QUANTUM CHEMICAL MODELING OF THE UNFOLDING OF β-SHEETS AIDED BY FIRST LAYER WATER MOLECULES

DÓRA PAPP

MASTER PROGRAM IN CHEMISTRY

MASTER'S THESIS

2013

SUPERVISORS:

PROF. ANDRÁS PERCZEL, DSC

TAMÁS BEKE-SOMFAI, PHD

EÖTVÖS LORÁND UNIVERSITY, FACULTY OF NATURAL SCIENCES

INSTITUTE OF CHEMISTRY, DEPARTMENT OF ORGANIC CHEMISTRY

CHALMERS UNIVERSITY OF TECHNOLOGY

DEPARTMENT OF CHEMICAL AND BIOLOGICAL ENGINEERING

DIVISION OF PHYSICAL CHEMISTRY

"Once I hoped

to seek the new and unknown"

/Muse – Explorers/

Statement

Name: Dóra Papp

Faculty: Faculty of Natural Sciences, Master Program in Chemistry

Neptun code: PLR45V

Title of the thesis: Quantum chemical modeling of the unfolding of β-sheets aided by first layer water molecules

As the author of the present thesis I declare under penalty of perjury that this thesis is based on my own work and it is my own intellectual product, therein the standard conventions of reference citation have been applied, it contains no parts written by others without proper citation.

Göteborg, 20/05/2013

Dóra Papp

Signature

Acknowledgement

I wish to thank, first and foremost, my supervisor Prof. Dr. András Perczel for suggesting this problem to deal with it, and for guiding my work with so much patience and attention. I would also like to thank him for his advices to help me find my way.

I am very thankful to my co-supervisor Dr. Tamás Beke-Somfai for his substantial and permanent help, and for his true optimism during my work. I am grateful that I could always count on him while staying in Sweden, and for the discussions on how to get on in scientific life.

I am indebted to Prof. Dr. Bengt Nordén for providing the opportunity to work in his group at Chalmers University of Technology in Göteborg.

My special thanks are extended to the members of MTA-ELTE Protein Modeling Group and Laboratory of Structural Chemistry and Biology, and the Division of Physical Chemistry at Chalmers.

I would like to thank my Friends, especially Bertalan Kovács for revising my thesis and for our memorable conversations; Fanni Bazsó for supporting me no matter what the situation is; Eszter Kárpáti, Zsófia Hargitai and Erika Dudás for being *four*; and Tamás Földes and János Sarka for being there when needed.

Thanks to Matt, Dom and Chris for flashing things so clearly and uniquely, and for everyday inspiration.

Last but not least, I would like to thank my Parents for always trusting me and supporting me, sometimes even without knowing the direction I am going toward; and for their selfless love.

Table of contents

1. Introduction

Nowadays, due to the recent advances in computer technology, quantum chemical methods become generally applied tools in chemical research. These methods have already been used successfully in designing drug molecules, finding alternative reaction pathways, or in spectroscopy for elucidating various spectra, as well as in studying catalysts of organometallic chemistry. It is important to point out, since chemistry remained basically an experimental science, that these methods can only be used in accordance with results based on experiments. With this in mind, quantum chemical modeling can be highly useful in supplementing, explaining and predicting such results.

This branch of chemistry has proven to be very successful in investigating biomolecules and biochemical processes as well. For instance, it is a commonly used tool in understanding the basics of protein folding, or in the case of this particular thesis, those of the protein unfolding process. Several proteins, especially those having extreme mechanical properties, have recently been subject of experiments performed by single molecule force spectroscopy (SMFS), which gave insight into the unfolding of individual molecules. The external pulling force applied through the cantilever of an atomic force microscope was recorded as a function of the extension, and in case of such proteins a typical sawtooth-like discontinuity appeared in these curves. Most of mechanically stable proteins feature β-sheet structural elements, the rupture of which was thought to be responsible for this characteristic curve shape. Despite numerous experimental and theoretical studies, the atomic level mechanism was still not fully understood until Beke-Somfai and Perczel have proposed a mechanism for the unfolding of small β-sheet models based on quantum chemical calculations, taking into account the role of first layer water molecules[1]. In order to gain a deeper understanding of the energetics and the subatomic-level background of this process, we further investigated the underlying basic steps and fundamental interactions. Since the changes in the hydrogen-bonding network proved to have the most significant effect on energetics, the detailed analysis of their formation and breakup will be presented here.

2. Literature review

2.1. Importance of β-sheet structures

Proteins, their structure and functions, as well as the mechanism of protein folding have been studied over the recent decades and the interest in this area continues to grow[2,3,4.] Since the energy landscape of the protein folding is yet unknown, both experimental[5,6,7,8,9] and theoretical[10,11,12,13,14] investigations are going on to explore that mysterious process. Proteins usually adopt their folded structure in nano-, or microseconds[15,16,17], depending on the size of the given protein[18], clearly without trying all possible pathways of folding[19,20]. To discover the mechanism of finding the proper path to fold, the investigation of the unfolding process is very expedient and proved to be useful[21,22].

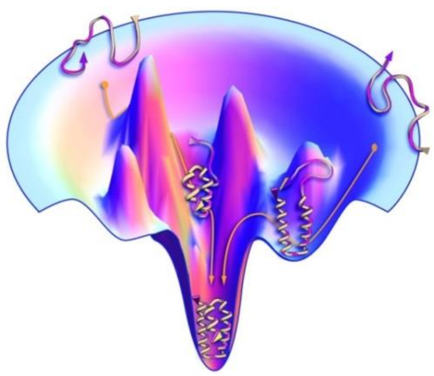

Figure 1 - Proteins have a funnel-shaped energy landscape with many high energy, unfolded structures and only a few low-energy, folded structures. Folding occurs via alternative microscopic trajectories[3].

As it has long been known, the 3D protein scaffold has a four-layer construction: the primary structure is the sequence of amino acid residues of the protein; the secondary structure is defined by the periodic arrangement of the backbone dihedral angles[23]. The spatial arrangement of the secondary structural elements with respect to each other defines the tertiary structure, that is, the native conformation mainly stabilized by weak interactions such as van der Waals forces. The quaternary structure describes the way how the different peptide chains are positioned compared to each other; it represents the entire conformation of the protein.

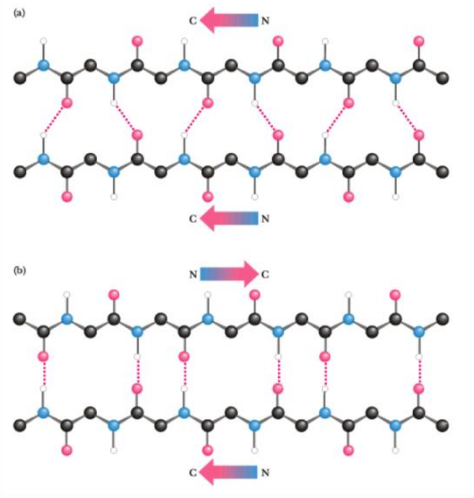

Figure 2 - Structure of β-sheets: (a): parallel - the orientation (arrows from N to C terminus) is the same in case of the two peptide strands, (b): antiparallel sheet – the orientation is opposite. Their H-bond network is therefore different: the parallel one has 12-membered so-called H-bonded pseudorings, whereas the antiparallel one has 14- and 10-membered ones, alternately[24].

Among secondary structural elements β-sheet was shown by theoretical calculations to be the most stable[25]. This stability is very important in the processes occurring in our cells, for instance it has a substantial role in protein aggregation[26,27], which role has been investigated in detail recently. Several neurodegenerative diseases[28], such as Alzheimer's disease, are caused by the formation of protein aggregates, so-called plaques consisting of amyloid fibrils. The structure of these fibrils was studied both by experiments[29,30] and theoretical investigations[25,31], and proved to be built up of mainly β-sheet elements[32,33].

β-sheets seem to be a typical element of silk proteins as well, which show extreme mechanical properties, such as the strength of silk materials exceeds that of steel[34,35]. Among other labs, the structural properties of silk-like models have been theoretically studied also in our group[36]. β-sheet is also a common motif in titin, a muscle protein with outstanding mechanical feature[37].

Another field where β-sheets are important is the structure and function of transmembrane proteins, for instance in β-barrel proteins[38]. β-barrels and β-propellers are also known as structural elements that try to protect proteins from aggregation through hiding the free β-edges by forming such motifs[39].

2.2. Experiments investigating the mechanical unfolding of β-sheets

Regarding the above outlined significance of β-sheets, their stability and structural properties have widely been investigated and still meet a broadening range of scientific interest. In the late 1990's, by when experimental techniques had been improved greatly, the first experiments were performed exploring the mechanical unfolding of proteins and other macromolecules.

By that time, single-molecule force microscopy using AFM (atomic force microscopy)[40,41] techniques had become a very innovative, unique and useful tool by means of which the investigation of atomic level structural properties of molecules turned into a feasible option. These kinds of experiments also gave insight into the energy landscape related to the unfolding and folding mechanism of β-sheet rich proteins.[42,43,44] In these experiments the mechanical properties of individual molecules is probed by exerting mechanical forces that induce conformational changes in proteins. The external pulling force applied through the cantilever of an AFM plays the role of a chemical/thermal denaturant, but gives insight into the unfolding experiment at a submolecular level. However, not only the AFM was used to perform such investigations, but optical and magnetic tweezers as well[45].

Titin has become an expansively studied molecule by methods inducing forced unfolding[46,47]. Two pioneer studies of the single molecule force spectroscopy (SMFS) experiments related to titin was published in 1997 by Rief et. al[42] and by Kellermayer et al.[48] (in the same issue of *Science*). They repeatedly stretched individual titin molecules, in the former case by means of a cantilever of an AFM, whereas in the latter case with laser tweezers, and recorded the force-extension curves.

Titin is a giant protein, spans the half-sarcomere and is an important example of proteins whose mechanical properties are essential for their biological function[49,50]. The passive tension developed by muscle sarcomeres when being stretched is largely due to the rubberlike properties of titin, which is also known as connectin[51]. The mechanically active region of titin in the sarcomeric I band[52] is assembled from immunoglobulin (Ig)-like domains arranged in tandem (tandem Ig) and a small part of nonmodular sequences rich in proline, glutamate, valine, and lysine (the PEVK region)[53]. It is suggested that the tandem Ig chain is an extensible chain that resists stretching at longer sarcomere lengths, whereas the PEVK region is increasingly extended under stronger forces[54,55]. The immunoglobulin domain consists of a 2-layer sandwich between antiparallel β-strands arranged in two β-sheets (**Figure 3**)[56,57].

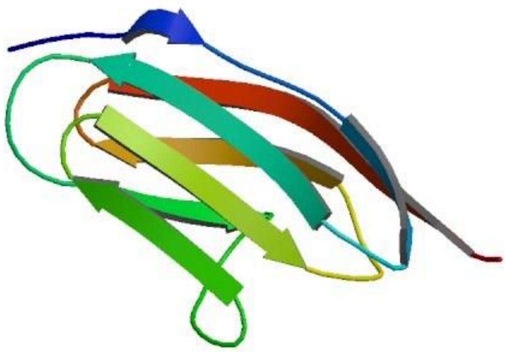

Figure 3 - Immunoglobulin-like module from titin I-band (in ribbon model representation)[58,59]

Rief et al.[43] adsorbed native titin molecules onto a gold surface from phosphate-buffer solution, then an AFM tip (Si_3N_4) was contacted to the surface to make a fraction of one of the proteins adsorb onto the tip. Then, extension curves (applied force vs. protein length extension) were recorded, and a sawtooth-like discontinuity was observed at larger extensions in the force curves. These sawtooth patterns exhibited periodic motifs: force-maxima appeared at extensions with differences between 25-28 nm. The maximum forces varied between 150 to 300 pN. The difference between a fully extended and a native immunoglobulin module is expected to be 31 nm, which was consistent to the periodicity observed (the Ig domains were probably not fully extended), so this indicated that the unraveling of a single titin molecule was measured, by unfolding of each domains. To prove this, they constructed titin fragments containing four and eight Ig-domains (Ig4 and Ig8, respectively). Then the SMFS experiment were repeated using these constructions, and extension curves were obtained that were superimposable and showed a similar sawtooth patterns as native titin, with the same periodical features.

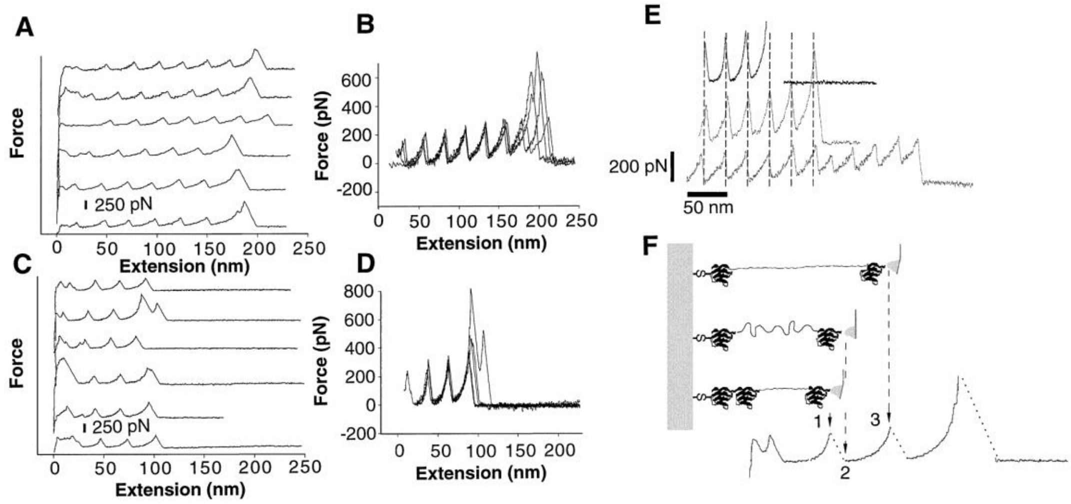

Figure 4 – Extension curves recorded during the forced unfolding of titin molecules by AFM. **A** and **B** show the curves of stretching Ig8 domains (**B**: superimposition). **C** and **D** show the curves of forced unfolding if Ig4 domains (**D**: superimposition). **E**: Force extension curves obtained from all three titin forms (from top to bottom: Ig4, Ig8, and native).**F**: A possible sequence of events during unfolding. (1) Before a domain unfolds, the extended polypeptide will be stretched until a holding force of 150 to 300 pN is reached and unfolding becomes highly probable. (2) Unfolding of an Ig domain abruptly reduces the holding force because of an increase in the length of the extended polypeptide. (3) Continued retraction of the AFM tip again stretches the extended polypeptide until a force is reached where the next Ig domain unfolds.[43]

They explained their results by modeling the unraveling process assuming that the extended protein behaves like a wormlike chain (WLC model)[60]. Using the equation

$$F(x) = \frac{kT}{b}\left(\frac{1}{4}\left(1 - \frac{x}{L}\right)\right)^{-2} - \frac{1}{4} + xL \qquad (1)$$

(where b denotes the persistence and L is the contour length, k stands for the Boltzmann's constant, x for the extension, and T is the absolute temperature). The experimentally obtained data and the computed ones appeared to be in a good agreement, as it can be seen in **Figure 5**.

From the observation that the peak maximum increases with increasing extension and due to the fact already shown that every peak reflects the unfolding of different Ig-domains, the conclusion arose that the domains unfold in a way that the weakest domain unfolds first and the strongest last.

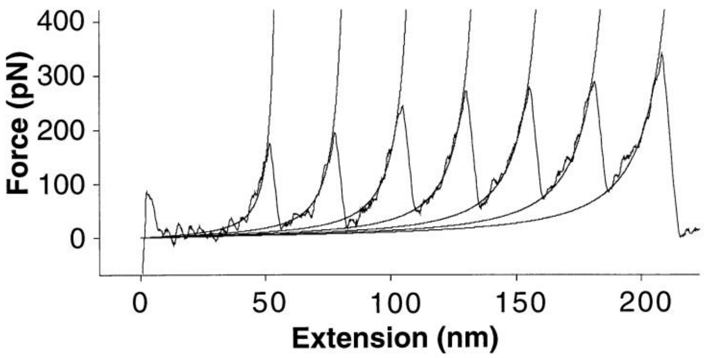

Figure 5 - The characteristic sawtooth pattern of unfolding explained as stepwise increases in the contour length of a polymer whose elastic properties are described by the wormlike chain model. The figure shows a force - extension curve obtained by stretching of a single Ig8 titin fragment modeled by the above equation (fitted curves).[43]

Kellermayer et al.[48] unfolded titin molecules by using laser tweezers technique. They stretched titin by attaching its ends to different latex beads; one was held by a movable micropipette and the other was trapped in force-measuring laser tweezers. Then the micropipette was moved with a constant rate, and the force generated in the molecule was monitored. When a maximum force was reached, the process was reversed.

To elucidate the results, they also used the WLC model proved to be successfully describing the force-extension curves obtained here as well. They recorded cycles (**Figure 6**) of unfolding and refolding of titin molecules and proposed a mechanism for such a process. The forced unfolding occurs in the following way according to their assumptions: 1) at the beginning a variable fraction of the molecule is already unfolded and the molecule behaves like a wormlike chain whose properties are dominated by the pre-unfolded fraction (WLC region). 2) As titin is stretched to high force (20-30 pN) transitions begin to occur inside the molecule. As it is further stretched, the mechanical denaturation of certain domains takes place; probably not every domain unfolds, when the maximum force (which was an instrumental limit) is reached. 3) Upon release of the unfolded titin, the molecule does not start to refold immediately but behaves instead as a WLC, with the properties of an unfolded polypeptide. 4) When the molecule is allowed to shorten to one-half of its release contour length, the refolding starts, which is also seen as a transition region in the extension curves. Here takes place the refolding of the Ig- and other domains as well. Then the contour length shortens significantly, the force drops, and a WLC-like behavior can be observed again.

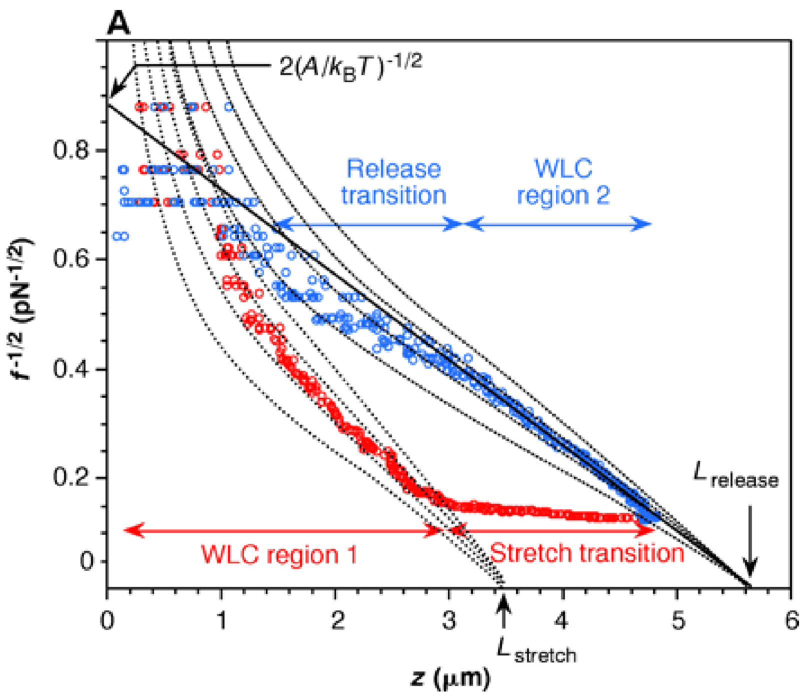

Figure 7 - The $f^{-1/2}$ vs. z (extension) plots for the stretch (red) and release (blue) of a triple titin tether at a rate of 64 nm/s. A linear fit to the force data at high extension is extrapolated to the z axis to obtain the contour length (L) and to the force to obtain the effective persistence length (A). Dotted lines from left to right are WLC curves (using the above equation[48]

It has been shown that the unfolding of titin occurs at a lower rate under mechanical force than in the presence of chemical denaturants, which indicates that during the unfolding titin must pass through some folding intermediates with high activation energies and pulling the molecule does not lower that barrier significantly. These intermediate states might be only accessible to force perpendicular to the pulling direction.

Comparing the behavior of several titin molecules under forced unfolding with the calculated curves of real muscle-response, the similarity indicated that titin is mainly responsible for the passive force response of muscle, and this process could be characterized by the unfolding-refolding kinetics of independent titin molecules.

Rief et al. later compared the mechanical stability of different types of Ig domains, constructed based on native titin structures from three different parts of the molecule, to each other[61]. Their results showed that domains involved in extensive protein-protein interactions, like those of the myosin-associated A-band region of titin, unfold under the lowest forces, possibly reflecting their stabilization by the binding forces of their respective interactions.

Mechanical stability of proteins was not only the subject of single molecule force spectroscopy but also was one of the achievements of it[62]. With SFMS the mechanical properties of a *de novo* designed protein were able to be tuned. It has already been known that the shear topology of two directly connected β-strands is a common feature of proteins having outstanding mechanical stability, but Sharma et al. have succeeded to create a new type of β-sheet topology with notable mechanical resistance. The structure contains non–connected β-strands but shows the mechanicals properties required, demonstrating that the connectivity of force-bearing β-strands is not mandatory for the extreme stability properties.

One of the most important of β-structures is the β-barrel that could not be out of SMFS experiments. Dietz and Rief showed that the mechanical unfolding of green fluorescent protein (GFP) proceeds by means of two subsequent intermediate states[63]. One is characterized by detachment of a seven-residue N-terminal α-helix from the β-barrel and though the detachment of this small helix completely destabilizes GFP, the β-barrel still remains intact. Mechanical stability of the GFP, however, was determined by the activation barrier of unfolding the β-barrel. Their AFM results revealed a second metastable mechanical intermediate with one complete β-strand detached from the barrel.

The rough energy landscape they reconstructed from the data obtained showed deep local minima reflecting largely structured, but metastable intermediates during the unfolding.

Sapra et al. have also studied the mechanical stability of a β-barrel, a transmembrane protein (OmpG) by pulling it also with AFM[64]. Determination of the unfolded protein stretches allowed assignment of the structural segment that unfolded, and it turned out that each of them consisted of two β-strands forming one β-hairpin, and each β-hairpin of the OmpG β-barrel was able to unfold individually or cooperatively with an adjacent β-hairpin.

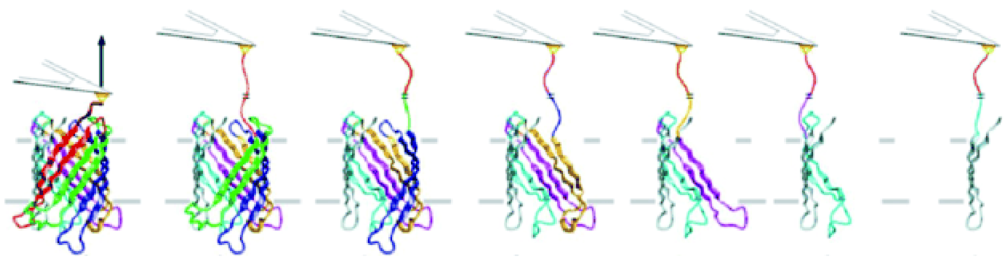

Figure 8 - The stepwise unfolding of OmpG β-barrel monitored by AFM (the cantilever of AFM is schemed in the top of the pictures)[64]

Their results showed that unfolding of single β-strands of OmpG requires forces of approximately 150–250 pN, and these forces, reflecting the interaction strengths stabilizing the β-strands, are much higher than those required to unfold the entire β-barrel protein GFP[63] at similar conditions. This indicated that the stepwise unfolding behavior of a transmembrane β-barrel protein is very different from the force-induced unfolding of the water-soluble β-barrel protein GFP. They suggested that when water-soluble proteins are unfolded, the exposure of their hydrophobic core to the hydrophilic aqueous solution is one of the driving forces leading to destabilization. Therefore, taking the effect of the environment's polarity into account is indispensable. The "natural" unfolding of β-barrel proteins by cellular machineries consuming ATP has also been studied recently[65].

Due to the importance of amyloid structures (about protein aggregation and related diseases) proved to be consisting mainly of β-sheet regions, their structure and mechanical features have also been studied by SMFS[66,67]. Kellermayer et al. found that β-sheets forming amyloid fibrils behave as elastic structures that can be "unzipped" from the fibril with constant forces, and the unzipping is fully reversible across a wide range of stretch rates suggesting that coupling, via the β-sheet, between bound and dissociated states is maintained[66]. The reversible rebinding of β-sheets at high loads and loading rates to the underlying fibril surface indicated that the associated state is strongly favored and a mechanically perturbed amyloid fibril is rapidly recovers by zipping together the β-sheets. Since the force-extension curves were different in cases of different amyloid peptides (probably because of dissimilarities in the arrangement and interactions of their β-sheets), they proposed that the repetitive force patterns could be utilized in the characterization of various amyloid fibrils.

Not only the unfolding characteristics of proteins were subjects of such examinations, but folding pathways have been investigated as well. Fernandez and Li used force-clamp atomic force microscopy to measure the end-to-end length of the small protein ubiquitin during its folding at the single-molecule level[68]. Ubiquitin was first unfolded and extended at a high force, then the stretching force was quenched and protein folding was observed. The folding trajectories obtained were continuous marked by several stages, which contradicts the generally held view that folding reactions correspond to transitions between well-defined discrete states.

The effect of the solvent on the unfolding process has also been investigated by Dougan et al. performing solvent substitution during SMFS experiments[69]. They characterized the unfolding transition state by two parameters: its activation energy, ΔG_u, and the elongation of the protein necessary to reach the transition state, Δx_u. They measured Δx_u for a human cardiac titin domain in the presence of water, glycerol and deuterium oxide. In aqueous media these values of Δx_u were comparable to the size of a water molecule, suggesting that water molecules are integral components of the unfolding transition state. They found that six to eight water molecules form part of the unfolding transition state structure, and that the presence of just one glycerol molecule in the transition state is enough to lengthen Δx_u. Changing the solution to D_2O the ΔG_u increased a bit (stronger hydrogen bonds were formed), but the solvent-exchange had no effect on the Δx_u value, it remained unchanged with respect to the one measured in water.

2.3 Theoretical modeling of the unfolding of β-sheet rich proteins

Mostly in combination with experimental data, or based on those, molecular dynamic (MD) simulations can give an atomic level insight to the unfolding (or folding) mechanisms of proteins that had been subjects of single molecule force spectroscopy.

Titin, probably the most widely studied protein by SMFS, was investigated in detail by molecular dynamic simulations as well. Marszalek et al. combined AFM measurements with MD calculations, and revealed the importance of terminal hydrogen bonds between two β-strands in the initiation of the unfolding process[70]. An abrupt extension observed during AFM experiments of each domain before the first unfolding event could be explained by the results of simulations showing that the rupture of a pair of hydrogen bonds near the amino terminus of the protein domain causes a significant extension. This was labeled a new intermediate that corresponds to an extension of a titin module by 15% of its resting length during unfolding. They proposed that this phenomenon is likely to be an important component of titin elasticity.

Fowler et al. have tried to further identify intermediates during unfolding of an immunoglobulin domain of titin, also combining experimental methods with MD modeling[71]. Their results indicated that the unfolding forces measured in an AFM experiment are required for the unfolding of the intermediate and not that of the native state. They suggested that this intermediate has a structure very similar to the native state; it remains stable under forced unfolding, and yet allows a significant lengthening of the titin molecule to be achieved, if the same unfolding mechanism applies to all Ig domains in it.

Esposito et al. have modeled, also by performing MD simulations, the cross-β-spine structure characteristic to the amyloid steric-zipper sequence, that has a high affinity for aggregation and therefore playing an important role in causing neurodegenerative diseases[32,72]. They suggested that steric zipper interactions at a β-sheet–β-sheet interface strongly contributed to the stability of these aggregates. Backbone structure showed significant variations during the simulations; and a significant twist of the β-strands is observed, which was in good agreement with the X-ray structure of amyloids determined earlier. Two pairs of sheets, while twisting, associated through stable peptide–peptide interactions. They found that even models composed by a limited number of strands could be rather stable.

Ackbarow et al. have investigated how individual protein building blocks respond to mechanical load, also using MD methods[73]. They studied the unfolding behavior of α-helical and β-sheet containing domains, and revealed that unfolding mechanism at fast pulling rates is sequential rupture of individual hydrogen bonds, whereas at slow pulling rates proceeds by simultaneous rupture of several H-bonds, in case of both secondary structural elements. This mechanism had previously been proposed by Bryant et al., who performed MD simulations on β-hairpin unfolding, and found that under conditions that generate low forces, the unfolding trajectory mimics an untethered, thermally accessible pathway, in which complete breakdown of backbone hydrogen bonds preceded dissociation of the hydrophobic cluster, whereas under more extreme conditions, the cluster and hydrogen bonds broke simultaneously[74].

It has been shown earlier that the α-helix to β-sheet transition is a universal deformation mechanism in α-helix rich protein under changes in cellular environment or mechanical load[75,76]. To model this phenomenon, Qin and Buehler have examined the behavior of α-helical coiled-coil proteins under stretch[77]. They found that the occurrence of the α-β transition is controlled by the length of constituting α-helical proteins. Proteins with greater lengths feature α-β transition leading to a significant increase in the protein's stiffness, strength, and energy dissipation capacity. Later, Zhmurov et al. characterized the α-β transition in human fibrinogen molecule using MD simulations of their forced elongation[78]. These computed force–extension profiles showed three distinct regimes: (1) the elastic regime, in which the coiled-coils act as entropic springs, (2) the constant-force plastic regime, characterized by a force-plateau, and (3) the nonlinear regime. In the plastic regime, the three-stranded α-helices undergo a noncooperative phase transition to form parallel three-stranded β-sheets. They proposed that this transition in coiled-coils might be a universal mechanism underlying mechanical properties of filamentous α-helical proteins.

The effect of the direction of the applied pulling forces has also been investigated, and it turned out that the energy landscape for mechanical unfolding is markedly anisotropic and the orientation of secondary structural elements relative to the pulling force vector has an important function[79].

Despite the several experimental and theoretical studies detailed above, the underlying atomic level mechanism and the energetic basis of the unfolding was still not fully understood. While, molecular dynamic simulations successfully described the unfolding process and identified several intermediates for various proteins, force fields have limited accuracy on hydrogen bonded system[80].

A quantum mechanics-based study of the forced unfolding of β-sheet models was performed by Beke-Somfai and Perczel[1]. Since the hyperfine resolution of the experimentally obtained sawtooth patterns is most likely related to the break-up of β-sheets (as it has been discussed in the above sections), more specifically the interstrand H-bonds conserved between the β-strands, they investigated the forced unfolding (stepwise separation) of two β-strands of trialanine residues by performing quantum chemical calculations scanning the potential energy (hyper)surface of the unfolding process. They also studied the unfolding aided by 1-3 explicit water molecules, due to the previously shown important role of solvent molecules in the stabilization of the transition states of mechanical unfolding[69].

They showed that the presence of water molecule facilitates the unfolding by reducing the activation energy barrier of the process by forming hydrogen-bonds with the peptide strands. Thus, the formation of these water-peptide H-bonds was responsible for the drops in the energy curves. Their results also indicated that there is a qualitatively different mechanism for parallel and antiparallel orientation in the initial phase of water-assisted forced sheet unfolding. In the case of parallel β-sheets, the presence of only a single water molecule could already be enough to stimulate rupture of consecutive backbone H-bonds along a β-sheet by stepping from one H-bond to the next one, similarly as a slider opens up a zipper. In contrast, the antiparallel strands were mechanically more stable during the unfolding, which was in accordance with their previously shown higher thermodynamic stability and natural abundance[25,81], only the models having three water molecules featured the zipper-like mechanism during unfolding.

2.4. Modeling of hydrogen-bonding in proteins

Modeling of non-covalent interactions in biomolecules has long been an object of interest. H-bonding is one of the most essential and most common interactions among those that hold together the secondary and tertiary structural elements of proteins, DNA, and other macromolecules having biological importance. Exploring the nature of their H-bonding network was primary in understanding the structure and function of such molecules. In addition, H-bonding affinity plays a substantial role in biological processes as well, for example enzyme reactions, inter- and intramolecular interactions, molecular recognition, aggregation, solvation, or in driving folding and unfolding events.

As it is detailed in the above chapters, H-bonds also proved to have basic function in the unfolding process of β-sheet containing proteins. To refine the picture provided by experiments and molecular dynamic simulations, and mainly to discover the energetics of this process the exploitation of the opportunities offered by quantum chemical methods is needed.

The backbone hydrogen-bonds of β-sheets have already been experimentally studied by individually replacing its backbone amides with esters[82]. Thermodynamic studies on these variants showed that the protein was most destabilized when H-bonds that had otherwise been enveloped by a hydrophobic cluster were perturbed; and kinetic studies indicated that native-like secondary structure formed in one of the protein's loops in the folding transition state, but the backbone was less ordered elsewhere in the sequence. This benchmarks the importance of backbone amide-amide H-bonds in the folding and stability of β-sheet structures.

By using quantum mechanical calculations, Morozov et al. have investigated orientation and distance dependence of hydrogen bonding energetics[80]. They found a remarkable agreement between the energy landscapes obtained from the quantum chemical calculations and the distributions of hydrogen bond geometries observed in protein structures.

The analysis of non-covalent interactions in biomolecules has set a challenge to theoreticians to create quantum chemical methods that can describe properly such effects[83,84]. With the appearance of DFT (density functional theory – will be explained further in *chapter 4.2.*) methods, large systems as biomolecules or a certain region of them can be accurately modeled, and the initial problems of cost-accuracy ratio seems to be solved.

In studying these interactions the role of "orbital-overlap" methods has been strengthened recently, for example that of Natural Bond Orbital analysis[85,86,87.] This method has already been widely applied in the investigation of biochemical systems featuring weak non-covalent interactions, and proved to be a very useful tool of this investigations by revealing the orbital-level background of such structure and behavior of such systems.[88,89,90]

Modeling of H-bonding on the basis of natural orbitals gives us the opportunity to display and handle this interaction in terms of overlaps of localized molecular orbitals, which stands very close to the "chemist's point of view"[91].

One of the early studies was in this field the one by Alabugin et al.[92]. They investigated the electronic basis of "improper" H-bonding suggesting that two interactions have important effect on the length of an X-H bond involved in an H-bond, namely hyperconjugation, and rehybridization.

A recent investigation on stabilizing interactions in proteins has been published by Bartlett et al.[88]. They analyzed an analogous interaction to usual H-bonds between amide groups in proteins, namely the so-called n_p-π^* interaction. This denotes the overlap of the p-rich lone pair of an amide oxygen (the s-rich one is involved in the usual H-bond) with the antibonding orbital of the C=O bond of the neighboring amide group. By performing quantum chemical calculations followed by Natural Bond Orbital Analysis along with searching protein structure databases, they have managed to reveal the orbital level background and significance of such an interaction.

Jakobsche et al., in contrast, have investigates the effect of the n)(π repulsion on protein stability. They performed Natural Steric Analysis[93] to reveal the magnitude of this interference and the dependence of it on the orientation of groups involved. For example, in the case of the above interaction also occurs this repulsion and weakens it. They suggested that the balance of the stabilizing overlap and the destabilizing repulsion determines the strength of an H-bond.

A weak interaction, termed as aliphatic C-H---S 3-center-4-electron interactions was discovered in an iron-containing (S bound to Fe) protein and verified also by using NBO analysis and by data obtained from NMR spectroscopy experiments[90]. This suggests that polarized sulfur atoms in proteins can engage in multiple weak interactions with surrounding aliphatic groups.

3. Aims

Several proteins featuring β-sheet structural elements have recently been subjects of single molecule force spectroscopy measurements, and these experiments gave insight into the mechanical properties and structure of individual protein molecules. Previous studies suggested that the rupture of β-sheets is responsible for the so-called sawtooth-like pattern of force-extension curves, which was obtained in every experiment. Despite the numerous experimental and theoretical studies, the atomic level details of these processes were not clear, until Beke-Somfai and Perczel proposed a mechanism for the unfolding of small β-sheet models, taking also into account the role of first layer water molecules, using quantum chemical (QC) methods[1].

Our aim, based on their results, was to further investigate the role of these water molecules and discover the subatomic level changes during the forced unfolding of β-sheets. Using their model structures, we performed additional QC calculations, and further analyzed the energetic and structural changes upon the unfolding. Since the changes in the hydrogen-bonding network proved to have the most significant effect on energetics, the detailed analysis of their formation and breakup, and its connection to the relative energy changes was a major goal of this study.

Besides, by using Natural Bond Orbital analysis we wanted to reveal the orbital level background of the unraveling of β-sheets, and decompose the relative energy gains into distinct localized orbital interactions. The differences in the contribution of such interactions to the total stabilization were to be explained as well. The quantum mechanical background of orbital interactions underlying H-bonds was also the subject of our interest.

4. Methods

4.1. Introduction to quantum chemistry

As it has been mentioned in the introduction section, nowadays quantum chemistry has become expansively used in solving chemical problems. Practically it is defined as the application of quantum mechanics as a physical model on chemical systems. Two branches of quantum chemistry are electronic structure and nuclear motion calculations, which partition can be made due to the Born-Oppenheimer approximation providing the concept of PES (potential energy surface).

The two basic principles in computing the electronic properties of a system are the variational and the perturbation procedures. The essence of the former is that the ground level energy of a system is searched by means of the variational principle, which states that it is equal to the exact energy only in the case of using the exact wavefunction of the system; otherwise the calculated energy value is an upper bound to the real (exact) one. The latter, perturbation method is based on the assumption that our system is only slightly different from an already known one, of which the Schrödinger equation can be calculated exactly, thus the energy of the system of interest can be calculated by perturbing (adding a correction term to) the Hamiltonian of the known system. In contrast to the variational method, this procedure does not provide an upper bound to the calculated energy even in the case of the exact wavefunction.

To be able to efficiently compute the electronic structure of molecules, the neglects to be made for optimizing the usage of resources and their extent must be precisely defined. For instance, the neglect of relativistic corrections does not have a notable effect in case of peptides, but the consideration of the electron correlation or the sort and size of the basis set on which the wavefunction is expanded (see under), the so-called *level of theory* has a substantial role in the successfulness of computations.

One of the first methods for computing electronic structure was the Hartree-Fock-Roothan method based on the variational principle: the wavefunction is expanded in a fixed basis set:

$$\Psi_i = \sum_{\mu=1}^{m} c_{i\mu}\chi_{\mu} , \tag{2}$$

thus in this equation only the coefficients remain unknown. Inserting this approximation on the variational principle, we just have to do the energy minimization with respect to the $c_{i\mu}$-s.

Nowadays most of the calculations are carried out either with post Hartree-Fock or DFT (density functional theory) methods, which all take into account in some wise the electron correlation that is neglected by the Hartree-Fock method. The formers are the Møller-Plesset/Many Body Perturbation Theory (MPn, MBPTn, where n is the so-called order of perturbation) , the CI (Configurational Interaction, based on the variational approach) and the CC (Coupled Cluster) methods . The DFT methods will be delineated further on.

The selection of the basis set, in which our wavefunction is expanded, is very important for the fastest possible reach of convergence in the so-called Self Consistent Field (SCF) procedure, the iterative algorithm applied in almost all calculations; and for the achievement of the best accuracy on a certain level of theory. Basis functions are generally chosen to be centered on the atoms to reflect the symmetry of the molecule, and can be Gaussians (applied mostly), Slater type orbitals or plane waves, for instance. In order to describe the structure more precisely, polarization and diffusion functions are used (even built-in versions); there are so-called split-valence, and correlation-consistent sets.

Due to the complexity of quantum chemical calculations, they cannot be carried out without computers. For that very reason, several program packages have been developed during the recent decades (for example the most commonly used is *Gaussian*) by help of which one can easily and efficiently perform such computations[94].

4.2. Density functional theory: the optimal level of theory for biomolecules

In the case of systems containing a number of electrons, for example biomolecules, there is no possibility of applying the most accurate level of theory and the largest basis set, because the more computational time invested does not provide significant improvement in accuracy. A level of theory, which is not too expensive but yields results being consistent with the experiments, is needed to apply. When studying such molecules, the aim generally is to obtain a picture, or a tendency, which qualitatively characterizes our system in a way that makes it utilizable, for instance, in the biological practice. To achieve this, we do not have to take the smaller digressions into account.

Regarding the above requirements, to investigate the properties of biological systems, the density functional theory (DFT) [95] proved to be a sufficient choice. It is based on the followings: the energy of the ground state is searched as a functional of the electron density of the ground state, using the variational principle. The electron density ($\rho(r)$) is a function depending on three spatial variables, and when integrated on the entire space it gives the number of electrons of the system (n):

$$\rho(r) = n \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \Psi^*(1,2,\dots,n)\Psi(1,2,\dots,n)dr_2\, dr_3 \dots dr_n, \tag{3}$$

where r_i $(i = 1,2,\dots,n)$ represent the three space coordinates corresponding to each electrons, respectively.

It is proven that an external potential ($V(r)$) defines unambiguously the electron density in a molecule and vice versa. According to the above, there is an assignment between the external potential and the wavefunction of the system and there is one between the wavefunction and the electron density, therefore this statement must be invertible, consequently there must be an unambiguous assignment between the electron density and the external potential. The straightforward consequence of this is the first Hohenberg-Kohn (Walter Kohn, Nobel Prize in 1998) theorem, which states that the energy of a system in ground state can be determined in knowledge of the electron density; that is, in a wider sense, a measurable **A** property of a system is an unambiguous functional of the ground state electron density.

The second Hohenberg-Kohn theorem provides the possibility of determining the electron density by stating that the ground state energy is a variational functional, which means that in the case of any approximate density ($\tilde{\rho}$) the following equation holds true:

$$E[\tilde{\rho}] \geq E_0[\rho_0] \tag{4}$$

where E_0 is the energy of the ground state, and ρ_0 is the electron density of the system in ground state. Using this, the minimum of the energy functional is sought considering the number of particles constant, as a side condition.

The expectation value of the energy can be decomposed into the following form:

$$\begin{aligned} E[\rho] = \langle \Psi[\rho] | \hat{H} \Psi[\rho] \rangle &= \langle \Psi[\rho] | \hat{T} \Psi[\rho] \rangle + \langle \Psi[\rho] | \hat{V}_{ee} \Psi[\rho] \rangle + \langle \Psi[\rho] | \hat{V}_{ne} \Psi[\rho] \rangle \\ &= T[\rho] + V_{ee}[\rho] + V_{ne}[\rho] \end{aligned} \tag{5}$$

where Ψ is the wavefunction of the system, $\hat{H}$ is the Hamiltonian, $\hat{T}$ is the kinetic energy operator, V_{ee} is the electron-electron repulsion potential, V_{ne} is the nucleus-electron attraction potential, and $T[\rho]$, $V_{ee}[\rho]$ és $V_{ne}[\rho]$ are the contributions of these to the global energy, respectively. Solving the variational problem outlined in the second Hohenberg-Kohn theorem is impeded by the fact that the form of $V_{ee}[\rho]$ is yet unknown. This term incorporates the classical repulsion and the non-classical exchange-correlation potential. The problem is caused by the latter, which cannot be calculated exactly.

With a procedure developed by Kohn and Sham one can operate with one-electron functions during the variational minimization like in Hartree-Fock (HF) method, hence the algorithms of the HF variational problem already integrated in many programs should not be changed in many respects and can be applied in DFT methods as well. This fact has widely increased the interest in DFT, and the accurate results that could be obtained with different approximate exchange-correlation functionals ensured a prominent place to it among the quantum chemical methods.

These approximate exchange-correlation functionals are mostly determined by fitting on experimental results or on data obtained from highly precise computations. Based on such procedures newer and newer functionals appear in literature; each is being adequate for different systems featuring different properties and interactions.

In the case of biological systems among the DFT functionals one should choose the one that describes most appropriately the properties and interactions of the system of interest. Comparing the DFT with other, *ab initio* (post Hartree-Fock or perturbation) methods in quantum chemistry, its outstanding cost-effectiveness ratio makes it the most commonly used tool in modeling of large systems. Its big disadvantage is that errors occurring during DFT calculations are not systematic, thus they are not predictable[94].

4.3. Theoretical background of the Natural Bond Orbital Analysis

The electron density function provides a complete description of the distribution of the electrons in the molecule, but it could be difficult to operate with it. It can be simplified by somehow dividing the charge cloud among the atoms, in this way generating a partial charge assigned to each atom. One way to achieve this is using the Mulliken population analysis, which divides the so-called overlap population equally among the basis functions.

A better procedure to generate partial charges is based on the eigenvalue equation (see below) of the first ordered reduced density operator, and is called Natural Bond Orbital Analysis[86, 91].

$$\boldsymbol{\Gamma}\varphi_i = p_k\varphi_i \tag{6}$$

where the p_k eigenvalues represent the population of the φ_i eigenorbitals. In fact, the density operator ($\boldsymbol{\Gamma}$) is the one-electron projection of the N-electron density distribution defined by the probability $|\Psi|^2$ ($\boldsymbol{\Gamma}$ is a Hermitian operator, a one dimensional projector), hence it is very suitable for characterizing the one-electron properties of the wavefunction.

In addition, as in the case of every Hermitian operator, the eigenfunctions of $\boldsymbol{\Gamma}$, the natural orbitals (NO), form a complete orthonormal set. The population (p_k) of any normalized trial orbital, and the orbital (φ_i) itself as well can be calculated as the expectation value of the density matrix. The Pauli exclusion principle ensures that $0 \leq p_k \leq 2$.

The Natural Bond Orbital analysis includes a series of transformations that generate localized (NAO), hybrid (NHO), bonding (NBO) or natural localized molecular orbitals (NLMO) from the initial functions of the selected basis set.

$$\chi_\mu \rightarrow \text{NAO} \rightarrow \text{NHO} \rightarrow \text{NBO, NLMO}$$

Besides, the basis functions can be transformed to delocalized natural orbitals (NO) or to canonical molecular orbitals (MO). Orbitals obtained in each step form an orthonormal set, i.e. they span the same dimensionality as the original basis set does; thus they can be used for generating the wavefunction and calculating the properties of the whole system alike.

The natural atomic orbitals (NAO) are one-centered localized orbitals that represent the „effective” natural orbitals of atom **A** in the molecule. The shape of NAOs is optimized for their effective atomic charge in the molecular environment, therefore if atom A has a more cationic character, then NAOs are more contracted; and if it is more anionic, the NAOs are more extended. In addition NAOs incorporate the proper nodal features due to steric (Pauli) confinement in the molecular environment.

Neglecting the mutual orthogonality between two atoms, the pre-orthogonal NAOs (PNAOs) can be obtained that can overlap with PNAO-s of other atoms. This feature enables us to estimate the strength of different interaction by calculating the stabilization energy coming from the overlap.

In consistence with our chemical intuitions, only the population of the core and valence NAOs matters in effect, whereas the Rydberg-type extra-valence (virtual) NAOs complete the span of the basis. By this partition, the original basis set can be reduced to the natural minimal basis (NMB) that includes the core and the valence NAOs. This neglect is appropriate due to the minor role that the virtual NAOs (natural Rydberg basis, NRB) play in NBO analysis, and causes a significant decrease in cost of computations

The natural bond orbitals (NBOs) are multiple-centered localized orbitals that provide a Lewis-type description of the chemical bond. The NBOs are composed of natural hybrid orbitals:

$$\Omega_{\mathrm{AB}} = a_A h_A + a_B h_B, \qquad h_A = \sum_k a_k \varphi_k{}^{(A)}, \tag{7}$$

where $a_A{}^2 + a_B{}^2 = 1$. Depending on the values of these coefficients the character of the NBO can range between covalent ($a_A = a_B$) or ionic ($a_A \gg a_B$) shape. The pre-orthogonal hybrid orbitals (PNHOs) can be derived from PNAOs using the same coefficients. They also form a complete basis set as well, which spans the dimensionality of the original basis set.

Natural Bond Orbital analysis can also provide a description standing close to Valence Bond theory, namely it calculates the "composition" of the above hybrid orbitals by using the conventional sp^{λ} notation, where λ expresses the percentage p-character (%-p) or s-character (%-s) of the hybrid; characterizing this way the corresponding NBOs.

Not surprisingly, the core NBOs have an almost total NAO character, whereas the nonbonding NBOs (a lone pair of electrons localized on one center) are identical to normalized NHOs($n_A = h_A$). All the NBOs that are built up of valence NHOs must be orthogonal to the antibonding NBOs, which are defined as follows:

$$\Omega_{AB}{}^{*} = a_B h_A - a_A h_B \tag{8}$$

The $\Omega_{AB}{}^{*}$-s are typically the most important non-Lewis type acceptor orbitals playing important role in supramolecular donor-acceptor interactions such as hydrogen bonding. Knowledge of these properties is essential in the understanding of a number of non-covalent and delocalization-related phenomena, which are beyond the concepts of Lewis-structures. With the neglect of Rydberg-type NBOs, the NBO set can also be reduced to the size of the NMB – in accordance with the chemical point of view.

From NBOs one can create the pre-orthogonal orbitals (as PNAOs are created from NAOs), the PNBOs that are able to overlap with each other. These overlaps can explain donor acceptor interactions like the formation of H-bonds, which is usually considered as the overlap between an antibonding orbital of an X-N bond (acceptor) and a lone pair of electrons of an atom with high electronegativity, such as F,O or N atoms.

In the NBO 5.9 program there is a built-in procedure, which provides estimation for such overlap energies using a second-order perturbation method. Interaction of unperturbed donor $\varphi_i^{(0)}$ (e.g. a lone pair) with acceptor $\varphi_{j^*}^{(0)}$ (e.g. an antibonding orbital) leads to the corresponding second-order $i \rightarrow j^*$ stabilization estimate

$$\Delta E_{i\rightarrow j^*}^{(2)} = -2\left\langle \varphi_i^{(0)} \middle| \hat{F} \middle| \varphi_{j^*}^{(0)} \right\rangle^2 \Big/ \left(\varepsilon_{j^*}^{(0)} - \varepsilon_i^{(0)}\right) \tag{9}$$

For the NBO analysis the only input is the wavefunction of the system, and since the natural orbitals are the (eigen)orbitals that strictly belong to Ψ, they are indeed intrinsic ("natural") to describe the electron density and other single-electron properties of Ψ.

A useful method based on the natural bond orbital theory is Natural Steric Analysis (NSA), which allows estimation of steric repulsion energy coming from the Pauli Exclusion Principle. Using the Fock (or Kohn–Sham) operator, we can evaluate the average energy of each occupied NBO and the associated pre-orthogonal PNBO in the usual manner.

$$F_{i,i}{}^{NBO} = \langle \Omega_i^{NBO} | \hat{F} | \Omega_i^{NBO} \rangle \tag{10}$$

$$F_{i,i}{}^{PNBO} = \langle \Omega_i^{PNBO} | \hat{F} | \Omega_i^{PNBO} \rangle \tag{11}$$

The PNBO energy ($F_{i,i}{}^{PNBO}$) corresponds to a Pauli-violating limit in which each constituent atomic orbital retains its isolated free-atom (PNAO) form. The energy difference between the Pauli-violating and the actual NBO energy ($F_{i,i}{}^{NBO}$) therefore measures the effect of increased steric pressure due to proximity of other atoms. The total NBO steric exchange energy (E_{steric}) is evaluated by summing such differences over all occupied NBOs:

$$E_{steric} = \sum_{i}^{occ.} \left(F_{i,i}{}^{NBO} - F_{i,i}{}^{PNBO} \right) \tag{12}$$

It might be necessary to decompose this energy approximately into contributions from the steric interactions of individual electron pairs. For each pair, the "partially deorthogonalized" PNBO/2 orbitals can be formed, which are non-orthogonal with respect to one another, but remain orthogonal with respect to all other NBOs. The partial contribution to steric exchange is therefore estimated as

$$E_{i,j}^{steric} = \left(F_{i,i}{}^{NBO} - F_{i,i}{}^{(PNBO/2)} \right) + \left(F_{j,j}{}^{NBO} - F_{j,j}{}^{(PNBO/2)} \right). \tag{13}$$

By summing these contributions over all distinct pairs, an alternative pairwise-additive estimate of the steric exchange energy can be obtained, which, however, is only an approximation to the more fundamental expression (eq. 12).

$$E_{pw}^{steric} = \sum_{i<j} E_{i,j}^{steric} \tag{14}$$

4.4 Applied methods, models and nomenclature

All computations were carried out using and Gaussian 09 software packages[96] with NBO 5.9 program[97] integrated. To imitate the unfolding of the model structures, previous relaxed potential energy surface scan calculations were performed at the B3LYP/6-311++G(d,p)//B3LYP-6-31G(d) level of theory[98,99] in vacuum, along the terminal C=O----H-N hydrogen bond distance[1].

After that, (related to this thesis) single point calculations were carried out both in water and in protein environment on geometries previously obtained in each scan step, more than 1200 overall, using IEFPCM[100] continuum model at the B3LYP/6-311++G(d,p) level of theory. Using this level of theory results in a higher level energy profile and makes the basis set superposition error (BSSE) negligible[101]. The apolar environment was specified by parameter $\varepsilon_{rel} = 4$, and by using 2,5 Å UFF radii value.[102]

To investigate the unfolding of sheet motifs partially accessed by water, both parallel (P) and antiparallel (AP) sheet models were investigated with 1, 2, and *3* explicit water molecules, which models were labeled as P1, P2, P3 and AP1, AP2 and AP3, respectively. Our models consisted of two trialanine strands originally bound in β-sheet conformation to imitate the protein backbone in a sheet. In these models the water molecules were initially coordinated on one side of the terminal H-bond. This way we tried to model the coordination of the first water molecules approaching to an otherwise buried sheet motif. Since it is supposed that the water can hardly approach the peptide-peptide H-bonds from the side chains, the initial position of the water molecules was chosen to be on the opposite side compared to the side chains.

The breakup of H-bonds was monitored in a 0.4 Å stepwise manner scanning the terminal C=O----H-N hydrogen bond distance.[1] A H-bond was considered to be present while the distance of the involved O----H bridge atoms was lower than 2.5 Å. This stepwise progression was performed until the interstrand H-bonds broke up and the two peptide strands became separated, usually at 35-40 Å.

Since the breakup and formation of the H-bonds were the most important events during the investigated forced unfolding processes, a nomenclature was introduced to differentiate the various H-bonds. The two peptide strands were labeled A and B, and the interstrand peptide-peptide H-bonds were numbered starting from the N-terminus of the B strand and named I1, I2, I3 and I4, respectively (**Figure 9**). The water molecules were labeled W1, W2 and W3 respectively, where W1 was usually coordinated with two H-bonds on peptide groups 1 and 2 on both strands in the initial (pre-optimized) structures of the scan. The peptide-water H-bonds were named as follows: $A1_OW1_H$, if the H atom of the W1 water molecule was bound to the carbonyl O atom of the first amide bond of strand A; and $B2_HW1_O$, if the O atom of W1 water molecule was bound to (the H atom of N-H bond in) the second peptide bond of strand B; the subscripts show the involved bridge atoms from each strand. (The numbering of amide bonds started from the N-terminus of strand B in each cases, as in the case of the interstrand H-bonds.)

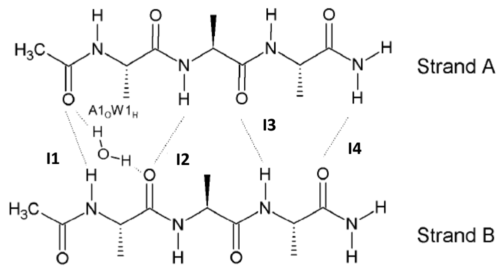

Figure 9 - Schematic representation of nomenclature on the antiparallel β-sheet model. (The H-bonds of the parallel sheet models were labeled in the same way as the antiparallel one.)

To investigate the orbital-level background of the hydrogen bonds playing a significant role in the unfolding process, Natural Bond Orbital overlap energies were computed using the built-in second order perturbation procedure in NBO 5.9 program, followed by Natural Steric Analysis computations that reveal the steric repulsion energies arising due to the Pauli Exclusion Principle between occupied orbitals, decomposed to pairwise repulsions. These computations were carried out in water (using IEFPCM solvent model) on the B3LYP/6-311++G(d,p) level of theory, by one Ångström steps.

The net overlap energies that effectively stabilize the H-bonds were estimated as differences between overlap energies and corresponding repulsion terms. The interstrand peptide-peptide and the water-peptide H-bonds were investigated in detail.

To describe H-bonds properly by using this method the following PNBOs and their pairwise overlap and contribution to steric repulsion were taken into account during our analysis. In the case of interstrand H-bonds the antibonding orbitals of amide N-H bonds, the lone pairs of carbonyl oxygen atoms, and the their overlap were considered, as well as the Pauli repulsion between the N-H bond orbital and the lone pairs of the carbonyl O-atom. Similarly, for water-peptide H-bonds the O-H antibonding orbitals of water and the lone pairs of the carbonyl O-atom, and since the water can act as donor and also as acceptor in H-bonds, the amide N-H antibonding orbitals and the lone pairs of the water molecules were taken into consideration, as well as their pairwise overlaps and repulsion terms, respectively.

The *s*-character of the hybrid orbitals of N-atoms in the N-H bonds, and that of the lone pairs of the carbonyl and water oxygen atoms were used for analysis in the case of the antiparallel model having one water molecule.

All structural and energetic changes were detailed in relation to the scanned terminal O---H distance.

5. Results and discussion

Here a further analysis is presented upon the previously[1] studied unfolding of trialanine dimer models aided by 1-3 water molecules. We have carried out single point energy calculations using implicit continuum models to consider bulk effects of both water and a buried state within aqueous and in protein environment on the geometries obtained in each step of the previously performed relaxed potential energy surface scan calculations, and examined the differences in the relative energy upon the unfolding process. Our results were elucidated having particular respect to the H-bonds forming and breaking up during the separation of the two peptide strands. Then we have computed the natural orbitals of the model systems and investigated the orbital-level background of the changes in H-bonding network in detail.

5.1. Energetic changes upon unfolding in aqueous and protein environment

We have further investigated the unfolding of trialanine models (see structures and nomenclature in *chapter 4.4.*) by performing single point calculations on the B3LYP/6-311++G(d,p) level of theory both in water and in protein environment (see details of solvent-models in chapter 4.4.) on the geometries obtained from the previously performed relaxed potential energy scan calculations on the B3LYP/6-31G(d) level of theory.

In accordance with our expectations the energy values showed significant differences under the two different conditions. Considering a polar aqueous media during the forced unfolding, the relative changes in energy (with respect to the energy of the first model, which is two peptide strands forming a β-sheet) appeared to be lower for almost every model than the energetic changes obtained in the apolar environment. Such apolarity may occur for instance in case of green fluorescent protein (GFP), which forms a β-barrel, where only one side of the β-sheets of the barrel is exposed to water, or similarly, when a β-sheet is buried inside of a large protein, and for membrane proteins containing β-sheet regions which are surrounded by the hydrophobic lipid bilayer[103].

In **Figure 10** the relative energy values are depicted as a function of the scanned terminal O---H distance.

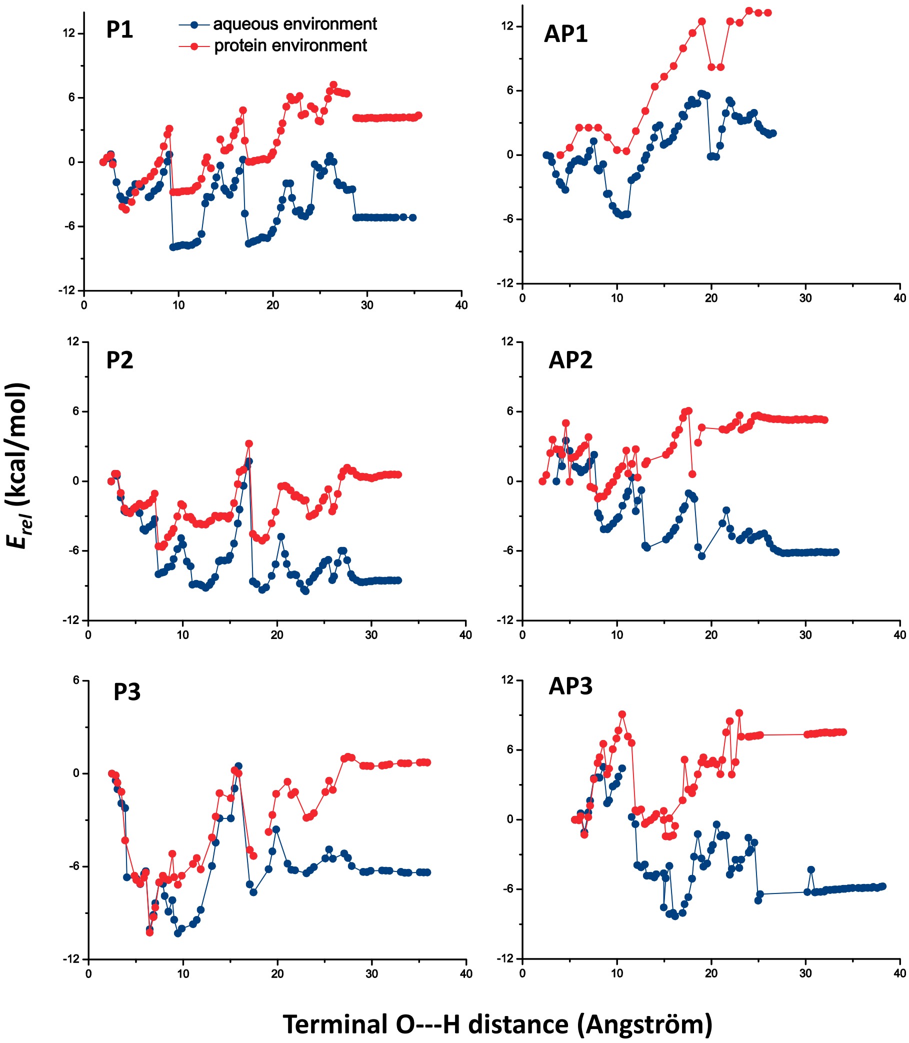

Figure 10 - Relative energy curve of parallel (LEFT) and antiparallel (RIGHT) sheet models, having one (TOP), two (MIDDLE) and three (BOTTOM) water molecules (ΔE with respect to the first model, which is two peptides strands bound in β-sheet conformation) as a function of the scanned terminal O---H distance in aqueous (polar, ε_{rel}=78.35, blue dots) and in protein (apolar, ε_{rel}=4.00, red dots) environment during the forced unfolding.

As it can be seen in **Figure 10** the rapid changes in relative energy values in protein environment (red dots) follow quite clearly that of the data obtained in water (blue dots), and the red curve run above the blue one for the most part. However, towards the stages of the unfolding process the differences become larger between energies coming from the two types of media. This can be explained by the greater stability of the backbone of a single peptide chain when surrounded by water (a polar media) than when surrounded by apolar environment.

The breakup of interstrand H-bonds is more favorable in aqueous environment due to the solvation of the polar groups which were involved in the H-bonds. This compensation is less pronounced in the protein environment, due to the much smaller relative permittivity of the media, thus the breakup of H-bonds is less favored.

That is why the energies in the final stage in case of protein environment are roughly equivalent to or higher than the one of the initial structure (ΔE is around zero, or positive), however, in the case of water they are lower than the initial ones, except for the Ap1 model (top left block of **Figure 10**). (This "deviation" is due to the low energy of the initial folded state, as for the P1 and AP1 models the single water molecule has several H-bonds formed already, while the additional waters in the P2, AP2, P3, and AP3 models are less "saturated by H-bonds" in the initial structure.) In addition, in the initial phase, β-strands are not as exposed to the surrounding media as they are in the subsequent steps of separation. In general it also can be said that the drops in relative energy, upon H-bond forming, prove to be larger in water then in apolar environment. (Changes of the H-bonding network will be detailed in the following chapter.)

The final averaged differences in the energy values, namely those between the two plateaus reflecting the fully separated state, is around 9 kcal/mol in the parallel case, and a 13 kcal/mol in the antiparallel one. This indicates that the unfolding of parallel sheets induces more energy gain upon unfolding (smaller ΔE in final stage). This reflects, on one hand, the fact that the quality of the solvent affects differently on the two types of β-sheet structures, and, on the other hand it indicates that the antiparallel sheets are more stable than the parallel ones, which has already been shown in previous studies[25,31].

Based on the above observation, one may conclude that aqueous (polar) environment stabilizes the unfolded structure for smaller β-sheets, whereas the separation of β-strands is definitely less favored when they are partially buried inside of a protein, amyloid fibrils, or inserted into a lipid bilayer.

5.2. The unfolding process from the viewpoint of H-bonding

As many experimental and theoretical studies has emphasized so far (for reference see chapter 2.) hydrogen bonding plays a substantial role in forming and holding together the secondary and tertiary structural elements of proteins, as well as β-sheets. Here a detailed analysis and explanations are presented of the effects of the breakup and formation of H-bonds on the relative energy changes of our model systems upon the forced unfolding process. For the nomenclature of H-bonds in our models see *chapter 4.4*.

As regards the modeled unfolding process, generally it can be said the in parallel models the water molecules are moving toward the interior of the sheet between the two stands, while being separated, by gradually forming hydrogen-bonds with free amide bonds. The formation of these water-peptide H-bonds induces a great drop in the relative energy change accompanying the process, in this way lowering the energy barrier of unfolding. When formed, water-peptide H-bonds stabilize the emerging intermediates, and this stabilization causes the decrease in energy. For antiparallel models this type of "water-walking" cannot be observed, water molecules are rather attached to the first or second amide bonds of one strand. Only in the case of AP3 model occurs that the water molecules help in the rupture of third and fourth interstrand H-bonds too. The more water molecules assist the unfolding, the more various intermediates can form, showing greater discontinuities in relative energy curves.

Here we detail the changes in the H-bonding network of three models: P1, AP2, P3, having 1, 2 and 3 water molecules to assist the forced unfolding process, respectively. In the case of P1 model the proceeding of "water-walking" can be followed nicely, and then, with AP2 we would like to demonstrate the behavior of an antiparallel model as well, and last, but not least, the P3 model with three water molecule showing the closest approach to explicit water model, and also proved to be less complex to elucidate than the AP3 model, will be presented.

P1 model

Firs let us demonstrate the changes in the hydrogen-bond pattern of the P1 model during the forced unfolding. **Figure 11** shows the initial structure of P1 model numbering the interstrand H-bonds. This structure had four inter- and two intrastrand H-bonds, on strand B, and two H-bonds were formed, $A1_OW_H$ and $B2_OW_H$, with the water molecule coordinated on the two strands at the first peptide bond. During the scan the four interstrand H-bonds broke up at approximately 2.5Å, 5.8Å, 17Å, and 24.6Å scanned distance, respectively.

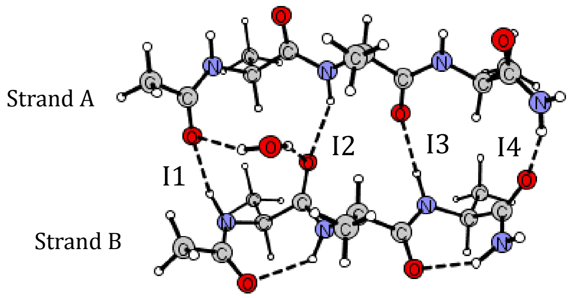

Figure 11 – The initial structure of P1 model. Labels IX, where X=1,2,3,4 represent the interstrand hydrogen bonds, respectively, with numbering starting from the N terminus of strand A (see nomenclature in chapter 4.4.). Dashed lines denote the H-bonds, which were considered to be present when the bond-length was lower than 2.5 Å.

Shortly after the breakup of the first interstrand C=O---H-N H-bond (I1), the water molecule moved in between the two peptide strands and a third peptide-water H-bond was formed ($B1_HW_O$, at ~2.8A scanned distance). Almost simultaneously with the second interstrand H-bond (I2), this peptide-water H-bond broke up at ~5.8A, where the rapid increase in ΔE shows that no new stabilizing effects occur.

At ~9.0Å due to the breakup of the A1oWH peptide-water H-bond and the movement of the water molecule towards the interior of the sheet to form an H-bond with the 2nd peptide bond on strand A, a significant ~7kcal/mol and 9 kcal/mol stabilization appeared, in the buried and aqueous environment, respectively. The decrease is also shown in average peptide-water H-bond length).

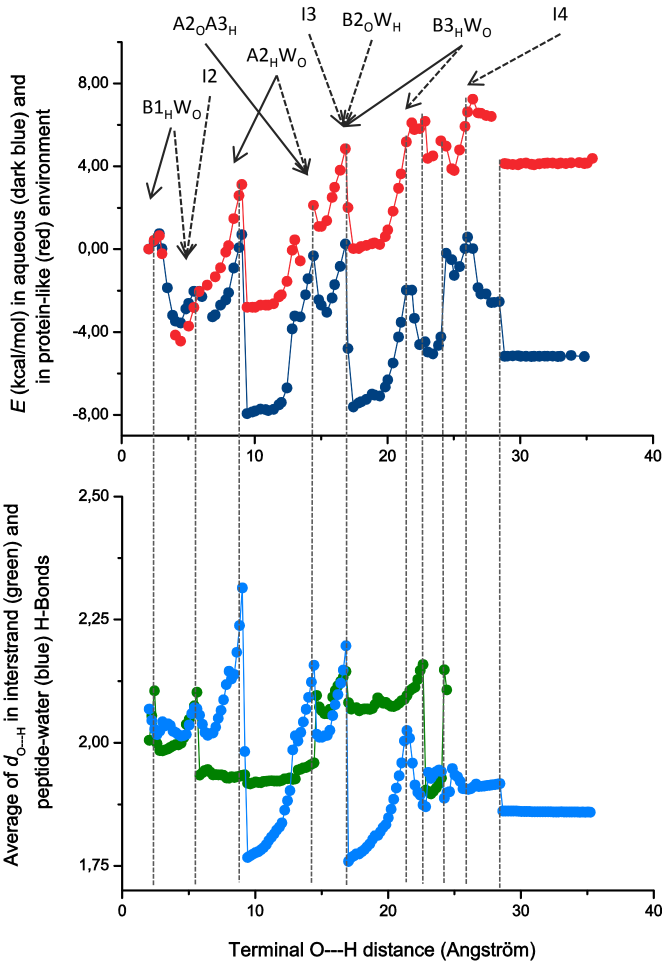

Figure 12– Changes in relative energy in aqueous (dark blue) and in protein (red) environment as a function of the terminal O---H distance during the unfolding process in the case of *P1 model*. H-bond forming and breaking events are labeled with arrows: solid lines denote forming events, and dashed lines denote breakups. (TOP) Averaged distances in interstrand peptide-peptide H-bonds (green) and in water-peptide H-bonds (blue). (BOTTOM). The vertical dashed lines mark the corresponding changes in the H-bonding network.

At ~13.2 Å, a third peptide-water H-bond formed, $A3_OW_H$, which was also associated with a minor energy decrease in apolar environment, but only a little change in polar media. At ~14.6Å the $A2_HW_O$ broke up and the water molecule coordinated on the I3 H-bond (increase in the distance of interstrand H-bonds, whereas decrease in average peptide-water H-bond length). At the same time the conformational change of the middle residue in strand A from β_L to γ_L formed a new intrastrand H-bond.

The breakup of $A2_HW_O$ resulted in a slightly less increase in ΔE due to the starting formation of the water-peptide H-bond, which also weakened I3, but caused some stability overall, and the forming of the intrastrand H-bond.

The breakup of the 3rd interstrand H-bond at 17Å happened simultaneously with two additional events: the $B2_OW_H$ peptide-water H-bond broke up, and a new peptide-water hydrogen bond formed, $B3_HW_O$, with the water molecule moving towards the C-terminus, all resulting in a ~8 kcal/mol stabilization in water, and a ~ 5 kcal/mol one in protein environment.

This latter H-bond broke up at ~21.6Å. It is suggested that the I4 H-bond was preserved by breakup of the I3 H-bond at ~23.0Å, where the γ_L conformation of the middle residue in strand A changed back to β_L. Due to this, parameters of the 4th interstrand H-bond relaxed back towards optimal values, which was the reason for the decrease in relative energy. Finally, at ~24.6Å I4 H-bond broke up as well. The A and B strands arrived into a full β_L and γ_L conformation, respectively. At ~29 Å the two strands became fully separated (marked by a drop in ΔE), and the plateau, which follows this point showing no change, refers to this dissociated state.

As it has already been mentioned above, in the case of P1 model, a systematic moving of the water molecule was observed, a kind of "walking" (**Figure 13**) between the two strands while assisting the breaking up of the interstrand H-bonds by forming water-peptide H-bonds. Usually the water molecule first coordinated on a peptide-peptide H-bond weakening it in this way, then, as the two strands were being separated the peptide-peptide connection was replaced by a water-peptide one, resulting in the restabilization of the system.

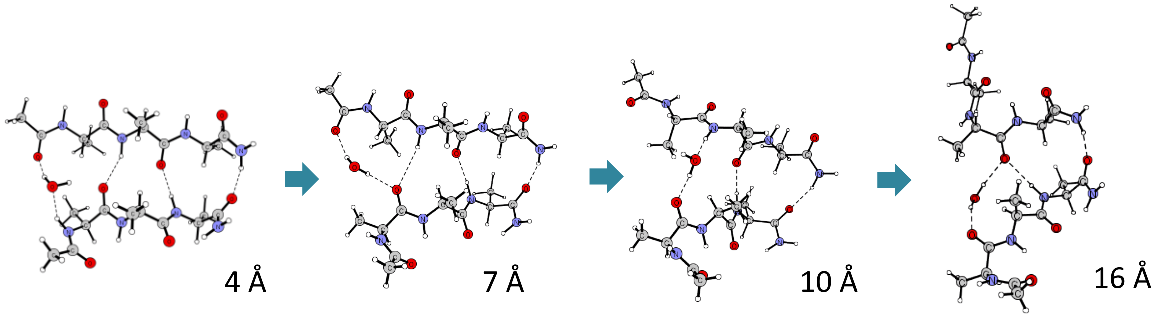

Figure 13 – "Water-walking" in P1 model, captured at scan steps at 4,7,10 an 12Å

Considering the drops and increases appeared in the relative energy curves, we can conclude that the formation of a water-peptide H-bond causes a remarkable stability (decrease in ΔE) each time it occurs, particularly in water. In most cases this stabilization exceeds the destabilizing effects of breakup of interstrand H-bonds and results significant decrease in energy.

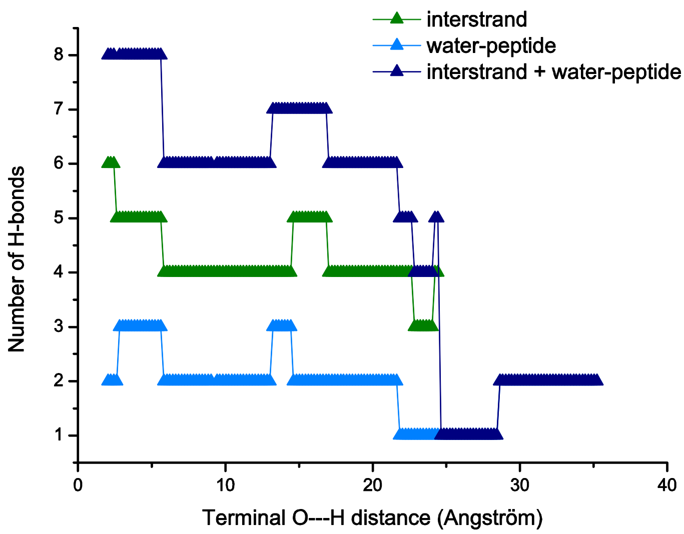

Figure 14 – Number of present interstrand (green) and water-peptide (blue) H-bonds at given distances, and the sum of them (dark blue).

The tendencies observed indicate that water-peptide H-bonding causes a greater energy gain then an H-bond between two amide bonds. What also can be seen from the energy curves is that the W_O bonds, when the oxygen atom of the water molecule is coordinated on the N-H bond of an amide residue, induce larger changes in energy, then the W_H ones, when an amide O is bound to the H-atom of the water molecule. The background of these phenomena will be further analyzed in *chapter 5.3.*.

To summarize the formation and rupture of H-bonds, on **Figure 14** we depicted the actual number of H-bonds, showing how the H-bonds alternate while unfolding.

AP2 model

Now let us demonstrate the unfolding assisted by two water molecules of AP2 model. The initial structure, which is shown in **Figure 15**, contained four interstrand H-bonds. The first water molecule, W1, was initially coordinated on the two peptide strands by forming one H-bond with each strand, $A1_OW1_H$ and $B1_OW1_H$. The second water molecule, W2, formed a hydrogen bond with W1 and was located closer to strand B.

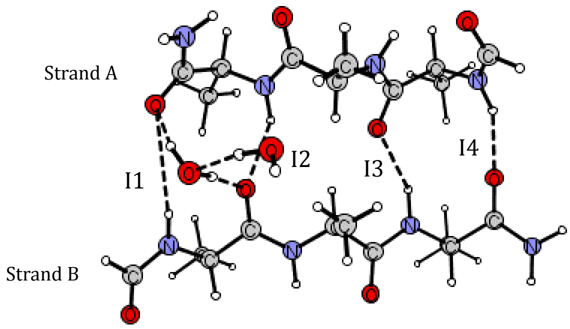

Figure 15 – The initial structure of AP2 model. Labels IX, where X=1,2,3,4 represent the interstrand hydrogen bonds, respectively, with numbering starting from the N terminus of strand A (see nomenclature in chapter 4.4.). Dashed lines denote the H-bonds, which were considered to be present when the bond-length was lower than 2.5 Å.

The interstrand H-bonds broke up at 2.5Å, 7.4Å, 17.8Å and at 18.8Å. After the breakup of $A1_OB1_H$, an intrastrand H-bond, $A1_HA2_O$ formed. This was shortly followed by the formation of the $B1_HW1_O$ peptide-water H-bond at 3.2Å. At the same time, an intrastrand H-bond formed, $A3_HA4_O$, at the other terminus of strand A. At 4.2Å the first residue in strand B shifted into γ_L as well, which formed $B1_OB2_H$. This was followed at the next scan step by the formation of $B3_OB4_H$. The forming of these intrastrand H-bonds resulted in the increasing of the average distance of interstrand H-bonds (**Figure 15**).

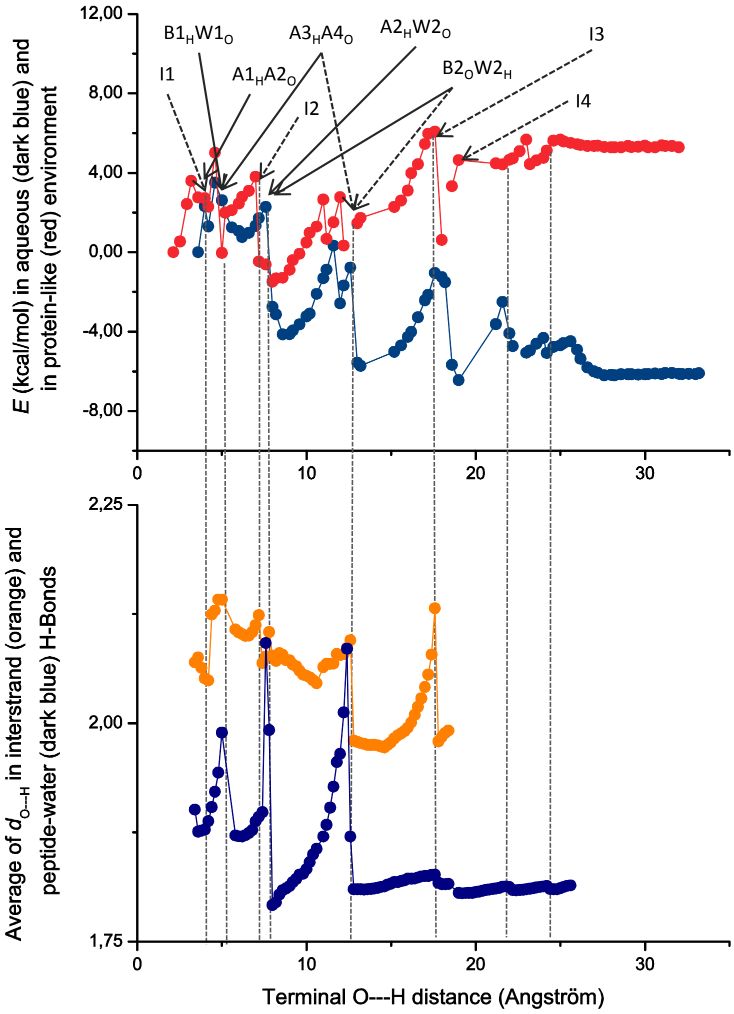

Figure 16 – Changes in relative energy in aqueous (blue) and in protein (red) environment as a function of the terminal O---H distance during the unfolding process in the case of *AP2 model*. H-bond forming and vanishing events are labeled with arrows: solid lines denote forming events, and dashed lines denote breakups. (TOP) Averaged distances in interstrand peptide-peptide H-bonds (orange) and in water-peptide H-bonds (dark blue). (BOTTOM). The vertical dashed lines mark the corresponding changes in the H-bonding network.

The $B1_HW1_O$ H-bond broke up at 5.4Å, clearly shown in **Figure 16** by increasing average length of peptide-water H-bonds, and until the breakup of the second interstrand H-bond (I2), a rapid increase in ΔE was observed. At 7.4Å, the I2 H-bond broke up followed by the formation of $A2_HW2_O$, as well as $B2_OW2_H$, where W2 bridged the two atoms of the broken H-bond, which caused a significant drop in relative energy (~6 kcal/mol in water, and ~3 kcal/mol in protein environment) and in peptide-water H-bond distances as well. However, in the next scan step $B2_OW1_H$ broke up, partially because the two water molecules formed a bifurcated H-bond with $B2_O$, which caused the increase of peptide-water H-bond distances. Meanwhile the interstrand H-bonds strengthened slightly, indicated by decreasing H-bond length, but the overall change in energy in this section was a monotone increase.

This tendency stopped at 12Å, when $B1_OB2_H$ broke up. This prevented the breakup of the remaining H-bonds connecting the two strands, which resulted in a drop in ΔE. At 12.6Å the $B2_OW2_H$ broke and from this point both water molecules were attached only to strand A, which is reflected in the large decrease in the average distance of peptide-water H-bonds.

In the next step the $A3_HA4_O$ intrastrand H-bond broke up, with the corresponding residue shifted from γ_L to β_L, which induced the strengthening of interstrand H-bonds (decrease distance). After this point, ΔE and the average distance of interstrand H-bonds increased monotonically and the remaining two interstrand H-bonds broke up nearly at the same scanning distance, i.e. at 17.8A and 18.8A.

When fully separated, the two water molecules were H-bonded to the first two peptide groups of strand A. Strand B had γ_L conformation for the first and third residues, β_L for the second, whereas strand A had γ_L conformation for the first and β_L for the remaining two residues.

Figure 17 now depicts the actual number of H-bonds that are present at a scan step. The changes in H-bonding network here also show clearly how peptide-water and interstrand peptide-peptide H-bonds alternate.

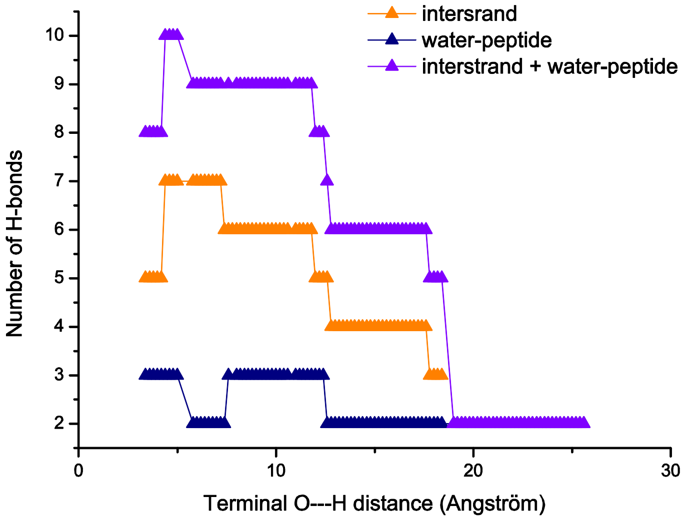

Figure 17 – Number of present interstrand (orange) and water-peptide (dark blue) H-bonds at given distances, and the sum of them (violet).

The consequences made in the case of the unfolding of P1 model still hold true here, for example that the W_O-type H-bonds cause larger stability, especially in water. Additionally, in both cases the strengthening of the remaining H-bonds can be observed when one break up.

P3 model

Finally let us analyze a model having three water molecules assisting the unfolding, let us consider the case of P3 model. The initial geometry (which is schemed in **Figure 18**) had four interstrand and two intrastrand H-bonds due to the γ_L conformation of the first and third residues in strand B. Two peptide-water H-bonds were present, $A1_OW1_H$ and $B2_OW1_H$. The second and third water molecules initially formed a single H-bond with the first water, $W1_OW2_H$ and $W1_OW3_H$, with position of W2 and W3 closer to strand A and B, respectively. The interstrand H-bonds were disrupted consecutively at ~2.5A, 6.3A, 16.7A and 25.5A.

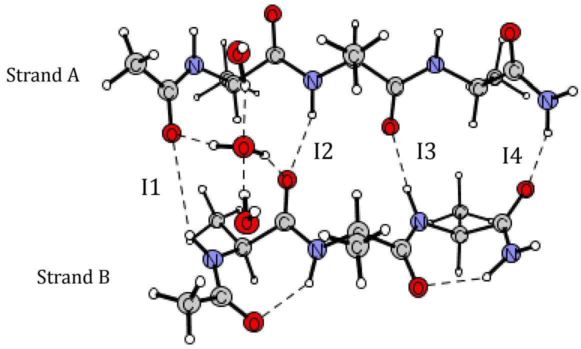

Figure 18 – The initial structure of P3 model. Labels IX, where X=1,2,3,4 represent the interstrand hydrogen bonds, respectively, with numbering starting from the N terminus of strand A (see nomenclature in chapter 4.4.). Dashed lines denote the H-bonds, which were considered to be present when the bond-length was lower than 2.5 Å.

At 3.8Å, W3 jumped into the hole between the bridge atoms of the first interstrand H-bond and formed hydrogen bonds with both of them: $A1_OW3_H$ and $B1_HW3_O$, which is reflected by a great decrease in relative energy and the shortening of the peptide-water H-bonds as well (**Figure 19**). The former broke up shortly after at 4.2Å because $A1_O$ had bifurcated hydrogen bonds with both $W1_H$ and $W3_H$.

The breakup of the second interstrand H-bond at 6.3Å is accompanied by the formation of $A2_HW2_O$, which also caused significant decrease in ΔE and in the average length of interstrand peptide-peptide bonds (the remaining H-bonds strengthened slightly). At 7.7Å the $A1_OW1_H$ broke up, and at 11.9Å the $A2_OA3_H$ intrastrand H-bond formed, with the conformational change of the second residue to γ_L.

At 14.7Å the $W1_OW3_H$ H-bond broke up and simultaneously W3 formed a $B2_OW3_H$, which is indicated by a decrease in peptide-water H-bond length, but the relative energy further increased after a temporary stabilization. This is due to the breakup of water-water H-bond, which is the strongest among all in our models, thus the formation of a peptide-water H-bond cannot compensate its rupture energetically.

At 15.8 Å W2 forms a new peptide-water H-bond, $A3_OW2_H$, by flipping its outer hydrogen towards the interior of the remaining sheet. This way the breakup of the third interstrand H-bond (I3) was prepared and took place at 16.7Å. In parallel,

$A2_HW2_O$ broke up, W2 moved toward the C-terminus and formed $B3_HW2_O$ which can be observed as a rapid decrease in ΔE, and so it is in the distances of interstrand and water-peptides H-bonds.

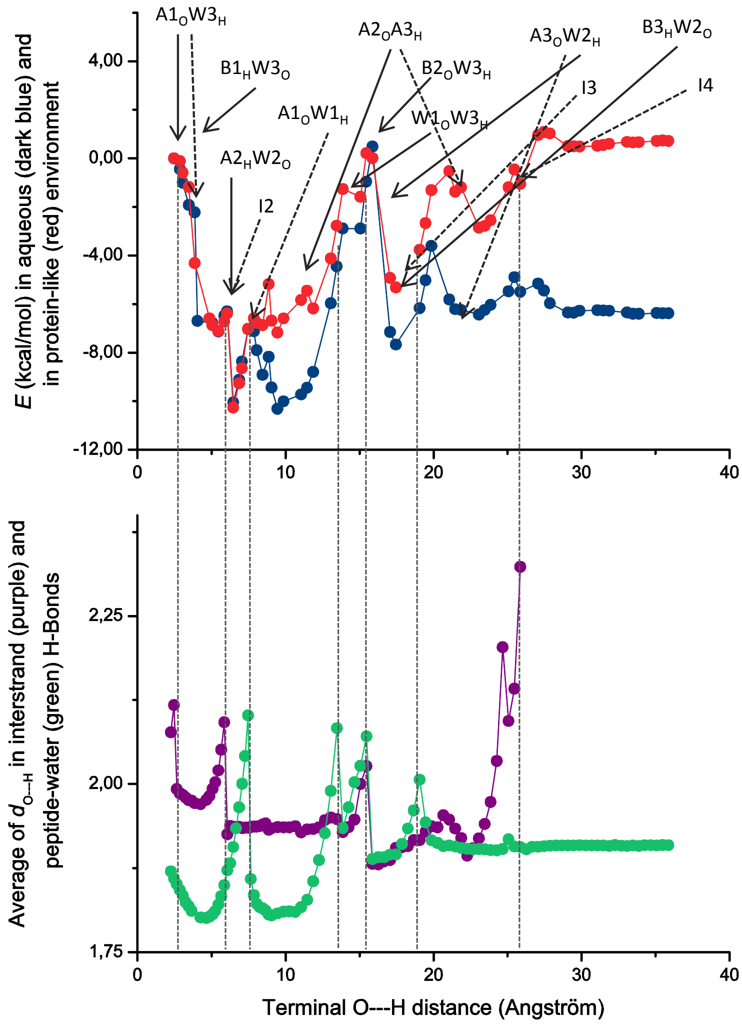

Figure 19 – Changes in relative energy in aqueous (blue) and in protein (red) environment as a function of the terminal O---H distance during the unfolding process in the case of *P3model*. H-bond forming and vanishing events are labeled with arrows: solid lines denote forming events, and dashed lines denote breakups. (TOP) Averaged distances in interstrand peptide-peptide H-bonds (purple) and in water-peptide H-bonds (green). (BOTTOM). The vertical dashed lines mark the corresponding changes in the H-bonding network.

At 20.3Å $A3_OW2_H$ broke up and every water molecule was separated from strand A. To prevent the breakup of the fourth interstrand H-bond, the γ_L conformation of the second residue in A changed to β_L at 22.7Å, which left strand A totally extended. Although this way one intrastrand H-bond broke up, the preservation of the more preferred interstrand H-bond still resulted in a decrease both in relative energy and in the length of the last interstrand H-bond (I4).

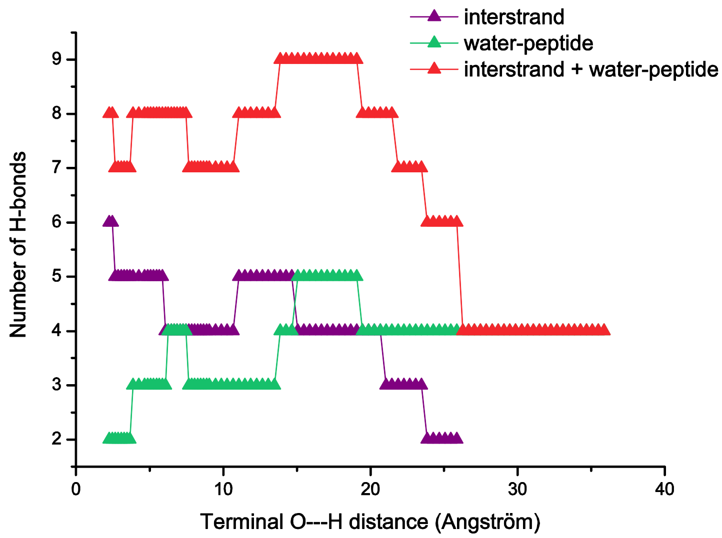

Figure 20 – Number of present interstrand (purple) and water-peptide (green) H-bonds at given distances, and the sum of them (red).

Finally, at 25.5A, the last interstrand H-bond broke up. The conformation of the final peptide strands was again β_L for strand A. For strand B the second residue remained in extended β_L conformation due to the two waters H-bonded to the neighboring amide groups, while the first and third residues were γ_L.

Figure 20 finally shows the actual number of the present H-bonds in each step. In this case, events are somewhat harder to follow, due to the three water molecules causing numerous changes.

5.3. Changes in the H-bonding network followed at the orbital-level

Analyzing the energy curves of the forced unfold in the previous section revealed that formation and breaking of H-bonds has a dominating role in the overall process. However, based on the atomic structure and the energies alone, the timing of

H-bond formation and breakup cannot be determined accurately and should rather be followed at a sub-atomic level. Therefore, in order to understand the background of H-bonding, Natural Bond Orbital calculations were performed and the nature of H-bonding in the level of molecular orbitals was investigated. Stabilizing and destabilizing interactions have been studied, represented by overlaps between donor-acceptor orbitals and repulsion between occupied ones, respectively. The characteristics of certain orbitals that are involved in H-bonds have been examined as well.

Since H-bonds seem to be the key actors in facilitating the unfolding by reducing the energy-barrier between the intermediates of the process, exploring the properties of the underlying interactions in terms of quantum mechanics can provide a much deeper understanding of their nature.

A hydrogen-bond is usually considered to occur as a result of mainly two interactions: on one hand the stabilizing overlap of the antibonding orbital of the X-H bond and the lone pair of electrons of an atom having high electronegativity, such as F, N or O ; and, on the other hand from the repulsive interference between the bonding orbital of the X-H bond and the same lone pair. In the former case a partial charge transfer occurs between the occupied donor (lone pair) and the unoccupied acceptor (antibonding) orbital, whereas in the latter case repulsion appears between the two occupied orbitals due to the Pauli Exclusion Principle. Although, other interactions can also take part in evolving an H-bond, the ones described above are the most fundamental.

Regarding forced unfolding, each interstrand peptide-peptide H-bonds were disrupted, step by step. While the terminal O---H distance was extended, various changes occurred in these bonds that cannot be revealed only by measuring the H-bond distances between strands. **Figure 21** shows the (pre-orthogonal) natural bond orbital representation of the two effect playing dominant role in producing interstrand H-bonds: the attractive and repulsive overlaps between corresponding orbitals.

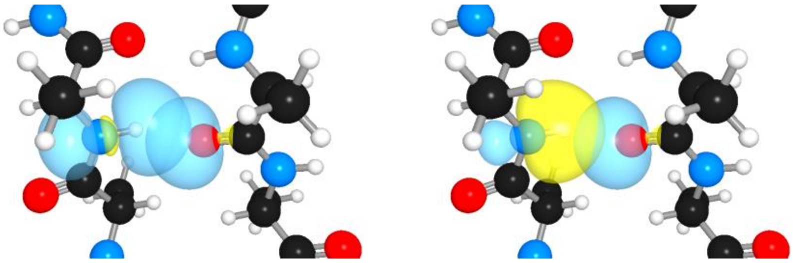

Figure 21 – Interstrand H-bond between the third amide bonds (I3) of the two strands in the initial structure of AP1 model: overlap of the (s-rich) lone pair of electrons of the carbonyl O with the antibonding orbital of the amide N-H bond (LEFT). Repulsive interaction between the same lone pair and the bonding orbital of the same amide N-H bond (RIGHT). All orbitals are represented in PNBO basis.

While separating the two peptide strands from each other, one can observe changes not only for the H-bond which is about to break up, but also in the strength of the remaining ones. This can be explained as an answer given by the system for the pulling force: it weakens not only the actually breaking H-bond, but the subsequent ones as well, and features changes that provide a holding force against the extension. When an interstrand H-bond is broken, a relaxation can be observed in the parameters of the remaining ones, and other favorable conformational changes may occur in the peptide strands.

On **Figure 22** the net stabilization energy of each interstrand H-bond is shown in cases of each model as a function of the scanned terminal O---H distance. This net energy was approximated as the difference between the stabilization energy of donor-acceptor orbital overlaps calculated by the second order perturbation method built in the NBO 5.9 program and the energy that comes from the steric (Pauli) repulsion of the two corresponding occupied orbitals. This comparison was made in aqueous environment.

Figure 22 shows mainly the stabilization energy of 2nd, 3rd and 4th interstrand bonds, because the first one either has already broken or has never been present because of the water molecules positioned initially on that H-bond. It can be seen that these “curves” are not continuous in every case as the H-bonds temporary rupture when repulsion overcomes the stabilization coming from favored overlaps.

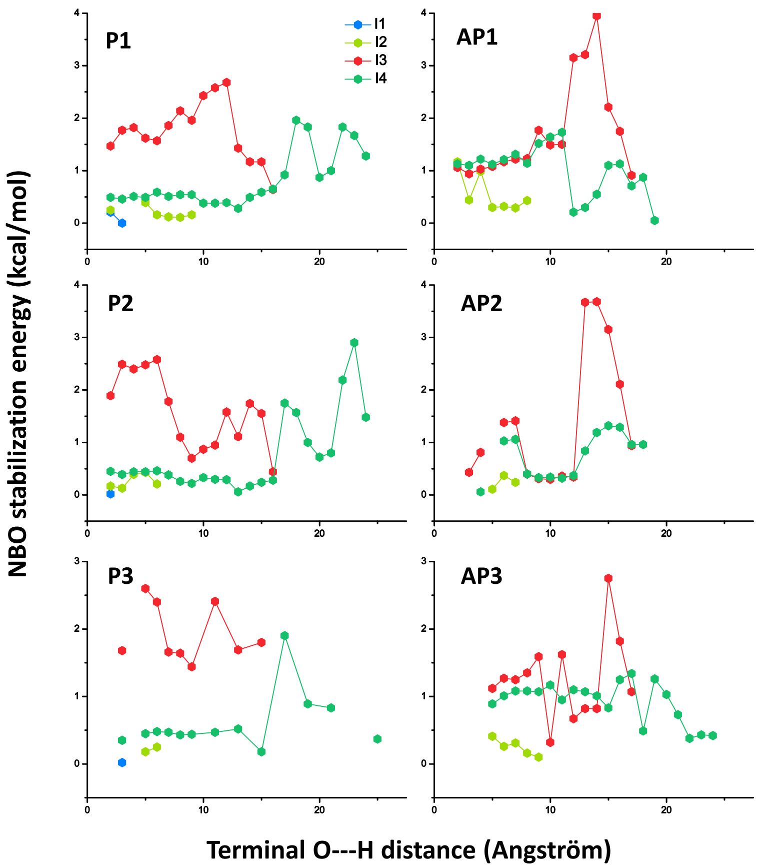

Figure 22 – The changes in stabilization energy of interstrand H-bonds for the parallel (LEFT COLUMN) and antiparallel models (RIGHT COLUMN) having 1 (TOP), 2 (MIDDLE) and 3 (BOTTOM) water molecules. Stabilization energy was estimated as the difference between the overlap energy obtained from a second order perturbation calculation and the repulsion energy caused by the Pauli exclusion principle (see the exact method of estimation in *chapter 4.3. and 4.4.*) as a function of the scanned terminal O---H distance. Blue hexagons denote the stabilization energies corresponding to I1, yellow-green ones denote that of I2, red ones denote that of I3 and sea-green ones denote that of I4, respectively. Data when repulsion overcame the overlap stabilization were not depicted.

The more water molecules are present in the structure, the more complicated are the changes, that is why the curves of models with 3 water molecule are harder to explain; one point cannot be assigned to just one event but to more and more. For this reason, we would like to highlight the important events, similarities or differences of the given models.

First, one can notice that for the parallel models in all cases a certain strengthening of the I4 H-bond can be observed when I3 breaks up, with approximately 1,5 kcal/mol stabilization energy gain. For every parallel structure during the rupture of the 3rd H-bond is assisted by one water molecule in a way that it coordinates first to the atoms involved in I3, then by systematically weakening the peptide-peptide H-bond. First the water-peptide H-bond is formed with both strands, then, when the H-bond broke up, the water molecule attaches to just one of them, forming either a W_H (P1 model) or a W_O (P2 and P3 model) type H-bond. The rupture of I3 is observed right after that the water molecule had got wedged in the peptide-peptide H-bond, which is followed by the formation of $B3_OB4_H$ in every case (structure shown in **Figure 23**). This is related to the strengthening of I4 after the breakup of I3 as the formation of this intrastrand bond makes the N-H of the 4th amide bond on strand B more polar hence the I4 H-bond becomes energetically more preferred.

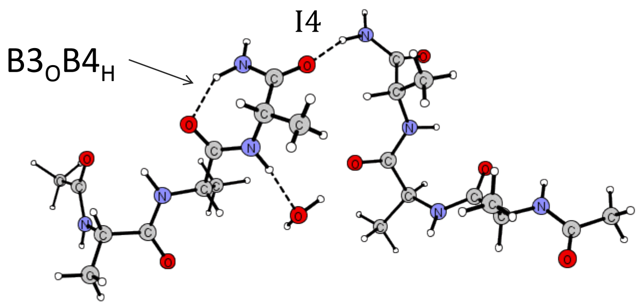

Figure 23 – Structure of P1 model at 17,4 Å, highlighting the I4 interstrand and the formed $B3_OB4_H$ intrastrand H-bond.

In antiparallel models a consistent 2.5-3 kcal/mol strengthening of I3 was observed around 12-15 Å scanning distance, which is caused by the shifting of the 3rd residue of strand A to ε_L conformation in each case. The 3rd interstrand H-bonds specifically become more stable because of the formation of a weak “improper” H-bond between the C-terminal carbonyl group of strand A and the methyl side chain of the 2nd residue of strand B (**Figure 24**), and this interaction provides a more favorable conformation and hence lower energy for I3.

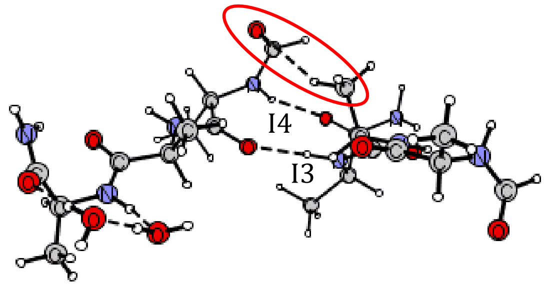

Figure 24– Structure of AP1 model at 12,2 Å, showing I3 and I4 interstrand H-bonds, and the formed intrastrand A3$_O$A4$_H$ as well.

These results also confirm the above described resistant behavior of the system: it exhibits such changes that can somehow stabilize the intermediates during unfolding, in this way trying to impede the unraveling of the structure. These effects can mainly be observed for I3 and I4, because these interstrand H-bonds represent best the features of the more buried part of a β-sheet. To understand better these phenomena, investigation of longer models will be performed in the future.

Now let us discuss the properties of water-peptide H-bonds, the main contributors to the stabilization of intermediates during unfolding, in terms of natural orbitals.

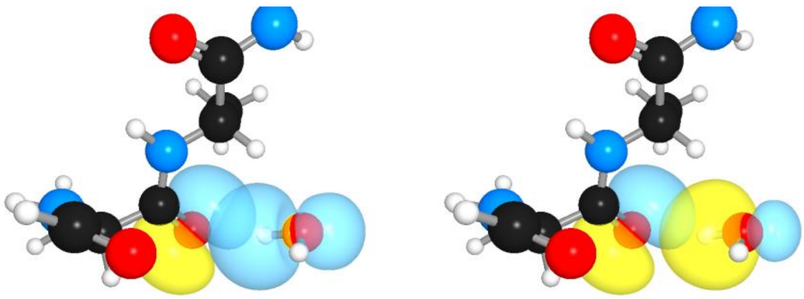

Figure 25 – H-bond between the (p-rich) lone pair of electrons localized on the second carbonyl oxygen and the antibonding orbital of the O-H bond in the water molecule when the water is coordinated on one strand in AP1 model at 21Å (LEFT). Repulsion between the same lone pair and the bonding orbital of the same O-H bond (RIGHT). All orbitals are represented in PNBO basis.

These interactions come up, since water can act in H-bonding both as a donor and an acceptor, either as the overlap of the lone pair of the water oxygen atom and the

N-H antibonding orbital (W_O type H-bond), or as the overlap of the O-H antibonding orbital of water and the lone pair of the carbonyl oxygen of an amide bond (W_H type H-bond). Consideration of repulsive terms is required here as well. The orbitals fundamentally involved in the W_H type interaction are shown in **Figure 25**.

On **Figure 26** the net stabilization energy of water-peptide H-bonds, estimated as in the case of interstrand H-bonds, is depicted as a function of the scanned terminal O---H distance. Due to the complexity of the models with three water molecules, analysis was carried out only for the models having one or two water molecules.

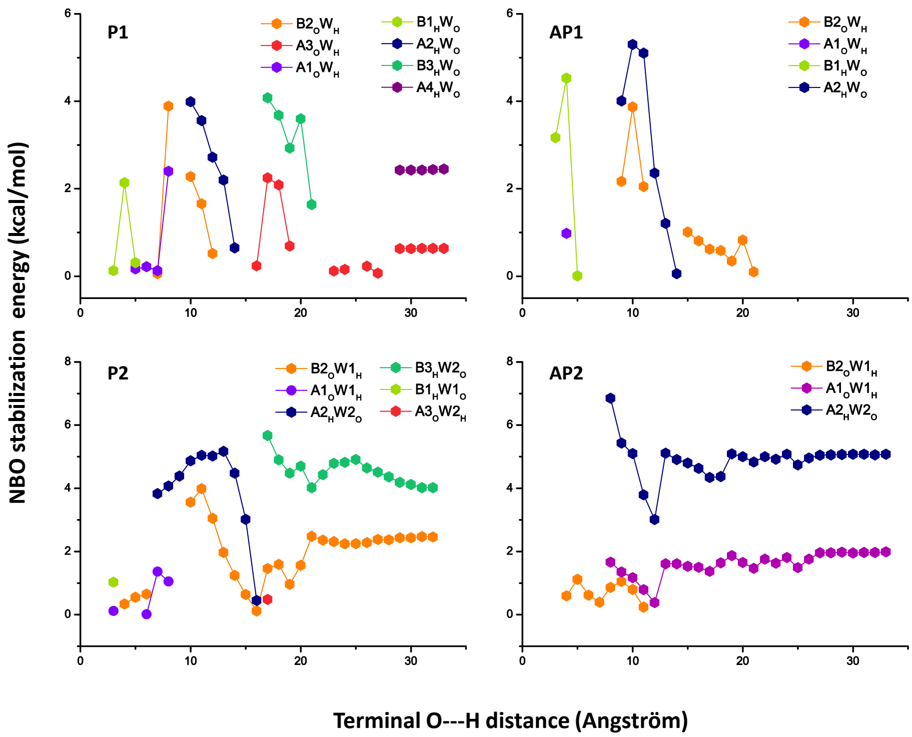

Figure 26 - The changes in stabilization energy in the case of water-peptide H-bonds as a function of the scanned terminal O---H distance. For calculation details see *chapters 4.3. and 4.4.*Labels of H-bonds are explained in *chapter 4.4.* Data when repulsion exceeds the overlap stabilization were not depicted.

Note, that H-bonds formed with the water oxygen as an electron donor (W_O) have systematically greater stability than those in which water plays the role of electron acceptor through the O-H antibonding orbital (W_H). This phenomenon is not surprising, because the O atom in an amide bond, is less polarized than the one in water, thus the former could only moderately donate electrons to any antibonding orbitals.

In the case of P1 and P2 model, the above described “water-walking” is very clearly seen, which means the moving of the water molecule(s) toward the interior of the sheet while assisting disruption of the interstrand H-bonds (**Figure 13**).

In the case of antiparallel models with 1 or 2 water molecules this cannot be observed (**Figure 26**). The water molecules do not assist the rupture of H-bond as much as for parallel models, they get soon attached to the second (AP1) or to the first and second (AP2) amide bonds of one strand. This could be due to the special structure of antiparallel sheets, namely that they consist of alternating 10-membered and 14-membered H-bond pseudorings.[31] Thus, the middle the 14-membered pseudoring has “larger separation” between amide bonds as in the parallel models with 12 atom pseudorings. In contrast, in the AP3 model such a “water-walking” occurs, which indicates that 3 water molecules are already sufficient to overcome the distance between amide bonds in a 14 atom pseudoring.

Events occurring during the forced unfolding of P1 model, are summarized on **Figure 27**. On the left hand side of this figure, the different events, namely the emerging and vanishing of interstrand (middle) and water-peptide (bottom) H-bonds demonstrated by the changes in net NBO energy, are correlated to the average distances of the two types of H-bonds as a function of the scanning distance. Major events strikingly coincide; for example formation of water-peptide H-bonds is clearly signed by the decrease in the average length of these bonds, suggesting that Natural Bond Orbital analysis provides a proper description of their presence or absence.

Most importantly, NBO analysis provides stabilization energy values decomposed to individual H-bonds. These values, which were estimated as a result of two fundamental interactions of purely quantum mechanical nature: Pauli repulsion and hyperconjugation, not only fit with H-bond lengths, but perfectly reflects the major changes in total relative energy obtained from DFT calculations. This is further strengthened by the fact that summing the energy contribution of these distinct interactions almost quantitatively reproduces the differences observed in relative energy changes of the entire model.

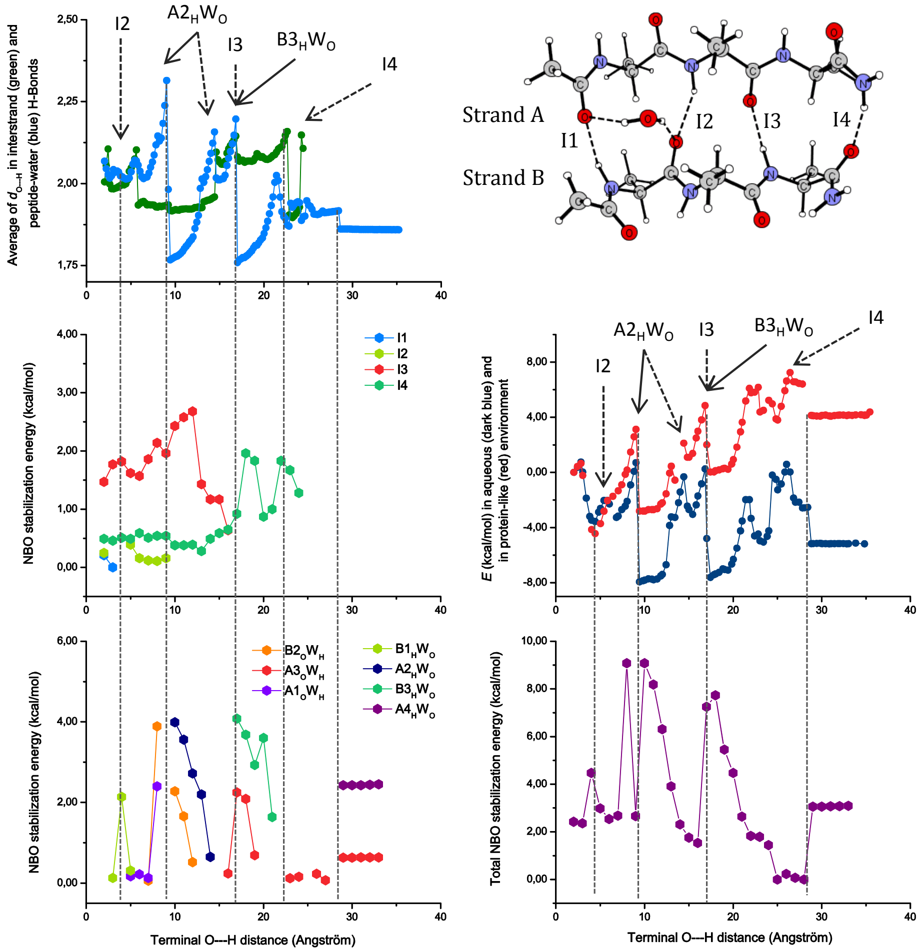

Figure 27 – Summary of unfolding in the P1 model. LEFT: The average length of interstrand peptide-peptide (green) and water-peptide (light blue) bonds (TOP), NBO stabilization energy of the interstrand (MIDDLE), and water-peptide H-bonds (BOTTOM). RIGHT: The P1 model (TOP), energetics of the unfolding obtained in water (blue) and in a protein environment (red) (MIDDLE), and the sum of peptide-peptide and water-peptide NBO stabilization energies (BOTTOM) Major changes in the H-bond pattern are labeled and marked by dashed lines.

This fact implies two important conclusions: one is that this particular proceeding, namely the forced unfolding of two trialanine strands originally bound in β-sheet conformation, can be described by decomposing the total energy to distinct interactions, in this special case to H-bonds; the other one is, that the proper consideration of effects that play the most significant roles in certain interactions is essential for the precise description of a chemical phenomenon.

The summation of the contributing interactions describes well the shape of the relative energy curves for the other models as well (**Figure 28**). NBO stabilization energies closely correlate with the changes in ΔE values during the unfolding of the investigated sheet models.

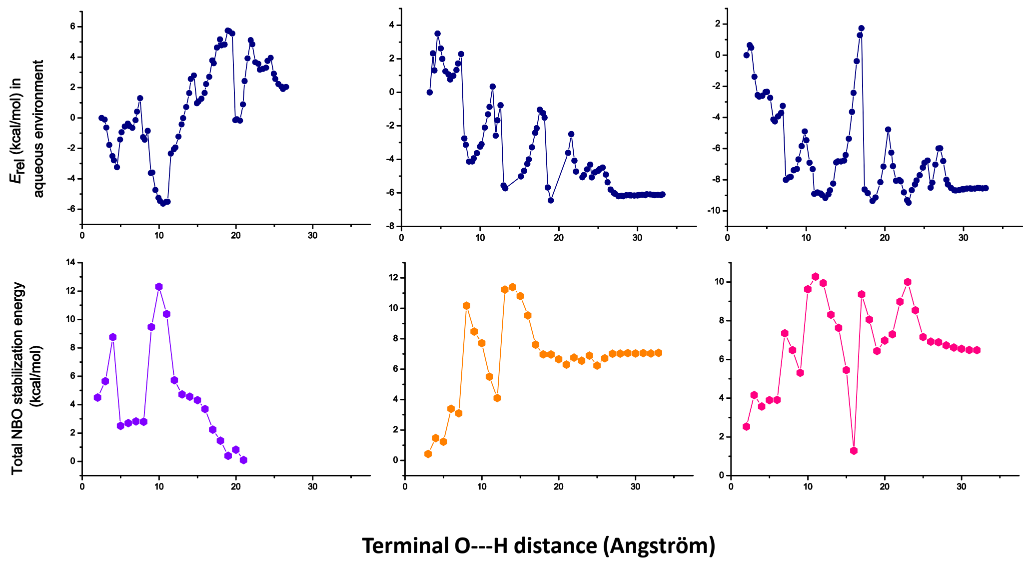

Figure 28 - On the top graphs the relative energy curves are shown obtained in aqueous environment, and on the bottom graphs the summation of stabilization energies corresponding to individual interstrand peptide-peptide and water-peptide H-bonds (obtained also in water) are depicted.

For AP1 model a good match can be observed until full separation of the two strands (~ 21 Å). In case of models having two water molecules, AP2 and P2, the two curves are also in a good agreement, but the NBO stabilization energies of intrastrand and water-water H-bonds were not considered. In the case of P2 model, despite the neglect of the above mentioned interactions, NBO energy coming from interstrand and peptide-water H-bonds matches well with the changes in ΔE.

These results indicate that the consideration of only the interstrand and the peptide-water H-bonds is appropriate and fairly describes the major part of the changes observed in relative energy. In addition, and of course somewhat straightforwardly, simultaneous consideration of the repulsive interferences and the stabilizing overlaps reproduces almost quantitatively (within ~2kcal/mol) the emerging relative energy differences during unfolding (**Figure 28**).

To further study the background of H-bonding, we wanted to reveal the changes in hybrid orbitals of the N atoms in amide N-H bonds during the unfolding, and also wanted to monitor those of the lone pairs of oxygen atoms, both the ones bound in carbonyl groups and those of the water molecule. For this, the s-character of these orbitals were monitored.

On **Figure 29** data, obtained from the same NBO calculations, are depicted for comparison with net NBO energies that represent events occurring during unfolding of AP1 model.

When an H-bond is disrupted, the s-character of N in N-H bond decreases. This is due to the higher polarity of N-H bond when taking no longer part in H-bonding, and a higher polarity of an X-H bond induces decrease in s-character in the hybrid orbital of an X atom[92].

Since p-character of hybrid orbitals provides the directionality of a bond, it is easy to explain the changes of lone pairs of O-atoms. When bound in H-bonds, their lone pairs undergo a distortion, as a consequence of which p-content increases, and s-character decreases. It is very clearly demonstrated in the case if I3 bond around 12-14Å scanning distance: while I3 bond is getting stronger, a significant decrease appears in s-character of the corresponding lone pair of amide oxygen; and in the case of water O-atom at ~5Å and ~13Å (**Figure 29**).

Considering the lone pair of the oxygen atom of the water molecule, the changes occurring are reflected by much greater differences. This is due to the greater flexibility of the lone pairs of an O-atom in water than that of those on an O-atom already bound in an amide bond having lone pairs partially delocalized in it. Changes in the s-character of the lone pair of water oxygen reflect mainly on the formation and rupture of W_O type H-bonds (5 and 12-13 Å, **Figure 29**), which is in consistence with our previous conclusion implying the greater stability of these types of H-bonds.

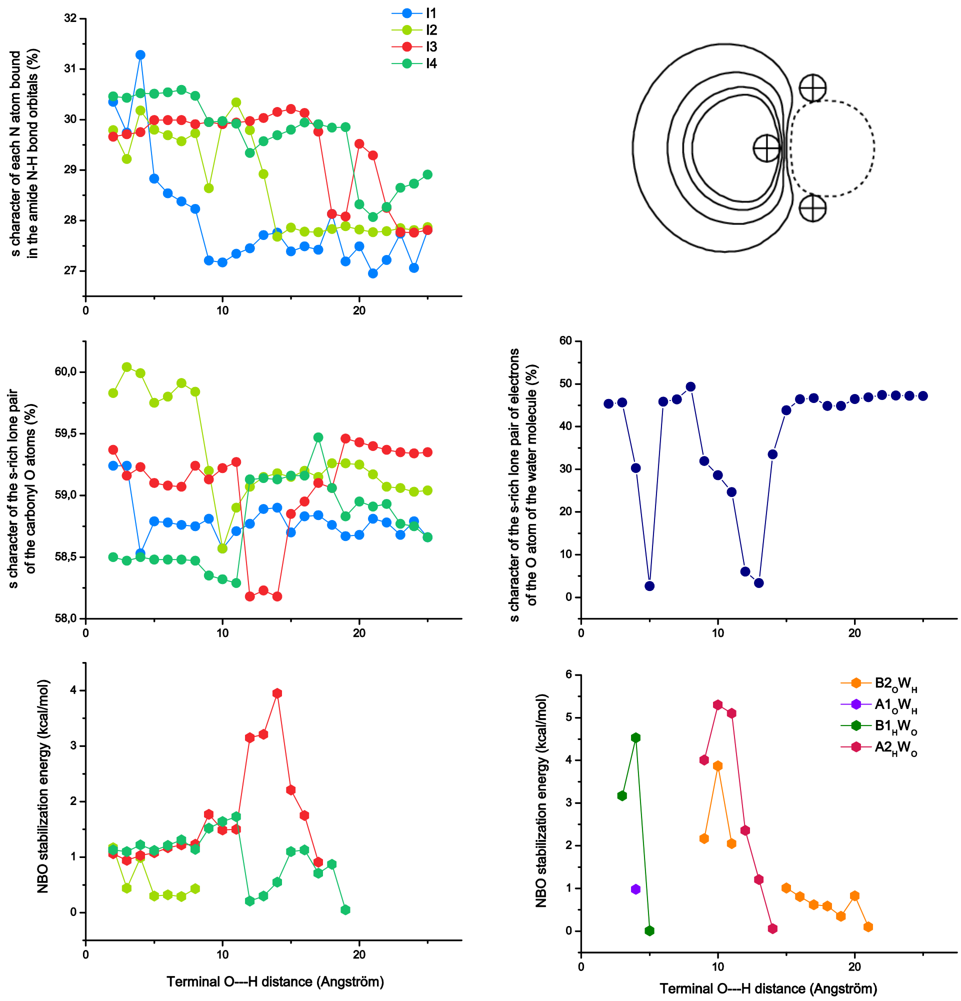

Figure 29 - Summary of unfolding in the AP1 model. LEFT: The changes in s character of amide N atoms in each N-H bond (TOP), changes in s character of the s-rich lone pair carbonyl O atoms (MIDDLE), and net NBO energies for interstrand H-bonds (BOTTOM). RIGHT: the s-rich lone pair of the water molecule in AP1 model (TOP), changes in s character of the lone pair of the water molecule (MIDDLE), and water-peptide NBO stabilization energies (BOTTOM).

As it is detailed in this chapter, using NBO methods gave insight to the orbital-level changes in the H-bonding network of our models during unfolding, and allowed a deeper investigation of the H-bonds in terms of hybrid orbitals.

6. Summary

We investigated the forced unfolding of small β-sheet models aided by first layer water molecules to understand the atomic-level details of this process, which had been widely studied both by experiments and theoretical simulations (for reference see *chapter 2.*). Since previous studies indicated that taking the effect of explicit solvent molecules into account is required[69], we used 1-3 water molecules to assist the unfolding of our models, which was combined with continuum solvent models to imitate the effect of the aqueous media and that of an apolar environment as well.

Our results showed[104], in accordance with experiments[64], that there is a significant difference in energy required to unfold a β-sheet in aqueous solution, than in apolar environment. These results explain quantitatively the different dynamic properties and behavior of membrane and water soluble β-sheet proteins and protein regions under external force, far from equilibrium. The mechanism of how first layer water molecules aid the unfolding was also shown: by forming H-bonds with the peptide backbone, the overall unfolding event is split up by several intermediate states, which are most likely separated in time. This way a significant energy gain occurs and the energy barriers of each step are lower than the total barrier without the help of explicit solvent, which may increase the overall rate of the process. The studied models reveal how water molecules stabilize the intermediate structures, and behave like "catalysts" of the unraveling of β-sheets.

Since the changes in backbone H-bonding network appeared to be the most important events during the unfolding, their formation and rupture was investigated in detail. Two kinds of H-bonds was selected for analysis, the intrastrand peptide-peptide H-bonds, which disrupt subsequently when separating the strands, and the water-peptide H-bonds, that are responsible for stabilization. We showed that the latter ones assist the breakup of the former ones, by replacing them during the unfolding. The changes in relative energy can be properly described by the alternation of H-bonds in the structure; the changes in length of the two types of H-bond correlate well with those in ΔE values.

Based on these results the decomposition of the relative energy into distinct terms assigned to individual H-bonds has become a goal of our investigations. For this, we used Natural Bond Orbital analysis, which represents interactions in terms of localized (natural) orbitals. Using those types of orbitals that are allowed to overlap with each other, we were able to estimate the stabilization energy coming from each H-bond (the previous two types were selected again). Besides the usual donor-acceptor overlaps, the consideration of repulsive interference proved also to be required. The net stabilization energy, estimated as the difference of this two interactions, corresponding to individual H-bonds provided new information about the cooperativity in H-bonding network. For example, it has been revealed, that under pulling force, the system exhibits changes in structure that generate a holding force against the pulling one; and these changes can be monitored by NBO parameters. The summation of the distinct stabilizations of individual H-bonds described remarkably well the shape and values of ΔE curves.

NBO analysis also gave us the opportunity to look into the "sub-orbital" level background of H-bonding, thus we monitored the changes of s-character of certain hybrid orbitals involved in H-bonds, which is shown to follow precisely the changes in such interactions.

It can be said, that we have investigated the unfolding of β-sheet models in three different levels: on the atomic level, while studying the formation and rupture of selected H-bonds; on the sub-atomic level by considering H-bonding in terms of orbital overlaps and repulsion interactions; and on the sub-orbital level, monitoring the changes in hybridization of selected orbitals that are involved in H-bonds.

Our results showed new and quantitative information explaining the different stability of water-soluble β-sheet regions of proteins and the ones inserted into a more hydrophobic environment, such as a β-barrel motif or a lipid bilayer. The decomposition of relative energy changes upon the forced unfolding into energy values corresponding to single H-bonds, allowed us to investigate their contribution to stabilization individually. The monitoring of changes in hybridization of certain natural orbitals in atoms involved in H-bonds provides an opportunity to follow the formation and breakup of certain H-bonds in larger molecules, helping in this way in revealing the dynamic properties of more complicated systems.

As indicated earlier[1], by estimating the length of solvent exposed and more buried β-sheet regions and using the current energetic results, a simple method can be developed to predict the mechanical stability of a particular protein with semi-quantitative accuracy. Further on, the use of NBO analysis to locate contribution of individual H-bonds to mechanical stability indicates that similar studies on more complex systems will allow localization of key stability points in proteins.

7. Összefoglaló

Számos fehérje mechanikai sajátságait tanulmányozták az utóbbi időben az egyes molekulák széthúzó erő hatására történő letekeredését vizsgálva, az úgynevezett Single Molecule Force Spectroscopy technikával[42,48]. Biológiailag fontos, főként β-redős szerkezetű fehérjék (például titin) feltekeredésének és felbomlásának energiaviszonyait is tanulmányozták ezzel a módszerrel, amelyet később kis β-redős modellszerkezetek oldószerben történő letekeredésének explicit vízmolekulákkal való modellezésével magyaráztak[1].

E folyamat mélyebb megértéséhez további kvantumkémiai számításokat végeztünk, mind poláris mind apoláris oldószermodellben, így modellezve a sejtbeli vizes közeget, illetve a sejtmembránban vagy pedig egy fehérje belsejében található β-redők apoláris környezetét. Eredményeink kvantitatív becslést adtak e két típusú környezetben előforduló fehérjék eltérő stabilitására, valamint lehetővé tették a letekeredésük folyamatának részletes magyarázatát. Ebben a folyamatban a vízmolekulák katalizátorként viselkednek: a peptid szálakkal H-kötéseket kialakítva, és ezekkel helyettesítve a felszakadó szálközi H-kötéseket, stabilizálják a megjelenő intermedier szerkezeteket, így csökkentve a letekeredés energiagátját. Mivel a H-kötésekben fellépő változások voltak a legnagyobb hatással a folyamat teljes energiaváltozásaira, ezek kialakulásának és felbomlásának részletes vizsgálatát is elvégeztük.

Ezután a Natural Bond Orbital analízis[86] segítségével a teljes energiaváltozást felbontottuk a rendszer természetes pályáinak különálló kölcsönhatásaira, valamint az ezek eredményeképpen előálló egyedi stabilizációs energiákra. Ezek összegzése, a legfontosabbnak ítélt kölcsönhatásokat figyelembe véve, mind alakjában, mind az energiaértékekben jól visszaadta a számított teljes energiaváltozást. Emellett a H-kötésben résztvevő természetes hibridpályák "összetételében" fellépő változásokat vizsgálva követni tudtuk ezen kölcsönhatások alakulását a folyamat során.

Munkánk során a β-redők külső erő hatására történő felbomlásának részleteit vizsgáltuk, és ehhez egy olyan összetett eljárást használtunk, amely a későbbiekben más, komplexebb biológia rendszerek és folyamatok diszkrét kölcsönhatásokon alapuló vizsgálatára lehet alkalmas.

8. References

[1] Beke-Somfai, T.; Perczel, A. *J Phys Chem Lett* **2010**, *1*, 1341.

[2] Zhang, Z. Q. *Acta Phys-Chim Sin* **2012**, *28*, 2381.

[3] Dill, K. A.; MacCallum, J. L. *Science* **2012**, *338*, 1042.

[4] Schaeffer, R. D.; Daggett, V. *Protein Eng Des Sel* **2011**, *24*, 11.

[5] Oliveberg, M.; Wolynes, P. G. *Q Rev Biophys* **2005**, *38*, 245.

[6] Englander, S. W. *Annu Rev Bioph Biom* **2000**, *29*, 213.

[7] Dyson, H. J.; Wright, P. E. *Chem Rev* **2004**, *104*, 3607.

[8] Michalet, X.; Weiss, S.; Jager, M. *Chem Rev* **2006**, *106*, 1785.

[9] Dyson, H. J.; Wright, P. E. *Annu Rev Phys Chem* **1996**, *47*, 369.

[10] Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. *Annu Rev Biophys* **2010**, *39*, 159.

[11] Chan, H. S. *Biophysical Journal* **2011**, *100*, 373.

[12] Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu Rev Phys Chem* **2007**, *58*, 57.

[13] Daggett, V. *Chem Rev* **2006**, *106*, 1898.

[14] Henry, E. R.; Eaton, W. A. *Chem Phys* **2004**, *307*, 163.

[15] Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *P Natl Acad Sci USA* **2000**, *97*, 13518.

[16] Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863.

[17] Gruebele, M. *Annu Rev Phys Chem* **1999**, *50*, 485.

[18] Naganathan, A. N.; Munoz, V. *Journal of the American Chemical Society* **2005**, *127*, 480.

[19] Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167.

[20] Guo, Z. Y.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.

[21] Rovo, P.; Straner, P.; Lang, A.; Bartha, I.; Huszar, K.; Nyitray, L.; Perczel, A. *Chem-Eur J* **2013**, *19*, 2628.

[22] Takano, K.; Okada, J. *Biophysical Journal* **2013**, *104*, 564a.

[23] Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J. L.; Marcoccia, J. F.; Csizmadia, I. G. *Journal of the American Chemical Society* **1991**, *113*, 6256.

[24] http://classconnection.s3.amazonaws.com/109/flashcards/1724109/png/b_sheet1344982320511.png

[25] Perczel, A.; Hudaky, P.; Palfi, V. K. *Journal of the American Chemical Society* **2007**, *129*, 14959.

[26] Dobson, C. M. *Semin Cell Dev Biol* **2004**, *15*, 3.

[27] Kopito, R. R. *Trends Cell Biol* **2000**, *10*, 524.

[28] Ross, C. A.; Poirier, M. A. *Nat Med* **2004**, *10*, S10.

[29] Benzinger, T. L. S.; Gregory, D. M.; Burkoth, T. S.; Miller-Auer, H.; Lynn, D. G.; Botto, R. E.; Meredith, S. C. *P Natl Acad Sci USA* **1998**, *95*, 13407.

[30] Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A.; Riekel, C.; Sambashivan, S.; Liu, Y.; Gingery, M.; Grothe, R. *Febs J* **2005**, *272*, 78.

[31] Pohl, G.; Jakli, I.; Csizmadia, I. G.; Papp, D.; Matias, G. F.; Perczel, A. *Phys Chem Chem Phys* **2012**, *14*, 1507.

[32] Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773.

[33] Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. O.; Riekel, C.; Eisenberg, D. *Nature* **2007**, *447*, 453.

[34] Kenney, J. M.; Knight, D.; Wise, M. J.; Vollrath, F. *Eur J Biochem* **2002**, *269*, 4159.

[35] Tirrell, D. A. *Science* **1996**, *271*, 39.

[36] Pohl, G.; Beke, T.; Borbely, J.; Perczel, A. *Journal of the American Chemical Society* **2006**, *128*, 14548.

[37] Wang, K. *Adv Biophys* **1996**, *33*, 123.

[38] Wimley, W. C. *Curr Opin Struc Biol* **2003**, *13*, 404.

[39] Richardson, J. S.; Richardson, D. C. *P Natl Acad Sci USA* **2002**, *99*, 2754.

[40] Ohnesorge, F.; Binnig, G. *Science* **1993**, *260*, 1451.

[41] Giessibl, F. J. *Rev Mod Phys* **2003**, *75*, 949.

[42] Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109.

[43] Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. *Science* **1997**, *275*, 1295

[44] Carrion-Vazquez, M.; Oberhauser, A. F.; Fisher, T. E.; Marszalek, P. E.; Li, H. B.; Fernandez, J. M. *Prog Biophys Mol Bio* **2000**, *74*, 63.

[45] Neuman, K. C.; Nagy, A. *Nat Methods* **2008**, *5*, 491.

[46] Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophysical Journal* **1998**, *75*, 3008.

[47] Oberhauser, A. F.; Hansma, P. K.; Carrion-Vazquez, M.; Fernandez, J. M. *P Natl Acad Sci USA* **2001**, *98*, 468.

[48] Kellermayer, M. S. Z.; Bustamante, C. *Science* **1997**, *277*, 1117.

[49] Wang, K.; Wright, J.; Ramirezmitchell, R. *Biophysical Journal* **1985**, *47*, A349.

[50] Horowits, R.; Kempner, E. S.; Bisher, M. E.; Podolsky, R. J. *Nature* **1986**, *323*, 160.

[51] Maruyama, K.; Matsubara, S.; Natori, R.; Nonomura, Y.; Kimura, S.; Ohashi, K.; Murakami, F.; Handa, S.; Eguchi, G. *J Biochem-Tokyo* **1977**, *82*, 317.

[52] Linke, W. A.; Rudy, D. E.; Centner, T.; Gautel, M.; Witt, C.; Labeit, S.; Gregorio, C. C. *Journal of Cell Biology* **1999**, *146*, 631.

[53] Labeit, S.; Kolmerer, B. *Science* **1995**, *270*, 293.

[54] Linke, W. A.; Ivemeyer, M.; Olivieri, N.; Kolmerer, B.; Ruegg, J. C.; Labeit, S. *Journal of Molecular Biology* **1996**, *261*, 62.

[55] Gautel, M.; Goulding, D. *Febs Lett* **1996**, *385*, 11.

[56] Bork, P.; Holm, L.; Sander, C. *Journal of Molecular Biology* **1994**, *242*, 309.

[57] Brummendorf, T.; Rathjen, F. G. *Protein Profile* **1995**, *2*, 963.

[58] Improta, S.; Politou, A. S.; Pastore, A. *Structure* **1996**, *4*, 323.

[59] http://www.rcsb.org/pdb/images/1tit_asym_r_500.jpg

[60] Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. *Science* **1994**, *265*, 1599.

[61] Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophysical Journal* **1998**, *75*, 3008.

[62] Sharma, D.; Perisic, O.; Peng, Q.; Cao, Y.; Lam, C.; Lu, H.; Li, H. B. *P Natl Acad Sci USA* **2007**, *104*, 9278.

[63] Dietz, H.; Rief, M. *P Natl Acad Sci USA* **2004**, *101*, 16192.

[64] Sapra, K. T.; Damaghi, M.; Koster, S.; Yildiz, O.; Kuhlbrandt, W.; Muller, D. J. *Angew Chem Int Edit* **2009**, *48*, 8306.

[65] Shin, Y.; Davis, J. H.; Brau, R. R.; Martin, A.; Kenniston, J. A.; Baker, T. A.; Sauer, R. T.; Lang, M. J. *P Natl Acad Sci USA* **2009**, *106*, 19340.

[66] Kellermayer, M. S. Z.; Grama, L.; Karsai, A.; Nagy, A.; Kahn, A.; Datki, Z. L.; Penke, B. *Journal of Biological Chemistry* **2005**, *280*, 8464.

[67] Karsai, A.; Martonfalvi, Z.; Nagy, A.; Grama, L.; Penke, B.; Kellermayer, M. S. Z. *J Struct Biol* **2006**, *155*, 316.

[68] Fernandez, J. M.; Li, H. B. *Science* **2004**, *303*, 1674.

[69] Dougan, L.; Feng, G.; Lu, H.; Fernandez, J. M. *P Natl Acad Sci USA* **2008**, *105*, 3185.

[70] Marszalek, P. E.; Lu, H.; Li, H. B.; Carrion-Vazquez, M.; Oberhauser, A. F.; Schulten, K.; Fernandez, J. M. *Nature* **1999**, *402*, 100.

[71] Fowler, S. B.; Best, R. B.; Herrera, J. L. T.; Rutherford, T. J.; Steward, A.; Paci, E.; Karplus, M.; Clarke, J. *Journal of Molecular Biology* **2002**, *322*, 841.

[72] Esposito, L.; Pedone, C.; Vitagliano, L. *P Natl Acad Sci USA* **2006**, *103*, 11533.

[73] Ackbarow, T.; Chen, X.; Keten, S.; Buehler, M. J. *P Natl Acad Sci USA* **2007**, *104*, 16410.

[74] Bryant, Z.; Pande, V. S.; Rokhsar, D. S. *Biophysical Journal* **2000**, *78*, 584.

[75] Meier, M.; Seelig, J. *Journal of the American Chemical Society* **2008**, *130*, 1017.

[76] Kreplak, L.; Doucet, J.; Dumas, P.; Briki, F. *Biophysical Journal* **2004**, *87*, 640.

[77] Qin, Z.; Buehler, M. J. *Phys Rev Lett* **2010**, *104*.

[78] Zhmurov, A.; Kononova, O.; Litvinov, R. I.; Dima, R. I.; Barsegov, V.; Weisel, J. W. *Journal of the American Chemical Society* **2012**, *134*, 20396.

[79] Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat Struct Biol* **2003**, *10*, 731.

[80] Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. *P Natl Acad Sci USA* **2004**, *101*, 6946.

[81] Perczel, A.; Gaspari, Z.; Csizmadia, I. G. *J Comput Chem* **2005**, *26*, 1155.

[82] Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. *Nature* **2004**, *430*, 101.

[83] Hobza, P.; Zahradnik, R.; Muller-Dethlefs, K. *Collect Czech Chem C* **2006**, *71*, 443.

[84] Cerny, J.; Hobza, P. *Phys Chem Chem Phys* **2007**, *9*, 5291.

[85] Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem Rev* **1988**, *88*, 899.

[86] Glendening, E. D.; Landis, C. R.; Weinhold, F. *Wires Comput Mol Sci* **2012**, *2*, 1.

[87] Weinhold, F. *J Comput Chem* **2012**, *33*, 2363.

[88] Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. *Nat Chem Biol* **2010**, *6*, 615.

[89] Jakobsche, C. E.; Choudhary, A.; Miller, S. J.; Raines, R. T. *Journal of the American Chemical Society* **2010**, *132*, 6651.

[90] Westler, W. M.; Lin, I. J.; Perczel, A.; Weinhold, F.; Markley, J. L. *Journal of the American Chemical Society* **2011**, *133*, 1310.

[91] Weinhold, F.; Landis C. R.; Valency and bonding - A Natural Bond Orbital Donor-Acceptor Perspective, Cambrigde University Press, Cambridge, UK, **2005**, Chapter 5.2

[92] Alabugin, I. V.; Manoharan, M.; Peabody, S.; Weinhold, F. *Journal of the American Chemical Society* **2003**, *125*, 5973.

[93] Badenhoop, J. K.; Weinhold, F. *J Chem Phys* **1997**, *107*, 5406

[94] Veszprémi Tamás, Fehér Miklós, A kvantumkémia alapjai és alkalmazása, Műszaki Könyvkiadó, **2002**.

[95] Kohn, W.; Becke, A. D.; Parr, R. G. *J Phys Chem-Us* **1996**, *100*, 12974.

[96] Gaussian 09, Revision B.01,

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr.,

J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2010.

[97] NBO 5.9. E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, and F. Weinhold (Theoretical Chemistry Institute, University of Wisconsin, Madison, WI, **2011**; http://www.chem.wisc.edu/~nbo5)

[98] Becke, A. D. *Phys Rev A* **1988**, *38*, 3098.

[99] Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys Rev B* **1988**, *37*, 785.

[100] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem Rev* **2005**, *105*, 2999.

[101] Beke, T.; Csizmadia, I. G.; Perczel, A. *Journal of the American Chemical Society* **2006**, *128*, 5158.

[102] Beke-Somfai, T.; Lincoln, P.; Norden, B. *P Natl Acad Sci USA* **2011**, *108*, 4828.

[103] Schulz, G. E. *Curr Opin Struc Biol* **2000**, *10*, 443.

[104] Papp, D.; Perczel, A.; Beke-Somfai, T,; manuscript in preparation