kunkel et al. 1987). A strategy for obtaining more extensive changes in a particular region of the encoding gene is cassette mutagenesis, where a cassette (typically 20-100 bp) of synthetic or digested DNA is used to replace a corresponding segment of a gene. This procedure is suitable for exchanging larger fragments, such as segments encoding entire secondary structure elements. The invention of the PCR technique by Kary Mullis (Mullis and Faloona 1987) was the beginning of a new era for recombinant DNA technology. When used as a tool in site directed mutagenesis, oligonucleotides containing the desired DNA sequence are typically used as primers for the amplification of PCR products containing the mutation.

The advances in DNA technology has transformed site directed mutagenesis to a rather straightforward technique, however it is significantly more challenging to predict the consequences that the changes will have on the protein. Knowledge about the protein structure, binding site or active site may greatly facilitate a successful outcome. In the past, residues within the hydrophobic core were often chosen in attempts to improve the thermodynamic stability of proteins (van den Burg and Eijsink 2002). However, the protein core is often already well packed and with increasing knowledge about protein structure and stability, focus has changed to mutations of surface residues, which have often turned out to give better results (van den Burg and Eijsink 2002). For example has improvements of electrostatic interactions between surface residues (Grimsley et al. 1999; Spector et al. 2000) and helix dipole N-cap and C-cap residues (Marshall et al. 2002) led to increased stability. A number of proteins have also been stabilized by reducing the entropy of the unfolded state, such as the introduction of prolines (Muslin et al. 2002) or disulfide bridges (Matsumura et al. 1989).

An important clinical example, where site-directed mutagenesis successfully has led to improvements of a protein is that of insulin. The rate at which insulin reaches the blood stream following an injection is in some cases not fast enough and there were thus considered necessary to speed up the process (Brange and Volund 1999; Vajo and Duckworth 2000). The observation of a lag phase in absorption after subcutaneous injection of insulin and the observation that insulin in high concentration form dimers and hexamers led to the idea that if the monomeric fraction could be increased, this would result in faster absorption. The surfaces involved in self association were identified and mutated (Brange et al. 1988; Brange 1997), which resulted in a faster acting insulin that is now in clinical use (Brange and Volund 1999; Vajo and Duckworth 2000).
In the late 1980s methods for introducing multiple mutations in various combinations were becoming more common (Merino et al. 1992). In such approaches the gene is simultaneously varied at several positions to create a pool of mutant genes, encoding a corresponding protein library of high diversity. The strategy depends on the purpose of the project, and can involve technology to address the randomization to specific positions (few or many, at each position using codon mixes for up to 20 amino acid (combinatorial mutagenesis)) or to spread them randomly over the entire gene. The most common methods used for this purpose utilize degenerate oligonucleotides, error prone PCR and/or DNA shuffling (Stemmer 1994; Arnold 1998; Cramer et al. 1998; Zaccolo and Gherardi 1999) and random insertion/deletion mutagenesis (Murakami et al. 2002).

To be able to find a member in the library with desired properties, the development of powerful screening and selection technologies has been essential. Methods like phage display (Smith 1985), ribosome display (Hanes and Plückthun 1997), and mRNA display (Roberts and Szostak 1997) are used for selection by providing methodology for the physical linking between genotype (the encoding nucleic acid) and phenotype (the protein) during the procedure. In most instances, the selection procedure is based on affinity for a specific target protein or molecule bound for example to a matrix. The library members that bind to the target protein will be separated from the rest of the library. The selection procedure can be repeated in several rounds. Depending on the selection method used, further mutagenesis between rounds are more or less easy to introduce.

The most commonly used method for display and selection of library members is phage display. The ability to display and select a peptide on the surface of filamentous phage was first demonstrated by George Smiths group (Smith 1985; Parmley and Smith 1988). The fact that the phage particle harboured the DNA encoding the displayed protein inside made it ideal for displaying combinatorial protein libraries. Shortly, three different groups simultaneously described its use in displaying libraries of randomized peptides from which ligands to monoclonal antibodies (Cwirla et al. 1990; Scott and Smith 1990)and to streptavidin (Devlin et al. 1990) could be selected. The same year McCafferty et. al displayed libraries of the variable domain of antibodies, single-chain Fv (McCafferty et al. 1990).

In phage display, each of the randomized genes is inserted into one of the phage genes coding for a coat protein. Usually protein pIII located with only a few copies at one of the tips of the phage...
particle, or protein pVIII present in a few thousand copies along the sides of the phage particle, are used as fusion protein. When the phage is assembled the proteins will be displayed on the surface of the phage particles. The variant with desired qualities can be selected and then enriched in a host cell. Since the genotype is linked to the phenotype, the amino acid sequence of the selected variant will be retrieved from DNA sequencing. Different *E. coli* phages are commonly used for phage display and the most common is the Ff filamentous phage family, which includes M13, fd and f1. One advantage with this phage is that it can respond to the increased genome size by becoming longer. A possible drawback might be if the modified protein has difficulties to pass through the plasma membrane of *E. coli* as the phage is assembled on its way out. One limitation of the phage display system compared to the other methods mentioned above is the need to transform the library into *E. coli* cells, which sets the limit for the library size.

### 5.2 Engineered binding proteins from alternative protein scaffolds

For a long time, antibodies have been used as binding proteins due to the possibility to generate such binders to such a vast number of foreign molecules. In consequence of the development of combinatorial protein engineering techniques in the nineties, naturally existing proteins adopted with completely new binding specificities are now generated as interesting alternatives to antibodies and other binding molecules. Even though antibodies and antibody fragments still are widely used as research tools as well as applications for diagnostics and medicine, their large size, often low expression yields, poor tissue penetration ability leads to certain disadvantages, where novel engineered proteins might offer an interesting alternative.

These engineered binding proteins use the framework of a chosen protein, denoted scaffold protein, as base for harboring a new binding site. Different binding sites are generated to the scaffold by combinatorial variation of amino acids within one or more regions of the scaffold protein followed by functional selection. According to Nygren and Skerra, scaffold proteins can be divided into three different groups depending on the architecture of the interface area (Nygren and Skerra 2004). (i) It can for example be made from variation of amino acids in a single continuous loop as has been demonstrated for the use of staphylococcal nuclease and thioredoxin where combinatorial strategies involving 16 or 20 randomized positions have been described (Colas *et al.* 1996; Fabbrizio *et al.* 1999; Norman *et al.* 1999). Even though higher affinities have been obtained with this concept compared to earlier attempts using
conformationally flexible epitope libraries, this method has not been able to fulfill the expectations of using the peptides either as isolated functional unit or as template for non-peptidic substances in drug development.

(ii) As often encountered for naturally occurring protein-protein complexes, the binding interface is built of non-contiguous regions of one or several secondary structural elements generating a patch on the surface of the protein. The class of binding proteins called affibody molecules provides a suitable example of this type of binding interface (further described in chapter 7), where several specific binding proteins have been selected with dissociation constants typically in the µM to nM range (Nord et al. 1995). Another example is the cellulose-binding domains (CBD), which adopt a triple-stranded anti-parallel β-sheet, stabilized by cysteine bridges (Smith et al. 1998; Lehtio et al. 2000). Here seven positions were randomized and binders to different targets in the low µM range have been selected.

(iii) The interface may also be formed by several loops, forming a continuous surface on the scaffold protein. The best example of this type of scaffold protein is probably exemplified by the immunoglobulin fold itself, where the binding site is constituted of three hypervariable loops in each domain. The VHH domains found in Cameloids (lama, camel, and dromedary), represent a natural group of antibodies devoided of light chain and due to its monomeric form make an interesting alternative as scaffold protein. This type of scaffold often forms extended conformations and appears to be suitable as inhibitors for enzyme active sites (Lauwereys et al. 1998; Muyldermans et al. 2001). The VHH domain has inspired to the construction of a camelized human V_{H} domain. CDR-3 was randomized and binders to different targets were selected. However, in some cases self association and low melting temperatures were seen suggesting that further engineering to optimize this camelized variants was needed (Davies and Riechmann 1995; Martin et al. 1997).

The method used for the construction of libraries, and the intended use of the selected variants set different criteria for the selection of a suitable scaffold. However, some general properties including a monomeric organization, a high solubility combined with a stable structure tolerable to multiple substitutions are most likely important. A small scaffold-protein that can be produced by chemical synthesis might facilitate incorporation of chemically modified residues or isotopes in selected variants after selection which can further facilitate their use in areas like diagnostics and in vivo imaging.
6 Affibody molecules

6.1 Protein A

Some pathogenic gram positive bacterial strains have been found to express cell wall anchored surface receptors, with repetitive domains on their surface. They bind to host proteins and are believed to have a role in the virulence of the bacterium. Protein A from Staphylococcus aureus (SPA) was described as a binder to the Fc region (Forsgren and Sjöquist 1966) of human IgG subclasses 1, 2 and 4 (Langone 1982; Eliasson et al. 1988) but only weakly to subclass 3 (Kronvall and Williams 1969; Ankerst et al. 1974). SPA was also found to bind to the heavy chain of a limited number of Fab fragments of human immunoglobulins (VH3) (Sasso et al. 1991; Potter et al. 1996). The full sequence of staphylococcal protein A was obtained by Uhlén and coworkers (Uhlen et al. 1984). Five homologous domains (E, D, A, B and C) of approximately 58 amino acids each were identified, which all binds individually to Fc (Moks et al. 1986). They also have μM affinities for the Fab region of certain antibodies (Jansson et al. 1998).

Protein A is widely used for purification and detection of antibodies (Langone 1982) and as affinity tag for recombinant proteins (Ståhl and Nygren 1997).

6.2 The Z domain

The Z domain was originally designed for use as a fusion protein to facilitate expression and purification of recombinant proteins. The B-domain of SPA was chosen as template since its sequence was closest to consensus of the five domains and it did not contain methionine residues, which made it resistant to cyanogen bromide cleavage. Also, data from a crystal structure of the B-domain in complex with Fc were available at the time (Deisenhofer 1981). Two amino acids were replaced, one for cloning reasons (A1V) and one to remove a hydroxyl amine cleavage site (G29A) (ref Nilsson et al. Prot. Eng. 1987). The B and Z-domain bind to Fc with similar affinity (KD of 10-70 μM) and on- off-kinetics (Jendeberg et al. 1995; Braisted and Wells 1996) but the Fab binding is very weak due to the G29A mutation (Jansson et al. 1998).
The three dimensional structures of the B, E and Z domains have been determined by NMR (Gouda et al. 1992; Starovasnik et al. 1996; Tashiro et al. 1997; Zheng et al. 2004). They all show similar three-helix bundle folds, however helix 1 of the B domain has a larger tilt angle with respect to the other two helices.

The crystal structure of the B-domain in complex with Fc₁ were used to identify amino acids in the binding surface (Deisenhofer 1981). Eleven amino acids in helices 1 and 2 from the B-domain and 9 residues from the C₁₁₂-C₁₁₃ interface of the Fc-region were suggested to be involved directly in binding. Similar binding epitopes of the Z and the B domain were indicated by a mutational analysis (Cedergren et al. 1993; Jendeberg et al. 1995). In one study, four of the residues proposed to be involved in binding (L17D, N28A, I31A and K35A) were replaced and the resulting mutants all showed reduced binding affinity for Fc. CD measurements further suggested that all mutants had similar secondary structure contents as the Z-domain.

6.3 Affibody molecules

Affibody molecules are based on combinatorial engineering of the Z domain, where 13 amino acids in the binding interface to Fc have been selected for randomization. Nine of eleven residues (Q9, Q10, N11, F13, Y14, L17, N28, Q32, K35) were selected due to their suggested roles in the complex between the B domain of protein A and Fc (Deisenhofer 1981). Also residues H18, E24, E25 and R27 were included based on their location (Nord et al. 1995).
A first naïve affibody molecule library (ZLib-1) was constructed, using degenerated oligonucleotide sequences that involved the NN(G/T) degeneracy code for the 13 positions, which covers all 20 amino acids as well as the TAG amber stop codon (Nord 1995). In a second library, (ZLib-2) a (C/A/G)NN codon degeneracy was used excluding stop codon at the expense of the aromatic amino acids (Trp, Tyr, Phe) (Nord et al. 1997). Each library had a size of approximately 4·10^7 transformants, while randomization of 13 residues to any of the 20 amino acids would theoretically result in 20^{13} different variants. Consequently, only a fraction of all possible variants will be represented in the libraries.

The libraries were constructed for monovalent phage display, using a phagemid vector. In a first trial the targets, Taq DNA polymerase, human insulin and a human apolipoprotein A-1 variant were used to prove the concept. Specific affibody molecules were selected from the phage-displayed libraries and were able to bind target with μM affinity (Nord et al. 1997). The affinity for a Taq DNA polymerase affibody molecule was further increased to 30-50 nM by affinity maturation (Gunneriusson et al. 1999). Affinity maturation of human factor VIII specific affibodies are described by (Nord et al. 2001).

Several other affibody molecules to a variety of targets have also been selected, some of which are high-affinity (nanomolar) binders (Hansson et al. 1999; Rönnmark et al. 2002; Sandström et al. 2003; Wikman et al. 2004) and a range of applications and follow-up engineering concepts has has been described such as surface display on S.carnosus (Gunneriusson et al. 1999) anti-idiotypic affibodiy molecules (Eklund et al. 2002) fusion to Fc (Rönnmark et al. 2002) protein purification
(Nord et al. 2000; Gräslund et al. 2002), and diagnostics (Andersson et al. 2003; Renberg et al. 2004) protein microarrays (Renberg et al. 2005)

A larger library with approximately $3 \cdot 10^9$ members was later constructed (Wikman et al. 2004) and a binder to the extra cellular domain of the human epidermal growth factor was selected. This affibody molecule showed specific binding to a tumor cell line expressing HER2/neu and has later been further affinity matured into a 22 pM binder, which holds promise for diagnostic and therapeutic applications (Orlova et al. 2006).
Present investigation

The work presented in this thesis focuses on the structure, stability and affinity of an engineered protein-protein complex, consisting of a member from a group of engineered binding proteins called affibody molecules and its protein target. At the start of this study there was no known structure of an affibody molecule, free or in complex with a target. The NMR structure of the complex between the investigated affibody molecule-target pair gives insight into the mechanism for how this affibody molecule recognizes and binds to its target. This affibody was further characterized in its free state by different biophysical methods in an attempt to understand how the mutations introduced in the parental scaffold to generate the binding surface affects the stability of the protein. An hypothesis on how the binding affinity could be increased was investigated in a thermodynamic study involving five disulfide-stabilized mutants. The results presented in this thesis are relevant for further design and selection of affibody molecules, but also for synthetic binding proteins in general since only a limited number of structural and biophysical studies have been performed on this group of proteins (Binz and Pluckthun 2005).

In a wider perspective, the affibody molecule-target protein complex is a good model system for structural and thermodynamical studies of protein-protein interactions to understand the basis for molecular recognition. The small size of both proteins involved in this complex makes this system ideal for studies by NMR. Also, the relative ease with which these proteins could be produced facilitated characterization with several biophysical methods including circular dichroism (CD) spectroscopy, fluorescence spectroscopy and isothermal titration calorimetry (ITC).
7 Structural analysis of an affibody-target pair (I, II)

7.1 The $Z_Z^{SPA-1}$ system

The NMR structure of a protein-protein complex consisting of the $Z_{SPA-1}$ affibody and its binding partner the Z domain was determined by nuclear magnetic resonance (NMR). With the objective to select for anti-idiotypic affibody binders, $Z_{SPA-1}$ was selected as a binder to Staphylococcal protein A (SPA) and recovered by competitive elution using human IgG (Eklund et al. 2002). The affibody were found to bind with µM affinity to all five domains of SPA, as well as to the Z-domain. This means that $Z_{SPA-1}$ have acquired a binding site capable of interaction with the protein serving as its ancestor.

The choice to study the $Z_Z^{SPA-1}$ complex was made for several reasons. First, no structure of an affibody was known at the time. The competitive elution with IgG during the selection procedure made it appear likely that at least part of the recognized binding surface on $Z$ overlapped with its IgG-binding interface. But an atomic structure of the complex was needed to give the full picture. The determination of the NMR structure of this complex, as well as the crystal structure solved by another group (Högbom et al. 2003) in the same journal gave the first view of an affibody binder in complex with a target protein (II). Second, the small size of $Z_Z^{SPA-1}$ (~15 kDa for the whole complex) makes this complex a good model system for structural studies by NMR of protein-protein interactions in general and for artificial binding proteins in particular. It can also be mentioned that the character of these small, and easily handled proteins have been highly appreciated during the course of this project.

The proteins were expressed intracellularly in E.coli using a variant of the pET-28a(+) vector denoted pT7 (Eklund et al. 2002), resulting in an extra methionine at their N-terminal. The unlabelled proteins were produced in Luria Broth medium, while isotope enriched proteins were expressed in minimal medium supplemented with $^{15}$N-$\text{(NH}_4\text{)}_2\text{SO}_4$, $^{13}$C-glucose or both (Jansson et al. 1996). Isotope labelling for NMR studies was necessary to be able to discriminate between the signals of the two proteins in the complex. One protein component at a time was analyzed, which was facilitated using NMR samples with one protein labelled and the other one unlabelled and vice versa. Both proteins were purified by affinity chromatography (IgG-sepharose and SPA-sepharose for Z and $Z_{SPA-1}$ respectively) followed by gel filtration (I). The pH was first chosen to

33
be 6.8 which would be inbetween the theoretical pI values for Z and Z\textsubscript{SPA-1}, according to expasy (www.expasy.org) but since Z\textsubscript{SPA-1} slowly precipitated in our samples at those conditions, the pH was changed to 5.6. The temperature for NMR experiments was chosen to 30 °C from the appearance of NMR spectra. NMR data were acquired on Bruker Avance 500 MHz and 600 MHz spectrometers and on a Varian Inova 800 MHz instrument equipped with 5 mm triple resonance \(^1\text{H}, ^{13}\text{C}, ^{15}\text{N}\) probes.

### 7.2 Structure analysis

The structure of the complex reveals a combination of side chain rearrangements (i.e. induced fit) of the Z-domain and coupled folding of the affibody molecule that occurs upon complex formation. In the complex, Z\textsubscript{SPA-1} is able to adopt a similar and compact three helix bundle scaffold to that of the Z-domain. Backbone rms deviation between Z and Z\textsubscript{SPA-1} is only 0.9 Å for residues 8-56 of the two proteins. The differences in secondary structure between the two proteins in the complex are small, restricted to helix 1 that is somewhat shorter for Z\textsubscript{SPA-1}, starting on residue 6 instead of 4 as in the Z-domain and the loop between helix 2 and 3 in Z that instead forms a 3\text{10}-helical conformation in Z\textsubscript{SPA-1}. (Fig 7.1)

Side chain conformations and packing arrangements within the core are very similar in the two proteins. The fact that the structural properties of the Z-domain scaffold were preserved in the complex-bound affibody molecule was not obvious, since 13 residues (out of 58) at the surface of the Z-domain have been changed (Nord \textit{et al.} 1995; Nord \textit{et al.} 1997; Eklund \textit{et al.} 2002). Recently, NMR structures of an affibody-affibody complex and the involved affibody molecules in their unbound states have been solved (Lendel \textit{et al.} 2006). Here the overall structure of the parental Z scaffold is retained in both affibody proteins, both in the complex and in the free state. Besides that affibody molecules have been successfully selected towards a range of targets with as low as pM affinities (Nord \textit{et al.} 1997; Gunneriusson \textit{et al.} 1999; Hansson \textit{et al.} 1999; Nord \textit{et al.} 2001; Rönnmark \textit{et al.} 2002; Sandström \textit{et al.} 2003; Wikman \textit{et al.} 2004; Orlova \textit{et al.} 2006; Wikman \textit{et al.} 2006) these structures further prove the excellent properties of the Z domain as scaffold protein, with the ability to preserve the overall structure for all three investigated affibody molecules in their complex states and also for two out of three in their unbound states.
The recognition interface of Z:ZSPA-1 is composed of helices 1 and 2 of both proteins while helix 3 extends on their backside (Fig 7.1). The size of the interface comprises 800-900 Å² of each subunit which is of standard size for a typical protein-protein complex (1600 +/- 400 Å² of total buried surface area) (Lo Conte et al. 1999). The binding site has a high steric complementarity consisting of a mainly nonpolar surface (64%) lined by polar residues along solvent exposed edges. The interaction surface of ZSPA-1 is dominated by two hydrophobic cavities and a large protrusion. (Amino acids in ZSPA-1 is denoted with a *). The larger groove is formed by *Leu-9, *Gly-13, *Val-17, *Phe-32, *Leu-34, *Trp-35 and the aliphatic part of *Arg-14 and interact with Phe-13, Leu-17, the aliphatic part of Arg-27, Ile-31 and Leu-34 in the Z domain (Fig 7.2a) The side chains of *Val-17, *Lys-27, *Lys-28, *Phe-32, *Ile-31 compose the smaller cavity and interact with Tyr-14 and His-18 in Z (Fig 7.2a). Tyr-14 donates a hydrogen bond to the backbone carbonyl of *Asp-24 (Fig 7.3a). The large protrusion is formed by the bulky side chain of *Trp-35 and plays a central role in the interaction interface (Fig 7.2a). *Trp-35 is encircled by six residues from helices 1 and 2 in the Z domain and substitution of the tryptophan with a lysine abolishes binding (data not shown). The edges of the interface contain several polar interactions and hydrogen bonds (Fig 7.3a,b) including a possible salt bridge between Glu-24 and *Lys-7.
Figure 7.2 The Z:Z_{SPA-1} interface. 

a) Interacting residues of helices 1 and 2 of the Z domain (blue) are illustrated on the Z_{SPA-1} affibody shown by a surface representation. 

b) Interacting residues of helices 1 and 2 of the Z_{SPA-1} affibody (red) are illustrated on the Z domain shown by a surface representation. The 13 residues varied in the Z_{SPA-1} affibody are coloured in yellow. The position of mutated residues, not directly involved in the interaction with Z is marked with a star.
All randomized amino acids except *Val-11 and Thr-18 are involved in binding, although *Asp-24 is only taking part through its backbone carbonyl.

Figure 7.3 Detailed views showing a) a selection of hydrogen bonds and b) how the *Trp-35 residue of Z_{SPA-1} (red) is buried in a deep cavity on the Z domain (surface).

An interesting observation is that the interface of Z_{SPA-1} in several ways mimics the surface of Fc involved in binding to SPA and Z (i.e. the natural binding partner of Z) by resembling the interaction surface of Fc and inducing similar side chain rearrangements in the Z-domain upon binding (Fig 7.4a-c). A possible reason for Z_{SPA-1} binding to the same surface on Z as Fc can be that the Z_{SPA-1} affibody clone was selected by competitive phage elution using IgG.

Figure 7.4 Comparison of the interacting surfaces of a) Z_{SPA-1} and b) Fc and their interaction with the Z (B) domain (blue). c) Illustration of similar conformational rearrangements of Tyr-14 and His-18 side chains for the free Z domain (red) (Tashiro et al. 1997); Z in complex with Z_{SPA-1} (blue); and Z (B) in complex with Fc (green) (Deisenhofer 1981)
The characteristic features of the interaction surface of the Z-domain involve two protrusions and one deep and one shallow cavity. The two protrusions are formed of Tyr-14, His-18 and Phe-13, Leu-17, Ile-31 respectively. The deep cavity is filled with *Trp-35. Overall, $Z_{SPA-1}$ contributes by direct contact from 15 residues of which 10 were randomized (Fig 7.2b). Some of the substitutions are understood when analyzing the complex. The F13G substitution replaces the large bulky side chain of phenylalanine with the small glycine which contributes to the major hydrophobic cavity of $Z_{SPA-1}$. Another example is the K35W mutation which allows tryptophan to bind into the deep nonpolar cavity on Z. The 10 hydrogen bonds found at the interface is an average value (Lo Conte et al. 1999).

A comparison to the crystal structure solved by Högbom and colleagues (Högbom et al. 2003) of the same complex shows a very similar conformation of the backbones, with a rms difference of 1.18 Å for residues 5-56, and also a similar side chain packing. There are a few differences however including alternative side chain rotamers of Gln-32 and Lys-35 at or near the binding surface of Z. This is not surprising considering the different salt concentration in the buffers.
8 Biophysical characterization of the affibody Z\textsubscript{SPA-1} (II, III)

Even though the three-helix scaffold of the Z domain is fully preserved in the bound state of Z\textsubscript{SPA-1} this is not true for the free affibody. Early structural studies revealed that Z\textsubscript{SPA-1} is more dynamic than expected and behave like a molten globule (II). A more thorough investigation was performed to elucidate the cause behind this behaviour (III).

The Z\textsubscript{SPA-1} affibody was selected from a phage library were 13 surface exposed amino acids of Z were allowed to vary randomly. In Z\textsubscript{SPA-1}, all 13 residues were found to be changed and structural studies further showed that ten of these are directly involved in binding to its target, the Z-domain itself (II). The extensive mutation of 13 out of 58 amino acids affects the rather delicate balance of stabilizing forces within the protein. Despite the fact that its characteristics most likely will be energetically unfavourable for the binding affinity it was still selected as a binder from the library. The phage-display library used for selection was small (4 x 10\textsuperscript{7} members) compared to a full sampling of all possible variants (~10\textsuperscript{17}), which indicates that a "better" Z domain-binding affibody molecule might exist, but was not present in the relatively small library. Later, other stronger Z domain binders have indeed been selected from larger affibody libraries. Taken together this strongly supports the idea that Z\textsubscript{SPA-1} was selected because it was the best binder in a library of rather small size.

8.1 Molten globule characteristics

Z\textsubscript{SPA-1} undergoes self-association at millimolar concentration as evidenced from both gel filtration and CD experiments. In gel filtration, the elution profile is concentration dependent, corresponding to a dimer at low millimolar concentrations and approaching the expected retention volume at micromolar concentrations (Fig 8.1a). The peak has a highly asymmetric shape. Both features are characteristic for self associating proteins (Hansson et al. 2001), where the reduced retention volume indicates a larger average molecular size and the tailing indicate a rapid equilibrium between molecules of different size. Interestingly a single point mutation of Phe-32 back to the original glutamine results in a symmetric peak with only slightly smaller retention volume than the Z domain when loaded at one millimolar concentration. Also the
mutation W35K results in a protein that shows a symmetric peak but still migrates as a larger protein as far as can be seen when monitoring the process at 215 nm (Fig 8.1b).

The CD spectra of $Z_{\text{SPA-1}}$ in low $\mu$M concentration show a characteristic $\alpha$-helical appearance with minima at 208 and 222 nm even if the mean residue ellipticity is lower than for the $Z$-domain. CD experiments further support self association as spectra undergo major changes approaching millimolar concentration. The mean residue ellipticity decreases with increasing concentration indicating an average reduction of $\alpha$-helical content upon self association (Fig 8.2). Consequently, the results are consistent with $Z_{\text{SPA-1}}$ undergoing self-association at mM concentration and that two of the mutated residues Phe-32 and Trp-35 most likely are involved in the process.

**Figure 8.1** Elution profiles from gel filtration experiments. **A)** The $Z$-domain (solid squares), $Z_{\text{SPA-1}}$ loaded at 2 mM (solid circles) and $Z_{\text{SPA-1}}$ loaded at 0.1 mM (open circles). As reference the $Z$-domain 6.6 kDa and ribonuclease A 13.7 kDa (open squares) was used. **B)** The $Z_{\text{SPA-1}}$ mutants V11N (open circles), F32Q (diamonds), W35K (triangles), and F32Q/W35K (stars) loaded at 1 mM concentration.
The biophysical properties of $Z_{SPA-1}$ was further characterized by NMR, thermal and chemical unfolding and by ANS binding. The $\text{H}^{15}\text{N}$ heteronuclear single quantum correlation NMR spectrum (Fig 8.3a) shows severe broadening of several resonances consistent with a protein in which interconversions between several conformations lead to averaging. Low concentration (50 µM) or different pH does not influence the spectrum much. A significant improvement of the spectrum is observed when adding unlabeled Z domain protein to the sample. This results in sharp resonances and improved dispersion (Fig 8.3b, c).

The thermal melting profile of the affibody molecule shows significantly lower thermal stability compared to the Z domain (40 °C compared to ~75 °C). A reduced cooperative melting transition further supports loss of packing of the hydrophobic core. Reduced stability and low cooperative denaturation was also seen at room temperature for chemical denaturation using
guanidine hydrochloride (GuHCl) as denaturant, monitored by CD at 222 nm (III) These properties taken together, i.e. aggregation tendency, high secondary structure content but improper packing and low stability are typical signs of a protein with molten globule characteristics (Semisotnov et al. 1991; Dobson 1994; Ptitsyn 1995; Regan 2003).

To further test the hypothesis if ZSPA-1 behaves like a molten globule its ability to bind the hydrophobic dye ANS with accompanying fluorescence enhancement was examined. Indeed, ZSPA-1 binds ANS with a significant increase in fluorescence. On the contrary, the ANS binding to the Z domain and the ZSPA-1:Z complex was very weak (Fig 8.4). With these results we conclude that ZSPA-1 behaves as a molten globule in its free state but that it folds into a well-defined three-helix bundle upon complex formation.

**Figure 8.4.** Fluorescence emission spectra of 50 µM ANS with equimolar amounts protein, ANS was exited at 396 nm.
8.2 Structural and dynamic studies

Considering the extensive mutagenesis, involving 13 positions out of 58, it might not seem surprising that the stability of the protein is affected. Still it was selected from a phage display library among $10^7$ variants as a binder to protein A, in spite that the coupled events of folding and binding most likely has an unfavourable contribution to the binding affinity. Structural studies are desired to, on a molecular level, understand why the mutations have resulted in this behaviour of this affibody compared with the parental Z-domain.

The poor quality of NMR spectra so far prevented further structural studies of $Z_{SPA-1}$ in the free state. Because several reports have shown that a group of small, naturally occurring, organic compounds, called osmolytes can have a stabilizing effect on the folded state of proteins, this was considered as an interesting alternative for $Z_{SPA-1}$ (Baskakov and Bolen 1998; Bolen and Baskakov 2001; Celinski and Scholtz 2002). The discovery that the osmolyte trimethylamine N-oxide (TMAO) stabilizes the affibody molecule opened an alternative way for structural studies. Adding 1-1.6 M of TMAO to the sample greatly improved the appearance of the 15N-HSQC NMR spectrum as seen by sharper resonances while the dispersion did not change (Fig 8.5c). The improvement made it possible to partly assign the 15N-HSQC of $Z_{SPA-1}$. The assignment of backbone atoms (N, H$_N$, C$_\alpha$, H$_\alpha$, C$\gamma$) provided the secondary shifts which can be used for secondary prediction (Wishart and Sykes 1994). With the reservation that some chemical shifts were missing the result was consistent with all three helices being present in the molten globule state of $Z_{SPA-1}$(III). Also typical $\alpha$-helical nuclear Overhauser effect (NOE) pattern and long range NOEs present in the complexed form of $Z_{SPA-1}$ were observed which further supports the notion of all three helices being present and that the tertiary structure of the molten globule resembles the structure in Z:$Z_{SPA-1}$ complex. From this it was concluded that the reduced helicity of $Z_{SPA-1}$ is not the result from one helix being absent, instead the results agree with a transient unfolding of helices 1 and 2. However it is not possible to determine if the NOE connectivities come from a single conformation or if they represent an average of several rapidly interconverting conformations. A quantification of the dynamics was investigated by steady-state NOE, peak intensity/integral ratios and saturation transfer experiments. These experiments suggests that poor core packing results in increased dynamics of residues in helices 1 and 2,
leading to transient unfolding of these (III). The low free energy difference of the folded versus unfolded state also results in an increased population of the unfolded state.

8.3 Characterisation of four $Z_{\text{SPA-1}}$ mutants

To analyse the effect of a few single amino acid substitutions, four positions were selected and changed back to their wild type sequence. Val-11 was changed to Asn based on the structural analysis of the complex. The valine residue does not seem to contribute to the binding of the Z domain, whereas it instead increases the hydrophobic surface, which could be unfavourable. The point mutations F32Q, W35K and the double mutant F32Q/W35K were chosen because the NMR characterization of $Z_{\text{SPA-1}}$ indicated increased dynamics in that region.

The mutation F32Q contributes favourably to the character of the free $Z_{\text{SPA-1}}$. A symmetric peak with a concentration independent elution volume in gel filtration suggests that this protein does not aggregate (Fig 8.1b). The retention volume approaches that of the Z-domain and the still somewhat larger volume can be explained either by a larger hydrodynamic ratio due to its molten globule character or by a rapid folding-unfolding equilibrium. Further the stability and $^{15}$N-HSQC spectrum is improved to some extent (Fig 8.3c) and ANS binding is reduced (III). This is consistent with a more structured protein even though the denaturation profile indicates that the folding process is not fully cooperative. However, the F32Q mutation has a negative effect on binding to Z, lowering the affinity by a factor of 10 (III).

For the mutants W35K and F32Q/W35K the structural improvements are smaller than for F32Q. However, no one shows signs of aggregation (Fig 8.1b) and the effect of ANS is reduced (III). The $^{15}$N-HSQC spectrum is improved for F32Q/W35K (data not shown), while W35K has an unstructured appearance (Fig 8.5d). The mutation of W35K completely abolishes the binding affinity for the Z domain as expected since the tryptophan residue seems to have a major role in the binding interface (data not shown).
The first hypothesis that the increased number of hydrophobic residues on the surface of $Z_{SPA-1}$ compared to wild type would be the main determinant for the low stability and molten globule characteristics can not be confirmed from the results of the mutants. Another possible explanation could be the inherited property of amino acids to support certain secondary structures (Combet et al. 2000). This is supported by several secondary prediction programs generating results that agree well with experimental data (III).

To summarize the results so far, $Z_{SPA-1}$ shows several signs of molten globular characteristics in its unbound state, NMR data are consistent with all helices being present, however the low stability of the affibody seems to lead to transient unfolding of helices 1 and 2. Mutational studies do not support an increased hydrophobicity as a cause for the problem to fold into a well defined structure, instead secondary prediction programs suggests the helix propensity as a possible explanation.

Figure 8.5 $^{15}$N-HSQC spectra recorded at 30 ºC of
A) $Z_{SPA-1}$, B) F32Q
C) $Z_{SPA-1}$ with 1M TMAO added D) W35K
9 Thermodynamic stabilization of the $Z:Z_{SPA-1}$ complex

Data obtained this far revealed a rather complex system were the $Z_{SPA-1}$ affibody only adopted a well-defined conformation within the complex. In its free form on the other hand, the native state was best described as an aggregation prone protein with molten globule characteristics, where helices 1 and 2 underwent transient unfolding. Due to the low stability of the native state (i.e. molten globule state) a significant population also existed in a completely unfolded form at room temperature.

![Figure 9.1 Schematic drawing illustrating the different states the $Z_{SPA-1}$ affibody can adopt.](image)

$Z_{SPA-1}$ binds to $Z$ with a $K_d = 0.5$ to 2 $\mu$M depending on temperature, buffer and pH (Dincbas-Renqvist et al. 2004). On the other hand, the natural binding partner IgG (Fc part), which uses the very same recognition site when binding to $Z$, binds with a $K_d$ of 20 to 70 nM (Cedergren et al. 1993; Jendeberg et al. 1995). As mentioned before, the two complexes share several similarities at their binding interface and induce similar rearrangements of side chains on the Z-domain. There could be several explanations to the suboptimal binding between $Z_{SPA-1}$ and $Z$; for example a low thermal stability of $Z_{SPA-1}$ leading to a significant fraction (~16% at pH 5.7 and 30 °C) of unfolded protein already at room temperature (Dincbas-Renqvist et al. 2004). Or simply, the binding interface might not be optimal. It could also be a consequence of conformational stabilization of $Z_{SPA-1}$, going from a flexible to a highly ordered structure upon complex formation resulting in an unfavourable entropy penalty.

46
9.1 The $Z:Z_{SPA-1}$ binding thermodynamics

In a previous study a detailed thermodynamic characterization of the binding energetic of the $Z:Z_{SPA-1}$ complex was performed (Dincbas-Renqvist et al. 2004) and a possible reason for $Z_{SPA-1}$ being an intermediate and not a high affinity binding protein was investigated. Because of the low thermal stability of free $Z_{SPA-1}$, a fraction of unfolded protein needs to fold into the molten globule state prior to stabilization and binding. This process influences the analysis of the binding event and hence its contribution needs to be estimated and separated from binding. Thermal unfolding thermodynamics were studied by a van’t Hoff analysis of free $Z_{SPA-1}$ yielding the temperature dependent values of $\Delta H^\circ_{\text{unfold}}$ and $K_{\text{unfold}}$. These values were then used to correct the binding thermodynamics measured by ITC. The conclusion from this analysis suggests that the thermal stability has only a marginal effect on binding and cannot alone be responsible for the reason to $Z_{SPA-1}$ being a moderate binder.

Further, in an attempt to identify any structural inappropriates of the binding surface, it was compared to structural statistics of 75 other known protein-protein complexes (Lo Conte et al. 1999). It was found that the protein interface is of standard size (1632 compared to 1600 ± 400 Å$^2$), with an average number of hydrogen bonds at the interface (10 compared to 10.1 ± 4.8). The fraction of nonpolar compared to polar residues (64 % in $Z:Z_{SPA-1}$ compared to 56 ± 6%) and number of interface atoms (186 compared to 211 ± 81) are all similar to average values. Consequently, there was no obvious correlation between the binding surface and the suboptimal binding affinity.

Finally, the contribution from conformational stabilization of the molten globule state to a well folded three helix bundle was estimated theoretically. Structurally based energetics calculations indicate that the stabilization of $Z_{SPA-1}$ upon complex formation is entropically unfavourable ($\Delta S^\circ_{\text{conf}}$) for binding. However, it is also presumed that the conformational entropy penalty is partly compensated by a favourable desolvation entropy.

The stability of $Z_{SPA-1}$ in the free molten globule state is low, $\Delta G^\circ_{\text{unfold}} = 1.7$ kcal/mol, $T_m = 42 ^\circ\text{C}$ and $\Delta C^\circ_{p,\text{unfold}}$ is small possibly 0-200 cal/(mol K). The heat capacity change in protein folding is closely related to changes in water accessible polar and nonpolar surface area resulting
from formation of a well packed hydrophobic core which is not accessible to water (Murphy and Freire 1992; Spolar et al. 1992; Makhatadze and Privalov 1993). These results are therefore consistent with previous observations, suggesting that the hydrophobic core of free $Z_{SPA-1}$ is not fully formed (II, III).

In a following publication (IV) it was investigated if the conformational entropy effect on binding affinity could be reduced by disulfide engineering of $Z_{SPA-1}$ and at the same time the thermodynamic basis for such a possible improvement.

9.2 Disulfide engineering (IV)

The NMR structure of the $Z:Z_{SPA-1}$ complex was used to select amino acids suitably for cysteine substitution in order to stabilize $Z_{SPA-1}$ in its bound and well defined conformation but without altering the binding surface. Six disulfide-bonded variants of $Z_{SPA-1}$ were designed. In each mutant, two amino acids were changed to cysteines in order to form a disulfide bridge.

We applied site-directed mutagenesis to introduce cysteines at specific positions into $Z_{SPA-1}$. Several studies demonstrate that disulfide bonds can be used to stabilize native structures of globular proteins (Wells and Powers 1986; Matsumura et al. 1989; Matsumura et al. 1989; Mansfeld et al. 1997; Starovasnik et al. 1997; Mårtensson et al. 2002). It is believed that a disulfide bridge decreases the conformational entropy of the unfolded state, and thereby increases the stability of the folded state (Schellman 1955; Flory 1956; Anfinsen and Scheraga 1975; Pace et al. 1988; Mansfeld et al. 1997; Abkevich and Shakhnovich 2000). But an enthalpic effect on stabilization has also been observed (Kuroki et al. 1992; Betz 1993; Vaz et al. 2006). Similar to this study, disulfide bridges have been used to improve binding by conformational stabilization (Reiter et al. 1996; Starovasnik et al. 1997). However, to our knowledge the thermodynamic basis for improvement of binding affinity through conformational stabilization has not previously been studied.

Five double-cysteine mutants of $Z_{SPA-1}$ were effectively produced intracellularly in *E.coli* and affinity-purified using protein-A sepharose. Mutant 6:A12C/S41C could not be purified in amounts detectable by SDS-PAGE. The reason could be that mutant 6 does not bind to protein A.
9.3 Biophysical properties of the cysteine-mutant 3:F5C/S39C

Gel filtration was used to separate the dominant monomeric fraction from oxidized dimeric and multimeric forms (data not shown). The elution profile shows an asymmetric peak with a smaller retention volume than expected, indicating self association as been described for Z_A. After gelfiltration only one band could be seen on a SDS-PAGE gel under both reducing and nonreducing conditions (data not shown).

Mass spectrometry was successfully used to verify disulphide bridge formation in monomeric form by measuring the difference in molecular weight between reduced and oxidized state, corresponding to 2 Da (data not shown). Unprocessed N-terminal methionine was present in 20-40 % of the proteins.

The helical content and the thermal stability were analyzed by CD for both oxidized and reduced protein. The α-helical content of mutant 3:F5C/S39C is similar to that of Z_A under oxidizing conditions at 10 °C. Under reducing conditions the helical content is lower, which is partly a consequence of significant thermal unfolding occurring at 10 °C (Fig 9.3)
Thermodynamic stabilization of the Z:Z$_{SPA-1}$ complex

Figure 9.3 CD spectra of Z and five Z cysteine mutants as indicated. A) Under air oxidizing conditions. B) Under reducing conditions with 1 mM TCEP added.

The melting curves were used to quantify the unfolding equilibrium (data shown in IV). The melting temperature and the unfolding enthalpy (van’t Hoff) in the transition midpoint $\Delta H_{\text{unfold}}^\circ (T_m)$ could be determined for the oxidized state but the unfolding heat capacity $\Delta C_p^\circ_{\text{unfold}}$ could not be determined due to thermal unfolding occurring already at 10 ºC. The oxidized mutant show an increase in $T_m$ of 12 ºC compared to Z$_{SPA-1}$, but it is still considerable lower than the Z-domain with a $T_m$~79 ºC (data shown in II).

$^{15}$N-HSQC NMR spectrum was used to monitor homogeneity and hydrogen bond stability of the mutants. The $^{15}$N-HSQC spectrum for the oxidized state of mutant 3:F5C/S39C is improved with more and sharper resonances compared to Z$_{SPA-1}$ suggesting a stabilization of the native state (molten globule state). The reduced protein shows increased line broadening and low dispersion indicating poorly packed or partly unfolded conformations (data shown in IV).
The binding thermodynamics were characterized under both oxidizing and reducing conditions by ITC and a full thermodynamic characterization of mutant 3:F5C/S39C was performed. Since it is not possible to separate the contribution coming from the binding equilibrium from the thermal unfolding-folding contribution in an ITC experiment, the calorimetric values of the binding enthalpy, \( \Delta H^0_{\text{ITC}} \), and the binding constant \( K_d^{\text{app}} \) was corrected as described above. This could not be done for the reduced condition because poor thermal stability prevented the calculation of fraction unfolded protein. The estimated correction for mutant 3:F5C/S39C under oxidized conditions is small. The binding affinity of mutant 3 was increased by more than a factor of ten (\( K_d = 130 \pm 20 \) nM) compared to \( Z_{\text{SPA-1}} \). Interestingly this increase correlated with an increase in binding enthalpy compared to \( Z_{\text{SPA-1}} \) and not to the binding entropy. Additional experiments at different temperatures between 10 and 30 °C were performed to obtain more information of the binding characteristics (see table 9.1). The temperature dependence of the uncorrected enthalpy is weakly nonlinear but become linear after correction for the temperature dependent folding-unfolding equilibrium. This is more pronounced for \( Z_{\text{SPA-1}} \) as expected, due to its lower thermal stability as mentioned above. The \( \Delta C^o_{p,\text{bind}} \) for mutant 3 (oxidized) is lower than for \( Z_{\text{SPA-1}} \) (-190 ± 30 cal/mol, K compared to -400 ± 30 cal/mol, K) (Fig 9.4.). The difference can be related to the change in exposed surface area for the two proteins upon binding. The lower \( \Delta C^o_{p,\text{bind}} \) of mutant 3:F5C/S39C reflects a smaller change and consequently a better packed protein before binding, which is consistent with other results mentioned above.

9.4 Results of all cysteine mutants

The effect of the other cysteine mutations varies but none can compare with mutant 3 in terms of binding affinity (table 9.1). The introduction of cysteine bridges were primarily chosen to covalently stabilize helices 1 and 2 since they represent the binding surface. One cysteine bridge was introduced, in mutant 5:N21C/A56C, which connects helix 2 to helix 3. This is the most thermostable mutant with a melting temperature above 70 °C but this does not lead to improvement of the affinity. Comparing all mutants no correlation could be seen between thermo stability and binding affinity. This is consistent with earlier results (Dincbas-Renqvist et al. 2004) discussed above, suggesting that the effect of thermal unfolding have minor effect on the binding thermodynamics. The HSQC spectrum of mutant 5 is improved compared to \( Z_{\text{SPA-1}} \) suggesting a stabilization of the conformation(data shown in IV). The cysteine bridge in mutant
Thermodynamic stabilization of the Z:ZSPA-1 complex

4:F5C/S41C probably introduces some conformational strain into the structure as seen by a better HSQC (data shown in IV) and higher binding affinity to Z under reducing compared to oxidizing conditions (table 9.1).

**Figure 9.4** Corrected binding enthalpies for ZSPA-1 (open circle) ([Dinebas-Renqvist et al. 2004]) and mutant 3:F5C/S39C (oxidized) (solid circle) as a function of temperature. The lines indicate linear fits to determine $\Delta C_p^{\text{bind}} = -400 \pm 30$ cal/(mol, K) for ZSPA-1 and $\Delta C_p^{\text{bind}} = -190 \pm 30$ cal/(mol, K) for 3:F5C/S39C.

**Figure 9.5** Correlations between $\Delta G^{\text{bind}}$ and $\Delta H^{\text{bind}}$ for ZSPA-1 and five ZSPA-1 mutants binding to the Z domain. Symbols represent ZSPA-1 (star), oxidized mutants (solid circle) and reduced mutant 1 and 4 (open circle).

**Figure 9.6** Enthalpy-entropy compensation of $\Delta H^{\text{bind}}$ versus $\Delta S^{\text{bind}}$ for ZSPA-1 and five ZSPA-1 mutants binding to Z domain. The slope of the solid line is $380 \pm 18$ K. All data can be found in table 9.1.
The stabilization of the cysteine mutants were expected to be entropically explained, even if stabilization effects with an enthalpic origin also has been reported, as discussed above. A purely entropic effect should be possible to observe by an unchanged or reduced value of the unfolding enthalpy. And actually, the van’t Hoff enthalpies for mutant 1 and 4 are smaller for the oxidized than for the reduced state in line with an entropic stabilization but the difficulty to determine \( \Delta C^\circ_{\text{p,unfold}} \) and the high uncertainties in the \( \Delta H^\circ_{\text{unfold}} \) values hinders a reliable conclusion. The analysis of binding thermodynamics for all mutants including the reduced state of mutant 1 and 4, for which the binding enthalpy could be determined, and \( Z_{\text{SPA-1}} \), shows that the improvements in binding affinity is correlated with an increase in binding enthalpy (Fig 9.5). We also see strong enthalpy-entropy compensation (Fig 9.6) (Dunitz 1995). Consequently, the data imply that the improvement in binding affinity is the result of an enthalpic rather than an entropic contribution.

These results might be explained in the following way. The optimal binding interface, where high electrostatic and surface complementarity as well as strong hydrogen bonds can be formed, is likely to be similar for all mutants because the same residues contribute to binding. What is changed is the ability to support this optimal interface conformation. Hence, for the best mutant (mutant 3:F5C/S39C) a larger population of the proteins is in a conformation favorable for binding. However, to maintain this well-defined conformation in the complex is entropically unfavorable. To investigate the importance of different contributors to the binding entropy a deconvolution of the dominating terms were performed (Baldwin 1986; Murphy et al. 1995).

\[
\Delta S^\circ_{\text{bind}}(T) = \Delta S^\circ_{\text{solv}}(T) + \Delta S^\circ_{\text{conf}} + \Delta S^\circ_{\text{rt}} \\
\Delta C^\circ_{\text{p,bind}}(T) \ln \left( \frac{T}{385} \right) + \Delta S^\circ_{\text{conf}} + \Delta S^\circ_{\text{rt}}
\]

Here the calculation of the desolvation term is based on the assumption that the entropies of desolvation of polar and nonpolar amino acids are zero at 385 K (Baker and Murphy 1998). The contribution from the rotational and translational entropy (\( \Delta S^\circ_{\text{rt}} \)) is uncertain but values of -8 to -24 were used to cover a large range (Amzel 1997; Yu et al. 1999). The results suggests a decrease in the conformational binding entropy (\( \Delta S^\circ_{\text{conf}} \)) as was expected, but it also suggest an equal decrease in desolvation entropy (\( \Delta S^\circ_{\text{solv}} \)) resulting in a netto cancellation of the two terms. The decrease in \( \Delta S^\circ_{\text{solv}} \) is most likely the result of the more well-packed conformation of the unbound form of mutant 3 in its oxidized state.
Table 9.1 Thermodynamics for binding of Z$_{SPA-1}$ and five Z$_{SPA-1}$ mutants to the Z domain. Data obtained at 21°C. a Calculated as $\Delta C^\circ_{\text{P,bind}} \times \ln(T/385)$. b From eq shown above c Alternative values represent uncertainties in $\Delta S^\circ_{\text{rt}}$ as discussed in the text, d Data from Dincbas-Renqvist et al. 2004, (ox) oxidized state and (red) reduced state. Table modified from table 3 in work (IV)
Conclusion and outlook

The work presented in this thesis consists of a structural analysis of the ZSPA-1 affibody:Z domain complex, followed by a thorough biophysical characterization of the free affibody. Mutants were designed based on these results and their effect on binding was verified by a thermodynamical analysis.

The structure of the affibody-target protein complex suggests the hydrophobic effect to be the major driving force for complex formation, while electrostatic interactions form a rim around the core and provides several specific contacts. The Z domain binds to the affibody molecule by using the same surface as in the complex with Fc and in a similar conformation, while the affibody molecule uses the randomized surface for interaction. This was not surprising due to competitive elution with Fc during selection of this affibody binder and the fact that other binding sites would increase the risk for self recognition due to the high sequence identity between the two proteins.

The hinge region of the Fc fragment interacts with four known naturally occurring proteins by using a common binding site, but with different conformational rearrangements of the interface to accommodate to the binding surface of each protein (DeLano et. al. 2000). This binding site was also the main target observed for random peptide libraries, excluding functional reasons for targeting the same site (DeLano et. al. 2000). Thus, it appears that this surface possesses an intrinsic property to bind other proteins and peptides. A comparison of Fc-complex crystal structures showed that the convergent binding site is highly accessible, adaptive and hydrophobic. The ZSPA-1 affibody has all those characteristics and it is fascinating to see how well the ZSPA-1 affibody is able to mimic the surface of the Fc receptor inducing similar rearrangements of Z domain side chains as in the Z:Fc complex.

Complex formation involves an induced fit of the Z-domain and coupled folding of the ZSPA-1 affibody molecule. Many intrinsically unstructured proteins show coupled folding and binding. This is believed to have an important regulatory function for some transient complexes, uncoupling the affinity from the specificity by a large conformational entropy penalty (Tompa 2002). The same effect can be seen for ZSPA-1 even though no regulatory reason is involved. The affinity of Z can be increased by introducing stabilizing mutations in the affibody scaffold. Other affibody molecules with significantly higher affinities have been selected towards the Z domain.
Conclusion and outlook

It would be interesting to study these selected affibody molecules to understand the difference in complex formation and binding.

The $Z_{SPA-1}$ affibody was selected by randomization of 13 out of 58 amino acids, which in this case had an unfavourable affect on the stability. Possible explanations for this effect are an increased hydrophobicity, the introduction of unfavourable charge interactions on the surface and a reduced helix propensity. Even though considerable progress has been made in understanding structure-stability relationships in proteins, the goal of a theoretical model that makes quantitative predictions of the energetics based on structure alone is still not achieved. Further research must therefore aim to provide experimental data on this matter so that the procedures to select stable high-affinity binding proteins can be improved in a rational way.
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANS</td>
<td>8-anilino-1-naphtalenesulfonic acid</td>
</tr>
<tr>
<td>BSE</td>
<td>Bovine spongiform encephalitis</td>
</tr>
<tr>
<td>CBD</td>
<td>Cellulose-binding domains</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>Fab</td>
<td>Fragment, antigen binding</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescent activated cell sorting</td>
</tr>
<tr>
<td>Fc</td>
<td>Fragment, crystallisable</td>
</tr>
<tr>
<td>GuHCl</td>
<td>Guanidine hydrochloride</td>
</tr>
<tr>
<td>hGH</td>
<td>Human growth hormone</td>
</tr>
<tr>
<td>hGHbp</td>
<td>Human growth hormone binding protein</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum correlation spectroscopy</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin(s)</td>
</tr>
<tr>
<td>IL-2α</td>
<td>Interleukin 2 alfa</td>
</tr>
<tr>
<td>IL-2Rα</td>
<td>Interleukin 2 alfa receptor</td>
</tr>
<tr>
<td>IUPs</td>
<td>Intrinsically unstructured proteins</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>K&lt;sub&gt;a&lt;/sub&gt;</td>
<td>Association constant</td>
</tr>
<tr>
<td>K&lt;sub&gt;d&lt;/sub&gt;</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser effect</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>SPA</td>
<td>Staphylococcal protein A</td>
</tr>
<tr>
<td>Taq</td>
<td><em>Thermus aquaticus</em></td>
</tr>
<tr>
<td>TMAO</td>
<td>Trimethylamine-N-oxide</td>
</tr>
<tr>
<td>Z&lt;sub&gt;SPA-1&lt;/sub&gt;</td>
<td>SPA binding affibody</td>
</tr>
</tbody>
</table>
12 Acknowledgements

I would like to acknowledge Torleif for being my supervisor, for given me the opportunity to be a part of the affibody project and for your excellent scientific guidance, Helena for all your support and generosity, I would never had made this without you and Per-Åke for your contribution both at the start and at the end of this work. Thank you also for valuable scientific support, for showing interest and for many ingenious questions.

I would also like to thank the members of the affibody team Christofer, Vildan, Jakob and Alex for fruitful collaboration and your invaluable contribution to this work.

To Esmeralda for being my roommate in Tällberg, and to the other people that used to belong to the NMR group for good times: Susanne, Inger, Niklas, Martin; Per-Åke, Magnus, Peter, Anders H, Henrik, Mange, Anja, Alexandra, Mange Anders Ö and John.

Big thanks to Johan for help with references and pictures and to Alex for valuable advices when writing the thesis.

Also big thanks to Lotta and the other people at floor 2 for letting us move in and making me feel welcomed.

Invaluable for this work has also been the help with articles provided by Cecilia at the library at Albanova, thank you.

I would also like to send my biggest gratitude’s to my parents Siv and Folke and my parents in law Signe and Sven for all help, especially with all baby sitting.

To Maria for being a good friend and to Rikard for your coaching when writing my thesis.

To the rest of my big family Per, Karin, Louise, Christian, Jonathan, Anna, Anna, Freja and Inge.

To Lisa for also going for a Ph.D.

Och till sist men allra mest till solstrålarna i mitt liv, Johan, Malin, Amanda och Gustav.
13 References


References


