PALACKÝ UNIVERSITY OLOMOUC

FACULTY OF SCIENCE

DEPARTMENT OF PHYSICAL CHEMISTRY

STABILIZATION OF PROTEIN THREE-DIMENSIONAL STRUCTURE

Ph.D. Thesis

Petr Sklenovský

Olomouc 2011

Palacký University Olomouc

Faculty of Science

Department of Physical Chemistry

Stabilization of Protein Three-Dimensional Structure

Ph.D. Thesis

Olomouc 2011

Author: Mgr. Petr Sklenovský[1]   Specialisation: Physical Chemistry

Supervisor: doc. RNDr. Michal Otyepka, Ph.D.[1]

---

[1] Department of Physical Chemistry, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

I declare that I have worked out this thesis by myself using the cited references. Neither the thesis nor any of its part was previously used for obtaining any academic degree.

In Olomouc, ...

. . . .
Signature

v

"A theory is something nobody believes, except the person who made it. An experiment is something everybody believes, except the person who made it."

Albert Einstein

Contents

Chapter 1

List of abbreviations

AR	Ankyrin Repeat
ARP	Ankyrin Repeat Protein
ATP	Adenosine 5'-TriPhosphate
CATH	Class Architecture Topology and Homology
CDK	Cyclin-Dependent Kinase
DARPin	Designed Ankyrin Repeat Protein
EDA	Essential Dynamics Analysis
ESP	Electrostatic Surface Potential
FSSP	Families of Structurally Similar Proteins
HA	Harmonic Approximation
MD	Molecular Dynamics
MM	Molecular Mechanics
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PME	Particle-Mesh Ewald
PPII	a left-handed PolyProline II helix
p18	Ankyrin Repeat Protein p18^{INK4c}
QM	Quantum Mechanics
RESP	Restrained ElectroStatic Potential fit
RMSD	Root-Mean-Square-Deviation
RNA	RiboNucleic Acid
SCOP	Structural Classification Of Proteins
X-ray	X-ray crystallography
vdW	van der Waals

Chapter 2

Motivation

Significant progress in the determination of three-dimensional protein structure has been achieved throughout the last decades. Since 1958, when the first X-ray protein structure was determined, the number of protein structures considerably increased, reaching almost 70,000 items according to the PDB database (1) at the beginning of 2011. Additionally, the quality of the structural data is substantially better (e.g. higher resolution) and the size of some systems for which the structure is known exceeds 30,000 atoms (e.g. chaperonins and RNA polymerase). However, understanding of how the protein structure is related to its function (so called structure-function relationship) remains still a big challenge.

Most proteins do not operate as rigid species but they utilize their intrinsic flexibility (encoded in their structure) during the mechanism of action. With protein flexibility are, for example, associated ubiquitous biological processes such as ligand binding, enzyme catalysis and protein transport. Thus, unequivocally, knowledge of protein flexibility is a significant benefit in understanding the relationship between protein structure and function. Unfortunately, protein flexibility is currently more difficult to study in comparison to a fairly routine determination of protein structure. Besides experimental methods, among which the most popular are nuclear magnetic resonance and fluorescence spectroscopy, protein flexibility can be studied theoretically, employing molecular dynamics simulations. Molecular dynamics simulations are computational (*in silico*) methods, based on the classical Newtonian mechanics, which have a unique feature of probing the conformational space and timescales simultaneously. Usually, the molecular dynamics simulations of proteins are carried out at atomic resolution, using X-ray structures as the starting coordinates, and on timescales from hundreds of nanoseconds to microseconds (with the sub-picosecond/picosecond resolution). Thus, molecular dynamics simulations, owing to their unique property of providing atomic insight into short- and medium-ranged protein motions, serve as a powerful tool complementing experimental observations about protein flexibility.

This Ph.D. thesis is focused on the structure, stability and flexibility of three unique biomolecules: an ankyrin repeat protein p18$^{\text{INK4c}}$, a sodium-potassium ATPase and an artificial peptide chignolin. The conformational behavior of these systems was systematically investigated *in silico*, using unrestrained classical molecular dynamics simulations in explicit solvent. Moreover, the structural dynamics of sodium-potassium ATPase is discussed in parallel with tryptophan fluo-

rescence measurements. Our simulation data effectively complement current (namely experimental) knowledge about the structure and/or biological function of investigated biomolecules.

Chapter 3

Theory

3.1 Protein structure

Proteins are biopolymers composed of 20 different types of amino acids (monomeric units of proteins) which are virtually involved in all living organisms. Simultaneously with nucleic acids, and polysaccharides, proteins belong to the most abundant organic forms of matter in nature. Proteins are somehow involved in all cellular functions. For instance, some proteins are responsible for catalysis of chemical reactions (named enzymes), other permit flow of ions across cellular membranes and thus reveal the transport function, and other, named fibrous proteins, are responsible for the structural function (e.g. they form "skeleton" of cells). Among each other, proteins significantly vary in the number of amino acids involved in polymeric chains (i.e. in size) as well as in geometry which they adopt in their functionally active (native, "operating") states. Since the molecular structure of proteins is fairly complex, various levels of proteins' structural organization have been defined to simplify their study. There are four levels of protein structure.

Primary protein structure. The sequence of covalently bonded amino acids in polymeric protein chains represents so called primary structure (Figure 1A). The number of amino-acids involved in one polymeric chain of the naturally occurring protein ranges from 70 (villin headpiece) to 27,000 (a fibrous protein titin). Conventionally, the primary sequence is written from the left to the right, starting with the N-terminus and ending with the C-terminus, respectively.

Secondary protein structure. Secondary structure is a part of the three-dimensional protein structure which adopts an α-helical, β-hairpin or U-turn topology. Of course, there can be found extra types of secondary structures in proteins (e.g. PPII and 3_{10} helices), but these occur in nature rarely (2).

Tertiary protein structure. Tertiary structure of proteins relates to their overall three-dimensional shape (Figure 1B). The tertiary structure provides insight into the mutual arrangement of amino acid residues, direction of peptide bonds (or polymeric chain(s)), relative distribution

of polar and non-polar amino acids between the protein core and surface, network of stabilizing tertiary interactions (namely H-bonds), etc.
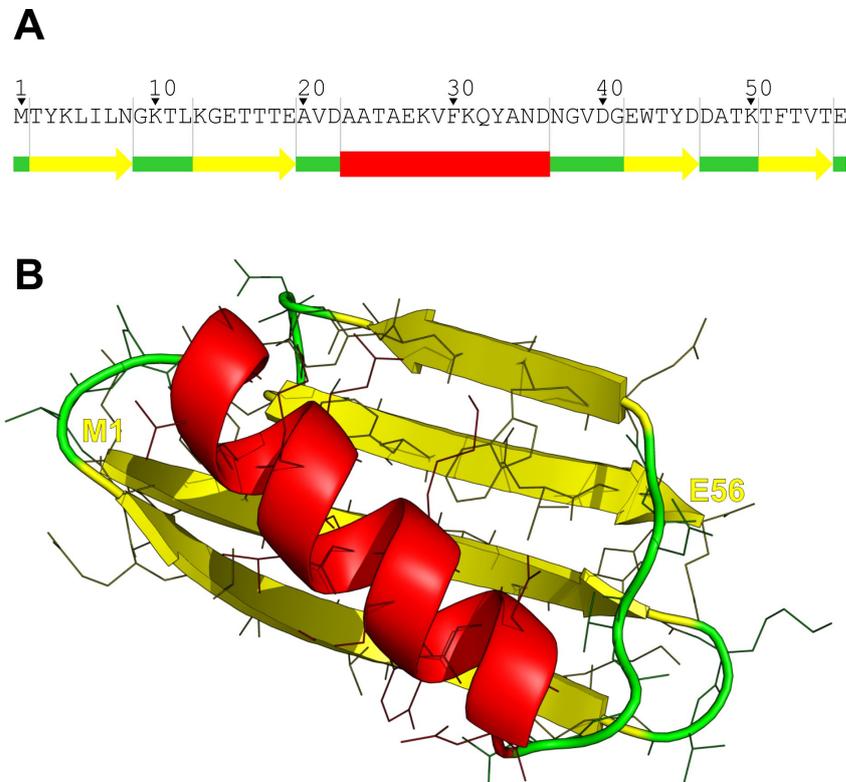
Figure 1 (A) Primary structure of protein G B1 domain (PDB ID 1PGA). The green, yellow and red bars mark residues with a random coil/loop, β- and α-helical structure, respectively. (B) Tertiary structure of the B1 domain with residues M1 and E56 highlighting the N- and C-terminus, respectively.

Quaternary protein structure.   Proteins are often composed from multiple folded polypeptide chains which interact witch each other only via non-covalent interactions (namely H-bonds, salt bridges, and vdW forces). Conventionally, this type of protein structure is called "quaternary protein structure" and was firstly observed for hemoglobin which is formed from four polypeptide chains (subunits): $2 \times \alpha$ and $2 \times \beta$ (3). Streptavidin, revealing an extraordinarily high affinity for biotin (vitamin B7), is another example of protein with a tetrameric quaternary structure (see Figure 2).

3.1.1   Classification of protein structures

Protein classification with respect to environmental conditions.   Proteins exist under various environmental conditions, which leave an obvious mark on their structures. The less water there is around proteins, the more important the hydrogen bonds (which stabilize the regular, periodic three-dimensional structures of the protein backbone) are, and the more regular the stable protein structure ought to be. Consequently, protein structure is divided according to the
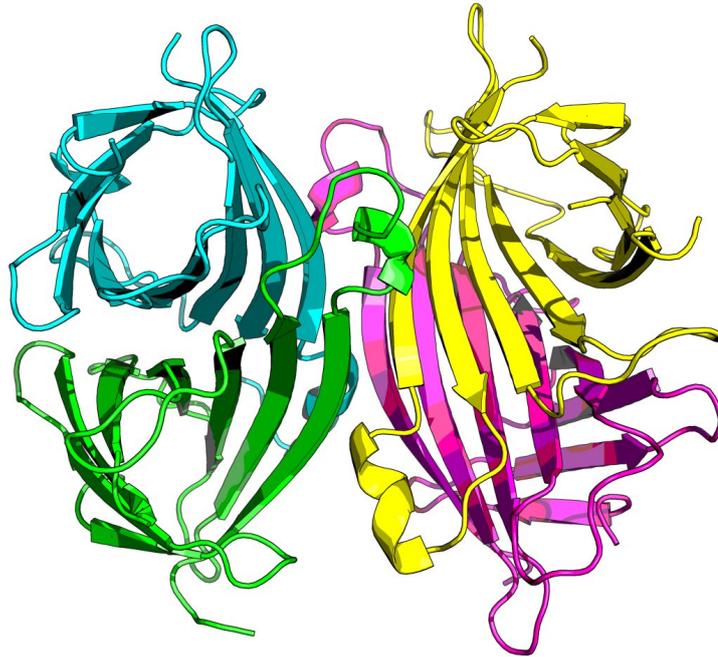
Figure 2 Quaternary structure of streptavidin composed of four subunits, each being approximately 160-residues-long and displaying the β-barrel architecture.

"environmental conditions" into three classes: (i) Fibrous proteins, (ii) Membrane proteins, and (iii) Water-soluble globular proteins (3).

Protein classification according to secondary element content. From the bioinformatics point of view, protein structures are classified according to their similarity in terms of the content and mutual arrangement of secondary structures (4). There are three most frequently used strategies of protein classification which among each other slightly differ in algorithmization: (i) FSSP (5), (ii) CATH (6), and (iii) SCOP (7). For instance, the CATH algorithm sorts out protein structures according to similarity in structural levels which are as follows: Class, Architecture, Topology and Homology. The Class informs us about the relative abundance of α- and β- secondary structures. Thereby, known are four types of protein structures in the Class level, involving only α, $\alpha+\beta$, only β and little/no secondary structures. The Architecture reveals similarity among proteins with respect to the orientation of secondary structure elements (loops are ignored). The Topology (also fold) describes the overall protein shape including relative arrangement of secondary structures and loops that connect them. Notably, the Topology level of classification focuses only on the inner part of proteins (protein cores) while the exposed structural moieties remain unnoticed. Finally, in the Homology level are proteins clustered with respect to their evolution (6).

Proteins are classified according to their structure, primarily, in order to fully understand the global relationships between their sequence, structure and function (8).

3.1.2 Protein flexibility

Proteins are not entirely rigid molecules. They possess an inherent flexibility that allows them to function through molecular interactions within the cell, among the cells and even between organisms. Flexibility allows proteins to respond to the presence of other molecules and/or to variations in the environment (9). Signal transduction, antigen recognition, protein transport and enzyme catalysis (to some extent) rely on proteins' ability to undergo conformational changes. In addition, protein flexibility bears a strong relationship to unfolding since proteins most likely start unfolding and finish folding at sites that are highly mobile (10).

Proteins exhibit two types of flexibility, systemic and segmental. Systemic flexibility is associated with small-scale (high frequency) fluctuations of main chain and side chain atoms in native conformations of proteins. In contrast, segmental flexibility can be described as a motion of one part with respect to another, which is of a large-scale and occurs on time scales significantly longer in comparison to the systemic flexibility. For the segmental flexibility is usually responsible small part of the protein structure which is often labeled hinge. Segmental flexibility is frequently related to protein function (11).

The timescale of the conformational events that underline protein flexibility spans 13 orders of magnitude (9) (see Figure 3). Bond vibrations and side chain rotations are the fastest motions occurring on the picosecond to nanosecond timescale. On the other hand, protein-ligand dissociation, and protein folding/unfolding are relatively slow processes observable on timescales from microseconds to hours (9, 12).
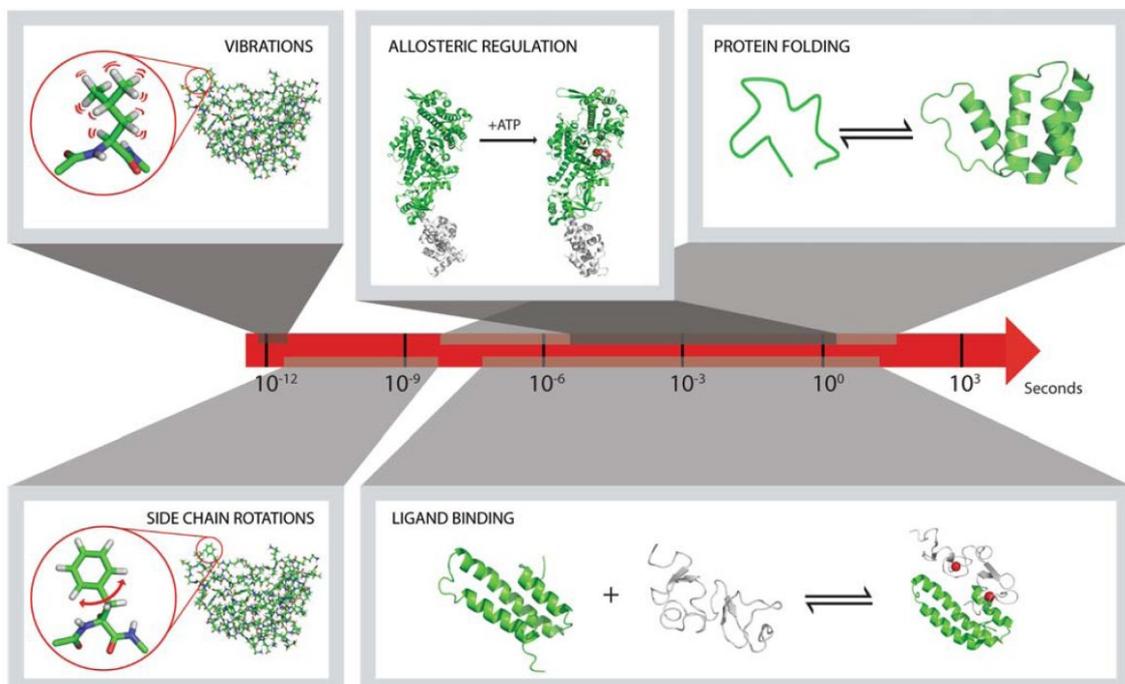
Figure 3 Timescales of different protein motions (processes). Bond vibrations and side chain rotations belong to the fastest motions while folding of proteins as well as ligand binding are slow processes which may happen with time constants of hours. The figure was adapted from ref. (9).

Protein flexibility can be studied using experimental as well as theoretical techniques. The most popular among the experimental methods is spectroscopy (e.g. NMR and fluorescence spectroscopy). Crystallographic B-factors are also used as flexibility reporters. However, one must take care during their interpretation since, e.g. crystal-packing interactions may affect them (13). In the field of theoretical studies of protein flexibility dominate molecular dynamics (MD) simulations and normal mode analysis (NMA). In NMA, the simple harmonic motions of the molecule about a local energy minimum are calculated (14).

3.2   Protein stability and folding energetics

Folding of proteins is accompanied by a subtle free energy change. In general, the free energy of a folded and biologically active protein is lower by only 5-15 kcal/mol in comparison to its unfolded (denatured) state (15). The marginal stability of the native functionally active protein over its unfolded state has a great biological importance. Living cells need globular proteins in correct quantities at appropriate times. So proteins must be easily degradable as well as foldable. Moreover, the catalytic activities of enzymes, and other important functions of proteins, generally require some structural flexibility, which would not be compatible with a rigidly stabilized structure (16). For the structure and stability of proteins are largely responsible non-covalent interactions. Non-covalent interactions are sufficiently strong to maintain a protein globule compact and stable under physiological conditions, however, their unique physico-chemical nature facilitates protein flexibility and reversibility of biological processes associated with protein molecular function (e.g. molecular recognition) (17).

3.2.1   Interactions stabilizing tertiary protein structure

Electrostatic, hydrogen-bonding (H-bonding), and van der Waals interactions fall into the most important non-covalent interactions responsible for protein stability. Next, to the protein stability contribute in rare cases the covalent disulfide bonds. Finally, to the folding free energy also significantly contribute intrinsic propensities and hydrophobic interactions, but these forces lack the "directionality" (i.e. are nonspecific and orientation independent) and thus cannot be treated as regular interactions (18, 19). Typically, the non-covalent interactions are classified to long-ranged and short-ranged, on the one hand, while local and nonlocal, on the other. Local interactions are those among chain segments that are connecting neighboring ($i \ldots i + 1$), or nearly neighboring residues in the sequence while nonlocal represent tertiary contacts which are sequentially distant (18).

Electrostatic interactions.   Long-ranged interactions in proteins, of a purely electromagnetic nature, which can be either attractive or repulsive, are called the electrostatic interactions. The extent of electrostatic interactions in a protein is dependent on pH and ionic strength. At low pH proteins are charged positively while at high pH they bear the negative net charge. At both pH extremes the electrostatic interactions are largely repulsive and native structure destabilizing because the charge density on the folded protein is greater than on the unfolded. Salt-bridge (ion pairing) is a specific electrostatic interaction in proteins which occurs when the opposite charges are in a close proximity. The salt-bridge is an attractive stabilizing interaction, which was for a long time considered to be the dominant force of protein folding (18). In fact, surface ion pairs only weakly contribute to protein stability. Interestingly, salt-bridges and their networks appear to constrain protein flexibility (11).

Hydrogen bonding.   A hydrogen bond occurs when a hydrogen atom is shared between generally two electronegative atoms. The dominant component of hydrogen bond is electrostatic but other

terms such as namely charge transfer and dispersion also contribute. In proteins, we distinguish two types of H-bonding interactions, that is, peptide and side-chain H-bonds. Peptide H-bonds are formed between backbone carbonyl (–CO) and amide (–NH) groups. Peptide H-bonds are believed to be the driving force of helix formation (18).

Van der Waals interactions. There are both attractive and repulsive vdW forces that control protein stability. The London dispersion represents the attractive interaction while the Pauli exclusion principle relates to the repulsive term. Both interactions are short-ranged with the repulsion acting on shorter distances (r^{-12} energy dependence) compared to the dispersion (r^{-6} energy dependence) (20). Frequently, in the literature (e.g. in the Dill's review (18)), dipole-dipole (or higher multipole-higher multipole) as well as induced dipole-dipole interactions are also termed vdW interactions, primarily because of their short-range. However, the dipole-dipole (or generally higher multipole-higher multipole) interactions from the physico-chemical point of view are rather of the electrostatic nature while induced dipole-dipole interactions (polarizable interactions) are classified as a special type of interactions (20). Therefore, in the next chapters of this thesis, vdW interactions will only be treated as the sum of London dispersion and Pauli repulsion. Although vdW forces are extremely weak, in comparison to other forces governing protein conformation, it is the huge number of such interactions that occur in large proteins that make them significantly contributing to protein stability and folding (see e.g. ref. (21)).

Disulfide bonds. For protein stability are also responsible covalent cross-links that connect two segments of the polypeptide backbone. A common cross-link is the disulfide bond (–S–S–) formed by mutual oxidation of two cysteines (their sulfhydryl groups) (19).

3.2.2   The enthalpic and entropic contributions to protein folding

From the thermodynamic point of view, the free energy of protein folding (ΔG_{fold}) is equal to the difference of two components: change of folding enthalpy (ΔH_{fold}) and change of folding entropy multiplied by the thermodynamic temperature ($T\Delta S_{fold}$), i.e.,

$$\Delta G_{fold} = \Delta H_{fold} - T\Delta S_{fold}. \tag{3.1}$$

In general, the changes of enthalpic as well as entropic terms are in their absolute values at least an order of magnitude larger than the overall free energy change (9, 22).

Among the enthalpic contributions (forces which are responsible for the change of enthalpy upon protein folding) belong the electrostatic interactions, vdW interactions, H-bonds and disulfide bonds (16).

The entropy term refers to the "conformational freedom" of a protein (its disorder). The change of backbone conformational entropy is the largest single factor in the energetics of protein folding that opposes the folding process (i.e. this factor has a large unfavorable value in all conditions) (19). The opposing force to backbone conformational entropy, which is at physiological temperature also entropic but thermodynamically favorable (18), is associated with hydrophobic interaction (also known as hydrophobic factor/effect). The hydrophobic interaction is associated

with low solubility of non-polar residues in water. In the unfolded state, proteins expose almost all non-polar residues to water. Around the non-polar residues (solutes), water is more organized and has lower entropy in comparison to the bulk water (20). Once the protein is folded, the non-polar residues are buried and in tight contact with each other preventing water to interact with them. In other words, the area (number) of non-polar sites decreases upon protein folding and water can be more disordered, coinciding to the increase of its entropy. The hydrophobic interaction simply accompanied with aversion of water of the non-polar residues is suggested to be the dominant force of protein folding and generally, assembly of hydrophobic particles (18, 19, 23).

3.3   Ankyrin repeat proteins

Ankyrin repeat proteins (ARPs) play critical regulatory roles *in vivo* since they mediate specific interactions with target biomolecules and thus order the sequence of events in diverse cellular processes (24, 25). For instance, the INK4 family of ARPs is responsible for specific inhibition (inactivation) of cycline-dependent kinases CDK4 and CDK6 between the cell cycle phases G1 and S (26). ARPs are virtually involved in organisms from all phyla. However, majority of them have been identified in eukaryotes (27). ARPs have a unique elongated tertiary structure composed of repeating motifs called ankyrin repeats arranged in tandem "quasi one-dimensional" arrays (Figure 4A). Ankyrin repeat (AR) is a ∼33-residue-long supersecondary structural unit formed of β-hairpin, helix-turn-helix and loop elements (Figure 4B). The number of ankyrin repeats in naturally occurring ARPs moves from 4 to ∼30. The vast majority of ARPs contains 5 to 6 ankyrin repeats. Interestingly, larger ARPs (ca. 10 and more ARs) adopt a C-shaped structure and have very interesting elastic properties, which allow them to act as nano-springs (28).

3.3.1   The minimal stable motif of ARPs.

So far, ARPs with less than 4 ankyrin repeats have not been found in nature. Nonetheless, Zhang and Peng expressed the truncated variant of the INK4 family protein p16$^{\text{INK4a}}$ (p16) containing only the last two C-terminal ARs of the wild type protein. This truncated p16 variant exhibited structure highly similar to that of the C-terminal part of the native p16 protein (29). In principle, the study of Zhang and Peng initiated the researchers in the field of protein engineering to design mutated ARPs (so called designed ankyrin repeat proteins/DARPins) containing only 2 or 3 ARs with interesting thermodynamic and binding properties (30). Moreover, the question concerning "What is the minimal stable unit (motif) of ARPs?" has been anticipated by Zhang and Peng. In many experimental as well as theoretical studies was suggested that a single AR is not able to fold to its ARP-like structure. The free energy of single AR folding is positive, what is a thermodynamically not favorable process (31). On the other hand, Zhang and Peng suggested that a pair of ankyrin repeats is stable. The most interesting results associated with this objective were published in 2005 by Ferreiro *et al.* (32) who concluded that the minimal folding motif of ARPs is composed of one complete AR connected with the β-hairpin to the N-terminal α-helix of the subsequent AR (Figure 5). Unfortunately, this interesting conclusion was supported by folding simulations based on the coarse-grained Gõ-type potential which has some approximations and limitations in comparison to the classical MD simulations (e.g. the Gõ-potential takes into account only the attractive contacts derived from the initial structure and thus not allows domain swapping). This fact prompted us to investigate the Ferreiro's hypothesis using classical MD simulations of multiple fragments of the ARP p18$^{\text{INK4c}}$ (p18) (see Synopsis of results).

3.3.2   ARPs vs. globular proteins.

The modular architecture of ARPs facilitates experimental and theoretical studies which are not possible to perform with standard globular proteins. The list of unique ARPs' features is as follows:
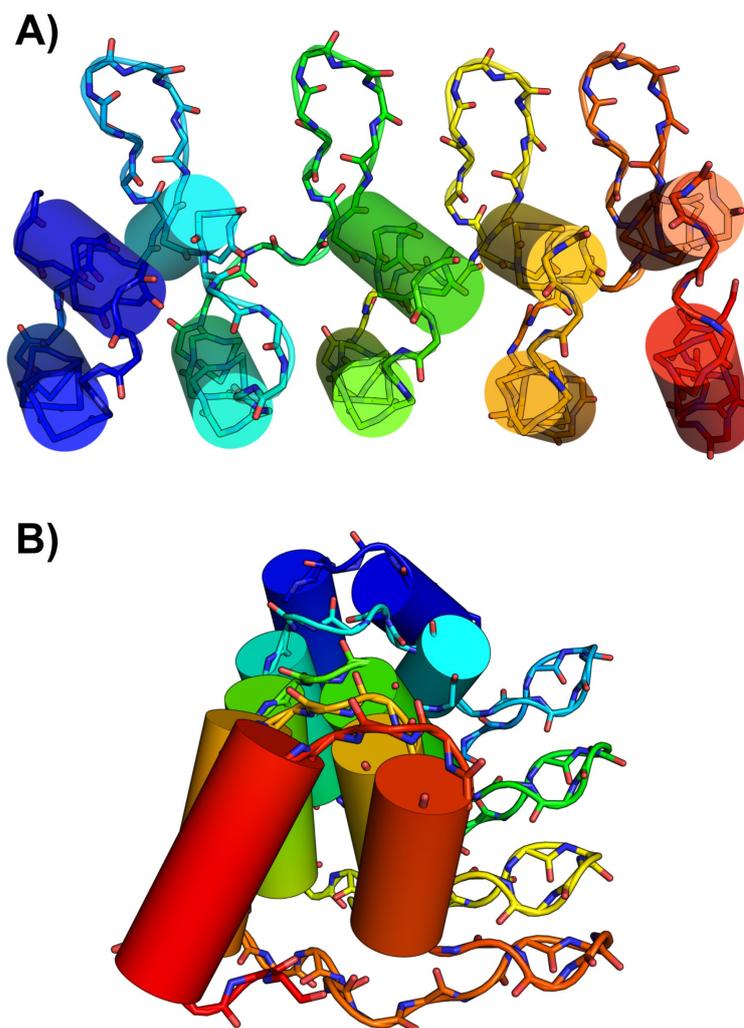
Figure 4 (A) Perspective view onto the structure of p18^{INK4c} with ARs I and V on the left and right, respectively. (B) Side view (from the C-terminal end) to the p18^{INK4c} structure showing the C-terminal AR V and perpendicular projection of loops from the ten-helical bundle to solvent.

(i) ARPs due to their elongated architecture lack long-range (sequentially distant) tertiary contacts. As a result, ARPs are ideal candidates for monitoring of the folding cooperativity. In classical globular proteins, folding cooperativity is not possible to study in such a detailed level due to the heterogeneous (irregular) and compact architecture of globular proteins (33).

(ii) ARPs tolerate insertions, deletions, and substitutions of ankyrin repeats. Thus, the evolution of ARPs was very likely not only accompanied by point mutations but also repeat deletions and insertions occurred. In many experimental studies, the effect of repeat deletion, substitution, and insertion on ARPs structure, folding and stability was investigated (see e.g. ref. (34)).

(iii) Consistently with the low density of long range contacts, ARPs' folding is faster in comparison to the globular proteins.
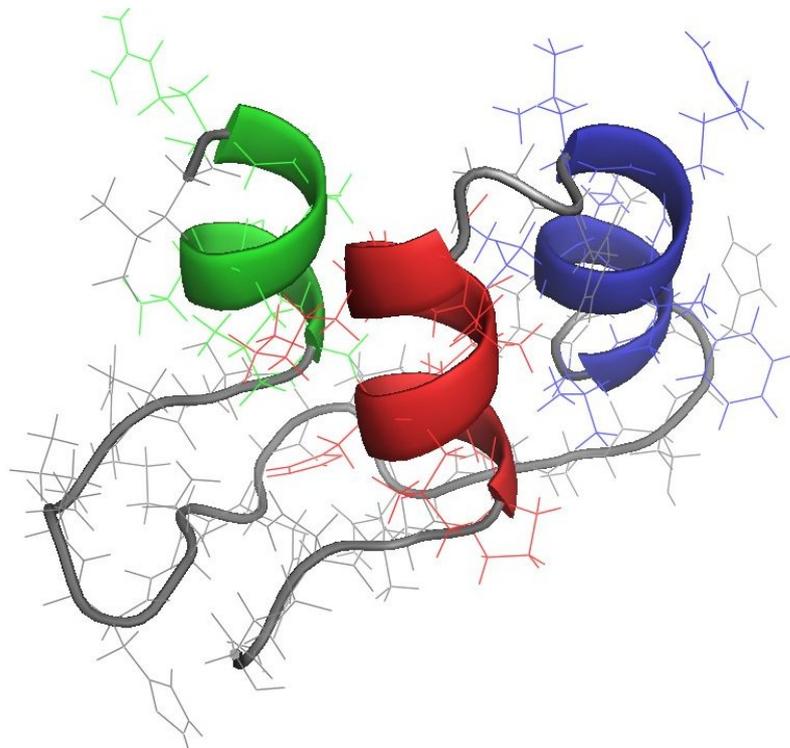
Figure 5 Three-dimensional structure of the Ferreiro's minimal folding motif of ARPs composed of one complete AR (red and blue helices with associated turns and loops) linked with β-hairpin to the N-terminal helix (green) of the consecutive AR.

(iv) Folding of ARPs frequently proceeds by multiple parallel pathways (35).

3.3.3   The ankyrin repeat protein p18$^{\text{INK4c}}$.

P18$^{\text{INK4c}}$ (p18) is a member of the INK4 family of proteins (for the INK4 *in vivo* role see above). There are in total four proteins in the INK4 family, all of which have modular (repetitive) architecture composed of four/five ankyrin repeats. P18 contains five ankyrin repeats in its ~160-residue-long polypeptide chain (26). In p18, all ankyrin repeats are stacked onto each other over their hydrophobic helix-turn-helix sides forming the ten-helical bundle with four β-hairpins orthogonally projecting to the solvent (see Figure 4B). The guanidine hydrochloride-induced denaturation at 293 K (two-state model fit) revealed that the p18 unfolding free energy (~3 kcal/mol) is relatively small in comparison to standard globular proteins (36). The stabilization free energy of the p18 structure (more exactly of the ten-helical core) was recently calculated in the theoretical study by Otyepka *et al.* (22). The core stabilization free energy based on empirical potential only was equal to 13.9 kcal/mol, while the scaled quantum-chemical methods yielded the free energy of -29.6 kcal/mol (22, 37).

3.4   Sodium-potassium pump

Active transport is the movement of a chemical compound against its concentration gradient (38). Active transport is used for concentration adjustment of various monovalent (e.g. Na^+, K^+) and divalent (Ca^{2+}, Cd^{2+}, Hg^{2+}) cations in the cytoplasm of most cells. If some energy for the translocation of the cellular stuff is needed (it can be obtained e.g. from ATP hydrolysis or light) we term it the primary active transport. Secondary active transporter uses the translocation of one solute downhill its electrochemical gradient for the translocation of other solute uphill its electrochemical gradient, e.g. the Na^+/Ca^{2+}-exchanger. Finally, passive transport does not require any energy for the translocation.

Na^+/K^+-ATPase is a membrane protein, being the member of P-type ATPases family. It is responsible for active transport of K^+ and Na^+ ions across the plasma membrane of most animal cells. The minimal functional unit consists of two subunits, of which the catalytic α-subunit is highly evolutionarily conserved over the entire P-type family. The α-subunit comprises three domains which are projected from the 10-helical transmembrane domain to the cytoplasm (see Figure 6A). The domains are labeled N (nucleotide-binding), P (phosphorylation), and A (actuator). The β-subunit is specific only for Na^+/K^+- and H^+/K^+-ATPases and it is essential for proper maturation of the enzyme into the plasma membrane (39) and probably also assists to the K^+-countertransport (40).

During the pumping cycle three Na^+ ions are translocated from the cytoplasm to the extracellular space while two K^+ ions are passed through the plasma membrane in the reverse direction. The energy for the active transport is derived from ATP which binds to the N-domain and subsequently phosphorylates the conserved aspartate located in the P-domain. The catalytic cycle of Na^+/K^+-ATPase is described in the Albers-Post scheme ((41, 42), see Figure 6B). The Albers-Post scheme postulates that Na^+/K^+-ATPase adopts two conformations during the catalytic cycle which are labeled E1 and E2. These two states differ in the affinity of the pump to Na^+, K^+ and ATP. In the E1 state, the cation binding sites, which are located in the transmembrane segment, are exposed to the cytoplasm, while in the E2 state toward the extracellular space. However, significant conformational changes between E1 and E2 states occur also in the cytoplasmic part, including domains N, P, and A (43).
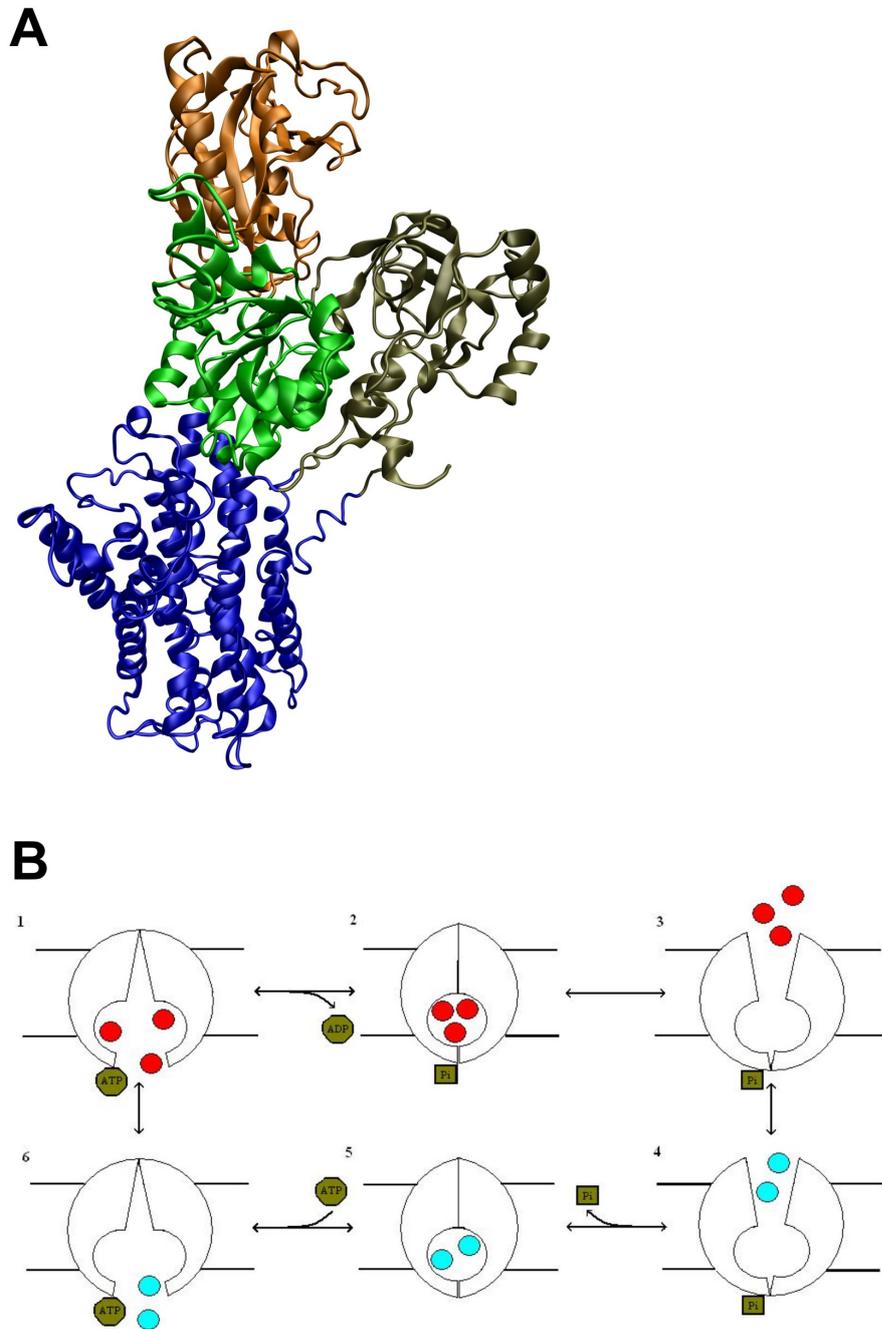
Figure 6 (A) Three-dimensional structure of the Na^+/K^+-ATPase α-subunit with the transmembrane domain colored in blue while N-, P- and A-domains in orange, green and black, respectively. (B) Albers-Post scheme illustrating individual pumping cycle steps of Na^+/K^+-ATPase. (1) Three Na^+ ions (red) bind to the pump which is in the E1 state and liganded with ATP. (2) Hydrolysis of ATP and subsequent phosphorylation of the conserved Asp occur. Then, the transition of the pump from E1 to E2 state is observed. (3) In the E2 state, the Na^+ ions are released to the extracellular part. (4) Two K^+ ions (cyan) bind to the pump which is in the E2 state and has K^+ binding sites exposed to the extracellular space. (5) The pump is dephosphorylated and assumes the E1 state which facilitates binding of ATP. (6) The K^+ ions are released to the cytoplasm and the pump with ATP liganded is ready to bind three Na^+ ions again. Thus, the cycle can be repeated. In each panel, the extracellular space is on the top while the intracellular on the bottom. The illustration was adapted from ref. (44) with author's permission.

3.5   Chignolin: a model of an ancient protein

So far, chignolin is the shortest amino acid sequence (–GYDPETGTWG–) which is able to spontaneously adopt a unique three-dimensional structure at *in vitro* conditions. Chignolin consists of only 10 residues which adopt a β-hairpin topology (Figure 7A). Chignolin was rationally designed in 2004 by Honda *et al.* according to the template scaffold taken from the protein G B1 domain (residues 45-52; (45)). Even though chignolin belongs to peptides, it is often labeled as mini-protein since its folding mechanism strikingly resembles that of the classical protein domains (see below). In fact, chignolin sequence is a consensus of ~10,000 naturally occurring β-hairpin segments. So, it is straightforward why Honda with co-workers, who are convinced that the protein evolution is divergent, call chignolin an ancient protein (more detailed explanation below) (45). Due to its small size and high foldability, chignolin is very often employed in theoretical studies assessing the accuracy of force fields for protein folding simulations (46).
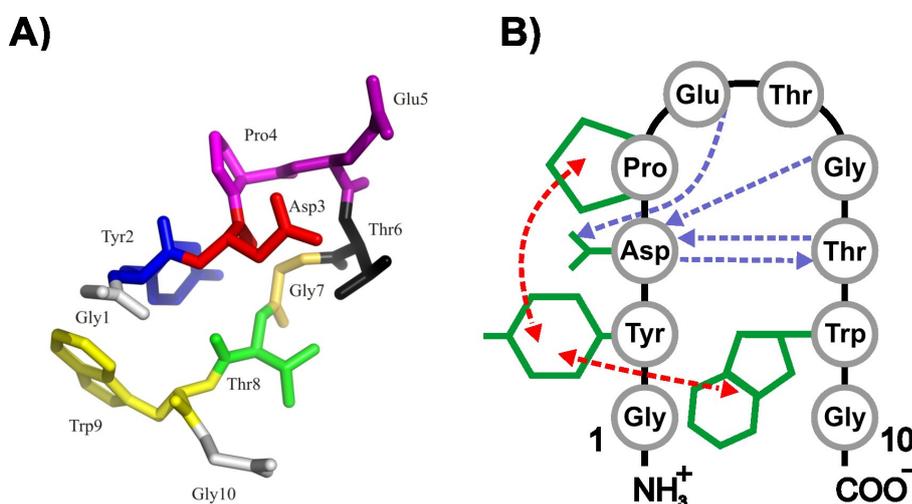
Figure 7 (A) Three-dimensional structure of β-hairpin peptide chignolin. (B) Dominant non-covalent interactions that stabilize the chignolin structure. Blue and red arrows mark H-bond and hydrophobic interactions, respectively.

3.5.1   Chignolin stability.

Chignolin is stabilized by cross-strand H-bonds between atoms Asp3(O) and Gly7(N) (the strongest), Asp3(N) and Thr8(O), and Asp3(O^{δ}) and Glu5(N) (see Figure 7B). The H-bonding pattern of chignolin, which is classified as [4:4] type (47), resembles that of the residues 45-52 of the template G B1 protein. In addition, chignolin is stabilized by hydrophobic interactions between aromatic side chains of Tyr2 and Trp9 (Ar-Ar T-shaped interaction) as well as Tyr2 and Pro4 (CH-π interaction) (Figure 7B). There is also electrostatic attraction between the negatively and positively charged C- and N-terminal ends, respectively (48). Among the most critical residues for the chignolin stability belong Asp3, Pro4, and Gly7. Residues Asp3 and Pro4 are responsible for arrangement of the backbone for H-bonding and initiation of the turn while Gly7, having the left-handed α-helical conformation (not favored by non-Gly residues), completes the turn (49).

3.5.2   Chignolin folding.

The unfolding free energy of chignolin equals to 1-2 kJ/mol ($\sim$0.25-0.5 kcal/mol), which is a considerably smaller value in comparison to regular proteins. However, the per-residue enthalpy and entropy changes upon chignolin unfolding (2.7 kJ/mol.residue and 8.8 J/mol.K.residue, respectively) are in a consistency with regular proteins (45). Moreover, in accord with classical protein domains is the concerted two-state mechanism of chignolin folding/unfolding as well. Chignolin folding time is estimated to be 1-2 μs (46). Interestingly, the native β-hairpin structure represents the global free energy minimum of chignolin (45).

3.5.3   The biological importance of chignolin.

Although chignolin is a synthetic biopolymer demonstrating not any *in vivo* function, its development has a significant impact on the hypotheses concerning the evolution and folding of proteins. Chignolin was designed from segments of naturally occurring proteins using the hairpin sequence of B1 domain of protein G (GB1 hairpin) as the template. Intrinsically, the 16-residue-long GB1 template is stable, having the ratio of folded to unfolded conformations equal to 50:50% at 300 K (50). However, the truncated variant of GB1 consisting of only 8 residues (the core loop and its flanking residues) is highly unstable in solution. Using the rational design, Honda with co-workers managed to find the GB1 hairpin variant (mutated variant) which is intrinsically foldable to β-hairpin bearing only 8 residues and called it the chignolin. In fact, chignolin corresponds to a consensus sequence of hundreds of naturally occurring hairpins when we consider the strategy of chignolin search employed by Honda *et al.* As a result, it is very likely, that many naturally occurring β-hairpins, which share the structure and sequence with chignolin significantly, are intrinsically also foldable to β-hairpin. Such autonomous folders may be *in vivo* responsible for initiation of protein folding ("folding triggers") and organizing of the overall native protein structures. Example of chignolin-like motif being the part of the biologically active protein was given recently by Terada *et al.* (49). The residues 342-349 of adrenodoxin reductase adopt a β-hairpin structure highly resembling the chignolin not only structurally but also sequentially since the chignolin residues Asp3, Pro4 and Gly7 critical for its stability are seen at corresponding positions in this chignolin-like segment.

Chapter 4

Computational methods

In this chapter, a brief theoretical background on the molecular mechanics is given. Since molecular mechanics is normally a platform for molecular simulations, the next part is focused on the standard molecular dynamics simulations being to date the most popular among the simulation methods. At the end of this chapter are summarized methods which belong to the most frequently used before the molecular dynamics simulation is initiated (homology modelling) and after the molecular dynamics data are collected (essential dynamics analysis). A majority of the aforementioned computational techniques was attempted in the studies which are summarized in Synopsis of results.

4.1 Molecular mechanics

Molecular mechanics (also empirical potential/force field) is a computational method which treats molecules as classical objects (4). Molecular mechanics (MM) is based on classical Newtonian physics and is therefore computationally significantly less expensive than the more accurate quantum mechanics. The purpose of molecular mechanics is to relate the potential energy of a given molecule with its three-dimensional configuration (its specific internal coordinates) (51, 52). The total potential energy (U_{pot}) is normally equal to the sum of three bonded and two non-bonded energy terms:

$$U_{pot} = U_{bond} + U_{angl} + U_{tors} + U_{elst} + U_{vdW}, \tag{4.1}$$

where U_{bond}, U_{angl}, and U_{tors} represent the potential energies of bond stretching, angle bending and dihedral angle torsion, respectively, and belong to the bonded terms of the total potential energy while the energies of electrostatic (U_{elst}) and van der Waals (U_{vdW}) interactions represent the non-bonded terms.

4.1.1 Treatment of bonded terms

Bond stretching and angle bending. A bond between the two atoms is in molecular mechanics modelled as a spring and its energy is described using the Hooke's law. The energy of angle bending is described using the Hooke's law as well. The use of the Hooke's law implies that both bond

stretching and angle bending are treated in molecular mechanics using the harmonic approximation. More accurately, bond stretching can be described using the Morse potential which is, however, computationally not as feasible as the harmonic potential (see Figure 8) (53).
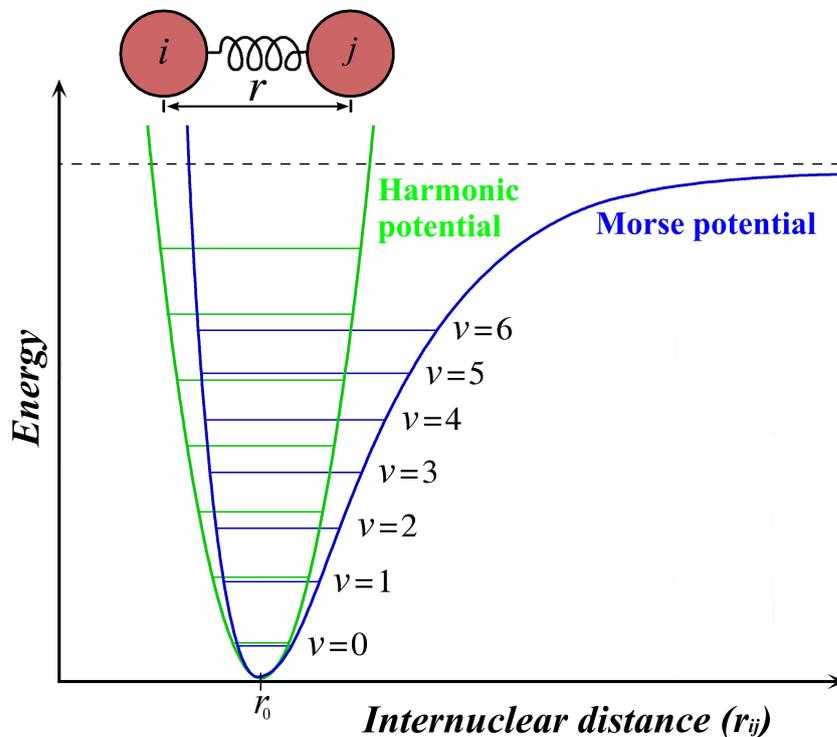
Figure 8 The harmonic oscillator potential (green) and the Morse potential (blue). The functional form for the Morse potential is $U_{bond} = D_r \left[1 - e^{-\alpha_r(r-r_0)} \right]^2$, where D_r is the dissociation energy of the bond and α_r is a fitting constant.

The analytic function for bond stretching, typically used in the majority of force fields, is as follows:

$$U_{bond} = K_r(r - r_{eq})^2, \tag{4.2}$$

where K_r is the force constant of bond stretching (bond rigidity) while r_{eq} with r represent the equilibrium bond length and a given bond length, respectively.

The analytic function for angle bending is similar to that for bond stretching and is as follows:

$$U_{angl} = K_\theta(\theta - \theta_{eq})^2, \tag{4.3}$$

where K_θ is the force constant of angle bending while θ_{eq} and θ represent the equilibrium angle and a given angle, respectively.

Torsions. If a molecule is composed of at least three consecutive covalent bonds, analytic function which describes rotation around a single bond (dihedral angle change) is present in MM calculations.

Torsions are in MM treated using a periodic function since they have normally multiple energy minima (Figure 9). The functional form for a molecular mechanic's torsion is as follows:

$$U_{tors} = \frac{V_n}{2} \left(1 + \cos\left[n\phi - \gamma\right]\right), \tag{4.4}$$

where V_n, n, and γ are the dihedral force constant, multiplicity and phase angle, respectively, while ϕ is a given dihedral angle.
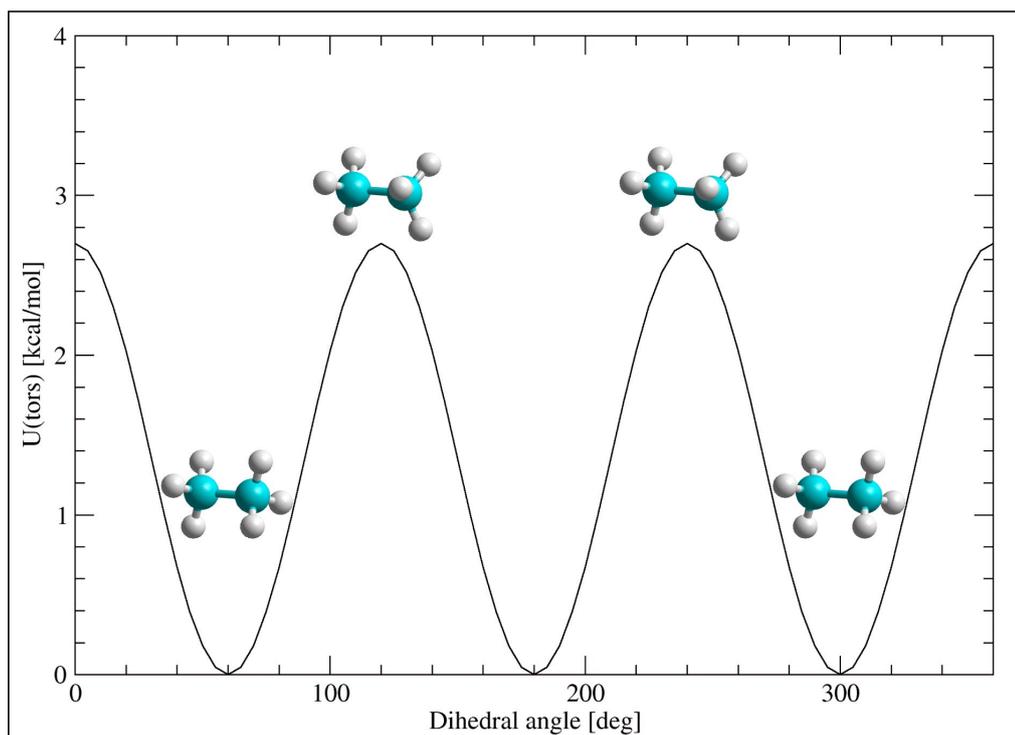
Figure 9 Energy profile of rotation around the C–C bond of ethane.

4.1.2 Treatment of non-bonded terms

In MM, atoms are modelled as Lennard-Jones spheres (hard-spheres) with fixed and centered point charges (51). Both the point charges and radii of the Lennard-Jones spheres (MM atoms) are parameters for the non-bonded electrostatic and vdW interactions, respectively.

The electrostatic interaction energy between the two atoms is calculated using the Coulomb's law:

$$U_{elst} = \frac{q_i q_j}{\varepsilon R_{ij}}, \tag{4.5}$$

where $q_i q_j$ is the square of atomic charges of interacting atoms i and j being at a distance of R_{ij} in the environment having the relative permittivity of ε.

The atomic charges are calculated using the quantum mechanical (QM) methods (often at the HF/6-31G* level) followed by the RESP (Restrained ElectroStatic Potential fit) procedure (54).

The purpose of the RESP analysis is to set the atomic charges in a studied molecule to represent most accurately the QM electrostatic potential.

The last term of the equation (4.1), potential energy of van der Waals (vdW) interaction, describes the dispersion and repulsion contributions between a pair of interacting atoms. The vdW term is calculated using a Lennard-Jones formalism which is typically written (53) as follows:

$$U_{vdW} = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^{6} \right], \tag{4.6}$$

where R_{ij} is the inter-atomic distance, σ_{ij} is the distance at which the repulsive and dispersion forces balance ($U_{vdW} = 0$ at this distance), and ε_{ij} is the Lennard-Jones well depth.

4.1.3   AMBER force field

The functional form of a pair-additive AMBER (55, 56) force field (e.g. *ff* 99 (54)) developed for molecular dynamics simulations of biomolecules is as follows (see equation (4.7)):

$$
\begin{aligned}
U(\mathbf{R}) &= \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} \left(1 + \cos\left[n\phi - \gamma \right] \right) + \\
&+ \sum_{i<j}^{atoms} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon R_{ij}},
\end{aligned} \tag{4.7}
$$

where $U(\mathbf{R})$ is the total potential energy of a given configuration of atoms and the forth term is a simple functional form for vdW interactions. The vdW term is treated using two terms of which the first is repulsive $A_{ij}/R_{ij}{}^{12}$ while the second is attractive ($B_{ij}/R_{ij}{}^{6}$ treats dispersion). The remaining terms of the equation (4.7) are described above (see equations (4.2) to (4.5)).

Summations in equation (4.7) indicate that each term is calculated for multiple atom pairs. The number of these atom pairs depends on a system topology and setup of the AMBER program. Typically, bond stretching, angle bending and torsions are calculated for covalently bonded atoms being separated by none, one ("1–2 interaction"), and two ("1–3 interaction") bonds, respectively. The electrostatic and vdW interactions include all possible atom pairs of the system (if the cutoff for non-bonded interactions is infinite), excluding only the "1–2" and "1–3" interactions. Notably, in the AMBER force field, the "1–4" vdW and electrostatic interactions are scaled (divided) by the factors of 2 and 1.2, respectively.

4.1.4   Force field limitation and accuracy

Implicit treatment of electrons.   Since electrons are not in MM treated explicitly, processes associated with the electron redistribution (e.g. chemical reactions) is not possible to study using force fields (57).

Harmonic approximation.   Within the harmonic approximation (HA), the dissociation of bonded atoms is not possible and the energies of the high-energy bond and angle vibrations

are described inaccurately (Figure 8). However, at the physiological temperature, the number of high-energy angle and bond vibrations is negligible (58) so that the HA treatment of angle bending and bond stretching using a simple quadratic function can be easily employed in molecular mechanics calculations at standard conditions (53).

Pair-wise additivity of force fields.   For instance, the 2nd generation Cornell *et al.* force field is pair-wise additive (in fact it is an effective two-body potential (59)), that is, it neglects all non-additive effects such as polarization of atoms in the electric field from the surroundings (51). The performance of pair-wise additive force fields is unsatisfactory for, e.g., inclusion of divalent cations, because their interactions lead to major polarization and charge-transfer effects which are neglected by the force field (52).

4.2   Classical molecular dynamics simulation

In connection to biomolecules, molecular dynamics (MD) simulation primarily provides insight into molecular motion at an atomic resolution. MD is a purely computational method evolving a multi-particle biological system with time to generate its trajectory (i.e. ensemble of conformations separated from each other by a regular short time-step). MD simulation serves as a useful bridge between the static structures of biological systems (e.g. X-ray, homology models) and their measurable properties of chemical and biological interest (60). The indispensable platform for classical MD simulations is a molecular mechanics force field (see above) since it defines forces (see equation (4.8)) acting among the atoms in the three-dimensional structure.

$$\boldsymbol{F}_i = -\nabla_i U_{pot}, \tag{4.8}$$

where $\boldsymbol{F}_i$ is the force which acts on the i-th particle of the system, calculated as the negative gradient of the potential energy U_{pot}.

The time evolution of the biological system is solved numerically by integrating the Newtonian equations of motions. To propagate the position and velocity vectors in a coupled fashion, a modification of Verlet's approach (61), called the leap-frog algorithm (known also as Velocity-Verlet algorithm), has been developed (53). The leap-frog computation scheme (considering particles' propagation under NVT conditions) consists of the following steps (see equations (4.9) to (4.12):

$$a_n = -\nabla_i U_{pot}/m_i, \tag{4.9}$$

$$v_{n+1/2} = v_{n-1/2} + a_n \Delta t, \tag{4.10}$$

$$v_{n+1/2} = v_{n+1/2}\lambda\left(T\right), \tag{4.11}$$

$$r_{n+1} = r_n + v_{n+1/2}\Delta t, \tag{4.12}$$

where m_i is atom mass, a_n is the acceleration, $v_{n+1/2}$ is the half-time-step-forward velocity, $v_{n-1/2}$ is the half-time-step-backward velocity (calculated at the simulation start from the Maxwell distribution), r_n is a given position vector (atom or particle position in the Cartesian space), r_{n+1} is the position vector after one integration step, T is the temperature, λ is velocity scaling factor, and Δt is the time-step length.

Typically, the time-step length Δt is one order of magnitude smaller than the highest frequency motion of the system. Usually, bond stretching vibrations limit time-step lengths to approximately 1 fs (53).

4.3   Homology modelling

Homology (comparative) modelling of proteins predicts three-dimensional atomic structure for a query sequence of the "target" protein when at least one experimentally determined protein structure ("template"), which is homologous to the query sequence, is available.

Normally, protein homology modelling consists of the following steps (62):

1. Identification of a suitable template (one or multiple templates can be used)

2. Sequence alignment (use of pair-wise/multiple alignment)

3. Refinement of the alignment (insertions and/or deletions and/or residue substitutions are performed)

4. Model construction (fragment assembly/segment matching/spatial-restraint model)

5. Model assessment (use of statistical potentials (or molecular mechanics energy calculations) for model refinement and validation)

Currently, the spatial-restraint model (63) is the most popular method for model generation.

The accuracy of protein structure prediction by homology modelling primarily depends on identity of the "query" and "template" sequences and quality of "template" protein structures (namely resolution, completeness, and R-factor) (62). Structures of compact protein regions (e.g. hydrophobic cores) are predicted more accurately than that of flexible loops.

Homology modeling has been intensively used for atomic structure prediction of membrane-bound proteins. The solved spatial structures of membrane proteins are few since it is fairly tricky to determine X-ray and NMR structures for systems with such a limited solubility (3). When there is available a three-dimensional structure for a membrane protein, it is often bearing numerous hydrophobic $\rightarrow$ hydrophilic mutations (to make the protein more soluble) and lacking hydrophobic regions (e.g. transmembrane anchors). As a result, membrane proteins represent a target frequently investigated by the homology modelling. Widely used was for example homology modelling in prediction of the atomic structure of microsomal cytochromes P450 (64) for which it was until 2000 (65) impossible to determine their structure experimentally. Also for another functionally very important membrane protein named Na^+/K^+-ATPase, the X-ray structure was not accessible for a long time. As a result, many studies which examined structure, flexibility and function of

Na^+/K^+-ATPase were based on homology modelling (see ref. (66) and studies in Appendices C and D).

Among the most popular programs for homology modelling belong MOE (Molecular Operating Environment), MODELLER (67) (based on spatial-restraint satisfaction), WHAT-IF (68), and YASARA. Homology modelling can be carried out using servers such as I-TASSER (`http://zhang.bioinformatics.ku.edu/I-TASSER/`, (69)) and ROBETTA (`http://robetta.bakerlab.org/`, (70)) as well.

4.4 Essential dynamics analysis

Essential dynamics analysis (EDA) is a computational method which separates the large-scale correlated motions from irrelevant fluctuations in a multi-particle system (e.g. a biomolecule). EDA is based on diagonalization of the covariance matrix of Cartesian atomic displacements (71). The covariance matrix is of the size $3N \times 3N$, where N is the number of atoms. The atomic displacements are calculated with respect to the average structure normally obtained from a trajectory and after removal of the translational and rotational motion. Upon matrix diagonalization, eigenvectors, representing the directions of the correlated motions, and eigenvalues representing the displacement variance of the respective eigenvectors, are obtained. In fact, EDA divides the conformational space of the analyzed system into two subspaces: (i) The first subset of eigenvectors represents the "essential" subspace characterized with a few degrees of freedom and large variance anharmonic fluctuations. (ii) The second subset of motions comprises fluctuations with a narrow and Gaussian-like distribution. Both subspaces are independent (orthogonal) with each other (71). The first few eigenvectors normally describe 90+% of the overall motion of a biomolecule and may be functionally relevant.

The EDA method was very useful for understanding of the relationship between the structure and function in proteins and nucleic acids. For example, employing EDA was derived and characterized the inter-domain motion of lysozyme (72, 73). Using a technique similar to EDA was carried out the analysis of general features of protein flexibility for more than 75% of all known protein folds (10) (see also the Dynameomics project (74)). Finally, the hinge-like (open $\leftrightarrow$ closed) flexibility of RNA kink-turns has been identified with EDA (75, 76).

EDA is available in AMBER (55, 56) and GROMACS (77) programs for molecular dynamics simulations and as a module of the WORDOM (78) program designed for efficient analysis of molecular dynamics trajectories. EDA is also possible to perform using the VMD plugin INTERACTIVE ESSENTIAL DYNAMICS (79).

Chapter 5

Results

5.1 Synopsis of results

This part of Ph.D. thesis is focused on the stability and/or flexibility of ankyrin repeat protein p18$^{\text{INK4c}}$, sodium-potassium pump and the synthetic peptide chignolin. In addition, this chapter includes a brief description of EK-peptide conformational behavior. Preferentially, discussed is the conformational behavior of the aforementioned systems as it was modeled using the classical unrestrained molecular dynamics simulations in explicit solvent at the Department of Physical Chemistry. The ATPase simulation data are discussed together with the fluorescence spectroscopy measurements carried out at the Department of Biophysics. The results presented in this chapter were published in total in five impacted scientific journals. Manuscripts at their full length (but without Supporting Information) are given in the Appendices A to E.

5.1.1 The minimal stable unit of ankyrin repeat protein p18$^{\text{INK4c}}$

The intrinsic stability and flexibility of the pair of ankyrin repeats IV and V of the protein p18$^{\text{INK4c}}$ (see Figure 10) was examined using classical all-atomic molecular dynamics simulations on 50-ns time scale. Moreover, a little bit non-traditionally, this ankyrin repeat pair was dissected into eight different structural moieties (fragments) each of which was simulated on the same time scale and using the same simulation protocol as the original full length structure. In all simulations (in total 9 MD productions) was employed the Cornell *et al.* force field *ff* 99 with TIP3P explicit water model.

Throughout the 50-ns time scale, only the pair of ankyrin repeats IV and V (i.e. the full length non-fragmented structure) and the pair of helix-turn-helix motifs α7-turn-α8 and α9-turn-α10 were capable of maintaining the native structure (see Figure 10). The remaining fragments were at the simulation ends highly distorted and lacked any signs of identity with the initial structure. Melting of single ankyrin repeat occurred on the nanosecond time scale. Melting of single α-helices occurred on a time scale of several hundred picoseconds.

Our results were highly consistent to the experimental observations by Zhang and Peng (29) who suggested that a single ankyrin repeat is intrinsically unstable in solution and the minimal
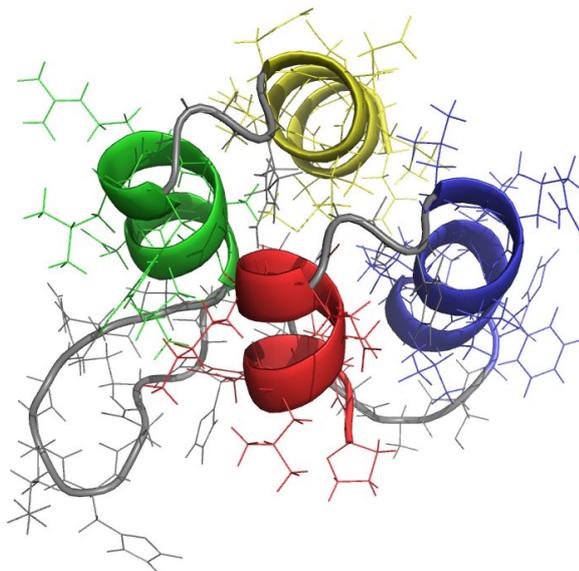
Figure 10 Three-dimensional structure of a pair of ankyrin repeats IV and V of protein p18$^{\text{INK4c}}$.

foldable unit of ankyrin repeat proteins is represented by a pair of ankyrin repeats. Since ankyrin repeat has two hydrophobic sides, it is very likely that at least one side must be covered by a neighboring repeat to make the ankyrin repeat thermodynamically stable. Otherwise, when both sides are uncovered, the entropy/enthalpy imbalance causes melting of ankyrin repeat.

The results of this study were published in the Journal of Molecular Modeling in 2008. The manuscript is available in Appendix A.

5.1.2   In silico analysis of the distribution of the folding stability throughout the ankyrin repeat protein p18$^{\text{INK4c}}$

In the next "p18 study" (see Appendix B) were carried out 50+ ns timescale MD simulations for 50 p18 fragments, including disconnected helix pairs, disconnected helix arrays, helix pairs connected by turns, full-length ankyrin repeats, etc. In addition, conformational behavior of the full-length p18 at room (298 K) and elevated (498 K) temperatures was investigated using MD simulations on the time scales of 50 and 20 ns, respectively. All simulations were run with the AMBER ff99 all-atomic force field with explicit water model treatment (TIP3P). The 498-K MD simulations (in total 4 parallel runs) were carried out under NVT while the remaining simulations under NpT conditions. Primarily, the study was focused on the stability of fragments, i.e. how well they are able to preserve their initial (p18-like) structure.

The simulations showed that all single helices, taken out of the p18 context, were not able to maintain their structure. The same was valid for larger fragments composed of pair of antiparallel helices, helix-turn-helix elements, four helix bundles and signle ankyrin repeats. On the other hand, two out of five fragments topologically coinciding with the minimal folding unit of Ferreiro *et al.* (32) preserved well their starting structure ($RMSD$ below 3.0 Å) and all pairs of ankyrin repeats

were stable along the 70-ns timescale. The Ferrerio's minimal folding units were characterized as lying at the "stability edge" between the single ankyrin repeats and the pair of ankyrin repeats. Very likely, using protein engineering, it would be possible to synthesize stable minimal structural untis of ARPs, coinciding with the motif proposed by Ferreiro.

Among the pairs of ankyrin repeats, the pairs bearing one internal and one capping repeat (pairs I-II and IV-V) were more stable. Capping repeats are more hydrophilic than their internal (buried) counterparts and are important for formation of a functionally active ARP (capping repeats prevent aggregation of ARPs). Frequently, capping repeats are targets of hydrophobic → hydrophilic mutations to facilitate expression of DARPins.

Next, our results indicated that the exposed loops of p18 were the most flexible parts of the protein and were less responsible for the p18 stability than the turns connecting the antiparallel helices. Very likely, the flexibility of loops is important for proper fit to target molecules during the complex formation, e.g. between p18 and CDK4/6.

Finally, the thermally induced unfolding simulations of p18 revealed melting of the system preferentially from its N-terminus (including ARs I and II) which was accompanied by melting of the remaining p18 parts (ARs III-V). The ensemble of structures with ARs I and II unfolded while ARs III-V keeping their native-like fold was an apparent unfolding intermediate of p18. The three-state unfolding mechanism of p18 was consistent with the experimental data on the unfolding of the p18 homolog p19$^{\text{INK4d}}$ (80).

This study was published in the Journal of Biomolecular Structure and Dynamics in 2010 and is presented in Appendix B.

5.1.3 Large-scale domain movement of the Na$^+$/K$^+$-ATPase cytoplasmic headpiece induced by ligand binding

The cytoplasmic segment of Na$^+$/K$^+$-ATPase between transmembrane helices M4 and M5 (hereafter C45) comprises full-length domains N and P. C45 contains binding sites for ATP (N-domain) as well as phosphorylation site of the conserved aspartate and Mg^{2+}-binding site (P-domain). So far, all the experiments indicate that the isolated C45 retains its tertiary structure identical with C45 embedded in a complete Na$^+$/K$^+$-ATPase (81), is able to bind ATP and TNP-ATP (82, 83), and our experiments described below suggest that it can undergo the same conformational changes that are expected for the corresponding part of the entire enzyme. Studying isolated C45 is popular due to its high solubility, making it an ideal model for laboratory experiments, and uncoupling from cation transport facilitates interpretation of experimental data.

In the study presented in Appendix C, conformational behavior of the isolated C45 upon binding of three different ligands was investigated using tryptophan fluorescence quenching and MD simulations. In total, nine single-tryptophan C45 mutants, distributed over the C45 surface, were expressed (see Figure 11). For each one-point C45 mutant, tryptophan fluorescence measurement was performed in the presence of Na$_2$ATP, Mg^{2+}, and MgATP ligands. Tryptophan fluorescence was also measured for unliganded (only Na$^+$) C45 mutants. The experimental measurements were complemented with MD simulations of all C45(wild type)-ligand variants as well as free C45. The simulations were conducted in Cornell *et al.* *ff*99 force field, with explicit TIP3P water and on

nanosecond time scales. As the starting structure for simulations was taken the C45 homology model built on the X-ray structure of Ca^{2+}-ATPase in closed (E2) conformation. In addition, a control 10-ns MD simulation was carried out with unliganded C45 X-ray structure (PDB ID 3B8E) assuming open (E1) conformation.

Our experimental and simulation data revealed that the isolated C45 opened upon binding of Na_2ATP. On the other hand, binding of Mg^{2+} and MgATP did not induce any dramatic C45 conformational change. The free (unliganded) C45 retained its initial closed conformation as well. Moreover, the 10-ns MD simulation of C45 X-ray structure confirmed the quality of the homology model since this structure (unliganded) exhibited irreversible open-close (E1-E2) transition in the simulation.

The importance of our data concerning C45 conformational behavior is twofold: (i) Our results indicate that the "ping-pong" E1-E2 conformational transition of C45 over pumping cycles is driven namely by cytoplasmic cofactors ATP and Mg^{2+}. (ii) Our data show that Mg^{2+} does not bind to C45 simultaneously with ATP (in the form of MgATP) as it is generally believed, but binding of ATP and Mg^{2+} is a sequential process. Firstly, ATP binds to C45 which is in closed (E2) conformation and displays a low-affinity ATP binding site. Then, C45 adopts the open E1 state and subsequently Mg^{2+} binds to it. Finally, autophosphorylation, presumably with the assistance of Mg^{2+} and A-domain, is initiated.

This study was published in Biochimica et Biophysica Acta in 2009 and is given in Appendix C.

5.1.4   Local changes in electrostatic surface potential of the Na^+/K^+-ATPase cytoplasmic headpiece induced by binding of ATP and/or Mg^{2+}

A novel method for monitoring of changes in electrostatic surface potential of Na^+/K^+-ATPase cytoplasmic headpiece C45 was introduced. The method was based on quenching of tryptophan fluorescence using two quenchers (acrylamide and iodide) differing between each other in charge and diffusion constant. Therefore, each quencher revealed different efficiency of fluorophor quenching. Since neutral acrylamide was sensitive only to steric accessibility of the fluorophor, negatively charged iodide carried also the information about the electrostatic surface potential in the fluorophor microenvironment. As a result, the electrostatic surface potential changes in the tryptophan environment could be easily extracted calculating the ratio of iodide and acrylamide Stern-Volmer quenching constants (labeled as the charge parameter). The higher was the charge parameter, the more positive was the environment around tryptophan reporter and vice versa.

In total, electrostatic surface potential (ESP) change in nine unique C45 microenvironments (Figure 11) was monitored upon binding of Na_2ATP, MgATP and Mg^{2+}. The data indicated that binding of Na_2ATP introduced the most dramatic ESP changes in four sites (Figure 11). While the N-domain side which is opposite to the ligand-binding region (three Trp reporters are at this side) demonstrated the positive shift of ESP upon Na_2ATP binding, the Asn642–Asn649 loop located in the bottom of the P-domain (Figure 11) exhibited shift to more negative values. In line with the experimental data, the last snapshot from MD simulation of C45 with Na_2ATP showed less negative ESP in the N-domain side being opposite to the ligand-binding. However, inconsistency

between the experimental and simulation data was identified for the entire P-domain including the Asn642–Asn649 loop region as well. Very likely, the simulation data did not fit the spectroscopic due to high flexibility of P-domain loops and missing 13 C-terminal residues in the C45 homology model.
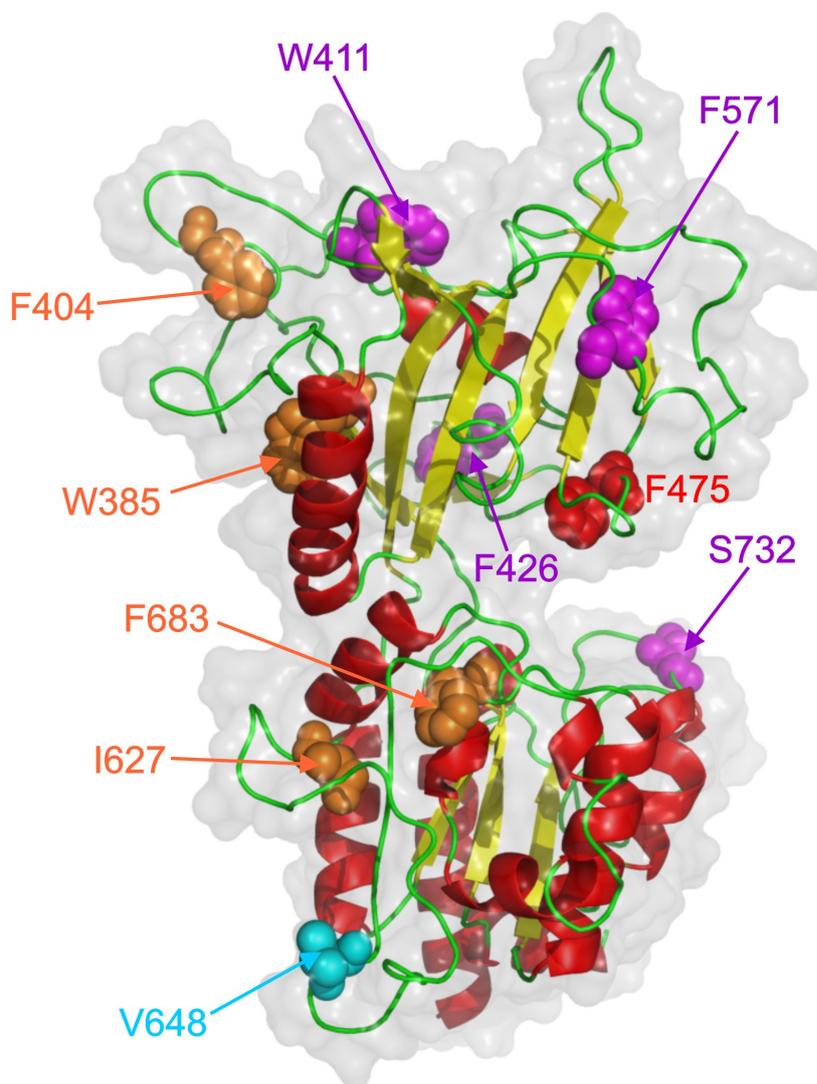
Figure 11 Positions of C45 residues (targets of single-point Trp/Phe mutations) which were used as reporters of the intrinsic tryptophan fluorescence. In orange and cyan are colored residues which underwent the change of ESP towards positive and negative values, respectively, upon binding of Na_2ATP. For residues in magenta the change of ESP was not observed upon Na_2ATP binding. The red F475 is located within the ATP binding site.

Altogether, our data suggested that ATP binding significantly affected the ESP potential in C45 parts distal from the ATP binding site and was not only restricted to the active site environment. This finding is of a great biological importance since it sheds light on the possible communication modus between the cytosolic C45 segment and transmembrane cation binding domain. The open C45 homology model with ATP bound demonstrates that its Asn642–Asn649

loop could easily contact a short cytoplasmic loop C67 in the full-length system (see Appendix D Figure 4). The C67 loop was already suspected by numerous authors to be the "bridge" responsible for coupling of events occurring in C45 with the transmembrane domain in P-type ATPases (84-88). Presumably, modulation of the electrostatic interaction between the Asn642–Asn649 and C67 loops by ATP binding may be a possible way how C45 transmits the information about ATP binding to cation binding sites in the transmembrane domain.

This study was published in the Biophysical Journal in 2009 and is given in Appendix D.

5.1.5 The effects of explicit solvent models on the conformational behavior and solvation of synthetic peptides

In the Appendix E study is explored the effect of the choice of AMBER force field and solvent model on the simulation behavior of peptides with either predominantly α (EK-peptide) or β (chignolin) structure. Using conventional molecular dynamics (MD) simulations on timescales from 50 ns to 1 μs, dynamics and solvation of both peptides is examined in three force fields (*ff*99, *ff*99SB, and *ff*03) and four water models (TIP3P, TIP4P, TIP5P, and SPC/E). In fact, both systems are simulated in all 12 possible combinations of solute force field and solvent model.

The MD simulations indicate that there is a dramatic effect of force field on the stability of each studied peptide. On the one hand, chignolin and EK-peptide reveal very high stability in *ff*99SB and *ff*03 force fields, respectively. On the other hand, chignolin readily melts in *ff*99 and the EK-peptide poorly preserves its α-helical structure in both *ff*99 and *ff*99SB force fields. The *ff*03 force field performs well for chignolin since the peptide stability in *ff*03 simulations is in a good agreement with experiment (ca. 60% of folded conformations at 300 K) (45).

In a sharp contrast to the force fields, the simulations do not demonstrate any significant effect of the water model on the stability of peptides. However, both systems turn out to be differently solvated in the water models. The TIP5P model is responsible for the highest specific solvation of both solutes. In contrast, TIP3P, being the model with the lowest viscosity, demonstrates the lowest extent of specific solvation. Different solvation of peptides by TIP5P causes that the salt bridges between Glu3...Lys6, Lys6...Glu9, and Glu9...Lys12 of the EK-peptide are more stable in TIP5P in comparison to other models. In addition, chignolin behavior seems to be affected by the high extent of TIP5P solvation in its unfolded state. While in TIP3P, TIP4P and SPC/E models combined with *ff*99 force field are predominantly populated unfolded chignolin structures with Gly7 moved apart from the native left-handed region, in TIP5P, despite the fact that chignolin is unfolded, Gly7 keeps permanently the left-handed conformation.

In sum, the study reveals that all examined water models perform similarly, considering only the conformational behavior of folded peptides. In contrast, each water model exhibits different extent of solute's specific solvation of which the best "solvating" model is TIP5P. The high degree of specific solvation of TIP5P is propagated to the conformational behavior of unfolded chignolin and has a stabilizing effect on the EK-peptide salt bridges. It is questionable if more realistic is the specific solvation of TIP3P, TIP4P, and SPC/E (demonstrating reduced solvation) or TIP5P (showing extensive solvation). As the TIP5P model very well represents the behavior of water solvent in comparison to other models, it is expected that the extensive specific solvation of TIP5P

is more physiological than the reduced one of TIP3P, TIP4P, and SPC/E. Consequently, parts of biomolecules which are exposed and considerably solvated should be simulated with the more realistic TIP5P model. In contrast, useful can be also the TIP3P model (although revealing the less realistic parameters) since it is computationally not expensive and facilitates sampling of proteins' conformational space. Therefore, in our study, TIP3P is recommended as the best model for simulations of structured proteins, in which the exposed parts (e.g. loops) are missing or are not of a particular interest.

This study was published in the Journal of Chemical Theory and Computation in 2010 and is given in the Appendix E.

Chapter 6

Summary

The presented Ph.D. thesis focuses on the structure, stability and structural dynamics of three entirely different proteins: an ankyrin repeat protein p18^{INK4c}, a Na$^+$/K$^+$-ATPase, and a synthetic β-hairpin chignolin. The structural dynamics and stability of all solutes was investigated with classical unrestrained MD simulations in explicit solvent (i.e. *in silico*). The majority of simulations was conducted with the 2nd generation Cornell *et al.* force field *ff*99 with extensive sampling on ten-nanosecond timescales.

In connection with p18^{INK4c} (p18) protein, we primarily paid attention to finding its minimal stable structural motif. Therefore, the p18 system was dissected into ~50 fragments, each of which was subsequently simulated at least on the 50-ns timescale in order to obtain information about relative stability of fragments. The simulations pointed out, that the fragment, bearing one complete ankyrin repeat capped with the N-terminal helix of the subsequent ankyrin repeat, may be *per se* an intrinsically stable structure. By far, experiments have suggested that the minimal stable motif of ankyrin repeat proteins consists of a pair of ankyrin repeats (i.e. structurally larger motif in comparison to that proposed by our simulations). We are convinced that a *de novo* synthesis of a minimal stable motif of ankyrin repeat proteins, coinciding with that predicted by our simulations, will be a great challenge for experimentalists in the near future.

The conformational behavior of Na$^+$/K$^+$-ATPase was assessed employing MD simulations as well as fluorescence spectroscopy. Both methods were primarily focused on the inter-domain movement of the C45 headpiece under various conditions (presence of Na$^+$, Mg^{2+} and ATP). According to fluorescence spectroscopy and MD simulation data, ATP and Mg^{2+} bind to C45 sequentially. First, ATP is responsible for triggering of the closed $\rightarrow$ open C45 conformational change. Subsequently, binding of Mg^{2+} initiates C45 transition in the opposite direction (i.e. open $\rightarrow$ closed movement of C45 domains).

In the case of chignolin, its structure and stability was examined using MD simulations started with different combinations solute force field – water model. In total, 12 combinations were attempted. The experimental chignolin stability was very well reproduced in *ff*03 force filed. The choice of water model did not significantly affect chignolin stability. In conclusion, MD simulation studies attempting to describe, e.g. chignolin folding (or β-hairpins resembling chignolin) should be carried out with *ff*03 force field. To enhance sampling of the conformational space over MD

simulations, when there is no need for careful treating of water solvent, use of TIP3P water model is recommended.

Altogether, the majority of MD simulations which are presented here matches well the experimental data concerning stability and/or flexibility of studied biomolecules. Moreover, the MD simulations complement the experimental data since they provide a unique atomic insight into the dynamics of studied systems. For example, the role of Mg^{2+} in the Na^+/K^+-ATPase's conformational behavior could be discussed within the atomic resolution by virtue of MD trajectories (see Appendix C). There is no doubt that MD simulations will be at least in the near future considered by researchers to be a valuable tool for elucidating the structural dynamics of proteins. Someday, hopefully, it will manage to entirely bridge the gap between protein structure and function. MD simulations will undoubtedly play an important role in such an excellent achievement.

Chapter 7

Shrnutí

Předkládaná Ph.D. práce se věnuje struktuře, stabilitě a vnitřní dynamice tří topologicky velmi odlišných biomolekul, jimiž jsou: ankyrin repeat protein p18$^{\text{INK4c}}$, sodno-draselná ATPáza a syntetický dekapeptid chignolin. Stabilita a flexibilita výše zmíněných systémů byla studována prostřednictvím klasických molekulově-dynamických (MD) simulací se zahrnutím explicitního modelu vody (tj. teoretickým *in silico* přístupem). Většina simulací byla provedena pomocí silového pole druhé generace *ff* 99, vyvinutého Cornellovou a kol., na časových škálách několika desítek nanosekund.

U proteinu p18$^{\text{INK4c}}$ (p18) jsme se zaměřili na identifikaci jeho nejmenšího stabilního strukturního motivu. P18 jsme rozdělili do cca 50 substruktur (fragmentů), jejichž stabilitu a konformační chování jsme studovali klasickými MD simulacemi na časových škálách minimálně 50 ns. Simulační data ukázala, že fragment složený z jednoho kompletního ankyrin repeat (AR) motivu a N-terminální α-šroubovice AR motivu sousedního (ve směru od N- k C-konci p18), je velmi pravděpodobně vnitřně stabilní struktura, která je schopna existence bez přítomnosti nativního kontextu (tj. okolních strukturních elementů proteinu p18). Experimentální data doposud nepotvrdily existenci ankyrin repeat proteinu o této velikosti. Doposud provedené experimenty naznačují, že minimální stabilní jednotkou ankyrin repeat proteinů je pár AR motivů, zatímco samotný AR motiv je vnitřně nestabilní. Naše MD simulace však ukazují, že je velmi pravděpodobně možné syntetizovat motiv složený z jednoho kompletního AR motivu a přibližně poloviny motivu sousedního. Pro experimentátory bude nicméně důležitým milníkem navrhnout sekvenci s dostatečně vysokou *foldabilitou* (schopností se složit), protože námi navržený minimální element ankyrin repeat proteinů vykazuje hraniční stabilitu, ležící mezi párem AR motivů a samotným AR motivem.

Konformační chování cytoplazmatického segmentu C45 sodno-draselné pumpy jsme studovali klasickými MD simulacemi společně se spektroskopickými fluorescenčními experimenty. Obě metody se primárně zaměřily na vzájemnou konformaci N- a P-domény po vazbě různých ligandů: Na$^+$, Mg^{2+} a ATP. Jak MD simulace, tak i fluorescenční data potvrdily, že po vazbě Na$_2$ATP dochází ke dramatické konformační změně C45, která vede ke zvětšení úhlu mezi oběma doménami, tj. k otevření C45. Mimo jiné, vazba Mg^{2+} k otevřenému C45 segmentu iniciuje jeho zavření. Na základě získaných dat se domníváme, že vazba Mg^{2+} a ATP k sodno-draselné pumpě může být popsána jako sled dvou kroků. Nejprve se k C45 segmentu váže ATP, které způsobí jeho otevření a následně po

vazbě Mg^{2+} dojde k zavření C45 spojenému s autofosforylací P-domény. Předpokládaný sekvenční mechanismus doplňuje současné poznatky o vazbě Mg^{2+} a ATP k sodno-draselné pumpě, kde se očekává, že se Mg^{2+} a ATP váží k otevřenému C45 segmentu simultánně ve formě MgATP.

Chignolin byl primárně použit jako testovací systém pro posouzení vlivu explicitního solventu na stabilitu a flexibilitu petidů/proteinů v klasických MD simulacích. Konformační chování (stabilita, solvatace) chignolinu na submikrosekundových časových škálách bylo sledováno celkem ve čtyřech explicitních modelech vody (TIP3P, TIP4P, TIP5P, SPC/E), z nichž každý model byl navíc zkombinován se třemi silovými poli (*ff* 99, *ff* 99SB, *ff* 03). Simulační data naznačují, že změna silového pole významně ovlivňuje stabilitu nativní β-hairpin stuktury chignolinu. Největší shody mezi teoreticky pozorovanou stabilitou chignolinu a experimentálními údaji bylo dosaženo se silovým polem *ff* 03. Naopak volba explicitního modelu vody nemá prakticky žádný dopad na stabilitu nativní struktury chignolinu v MD simulacích. Pro účely studií zabývajících se *ab initio* skládáním chignolinu a jemu podobných struktur, tj. systémů s vysokou mírou tzv. *β-hairpin propnesity*, je vhodné použití silového pole *ff* 03. Z provedené studie mimo jiné plyne, že TIP3P explicitní model je dobrým kandidátem pro simulace biomolekulárních systémů, v nichž nehrají významnou roli strukturní vody a které nemají mimořádně solvatované regiony. Ačkoli jde o model s nejméně realistickými parametry, jeho vysoká viskozita dovoluje rozsáhlejší *sampling* konfiguračního prostoru biomolekul v MD simulacích.

Prezentovaná Ph.D. práce dokumentuje, že diskutovaná simulační data se většinou dobře shodují s experimentálními daty jak v otázce stability, tak i flexibility (vnitřní strukturní dynamiky) vyšetřovaných biomolekul. Díky unikátní vlastnosti MD simulací podat atomární náhled na dynamické procesy probíhající v biomolekulách, jsou bez nadsázky rozšířeny současné poznatky o konformačním chování studovaných systémů. Vliv Mg^{2+} na konformační chování sodno-draselné ATPázy byl například díky simulacím vůbec poprvé diskutován na atomární úrovni (viz Appendix C). Není pochyb, že MD simulace budou i nadále užitečným nástrojem pro popis strukturní dynamiky proteinů a dalších biomolekul. Výhledově lze očekávat aplikaci simulačních metod na systémech dosahujících velikosti buněčných kompartmentů ba i dokonce celých buněk. Pokud se jednoho dne dočkáme kompletního pochopení vztahu mezi strukturou a funkcí proteinů, podíl MD simulací k završení takto mimořádně důležitého úkolu bude zajisté velmi významný.

Chapter 8

List of publications

1. M. Otyepka, **P. Sklenovský**, D. Horinek, T. Kubař, P. Hobza. How the Stabilization of INK4 Tumor Suppressor 3D Structure Evaluated by Quantum Chemical and Molecular Mechanics Calculations Corresponds Well with Experimental Results: Interplay of Association Enthalpy, Entropy, and Solvation Effects. *J Phys Chem B* 9, 4423–4429 (2006).

2. **P. Sklenovský**, P. Banáš, M. Otyepka. Two C-terminal ankyrin repeats form the minimal stable unit of the ankyrin repeat protein p18$^{\text{INK4c}}$. *J Mol Model* 14, 747–759 (2008).
   Attached as Appendix A

3. **P. Sklenovský**, M. Otyepka. In Silico Structural and Functional Analysis of Fragments of the Ankyrin Repeat Protein p18$^{\text{INK4c}}$. *J Biomol Struct Dyn* 27, 521–540 (2010).
   Attached as Appendix B

4. L. Gryčová, **P. Sklenovský**, Z. Lánský, M. Janovská, M. Otyepka, E. Amler, J. Teisinger, M. Kubala. ATP and magnesium drive conformational changes of the Na$^+$/K$^+$-ATPase cytoplasmic headpiece. *Biochem Biophys Acta-Biomembr* 1788, 1081–1091 (2009).
   Attached as Appendix C

5. M. Kubala, L. Gryčová, Z. Lánský, **P. Sklenovský**, M. Janovská, M. Otyepka, J. Teisinger. Changes in Electrostatic Surface Potential of Na$^+$/K$^+$-ATPase Cytoplasmic Headpiece Induced by Cytoplasmic Ligand(s) Binding. *Biophys J* 97, 1756–1764 (2009).
   Attached as Appendix D

6. P. Florová, **P. Sklenovský**, P. Banáš, M. Otyepka. Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact. *J Chem Theory Comput* 6, 3569–3579 (2010).
   Attached as Appendix E

7. **P. Sklenovský**, P. Florová, P. Banáš, K. Réblová, F. Lankaš, M. Otyepka, J. Šponer. Understanding the RNA flexibility using explicit solvent simulations: The ribosomal and group I intron reverse kink-turn motifs. *J Chem Theory Comput*, submitted.

Chapter 9

References

(1) H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. *Nucleic Acids Res* 28, 235–242 (2000).

(2) T. Schlick. Molecular Modeling and Simulation: An Interdisciplinary Guide. Springer-Verlag, New York (2002).

(3) A. V. Finkelstein, O. Ptitsyn. Protein Physics: A Course of Lectures. Academic Press, London (2002).

(4) A. Tramontano. Protein Structure Prediction: Concepts and Applications. WILEY-VCH, Weinheim (2006).

(5) L. Holm, C. Sander. *Nucleic Acids Res* 22, 3600–3609 (1994).

(6) L. H. Greene, T. E. Lewis, S. Addou, A. Cuff, T. Dallman, M. Dibley, O. Redfern, F. Pearl, R. Nambudiry, A. Reid, I. Sillitoe, C. Yeats, J. M. Thornton, C. A. Orengo. *Nucleic Acids Res* 35, D291–297 (2007).

(7) A. Andreeva, D. Howorth, S. E. Brenner, T. J. Hubbard, C. Chothia, A. G. Murzin. *Nucleic Acids Res* 32, D226–229 (2004).

(8) A. E. Todd, C. A. Orengo, J. M. Thornton. *J Mol Biol* 307, 1113–1143 (2001).

(9) K. Teilum, J. G. Olsen, B. B. Kragelund. *Cell Mol Life Sci* 66, 2231–2247 (2009).

(10) N. C. Benson, V. Daggett. *Protein Sci* 17, 2038–2050 (2008).

(11) S. Kumar, R. Nussinov. *Chembiochem* 3, 604–617 (2002).

(12) V. Daggett. *Chem Rev* 106, 1898–1916 (2006).

(13) G. Dodson, C. S. Verma. *Cell Mol Life Sci* 63, 207–219 (2006).

(14) G. G. Dodson, D. P. Lane, C. S. Verma. *EMBO Rep* 9, 144–150 (2008).

(15) C. N. Pace. *Trends Biochem Sci* 15, 14–17 (1990).

(16) C. Branden, J. Tooze. Introduction to Protein Structure. Garland Publishing, Inc., New York (1999).

(17) P. Hobza, K. Müller-Dethlefs. Non-Covalent Interactions: Theory and Experiment. RSCPublishing, Cambridge (2010).

(18) K. A. Dill. *Biochemistry* 29, 7133–7155 (1990).

(19) R. L. Baldwin. *J Mol Biol* 371, 283–301 (2007).

(20) P. Hobza, R. Zahradník. Mezimolekulové komplexy: Úloha van der Waalsových systémů ve fyzikální chemii a v biodisciplínách. Academia, Praha (1988).

(21) J. Vondrasek, L. Bendova, V. Klusak, P. Hobza. *J Am Chem Soc* 127, 2615–2619 (2005).

(22) M. Otyepka, P. Sklenovsky, D. Horinek, T. Kubar, P. Hobza. *J Phys Chem B* 110, 4423–4429 (2006).

(23) D. Chandler. *Nature* 437, 640–647 (2005).

(24) S. G. Sedgwick, S. J. Smerdon. *Trends Biochem Sci* 24, 311–316 (1999).

(25) P. Bork. *Proteins* 17, 363–374 (1993).

(26) R. Venkataramani, K. Swaminathan, R. Marmorstein. *Nat Struct Biol* 5, 74–81 (1998).

(27) S. Al-Khodor, C. T. Price, A. Kalia, Y. Abu Kwaik. *Trends Microbiol* 18, 132–139 (2010).

(28) M. Sotomayor, D. P. Corey, K. Schulten. *Structure* 13, 669–682 (2005).

(29) B. Zhang, Z. Y. Peng. *J Mol Biol* 299, 1121–1132 (2000).

(30) M. T. Stumpp, H. K. Binz, P. Amstutz. *Drug Discov Today* 13, 695–701 (2008).

(31) C. C. Mello, D. Barrick. *Proc Natl Acad Sci U S A* 101, 14102–14107 (2004).

(32) D. U. Ferreiro, S. S. Cho, E. A. Komives, P. G. Wolynes. *J Mol Biol* 354, 679–692 (2005).

(33) E. Kloss, N. Courtemanche, D. Barrick. *Arch Biochem Biophys* 469, 83–99 (2008).

(34) K. W. Tripp, D. Barrick. *J Mol Biol* 344, 169–178 (2004).

(35) E. R. G. Main, A. R. Lowe, S. G. J. Mochrie, S. E. Jackson, L. Regan. *Curr Opin Struct Biol* 15, 464–471 (2005).

(36) C. Yuan, J. Li, T. L. Selby, I. J. Byeon, M. D. Tsai. *J Mol Biol* 294, 201–211 (1999).

(37) P. Sklenovský. Teoretický výpočet interakčních energií a jejích komponent v proteinech. Diplomová práce, Olomouc (2006).

(38) B. Alberts. Essential Cell Biology. Garland Science, New York and London (2010).

(39) K. Geering. *J Bioenerg Biomembr* 33, 425–438 (2001).

(40) J. P. Morth, B. P. Pedersen, M. S. Toustrup-Jensen, T. L. Sorensen, J. Petersen, J. P. Andersen, B. Vilsen, P. Nissen. *Nature* 450, 1043–1049 (2007).

(41) R. W. Albers. *Annu Rev Biochem* 36, 727–756 (1967).

(42) R. L. Post, C. Hegyvary, S. Kume. *J Biol Chem* 247, 6530–6540 (1972).

(43) C. Weidemuller, K. Hauser. *Biochim Biophys Acta* 1787, 721–729 (2009).

(44) M. Huličiak. Detekcia interakcií Na+/K+-ATPázy s biologicky aktívnymi látkami metódami fluorescenčnej spektroskopie. Diplomová práca, Olomouc (2010).

(45) S. Honda, K. Yamasaki, Y. Sawada, H. Morii. *Structure* 12, 1507–1518 (2004).

(46) M. M. Seibert, A. Patriksson, B. Hess, D. van der Spoel. *J Mol Biol* 354, 173–183 (2005).

(47) B. L. Sibanda, J. M. Thornton. *Methods Enzymol* 202, 59–82 (1991).

(48) M. P. Hatfield, R. F. Murphy, S. Lovas. *J Phys Chem B* 114, 3028–3037 (2010).

(49) T. Terada, D. Satoh, T. Mikawa, Y. Ito, K. Shimizu. *Proteins* 73, 621–631 (2008).

(50) R. B. Best, J. Mittal. *J Phys Chem B* 114, 8790–8798 (2010).

(51) P. Banas, P. Jurecka, N. G. Walter, J. Sponer, M. Otyepka. *Methods* 49, 202–216 (2009).

(52) M. A. Ditzler, M. Otyepka, J. Sponer, N. G. Walter. *Acc Chem Res* 43, 40–47 (2010).

(53) C. J. Cramer. Essentials of Computational Chemistry: Theories and Models. John Wiley and Sons, Ltd, Chichester (2004).

(54) J. M. Wang, P. Cieplak, P. A. Kollman. *J Comput Chem* 21, 1049–1074 (2000).

(55) D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman. *Comput Phys Commun* 91, 1–41 (1995).

(56) D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R. J. Woods. *J Comput Chem* 26, 1668–1688 (2005).

(57) K. I. Ramachandran, G. Deepa, K. Namboori. Computational Chemistry and Molecular Modeling: Principles and Applications. Springer-Verlag, Berlin and Heidelberg (2008).

(58) M. Otyepka, J. Vřešťál. Statistická termodynamika: Stručný úvod. Masarykova univerzita, Brno (2004).

(59) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. *J Am Chem Soc* 117, 5179–5197 (1995).

(60) M. Remko. Molekulové modelovanie: Princípy a aplikácie. Slovak Academic Press, Bratislava (2000).

(61) L. Verlet. *Phys Rev* 159, 98–103 (1967).

(62) H. Venselaar, R. P. Joosten, B. Vroling, C. A. Baakman, M. L. Hekkelman, E. Krieger, G. Vriend. *Eur Biophys J* 39, 551–563 (2010).

(63) A. Sali, T. L. Blundell. *J Mol Biol* 234, 779–815 (1993).

(64) F. K. Friedman, R. C. Robinson, R. Dai. *Front Biosci* 9, 2796–2806 (2004).

(65) P. A. Williams, J. Cosme, V. Sridhar, E. F. Johnson, D. E. McRee. *J Inorg Biochem* 81, 183–190 (2000).

(66) H. Ogawa, C. Toyoshima. *Proc Natl Acad Sci U S A* 99, 15977–15982 (2002).

(67) N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, A. Sali. *Curr Protoc Bioinformatics* Chapter 5, Unit 5 6 (2006).

(68) G. Vriend. *J Mol Graph* 8, 52–56, 29 (1990).

(69) Y. Zhang. *BMC Bioinformatics* 9, 40 (2008).

(70) D. Chivian, D. E. Kim, L. Malmstrom, P. Bradley, T. Robertson, P. Murphy, C. E. Strauss, R. Bonneau, C. A. Rohl, D. Baker. *Proteins* 53 Suppl 6, 524–533 (2003).

(71) A. Amadei, A. B. M. Linssen, H. J. C. Berendsen. *Proteins* 17, 412–425 (1993).

(72) B. L. de Groot, S. Hayward, D. M. van Aalten, A. Amadei, H. J. Berendsen. *Proteins* 31, 116–127 (1998).

(73) S. Hayward, A. Kitao, H. J. Berendsen. *Proteins* 27, 425–437 (1997).

(74) D. A. Beck, A. L. Jonsson, R. D. Schaeffer, K. A. Scott, R. Day, R. D. Toofanny, D. O. Alonso, V. Daggett. *Protein Eng Des Sel* 21, 353–368 (2008).

(75) F. Razga, J. Koca, J. Sponer, N. B. Leontis. *Biophys J* 88, 3466–3485 (2005).

(76) K. Reblova, F. Razga, W. Li, H. Gao, J. Frank, J. Sponer. *Nucleic Acids Res* 38, 1325–1340 (2010).

(77) D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen. *J Comput Chem* 26, 1701–1718 (2005).

(78) M. Seeber, M. Cecchini, F. Rao, G. Settanni, A. Caflisch. *Bioinformatics* 23, 2625–2627 (2007).

(79) J. Mongan. *J Comput Aided Mol Des* 18, 433–436 (2004).

(80) C. Low, U. Weininger, M. Zeeb, W. Zhang, E. D. Laue, F. X. Schmid, J. Balbach. *J Mol Biol* 373, 219–231 (2007).

(81) C. Gatto, A. X. Wang, J. H. Kaplan. *J Biol Chem* 273, 10578–10585 (1998).

(82) M. Kubala, J. Plasek, E. Amler. *Eur Biophys J* 32, 363–369 (2003).

(83) M. Kubala, J. Plasek, E. Amler. *Physiol Res* 53, 109–113 (2004).

(84) P. L. Jorgensen, K. O. Hakansson, S. J. Karlish. *Annu Rev Physiol* 65, 817–849 (2003).

(85) G. Xu, D. J. Kane, L. D. Faller, R. A. Farley. *J Biol Chem* 279, 45594–45602 (2004).

(86) Z. Zhang, D. Lewis, C. Sumbilla, G. Inesi, C. Toyoshima. *J Biol Chem* 276, 15232–15239 (2001).

(87) G. Lenoir, M. Picard, J. V. Moller, M. le Maire, P. Champeil, P. Falson. *J Biol Chem* 279, 32125–32133 (2004).

(88) F. Corre, C. Jaxel, J. Fuentes, T. Menguy, P. Falson, B. A. Levine, J. V. Moller, M. le Maire. *Ann NY Acad Sci* 986, 90–95 (2003).

Chapter 10

Results — Appendix

10.1 Appendix A

Two C-terminal ankyrin repeats form the minimal stable unit of the ankyrin repeat protein p18$^{\text{INK4c}}$

Petr Sklenovský • Pavel Banáš • Michal Otyepka

Abstract

Ankyrin repeat proteins (ARPs) appear to be abundant in organisms from all phyla, and play critical regulatory roles, mediating specific interactions with target biomolecules and thus ordering the sequence of events in diverse cellular processes. ARPs possess a non-globular scaffold consisting of repeating motifs named ankyrin (ANK) repeats, which stack on each other. The modular architecture of ARPs provides a new paradigm for understanding protein stability and folding mechanisms. In the present study, the stability of various C-terminal fragments of the ARP p18$^{\text{INK4c}}$ was investigated by all-atomic 450 ns molecular dynamics (MD) simulations in explicit water solvent. Only motifs with at least two ANK repeats made stable systems in the available timescale. All smaller fragments were unstable, readily losing their native fold and α-helical content. Since each non-terminal ANK repeat has two hydrophobic sides, we may hypothesize that at least one hydrophobic side must be fully covered and shielded from the water as a necessary, but not sufficient, condition to maintain ANK repeat stability. Consequently, at least two ANK repeats are required to make a stable ARP.

Keywords

Ankyrin repeat, p18$^{\text{INK4c}}$, Minimal stable unit, Fragmentation, Molecular dynamics

Introduction

Ankyrin repeat proteins (ARPs) are non-globular biomolecules containing repeating units named ankyrin (ANK) motifs arranged in linear, tandem arrays (Fig. 1). The ANK repeat is a protein structural unit that is $\sim$33 residues long and consists of two β-strands and a pair of α-helices assembled in an antiparallel fashion and connected by a short loop (often formed by three residues). Two consensus sequences of the ANK repeat have been defined: the Consensus 1 ANK repeat comprising the $\beta2\alpha2$ ($\beta\beta\alpha\alpha$) motif, and Consensus 2 ANK repeat comprising the $\beta\alpha2\beta$ ($\beta\alpha\alpha\beta$) motif [1] (Fig. 2). Members of the ARP family co-localize and interact with various membrane and cytoplasmic proteins. The specificity of ARP–protein interactions is likely to be conferred by nonconserved residues flanking each ANK repeat, located at the tips of exposed loops [2].

The number of ANK repeats in naturally occurring ARPs varies from four [3, 4] to 29 [5]. Until 2000, only ARPs containing at least four ANK motifs were considered to be stable proteins [6].

However, a pioneering determination of a minimal stable motif of ARPs by Zhang and coworkers [7], which was based on proteolytic experiments and prediction algorithms, showed that two C-terminal ANK repeats of the naturally occurring p16[INK4a] [3] protein stacked on each other were stable and could fold independently of the rest of the protein. In addition, numerous artificial ARPs with enhanced thermostability and affinity have been designed recently [8–18], the smallest of which (2ANK) contained two identical ANK repeats [17]. Furthermore, on the basis of proteolytic experiments and X-ray crystallographic studies, it was concluded that single ANK repeats cannot form stable native structures under physiological conditions [7, 8], so at least two ANK repeats appear to be required to form stable ARPs, and thus comprise a minimal folding unit. On the other hand, Ferreiro and coworkers [19] suggested from Gō-type simulations that the minimal topological folding module of ARPs may be even smaller, consisting of one fully folded ANK repeat followed by the C-terminal helix of the neighboring repeat.

Protein p18[INK4c] [20] (p18) is a member of the cyclin-dependent kinase inhibitor (INK) tumor suppressor family and consists of five ANK repeat modules (Fig. 1). In the present study, the stability of various p18 C-terminal fragments was investigated using all-atomic molecular dynamics (MD) simulations in explicit water solvent at the tens of nanoseconds timescale, totaling 450 ns. Our results show that the minimal folding unit proposed by Ferreiro and coworkers [19] is unstable, and that the minimal stable unit of the p18 protein, at the studied timescale, consists of two C-terminal ANK repeats.
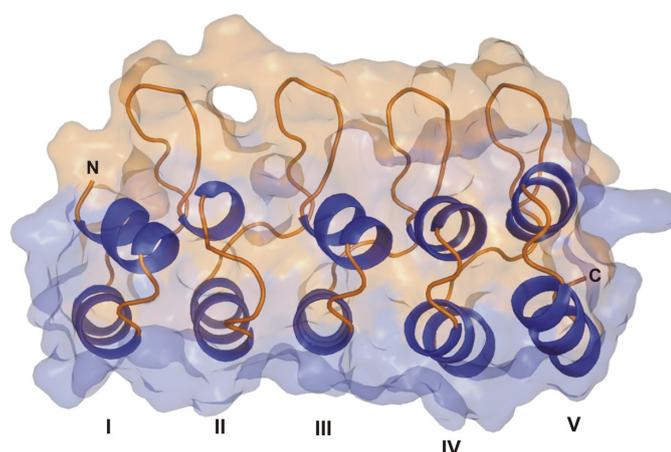

Figure 1: Structure of the p18[INK4c] protein (PDB entry 1IHB, chain B), which is a member of the cyclin-dependent kinase inhibitor (INK) tumor suppressor family with five ankyrin (ANK) repeat modules. The figure was generated by PyMol [30], http://pymol.sourceforge.net/
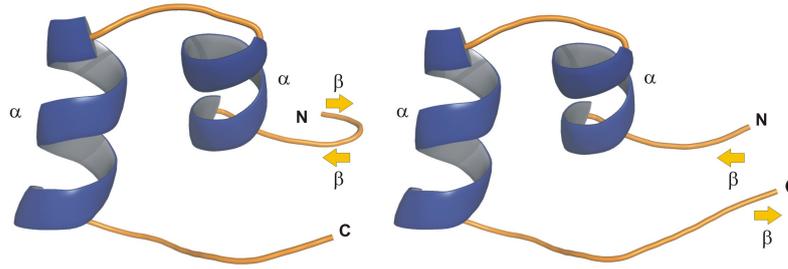
Figure 2: Illustrations of the ankyrin (ANK) repeat motif: *left* Consensus 1, $\beta2\alpha2$ motif (β-hairpin perpendicular to antiparallel α-helices linked by a short loop); *right* Consensus 2, β-strand-helix-loop-helix-β-strand sequence, for definitions, see [1]

Methods

Studied fragments

The crystal structure of the p18 protein [20] (PBD entry 1IHB, chain B) with optimized positions of hydrogen atoms (obtained using the Sander module of AMBER [21]; *parm99* force field [22, 23]) was used as a template for studying all of the following p18 fragments in MD simulations (Table 1). The N-terminus of each fragment was acetylated and the C-terminus was capped with an *N*-methyl group:

- Two pairs of α-helices: $\alpha7+\alpha8$ (residues 106–112 + 116–125) and $\alpha9+\alpha10$ (residues 140–146 + 150–159)

- Two helix-turn-helix motifs: $\alpha7$-turn-$\alpha8$ and $\alpha9$-turn-$\alpha10$ (residues 106–125 and 140–159)

- Consensus 2 ANK repeat IV (residues 102–135)

- The hypothetical minimal folding unit suggested by Ferreiro et al. [19] (truncated form with six N-terminal β-hairpin residues 100–105 missing) $\alpha7$-turn-$\alpha8$-loop-$\alpha9$ motif (residues 106–146)

- Consensus 2-based hypothetical minimal folding unit [19]: ANK IV-loop-$\alpha9$ (residues 102–146)

- Pair of helix-turn-helix motifs: $\alpha7$-turn-$\alpha8+\alpha9$-turn-$\alpha10$ (residues 106–125 + 140–159)

- C-terminal fragment $\alpha7$-turn-$\alpha8$-loop-$\alpha9$-turn-$\alpha10$ (residues 106–159)

Molecular dynamics simulations

MD simulations of the p18 fragments were carried out using the AMBER [21] suite of programs with the *parm99* force field [22, 23] and the following simulation protocol [24–26]: each system was neutralized by adding counter ions (either Cl^- or Na^+ ions according to solute charge) and immersed in a rectangular water box (TIP3P [27]) with a minimum distance between the solute and the box wall of 10 Å. Then, each system was minimized prior to the production phase of the

MD run, as follows. The protein was constrained and the solvent molecules with counter ions were allowed to move during a 1,000-step minimization followed by a 10-ps-long MD runs under $[NpT]$ conditions (p=1 atm, T=298.15 K). The side chains were then relaxed by several minimizations, with decreasing force constants applied to the backbone atoms. After the relaxation, each system was heated to 298.15 K as follows: from 10 K to 50 K for 20 ps, then from 50 K to 250 K for 20 ps, and finally from 250 K to 298.15 K for 30 ps. The particle-mesh Ewald (PME) method for treating electrostatic interactions was used, and all simulations were performed under periodic boundary conditions in the $[NpT]$ ensemble at 298.15 K and 1 atm using a 2 fs integration step. The SHAKE algorithm with a tolerance of 10^{-5} Å was used to fix positions of all hydrogens, a 9.0 Å cutoff was applied to nonbonding interactions and coordinates were stored every picosecond. Totally, nine independent simulations each 50 ns long were performed. Thus, the cumulative production time amounted to 450 ns.

Table 1 Basic features of studied fragments

Fragment	Charge (e)[a]	Number of residues	Number of heavy atoms	% of hydrophobic residues	Periodic box dimensions (Å)[b]
α7+α8	2	17	139	64.7	40.0×48.9×34.6
α9+α10	−1	17	126	58.8	41.2×44.6×37.4
α7-turn-α8	1	20	162	55.0	42.3×48.1×37.1
α9-turn-α10	0	20	153	50.0	42.4×46.4×38.2
ANK repeat IV	1	34	267	41.2	45.4×51.2×39.4
α7-turn-α8-loop-α9	2	41	317	43.9	45.1×54.3×43.6
ANK IV-loop-α9	1	45	346	42.2	45.7×54.6×45.1
α7-turn-α8+α9-turn-α10	1	40	315	52.5	44.2×49.7×44.6
α7-turn-α8-loop-α9-turn-α10	2	54	419	42.6	49.7×54.6×46.3

[a] The solute's total charge under physiological conditions

[b] Mean periodic box dimensions calculated from the last 5 ns of each molecular dynamic (MD) simulation

The stability and structural properties of the p18 fragments were evaluated by calculating root-mean-square-deviations (RMSDs) for backbone atoms from the initial structure, radius of gyration (R_g), secondary structure elements obtained from the DSSP program [28], and native contacts obtained using the *in-house* program RESDIST (P. B., unpublished software). The RESDIST program calculates a contact map (map of distances) among all residues represented by centers of masses of side chains. Contacts between $i \ldots i + 4$ residues and higher are considered as native (excluding contacts between $i \ldots i + 1$, $i \ldots i + 2$ and $i \ldots i + 3$ residues) if the distance between two residues is smaller than or equal to 6.0 Å. The program analyzes the number (percentage) of saved native contacts during the MD simulation. The mean values of all structure parameters are listed in Table 2.

Table 2 Summary of trajectories characteristics

Fragment	t (ns)[a]	Rg (MD)[b]	RMSD (Å)[c]	Native contacts (%)[d]	α-helicity (%)[e]
$\alpha7+\alpha8$	50	10.1 ± 3.4	6.1 ± 2.7	5.0 ± 6.7	14.0
$\alpha9+\alpha10$	50	8.5 ± 2.6	5.6 ± 1.9	0.7 ± 2.6	5.8
$\alpha7$-turn-$\alpha8$	50	7.7 ± 0.5	3.7 ± 0.3	22.3 ± 6.3	19.7
$\alpha9$-turn-$\alpha10$	50	9.0 ± 0.4	5.2 ± 0.6	12.6 ± 5.9	33.6
ANK repeat IV	50	8.3 ± 0.1	2.8 ± 0.1	24.3 ± 3.0	39.4
$\alpha7$-turn-$\alpha8$-loop-$\alpha9$	50	11.9 ± 0.5	10.0 ± 0.7	10.7 ± 3.1	34.6
ANK IV-loop-$\alpha9$	50	9.1 ± 0.1	3.2 ± 0.2	31.6 ± 3.4	54.0
$\alpha7$-turn-$\alpha8+\alpha9$-turn-$\alpha10$	50	8.5 ± 0.1	1.2 ± 0.3	75.3 ± 6.2	89.1
$\alpha7$-turn-$\alpha8$-loop-$\alpha9$-turn-$\alpha10$	50	9.4 ± 0.1	1.7 ± 0.3	53.5 ± 5.3	83.7

[a] The time evolution of presented structure characteristics are shown in the Supplementary material

[b] Mean radius of gyration $\pm$ SD of main chain atoms calculated from the last 5 ns of each simulation

[c] Root-mean-square-deviation of helices main chain atoms compared with the initial structure calculated as the mean $\pm$ SD from the last 5 ns of each MD simulation

[d] Mean number of native contacts $\pm$ SD calculated from the last 5 ns of each MD simulation

[e] Mean native α-helical content averaged along the MD simulation. The native α-helical contents are normalized to the initial structure, for which the α-helical content equals 100 %

Results

Helix pair $\alpha7+\alpha8$

Several hydrophobic contacts form between residues in the $\alpha7+\alpha8$ pair of α-helices (Leu122 with Pro106 and Leu107; Leu107 with Val123; and Ala110 with Val118 and Val119, see Fig. 3a), but this fragment is highly unstable and both helices rapidly lose their initial α-helical structures. The unfolding of the $\alpha7$ helix began from the N-terminus at $\sim$200 ps. Shortly thereafter ($\sim$300 ps) unfolding of $\alpha8$ started, again from the N-terminus. Then, at $\sim$600 ps, the middle turn of $\alpha8$ was reestablished and remained stable until Ala110 lost its contact with Val119 ($\sim$800 ps), which also initiated separation of the $\alpha7$ C-terminus from the $\alpha8$ N-terminus, accompanied by a significant lowering of the number of native contacts and α-helical content (cf. Fig. 4). The helices were then held together via hydrophobic contacts of Leu107 with Val119 and the hydrophobic portion of Glu120, and via an H-bond that formed between the Glu120 carboxyl oxygen and the His108 side chain (Fig. 3b). Those interactions were sufficiently strong to prevent dissociation of the two helices. Even at 1.5 ns, α-helices were still oriented in the antiparallel arrangement due to hydrophobic interaction among Leu107, Leu122 and Val119 (Fig. 3b). From this time until $\sim$9.5 ns, the $\alpha7$ helix refolded repeatedly, whereas the $\alpha8$ secondary structure remained in the non-helical conformation. Thereafter, refolding of the $\alpha8$ was initiated from the C-terminus and assisted by the formation of a hydrophobic cluster consisting of Leu107, Val118, Val119, Leu122 and Val123 (Fig. 3b). The reestablished α-helical structure of $\alpha8$ was maintained for $\sim$3 ns. Subsequently (at $\sim$12.5 ns), both α-helices lost their α-helical content (cf. Fig. 4). The dissociation of both α-helices appeared at $\sim$32 ns and no other refolding event was observed during the rest of the simulation.
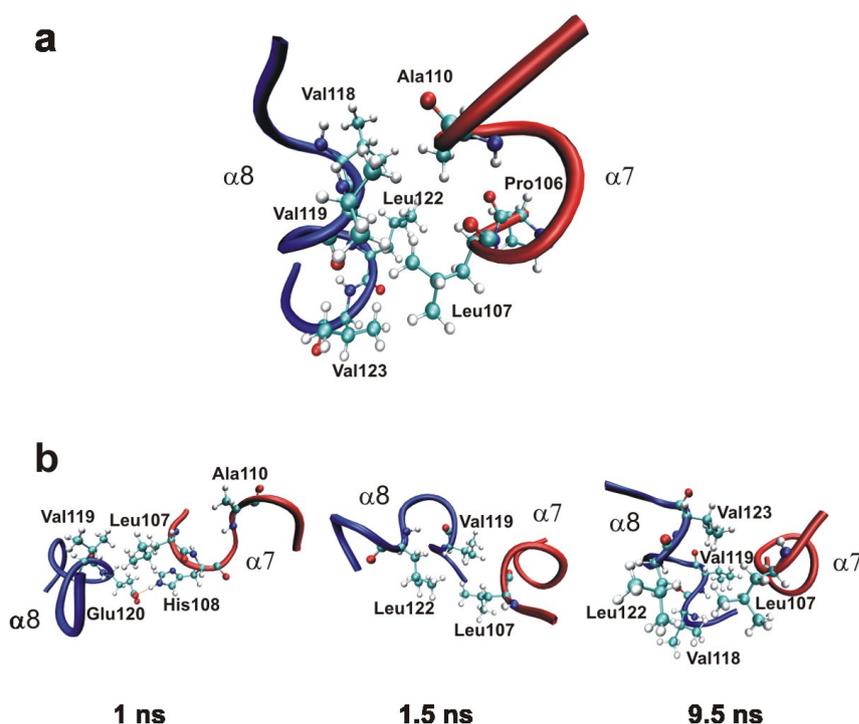
Figure 3: (a) The $\alpha7$ (*red*) and $\alpha8$ (*blue*) helix pair in its native conformation. (b) Snapshots from the $\alpha7+\alpha8$ fragment molecular dynamics (MD) simulation taken at various times

Helix pair $\alpha9+\alpha10$

The hydrophobic face of the $\alpha9$ helix consists of Cys141 and Ala144, while the hydrophobic face of $\alpha10$ consists of Val152, Val153 and Met156 (Fig. 5a). As in the previous case, this structural motif is highly unstable when taken out of its native context, readily losing its initial α-helical structure. Firstly, the $\alpha9$ unfolded simultaneously (from both termini) at $\sim$200 ps (Fig. 5b). This process was accompanied by a $\sim$20% reduction in native contacts. Then, at $\sim$350 ps, $\alpha10$ unfolded from the C-terminus and an additional $\sim$20% decrease in native contacts was observed (Fig. 5b). The $\alpha9+\alpha10$ pair dissociated at $\sim$6 ns (Fig. 5b) and, simultaneously, the α-helical content significantly decreased (cf. Fig. 4). Notably, except for a reformation of the N-terminal turn of the $\alpha10$ helix, which appeared at $\sim$40 ns, no other refolding events were observed.

$\alpha7$-turn-$\alpha8$ motif

Three residues (Glu113, Gly114 and His115) with side chains exposed to water make the turn connecting $\alpha7$ with $\alpha8$. The $\alpha7$-turn-$\alpha8$ fragment proved to be less dynamic than the unconnected variant (i.e., the pair of helices); the presence of the turn significantly restricts the conformational space of the fragment, thereby slowing kinetic processes such as unfolding and structural rearrangements.

At $\sim$2.2 ns, $\alpha7$ unfolded from the C-terminus, an event initiated by loss of the H-bond between the backbone amide hydrogen of His108 and the backbone carbonyl group of Lys112. By $\sim$2.5 ns, all native α-helical H-bonds were broken and the entire $\alpha7$ helix was fully unfolded. After that, the

main chain axis of the unfolded $\alpha 7$ flipped out of the helix-turn-helix plane by $\sim 90°$ (Fig. 6). The $\alpha 8$ helix then unfolded from the C-terminus at ~ 3.2 ns. Simultaneously, the distance between the N-terminus of $\alpha 7$ and C-terminus of $\alpha 8$ increased while the number of native contacts decreased (Fig. 6). Loss of the H-bond between the carbonyl group of Leu116 and the amide hydrogen of Glu120 probably initiated $\alpha 8$ melting. The middle turn of $\alpha 8$ adopted a 3_{10}-helix conformation, which was well maintained during the following simulation run. At ~ 3.5 ns, the N-terminal helix turn again reformed, and was maintained for ~ 1 ns, then, at 5.2 ns, a second refolding of the $\alpha 7$ helix occurred, which was propagated from the C-terminus and maintained for ~ 4 ns. Reestablishment of the $\alpha 7$ helical structure was initiated by the formation of hydrophobic contacts among the side chain of Leu109, the hydrophobic moiety of Lys112 and the hydrophobic moiety of Glu113 (Fig. 6). At ~ 6 ns, both helices again rearranged into the antiparallel conformation and the number of native contacts increased. At ~ 11.5 ns, the last refolding event of $\alpha 7$ occurred, beginning from the N-terminus, and the helical structure was maintained until ~ 19.5 ns. Then, the C-terminal turn of $\alpha 8$ refolded twice (at ~ 20 ns and ~ 23.5 ns), for ~ 2 ns on each occasion. No other refolding events were observed during the rest of the simulation. Finally, at ~ 46 ns, the opposite termini of individual helices moved away (Fig. 6). This extended structure of the $\alpha 7$-turn-$\alpha 8$ fragment was retained until the end of the simulation.

$\alpha 9$-turn-$\alpha 10$ motif

The turn here consists of Tyr147, Gly148 and Arg149 residues, all of which have water-exposed side chains. The mean α-helical content of the $\alpha 9$-turn-$\alpha 10$ fragment is almost six times higher than that of the $\alpha 9+\alpha 10$ pair (cf. Table 2), i.e., the two helices without the covalent linker. In the case of both $\alpha 7+\alpha 8$ vs $\alpha 7$-turn-$\alpha 8$ and $\alpha 9+\alpha 10$ vs $\alpha 9$-turn-$\alpha 10$, the turns restrict the conformational space and significantly enhance the α-helical propensity of the respective fragments.

Unfolding of the $\alpha 10$ helix started from the N-terminus at ~ 0.7 ns and was initiated by formation of an H-bond between the backbone carbonyl group of Asn150 and the side chain hydroxyl of Ser154, thereby disrupting the H-bond between the backbone carbonyl of Asn150 and the amide hydrogen of Ser154. This was followed by loss of the H-bond between the backbone carbonyl moiety of Leu155 and backbone amide hydrogen of Asn159, and consequent unfolding of the C-terminal and middle turn of $\alpha 10$. After that, formation of an H-bond between the carbonyl group of Val153 and side chain of Gln157 mediated disruption of the H-bond between the carbonyl group of Val153 and backbone amide hydrogen of Gln157 (Fig. 7). At ~ 6.2 ns, the $\alpha 9$ helix flipped out of the helix-turn-helix plane by almost $\sim 90°$ (T-shape arrangement, see Fig. 7). Then, at ~ 15 ns, the most significant refolding of the $\alpha 10$ middle turn appeared, assisted by formation of an H-bond between the carbonyl group of Val153 and the side chain of Gln157. The reformed middle turn of the $\alpha 10$ helix was stable for ~ 2 ns. Then, at ~ 17 ns, the hydroxyl group of Tyr147 (turn residue) formed an H-bond with the backbone carbonyl oxygen of Asn159 (Fig. 7). Simultaneously, the middle turn of the $\alpha 10$ again refolded due to formation of an H-bond between the backbone carbonyl of Ser154 and side chain of Gln157. After that (at ~ 20 ns), the distance between the two helices increased (the opposite helical termini moved away), accompanied by a reduction in native contacts. Simultaneously, $\alpha 9$ unfolded from its N-terminus. It is likely that the loss of
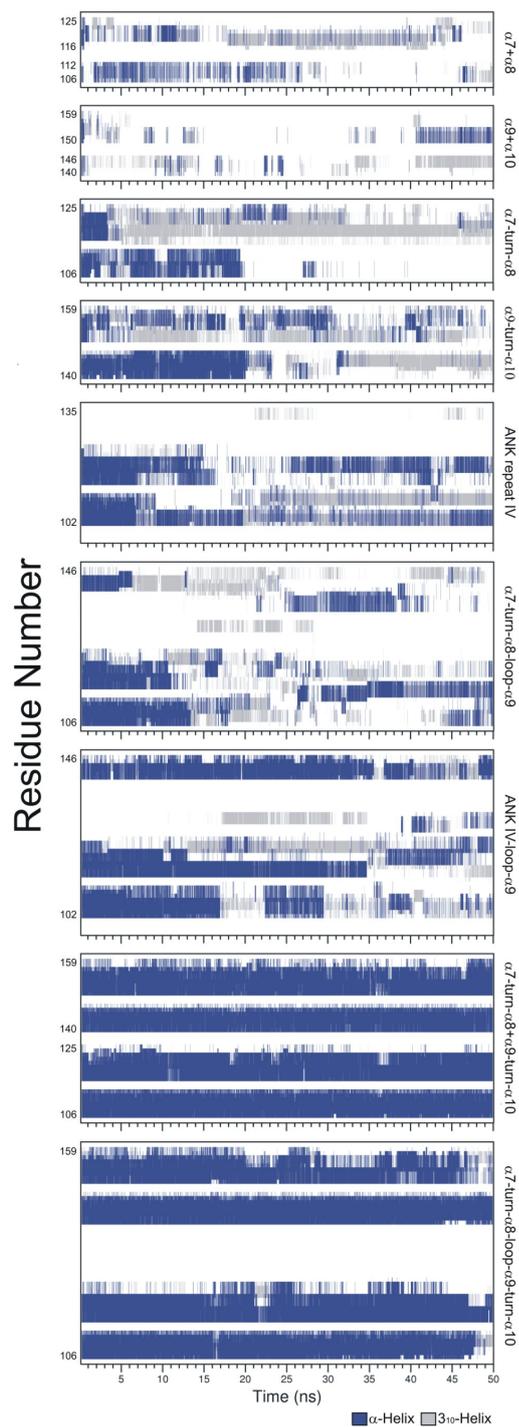
Figure 4: Plot of the secondary structure elements calculated by DSSP software as a function of time for all simulated p18 fragments. *Blue bars* residues with α-helical conformation, *gray bars* residues with 3_{10}-helical conformation. For the sake of simplicity, other secondary structure elements are not depicted
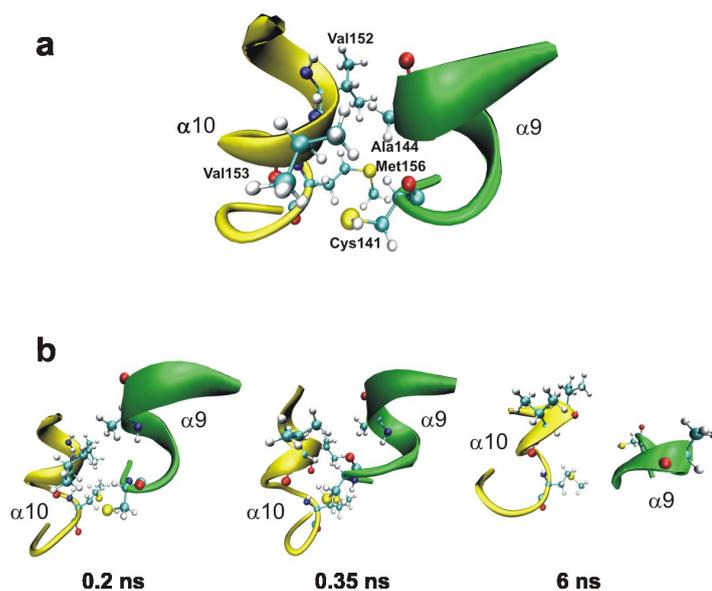
Figure 5: (a) The α9 (*green*) and α10 (*yellow*) helix pair in its native conformation. (b) Snapshots from the α9+α10 fragment MD simulation taken at various times
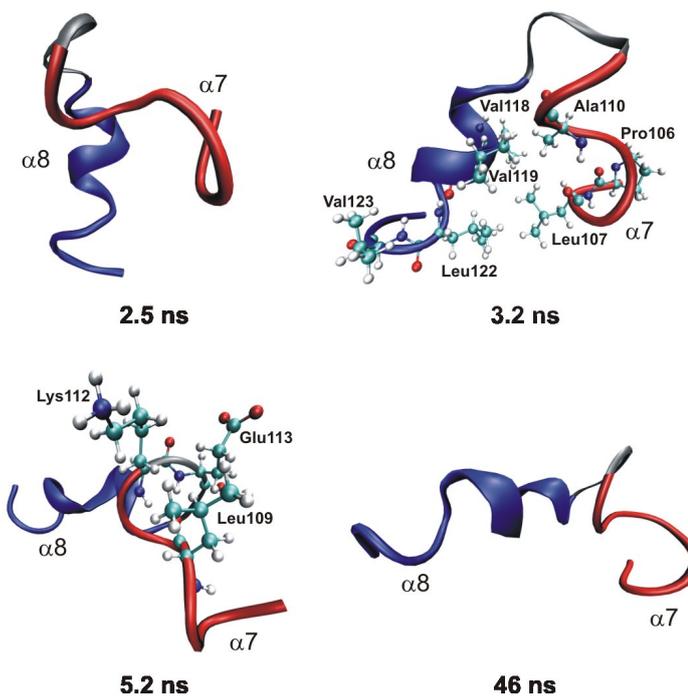
Figure 6: Snapshots taken from the α7-turn-α8 fragment MD simulation at various times; *red* α7, *gray* turn, *blue* α8

hydrophobic contacts among Leu143, Ala145 and Val153 initiated α9 unfolding. At $\sim$22.5 ns, the α9 helix briefly (for $\sim$1 ns) refolded, mediated by reformation of the hydrophobic cluster consisting of Leu143, Ala145 and Val153. Between $\sim$24.3 and $\sim$24.8 ns, the fragment adopted a parallel

arrangement (Fig. 7). After that, the T-shape arrangement of the fragment again appeared. At
∼31.1 ns, the last refolding of α9 was observed, and the helical structure was maintained for
∼0.9 ns. From this point onwards, no other refolding of the α9 helical structure was observed,
although α10 refolded several times in the remaining simulation time. At ∼31.3 ns, the antiparallel
arrangement of both helices was reestablished and maintained until ∼38.5 ns. Thereafter, the
T-shape conformation again appeared. Finally, at ∼42 ns, opposite termini of the two helices
moved away and the fragment adopted an extended structure, which was maintained until the end
of the simulation.
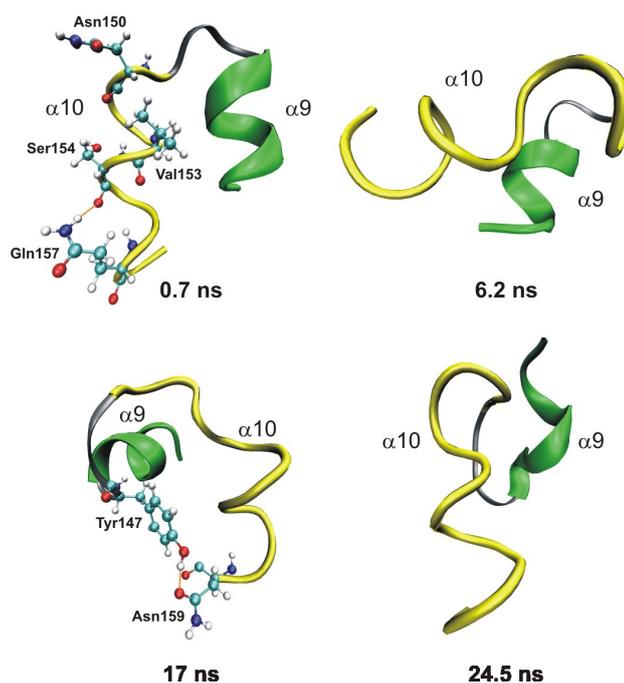
Figure 7: Snapshots taken from the α9-turn-α10 fragment simulation at various times; *green* α9, *gray*
turn, *yellow* α10

ANK repeat IV

The four C-terminal residues of α8 quickly (at ∼0.4 ns) adopted a flexible loop conformation.
Then, at ∼3.4 ns, all H-bonds between the N-terminal loop (residues 102–105) and the C-terminal
loop (residues 126–135) disrupted, and the flexibility of the C-terminal loop then rapidly increased
(Fig. 8). After that, at ∼6.5 ns, losses of H-bonds between the carbonyl group of Ala110 and the
backbone amide hydrogen of His115, and between the carbonyl moiety of Ala111 and backbone
amide hydrogen of Gly114 initiated unfolding of α7 from the C-terminus (Fig. 8). The α8 helix
unfolded at ∼8.5 ns from its N-terminus. During the following simulation run, the N-terminal
turn of α7 and the middle turn of α8 refolded several times. At ∼9 ns, the N-terminus of α7
was reestablished and remained stable until ∼13.5 ns. Between ∼11 ns and 14.5 ns, the middle
turn of α8 reformed, and this process was followed by propagation of the helical structure to the
N-terminus. The two helices moved away at ∼17.4 ns and the helix-turn-helix motif adopted a

"semicircular" arrangement (Fig. 8). At ∼18.8 ns, an H-bond between the His108 side chain and Asn129 carbonyl group initiated attachment of the C-terminal loop to the helix-turn-helix motif. At 27.5 ns, the ANK repeat IV adopted a globular arrangement, which remained stable until the end of the simulation. In this arrangement, the $\alpha7$ was kinked and connected via an H-bond between the Ala111 amide hydrogen and the Val130 carbonyl group (Fig. 8) to the C-terminal loop. The core of this globular fold consisted of a hydrophobic cluster of residues Ala111, Val119, Val123 and Val130.
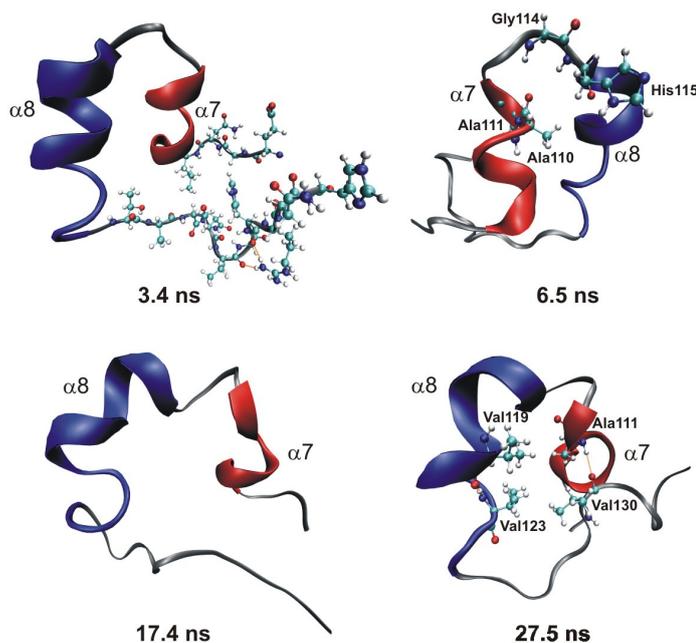
Figure 8: Snapshots taken from the ANK repeat IV fragment simulation at various times; *red* $\alpha7$, *gray* turn and loop, *blue* $\alpha8$

α7-turn-α8-loop-α9 motif

This fragment corresponds to the hypothetical folding unit suggested by Ferreiro et al. [19], the hydrophobic core of which consists of the following residues: Leu107, Ala110, and Ala111 in $\alpha7$; Val118, Val119, Leu122, and Val123 in $\alpha8$; and Ala140, and Ala144 in $\alpha9$. There is a salt bridge between the side chains of Lys112 and Glu113, which presumably stabilizes the C-terminal turn of $\alpha7$, and a salt bridge between Glu120 and Lys124 that stabilizes the helical structure of $\alpha8$ in the native conformation. The unstructured loop (residues 126–139) involves two H-bonds: one formed between the carbonyl oxygen of Asn129 and amide hydrogen of His132, and another between the amide hydrogen of Asn134 and carbonyl oxygen of Asp138 (Fig. 9a). The entire system is positively charged (+2e), since it includes three Lys, three Arg, two Asp and two Glu residues.

Apart from the increased flexibility of the $\alpha8$ and $\alpha9$ C-terminal turns, no other structural changes were observed in the first ∼6.5 ns, during which the $\alpha7$-turn-$\alpha8$-loop-$\alpha9$ fragment fold remained stable. Then, formation of a salt bridge among Arg145, Asp142 and Asp138 caused $\alpha9$ to adopt a 3_{10}-helix conformation (Fig. 9b). After a further 1 ns (at ∼7.5 ns), $\alpha9$ shifted towards the
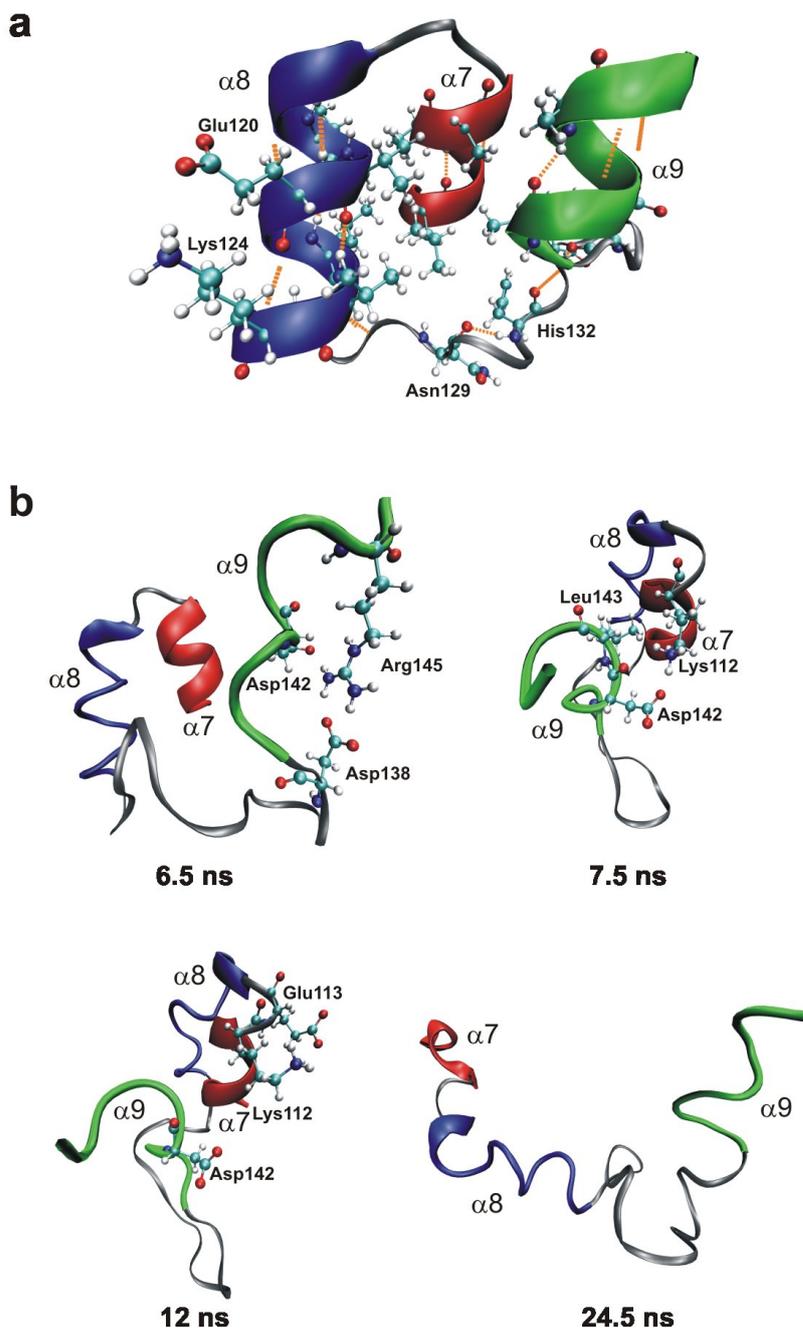
Figure 9: (a) Native structure of the α7-turn-α8-loop-α9 fragment (*red* α7, *gray* turn and loop, *blue* α8, *green* α9). (b) Snapshots from the α7-turn-α8-loop-α9 fragment MD simulation taken at various times

solvent-exposed side of α7. This motion was presumably directed by the formation of a salt bridge between Lys112 and Asp142 and a hydrophobic contact between Leu143 and the nonpolar moiety of Lys112 (Fig. 9b). After several nanoseconds (at ∼11 ns), α8 unfolded from the C-terminus and it is assumed that the bifurcated H-bond between the backbone carbonyl oxygen of His115 and the amide hydrogens of Val118 and Val119 initiated the unfolding process. The assembly of α7,

$\alpha 8$ and $\alpha 9$ formed at ~ 7.5 ns was broken at ~ 12 ns because the interaction between Lys112 and Asp142 vanished, due to the formation of a new salt bridge between Lys112 and Glu113 (Fig. 9b). After that (~ 13 ns), $\alpha 9$ moved away from $\alpha 7$ and $\alpha 8$, and losses of native hydrophobic contacts between $\alpha 7$ and $\alpha 9$ as well as between $\alpha 8$ and $\alpha 9$ were observed. Consequently, $\alpha 7$ unfolded from the C-terminus. The three-dimensional structure of this fragment was then extended and its core residues were significantly exposed to the water solvent. At ~ 24.5 ns, the antiparallel arrangement of the $\alpha 7$-turn-$\alpha 8$ motif was broken. Some hydrophobic residues became exposed to the solvent and the others formed non-native contacts that were not observed in the native fold (Fig. 9b), some of which assisted reformation of non-native α-helices spanning residues 113–117 and 139–143 (cf. Fig. 4).

ANK IV-loop-$\alpha 9$ motif

The Consensus 2 hypothetical minimal folding unit of the p18 protein consists of one ANK repeat motif (Consensus 2) followed by a C-terminal α-helix (Fig. 10a). The residues forming the hydrophobic core of the $\alpha 7$-turn-$\alpha 8$-loop-$\alpha 9$ fragment are listed above, and the H-bond network is similar to that described for the $\alpha 7$-turn-$\alpha 8$-loop-$\alpha 9$ fragment, except that additional H-bonds are formed between the backbone carbonyl oxygen and amide hydrogen of Glu102 and His135, and between the Asn104 side chain carbonyl oxygen and side chain amide hydrogen of Asn134 (Fig. 10a). The formal charge of the respective fragment is $+1e$. Three Lys, three Arg, three Glu and two Asp are involved in the fragment structure.

The conformation of the C-terminus of the $\alpha 9$ changed to a turn, stabilized by H-bonds formed among the main chain carbonyls of Asp142 or Leu143 and backbone amide hydrogens of Arg145, Leu146 and N-methyl. At ~ 2.5 ns, the C-terminus of $\alpha 8$ lost its helical conformation and also adopted a turn conformation. Loss of the $\alpha 8$ C-terminal helical turn then presumably initiated a shift (lasting ~ 5 ns) of $\alpha 8$ towards $\alpha 9$, to the position where $\alpha 10$ (which is missing in this fragment) occurs in the p18 structure (Fig. 10b). The C-terminal turn of $\alpha 8$ adopted a 3_{10}-helix conformation at ~ 13 ns. Then, the $\alpha 7$ unfolded from both termini at the same time, at ~ 17.3 ns, and the C-terminus of the $\alpha 7$ changed to a turn conformation, which was stabilized mainly by the formation of a bifurcated H-bond between the backbone carbonyl of Ala110 and both the amide hydrogen of Glu113 and His115 (Fig. 10b). At ~ 35 ns, $\alpha 8$ completely unfolded, while $\alpha 9$ unfolded from the C-terminus (Fig. 10b). From this time onward, all of the helices were able to refold to α-helical conformations, but these secondary structure elements were short-lived, appearing for periods ranging from approximately 1 to 8 ns. The major refolding events appeared in the $\alpha 8$ and $\alpha 9$ helices. The $\alpha 8$ helix reestablished its middle turn at ~ 37 ns and maintained it until ~ 45 ns. After that, the C-terminal turn of the $\alpha 8$ was quickly reformed and remained stable during the next simulation run. At ~ 48 ns, the helical structure of $\alpha 9$ fully reformed, and was maintained until the end of the simulation. It is likely that the hydrophobic contacts among Val119, Leu122, Val123, together with the hydrophobic moiety of Arg133, Cys141 and Ala144 stabilized the C-terminus of $\alpha 8$ as well as the helical structure of the $\alpha 9$ helix (Fig. 10b).
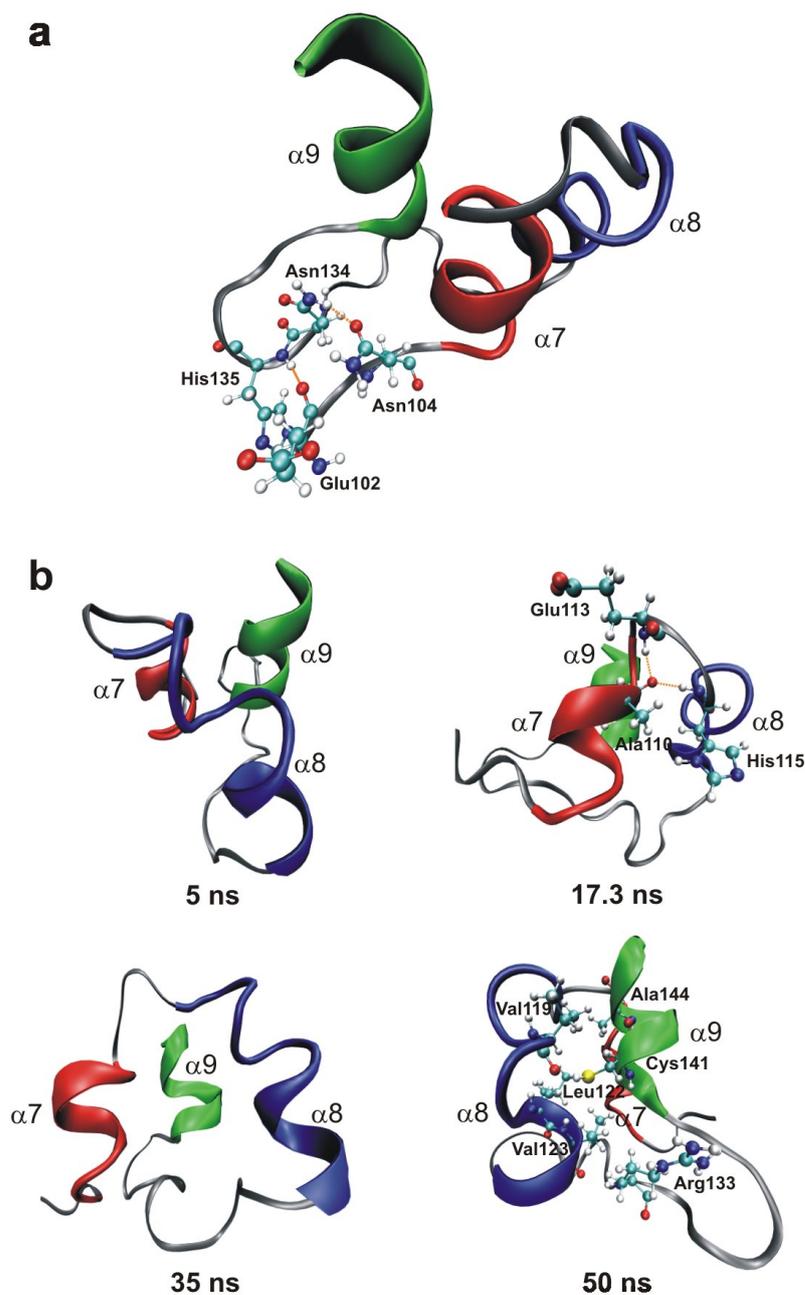
Figure 10: (a) Native structure of the ANK IV-loop-α9 fragment (*red* α7, *gray* turn and loop, *blue* α8, *green* α9). (b) Snapshots from the ANK IV-loop-α9 fragment MD simulation taken at various times

Four-helical bundles

Four-helical bundles (i.e., α7-turn-α8+α9-turn-α10 and α7-turn-α8-loop-α9-turn-α10 fragments) were significantly more stable than the smaller fragments, and their α-helical contents were well maintained in the simulation timescale (cf. Fig. 4). The α7 and α9 helices were generally more rigid than the α8 and α10 helices (Fig. 11), and the average α-helical contents of the α7-turn-α8+α9-turn-α10 and α7-turn-α8-loop-α9-turn-α10 fragment were 89% and 84%, respectively. Thus, the

four-helical bundles are slightly less stable when the loop connecting the two helix-turn-helix motifs is present than when it is absent (i.e., in the α7-turn-α8+α9-turn-α10 fragment, see Fig. 11). This is because of the greater flexibility of the α8-α9 loop (the loop connecting α8 with α9 in the α7- turn-α8-loop-α9-turn-α10 fragment), which readily makes contacts with the helices. It has substantially higher flexibility in simulations of these bundles, since it lacks inter-strand hydrogen bonds to the N-terminal hairpin of ANK IV, than in MD simulations of the ANK IV-loop-α9 fragment, where it does make these inter-strand hydrogen bonds.
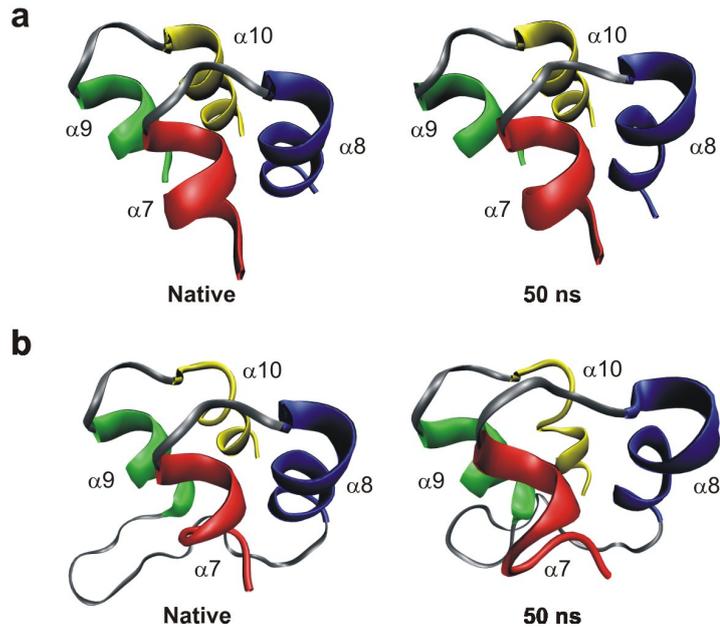
Figure 11: (a) Initial and last structures from the MD simulation of α7-turn-α8+α9-turn-α10 fragment (*red* α7, *gray* turns, *blue* α8, *green* α9, *yellow* α10). (b) Initial and last structures from the simulation of α7-turn-α8-loop-α9-turn-α10 fragment (*red* α7, *gray* turns and loop, *blue* α8, *green* α9, *yellow* α10)

Discussion

Here, we present all-atomic MD simulations, 450 ns long in total, in explicit solvent of various different C-terminal fragments of the p18 protein (Table 1). The main aim of this study was to test the hypothesis of Ferreiro et al. [19], who proposed from Gõ-type simulations that the hypothetical minimal folding module of ARP consists of one complete ANK repeat motif and the first helix of the following ANK repeat. According to this hypothesis, the minimal folding module of ARP contains three helices connected by a turn and loop. We selected nine C-terminal fragments of p18 to test this hypothesis. The smallest studied systems consisted of pairs of helices (two) while the largest contained two ANK repeats, coinciding with the minimal folding unit of ARP suggested by Zhang and Peng based on results obtained from proteolytic analyses and prediction algorithm [7].

Separations of the helices, and resulting elongation of the structures, were observed after $\sim$3.2 ns and $\sim$20 ns in simulations of the helix-turn-helix motifs of the α7-turn-α8 and α9-turn-α10 fragments, respectively. Both of these fragments were unstable, with mean α-helical contents

equal to 19.7% and 33.6%, respectively, and the contacts among hydrophobic residues in them were insufficiently strong to maintain their native folds. These observations are in agreement with findings of Du and Gai [29], who concluded that stability of the helix-turn-helix (α-hairpin) motif is enhanced by strong inter-strand hydrophobic clusters. The two helix pairs ($\alpha7+\alpha8$ and $\alpha9+\alpha10$) were the most unstable systems of all studied fragments, and both displayed similar behavior. Their native α-helical content quickly decreased due to unfolding of individual helices, accompanied by the loss of many native contacts, then several refolding events were observed until the individual helices dissociated, after which no further refolding of the α-helices was observed.

The C-terminal ANK repeat IV did not maintain its native fold and, after several nanoseconds, collapsed to a globular structure (cf. Fig. 8), which was likely stabilized by hydrophobic contacts and H-bonds between α-helical residues and residues located in the C-terminal loop. The mean α-helical content of the ANK repeat IV amounted to 39% and the mean percentage of preserved native contacts of the final structures was $\sim$24%.

The three helical bundles occurring in the ANK IV-loop- $\alpha9$ and $\alpha7$-turn-$\alpha8$-loop-$\alpha9$ fragments were not stable during the observed timescale. The mean α-helical content of the ANK IV-loop-$\alpha9$ motif was 54%, and the content of the $\alpha7$-turn-$\alpha8$-loop-$\alpha9$ motif was just 35%. The C-terminal region of the $\alpha7$-turn-$\alpha8$-loop-$\alpha9$ fragment was more flexible than the ANK IV-loop-$\alpha9$ system. Lack of the four N-terminal β-hairpin residues in the $\alpha7$-turn-$\alpha8$- loop-$\alpha9$ fragment prevents H-bonding of this moiety with the loop connecting $\alpha8$ with $\alpha9$. Therefore, the C-terminal region of the $\alpha7$-turn-$\alpha8$-loop-$\alpha9$ fragment gained higher flexibility and, early in the simulation (at 24.5 ns), the fragment adopted an extended structure.

The largest fragments, containing two pairs of helix-turn- helix motifs (i.e., four helical bundles), were the most stable systems among those simulated (Fig. 11), as indicated by their high mean α-helical contents of 89% ($\alpha7$-turn-$\alpha8+\alpha9$-turn-$\alpha10$) and 84% ($\alpha7$-turn-$\alpha8$-loop-$\alpha9$-turn-$\alpha10$). These values, together with the relatively low associated and convergent RMSDs and R_gs values, and high percentage of preserved native contacts (see Table 2), indicate that these fragments are stable in the 50 ns long simulation timescale applied here.

Conclusions

The all-atomic MD simulations in explicit solvents presented here show that the hypothetical minimal folding module of ARP p18 comprising one ANK repeat and the first helix of the following ANK repeat (containing three connected helices) proposed by Ferreiro et al. [19], is not stable at a 50 ns timescale and should not be regarded as the minimal folding unit of p18. On the other hand, the minimal stable unit of ARP, consisting of two ANK repeats stacked with each other (containing four connected helices), as suggested by Zhang and Peng [7], was considerably more stable at the same timescale. The lower stability of the hypothetical folding module is attributable to the larger exposure of the hydrophobic core residues to water. Since each non-terminal ANK repeat has two hydrophobic sides, we may further hypothesize that at least one hydrophobic side must be fully covered, shielded from water, to maintain ANK repeat stability. Thus, at least two ANK repeats are required to make a stable ARP.

Acknowledgments

References

1. Sedgwick SG, Smerdon SJ (1999) Trends Biochem Sci 24:311–316

2. Mohler PJ, Gramolini AO, Bennett V (2002) J Cell Sci 115:1565–1566

3. Serrano M, Hannon GJ, Beach D (1993) Nature 366:704–707

4. Mosavi LK, Williams S, Peng ZY (2002) J Mol Biol 320:165–170

5. Walker RG, Willingham AT, Zuker CS (2000) Science 287:2229–2234

6. Michaely P, Bennett V (1993) J Biol Chem 268:22703–22709

7. Zhang B, Peng ZY (2000) J Mol Biol 299:1121–1132

8. Mosavi LK, Minor DL, Peng ZY (2002) Proc Nat Acad Sci USA 99:16029–16034

9. Tripp KW, Barrick D (2007) J Mol Biol 365:1187–1200

10. Ferreiro DU, Cervantes CF, Truhlar SME, Cho SS, Wolynes PG, Komives EA (2007) J Mol Biol 365:1201–1216

11. Binz HK, Kohl A, Pluckthun A, Grutter MG (2006) Proteins: Struct Funct Bioinf 65:280–284

12. Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grutter MG, Pluckthun A (2004) Nat Biotechnol 22:575–582

13. Binz HK, Stumpp MT, Forrer P, Amstutz P, Pluckthun A (2003) J Mol Biol 332:489–503

14. Devi VS, Binz HK, Stumpp MT, Pluckthun A, Bosshard HR, Jelesarov I (2004) Protein Sci 13:2864–2870

15. Zahnd C, Wyler E, Schwenk JM, Steiner D, Lawrence MC, McKern NM, Pecorari F, Ward CW, Joos TO, Pluckthun A (2007) J Mol Biol 369:1015–1028

16. Kohl A, Binz HK, Forrer P, Stumpp MT, Pluckthun A, Grutter MG (2003) Proc Nat Acad Sci USA 100:1700–1705

17. Mosavi LK, Peng ZY (2003) Protein Eng 16:739–745

18. Main ERG, Lowe AR, Mochrie SGJ, Jackson SE, Regan L (2005) Curr Opin Struct Biol 15:464–471

19. Ferreiro DU, Cho SS, Komives EA, Wolynes PG (2005) J Mol Biol 354:679–692

20. Venkataramani R, Swaminathan K, Marmorstein R (1998) Nat Struct Biol 5:74–81

21. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P (1995) Comput Phys Commun 91:1–41

22. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197

23. Wang JM, Cieplak P, Kollman PA (2000) J Comput Chem 21:1049–1074

24. Bartova I, Otyepka M, Kriz Z, Koca J (2004) Protein Sci 13:1449–1457

25. Otyepka M, Bartova I, Kriz Z, Koca J (2006) J Biol Chem 281:7271–7281

26. Bartova I, Otyepka M, Kriz Z, Koca J (2005) Protein Sci 14:445–451

27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935

28. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637

29. Du DG, Gai F (2006) Biochemistry 45:13131–13139

30. DeLano WL (2002) DeLano Scientific. `http://pymol.sourceforge.net/`

10.2   Appendix B

In Silico Structural and Functional Analysis of Fragments of the Ankyrin Repeat Protein p18$^{\text{INK4c}}$

Petr Sklenovský • Michal Otyepka

Abstract

Ankyrin repeat proteins (ARPs) are ubiquitous proteins that play critical regulatory roles in organisms and consist of repeating motifs (ankyrin repeats) stacked in non-globular, almost linear, "quasi one-dimensional" configurations. They also have highly unusual mechanical properties, notably ARPs can behave as nano-springs. Both their essential cellular functions and distinctive nano-mechanical properties have aroused interest in ARPs for potential applications in medicine and nanotechnology. Further, the modular architecture of ARPs, which lack the long-range contacts that typically stabilize globular proteins, provides a new paradigm for understanding protein stability and folding mechanisms of proteins. In the present study, the stability of ARP p18INK4c (p18) and fifty p18 fragments was investigated by all-atomic molecular dynamics (MD) simulations in explicit water on a ∼3.3 micro-seconds timescale. The fragment simulations indicate that p18 α-helices are significantly stabilized by tertiary interactions, because in the absence of their native context they readily melt. All single p18 ARs and their structural elements are also unstable outside their native context. The minimal stable motifs are pairs of ARs, implying that inter-repeat contacts are essential for AR stability. Further, pairs of internal ARs are less stable than pairs that include a native capping AR. The MD simulations also provide indications of the functional roles of p18 turns and loops; the turns appear to be essential for the stability of the protein, while the loops both help to stabilize the p18 structure and are involved in recognition processes. Temperature-induced unfolding analysis shows that the p18 melts from the N-terminus to the C-terminus.

Keywords

Molecular dynamics simulation; Ankyrin repeat proteins; p18 fragments; Minimal folding unit; Protein unfolding; Protein stability.

Introduction

Ankyrin repeat proteins (ARPs) are ubiquitous proteins that have a non-globular, symmetrical structure consisting of four to tens of copies of repeating units called ankyrin repeats (ARs) (1, 2). The AR is a superhelical structural motif that is ∼33 residues long and composed of two β-strands and a pair of antiparallel α-helices connected by a short turn linkage (3). The ARs stack together,

so the overall structure of ARPs is roughly linear (or C-shaped for longer ARPs), with an extended α-helical hydrophobic core and β-hairpins (here named loops) exposed to the surrounding medium, as illustrated in (Figure 1) (4). The three-dimensional structure of ARPs is tightly connected with their biological functions, (1, 3, 5-7) which include participation in the regulation of diverse cellular processes by specifically binding to target molecules (8). Further, it was recently discovered that larger ARPs have interesting elastic properties, allowing them to act as nano-springs. Hence, they have very promising applications in nanomaterial science (9-11). Designed ARPs (DARPins) have also been identified as a novel class of binding molecules with broad potential as next-generation protein therapeutics in medicine (12, 13).

Unlike globular proteins, due to their near-linear or "quasi one-dimensional" architecture, (14) ARPs lack long-range interactions and have low contact order (15). Hence, analysis of the structure of ARPs avoids the "dissection" and "comparison" that often complicate studies of the stability of secondary structural elements and energy landscape mapping of more structurally complex, globular proteins (16). Thus, the modular architecture of ARPs provides an ideal model for exploring various general aspects of protein stability and folding mechanisms. In addition, insights regarding the stability of ARPs' building blocks and the interactions both amongst them and with other molecules should greatly facilitate the design of DARPins that could have wide nanotechnological and medical applications, as mentioned above.

Despite their non-globular structure, some ARPs fold via a two-state mechanism, implying that there is a high degree of cooperativity among repeats during folding (17). However, ARPs with larger numbers of repeats and/or varying degrees of stability among their repeats appear to fold via a multistate (sequential) mechanism (17, 18). The variations in folding stability among ankyrin repeats in ARPs and their folding mechanisms have been explored in several experimental analyses of various ARPs and their fragments and computer simulations (19-24). The structural features of internal and capping ARs, and the differences between them, have attracted particular attention since understanding the roles of these structural elements could facilitate attempts to design stable, functional DARPins (25, 26).

Another important objective is to identify a minimal stable ARP. Zhang and Peng (27) showed that two C-terminal ARs of the protein p16$^{\text{INK4a}}$ can fold in the absence of the native context, suggesting that pairs of ARs may form stable, foldable units of ARPs, although they observed no folding of a fragment consisting of the two N-terminal repeats. All-atomic molecular dynamics simulations have also recently shown that two C-terminal ARs of the protein p18INK4c can form a stable unit (28) and a stable artificial DARPin, consisting of just two ARs, has been produced by protein engineering (29). In contrast, experimental and theoretical studies have consistently indicated that a single AR would not be stable in solution, (19, 27, 28, 30) implying that intra-repeat stabilizing interactions are insufficient to stabilize single repeats. This finding was recently confirmed in quantitative experiments by Mello and Barrick (31) and theoretical calculations by Otyepka et al., (32). However, a coarse-grained simulation by Ferreiro and co-workers (33) suggested that the minimal stable module of ARPs may be "intermediate" between a single ARP and a pair of repeats, consisting of a complete AR and the first α-helix of the neighboring repeat from its C-terminus. It was also proposed that this motif can trigger the folding of the remaining

repeats. However, whether or not such a three-helical module can exist in a folded state in solution independently of its native context remains to be confirmed.
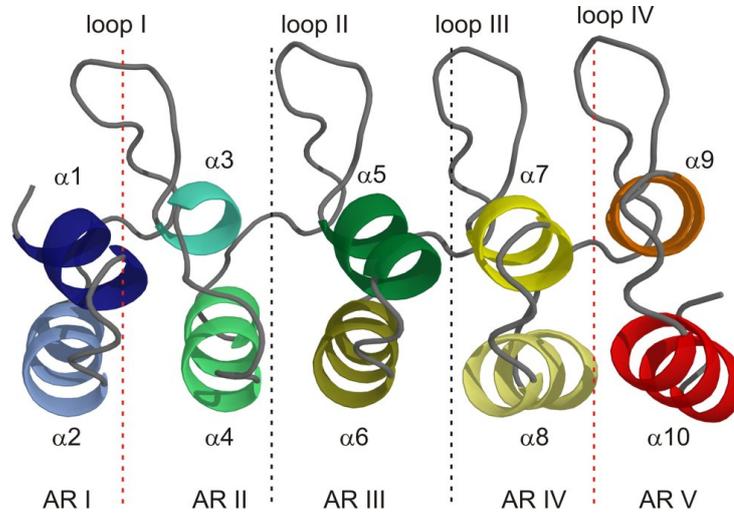
Figure 1: Crystal structure of p18$^{\text{INK4c}}$ ankyrin repeat protein (PDB ID 1IHB, chain B) with the following color coding of α-helices: α1 dark blue, α2 light blue, α3 cyan, α4 light green, α5 dark green, α6 dark olive, α7 yellow, α8 pale yellow, α9 orange, and α10 red. Turns and loops are colored gray. Dashed gray lines indicate the interfaces between internal ankyrin repeats while the red lines indicate the interfaces between terminal and adjacent internal repeats.

Clearly, much remains to be learnt about the structure and functional roles of ARPs and their modular units. In the study presented here a naturally occurring protein of the ARP family, p18$^{\text{INK4c}}$ (p18, Figure 1), was selected to: elucidate the role of native tertiary contacts in the stability of p18 α-helices; determine the structure of the minimal stable unit of the protein; assess the functional importance of turns and loops in its structure; and explore the roles of internal and capping ARs on the stability of ARPs. For these purposes the p18 protein was dissected into fifty fragments, differing in terms of topology, size and/or position in the intact protein (Table SI; Figure 2). Each fragment and the complete p18 protein were then modeled in a room-temperature (298 K), 50+ ns timescale, all-atom MD simulation in explicit water. The total simulation time was $\sim$3.2 μs.

The room-temperature simulations provide indications of the relative stability of all p18 fragments (structural elements) and the distribution of folding stability throughout the p18 protein. They also provide information about: the structure of the minimal topological unit of p18 that can retain a native-like fold when taken out of its native context; the role of native tertiary contacts in the stability of p18 α-helices; and the functions of p18 turns and loops at an all-atomic level of accuracy. In addition, simulations at elevated temperature (498 K) provide indications of the unfolding mechanism of the p18 protein.

Methods

Studied Fragments: p18 is a member of the INK4 family (34, 35) of proteins (inhibitors of cyclin-dependent kinases), which controls the G1-S phase of the cell cycle by specifically inhibiting
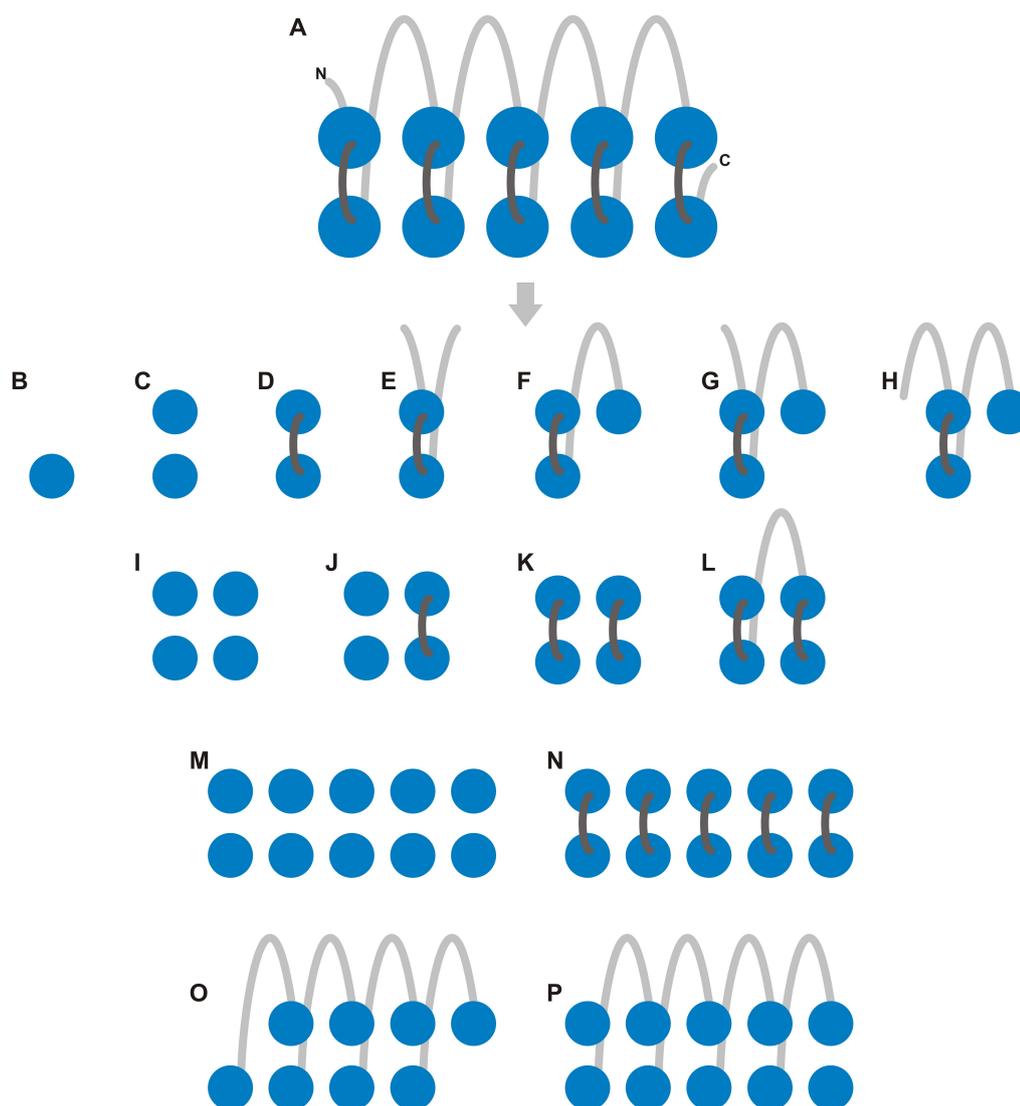
Figure 2: (A-P) Scheme showing the topology of the p18$^{\text{INK4c}}$ protein and its fragments studied by MD simulations. Black circles represent α-helices, light gray lines loops and dark gray lines turns. Studied fragments are listed in Table SI.

CDK4-cyclin D and CDK6-cyclin D complexes (36-38). The crystal structure of the p18 protein (39) (PBD entry 1IHB, chain B) with optimized positions of hydrogen atoms (obtained using the Sander module of AMBER (40), *parm99* force field (41, 42)) was used as a template for studying 50 p18 fragments in MD simulations (Table SI). In this study, we used the originally described sequence of p18 (39), as a template to generate consensus 2 ankyrin repeats. The N-terminus of each fragment was acetylated and the C-terminus was capped with an *N*-methylamine group. In addition, for simplicity, the β-hairpin portions of each AR were approximated as loops in this study. In other words, a loop represents the covalent linkage between adjacent helix-turn-helix motifs.

Molecular Dynamics Simulations at Room Temperature (298 K): MD simulations of the entire p18 protein and p18 fragments were carried out using the AMBER (40) suite of programs with the *parm99* force field (41, 42), and the following simulation protocol (43-45). Each system was neutralized by adding counter ions (either Cl^- or Na^+, according to solute charge) and immersed in a rectangular water box (TIP3P) (46) with a minimum distance between the solute and the box wall of 10 Å. Then, each system was minimized prior to the production phase of the molecular dynamics run, as follows. The protein was constrained and the solvent molecules with counter ions were allowed to move during a 1000-step minimization followed by 10 ps long molecular dynamics runs under $[NpT]$ conditions ($p = 1$ atm, $T = 298.15$ K). The sidechains were then relaxed by several minimizations, with decreasing force constants applied to the backbone atoms. After the relaxation, each system was heated to 298.15 K as follows: from 10 K to 50 K for 20 ps, then from 50 K to 250 K for 20 ps and finally from 250 K to 298.15 K for 30 ps. The particle-mesh Ewald (PME) method for treating electrostatic interactions was used, and all simulations were performed under periodic boundary conditions in the $[NpT]$ ensemble at 298.15 K and 1 atm using a 2 fs integration step. The SHAKE algorithm with a tolerance of 10–5 Å was used to fix positions of all hydrogens, a 9.0 Å cutoff was applied to nonbonding interactions and coordinates were stored every picosecond. In total, 51 independent simulations each 50+ ns long were performed. The cumulative production time of room temperature simulations amounts to $\sim$3.2 μs.

Molecular Dynamics Simulations at High Temperature (498 K): MD simulations at high temperature were used to study thermal unfolding of the p18 system, using the well established protocol developed by Daggett and co-workers (47). Four independent MD runs were produced to enhance sampling. The starting structures for the four independent runs were taken from classical MD simulations at room temperature (snapshots taken at 200, 400, 600 and 800 ps). Each system was equilibrated so that the density was equal to 0.823 g/ml (corresponding to the density of water steam at 498 K and 22 atm) (47) then the $[NVT]$ simulations were produced. Each high temperature simulation was run for $\sim$20 ns.

Analyses of MD Simulations: The stability and structural properties of p18 and its fragments were evaluated by calculating root-mean-square-deviations (RMSDs) for backbone atoms from the initial structure, radius of gyration (R_g), secondary structure elements obtained from the DSSP program (48) and native contacts obtained using the *in-house* program RESDIST (Pavel Banáš, unpublished software). The regions of enhanced flexibility of the p18 were assessed from the temperature B-factors calculated from the simulation by the ptraj module of the AMBER package. Essential dynamics analysis of the p18 trajectory was carried out using the GROMACS suite of programs (49). The RMSD vs. R_g plot of the p18 thermally induced unfolding was depicted as a density plot using the *in-house* script. The RESDIST program calculates a contact map (map of distances) among all residues represented by centers of masses of sidechains. Contacts between $i \ldots i + 4$ residues and higher are considered to be native (excluding contacts between $i \ldots i + 1$, $i \ldots i + 2$ and $i \ldots i + 3$ residues) if the distance between two residues is smaller than or equal to 6.0 Å. The program analyzes the number (percentage) of saved native contacts during the MD simulation.

Results

Stability of p18 Fragments

Single α-helices: None of the single α-helices (Figure 2B) taken out of their native context are able to maintain their native α-helical structure during entire 100 ns simulations, indeed unfolding events occur very quickly, within the first several hundred picoseconds (Supplemental Figure S1). In the unfolded state, the α-helices display dynamically equilibrating random coil, turn and 3_10-helix conformations, although some of them show a propensity to transiently re-adopt native α-helical structures, with lifetimes ranging from several hundred picoseconds to several nanoseconds. The longer ($\sim$10 residues) even-numbered α-helices have a higher propensity to refold than the shorter ($\sim$7 residues), odd-numbered α-helices (Supplemental Figure S1; Table I).

Pairs of α-helices: No pairs of α-helices (Figure 2C) are stable in the 50 ns timescale, and all pairs except $\alpha5+\alpha6$ dissociate (Table II) The $\alpha1+\alpha2$, $\alpha3+\alpha4$, and $\alpha9+\alpha10$ pairs dissociate early, at 1.4, 2.7 and 6.3 ns, respectively. The α-helical structure melts immediately after dissociation of the α-helices — clearly indicating that tertiary contacts between them are essential for maintenance of their α-helicity — as illustrated by the results of secondary element analyses shown in Supplemental Figure S2. The $\alpha5+\alpha6$ fragment does not dissociate, but disruption and reformation of the $\alpha5+\alpha6$ antiparallel conformation occur several times during the simulation, which reduces its number of native contacts and α-helical content.

Table I Properties of p18 single α-helices.

Helix	Number of residues	% of hydrophobic residues	t (ns)	Sequence	Hel-last[a] (%)	Helicity[b] (%)
$\alpha1$	7	71.4	100	ELASAAA	7.0	6.5
$\alpha2$	10	40.0	100	LEQLTSLLQN	10.8	17.1
$\alpha3$	8	62.5	100	TALQVMKL	6.2	15.6
$\alpha4$	9	66.7	100	PEIARRLLL	5.8	12.2
$\alpha5$	8	62.5	100	VIHDAARA	7.2	13.3
$\alpha6$	10	50.0	100	LDTLQTLLEF	19.2	26.4
$\alpha7$	7	71.4	100	PLHLAAK	0.1	6.6
$\alpha8$	10	60.0	100	LRVVEFLVKH	7.7	12.2
$\alpha9$	7	71.4	100	ACDLARL	2.2	6.1
$\alpha10$	10	50.0	100	NEVVSLMQAN	10.5	14.1

[a] The mean α-helical content calculated from the last 5 ns of the MD simulation. The α-helical content is normalized to the α-helical content of the initial structure, for which the α-helical content equals 100%.

[b] The α-helical content averaged over the entire simulation.

Helix-turn-helix Motifs: Helix-turn-helix motifs (Figure 2D) are also incapable of maintaining their native structure during the 50-ns MD simulations (Table II). The unfolding of all of the α-helices occurs during the first few nanoseconds, except for $\alpha9$, which melts after $\sim$20 ns. The antiparallel arrangement of α-helices is not preserved in any simulation of these fragments during

the 50 ns. Generally, the orientation of the α-helices shifts from an antiparallel to a perpendicular configuration (Supplemental Figure S3), accompanied by an abrupt fall in the number of native contacts, then the α-helices melt.

Four-helical Bundles: None of the four-helical fragments (Figure 2I) are stable (Table II), and the α-helices change their orientations with respect to each other within a short time, during the 50-ns MD simulations. In the simulation of the $\alpha1+\alpha2+\alpha3+\alpha4$ bundle, the $\alpha2$, $\alpha3$ and $\alpha4$ helices reorient (and melt) at $\sim$500 ps. The $\alpha3+\alpha4+\alpha5+\alpha6$ fragment loses its native-like arrangement at $\sim$1 ns, due to a shift and melting of $\alpha3$. The $\alpha5+\alpha6+\alpha7+\alpha8$ fragment only remains in a native-like conformation for 200 ps, then $\alpha6$ loses its contact with the rest of the fragment and melts simultaneously. Later ($\sim$22.2 ns) the $\alpha5+\alpha6+\alpha7+\alpha8$ bundle dissociates. In the simulation of the $\alpha7+\alpha8+\alpha9+\alpha10$ fragment, the native-like arrangement of α-helices is preserved for $\sim$400 ps, then the C-terminal residues of the $\alpha8$ helix lose their contacts with the $\alpha7$ and $\alpha10$ helices and the $\alpha8$ helix disrupts from the C-terminus. Simultaneously with $\alpha8$ disruption, $\alpha9$ moves slightly away from the $\alpha7$ and $\alpha10$ helices and subsequently melts. The reorientation of α-helices and their dissociation from bundles lead to reductions in the number of native contacts and simultaneous α-helix melting (Supplemental Figure S4).

Table II Properties of all systems simulated at room temperature (298 K).

Fragment	t (ns)	t_{dis}^a (ns)	R_g-MD^b (Å)	RMSD^c (Å)	Native contacts^d (%)	Hel-last (%)	Helicity (%)
$\alpha1+\alpha2$	50	1.4				9.7	16.0
$\alpha3+\alpha4$	50	2.7				33.2	20.4
$\alpha5+\alpha6$	50		6.9 ± 0.3	2.5 ± 0.2	31.0 ± 8.5	20.8	20.9
$\alpha7+\alpha8$	50	31.6				7.3	14.0
$\alpha9+\alpha10$	50	6.3				15.0	5.8
$\alpha1$-t-$\alpha2$	50		7.9 ± 0.2	3.1 ± 0.5	34.9 ± 9.1	34.5	34.2
$\alpha3$-t-$\alpha4$	50		6.8 ± 0.1	3.8 ± 0.1	15.5 ± 4.3	11.8	20.0
$\alpha5$-t-$\alpha6$	50		6.7 ± 0.3	3.6 ± 0.5	20.2 ± 8.1	10.9	27.3
$\alpha7$-t-$\alpha8$	50		7.7 ± 0.5	3.8 ± 0.3	22.3 ± 6.3	3.5	19.9
$\alpha9$-t-$\alpha10$	50		8.3 ± 0.5	5.2 ± 0.6	12.6 ± 5.9	7.6	33.6
$\alpha1+\alpha2+\alpha3+\alpha4$	50		8.7 ± 0.2	7.0 ± 0.3	21.9 ± 4.5	16.0	25.7
$\alpha3+\alpha4+\alpha5+\alpha6$	50		8.7 ± 0.2	6.2 ± 0.2	35.0 ± 5.3	18.8	29.9
$\alpha5+\alpha6+\alpha7+\alpha8$	50	22.2				20.6	32.2
$\alpha7+\alpha8+\alpha9+\alpha10$	50		9.7 ± 0.3	7.2 ± 0.4	9.6 ± 2.6	10.8	13.8
$\alpha1+\alpha2+\alpha3$-t-$\alpha4$	50		9.1 ± 0.3	8.1 ± 0.4	19.0 ± 3.1	34.9	33.5
$\alpha3+\alpha4+\alpha5$-t-$\alpha6$	50		9.1 ± 0.1	7.5 ± 0.3	15.0 ± 4.0	44.2	49.1
$\alpha5+\alpha6+\alpha7$-t-$\alpha8$	50		8.7 ± 0.1	6.6 ± 0.4	23.1 ± 4.3	19.6	38.2
AR I	50		10.1 ± 1.4	6.1 ± 1.8	21.2 ± 9.1	43.5	45.1
AR II	50		8.4 ± 0.3	5.8 ± 0.3	17.8 ± 5.8	67.3	43.4
AR III	50		8.4 ± 0.2	4.2 ± 0.2	19.5 ± 2.9	63.8	54.0
AR IV	50		8.3 ± 0.1	4.9 ± 0.2	24.3 ± 3.0	35.1	39.5
AR V	50		8.1 ± 0.3	3.2 ± 0.5	15.3 ± 7.6	23.5	40.4
$\alpha5$-t-$\alpha6$-loop-$\alpha7$	50		9.0 ± 0.1	2.9 ± 0.2	51.4 ± 5.5	75.8	82.0
$\alpha7$-t-$\alpha8$-loop-$\alpha9$	50		11.9 ± 0.5	10.2 ± 0.6	10.7 ± 3.1	28.8	34.6
AR III-loop-$\alpha7$	50		9.5 ± 0.2	4.1 ± 0.2	26.5 ± 3.7	43.8	53.8
AR IV-loop-$\alpha9$	50		9.1 ± 0.1	5.0 ± 0.3	31.6 ± 3.4	32.0	54.0

Table II (Continued)

Fragment	t (ns)	t_{dis}^a (ns)	R_g-MD[b] (Å)	RMSD[c] (Å)	Native contacts[d] (%)	Hel-last (%)	Helicity (%)
mfu-7-8-9	50		8.7 ± 0.1	3.3 ± 0.1	40.3 ± 4.3	82.8	81.5
α1-t-α2+α3-t-α4	70		8.7 ± 0.1	2.5 ± 0.2	56.4 ± 4.6	66.9	69.7
α3-t-α4+α5-t-α6	70		9.3 ± 0.1	4.8 ± 0.2	27.2 ± 3.1	23.1	44.6
α5-t-α6+α7-t-α8	70		9.8 ± 0.3	4.6 ± 0.8	23.8 ± 4.9	37.6	61.8
α7-t-α8+α9-t-α10	70		8.4 ± 0.1	1.5 ± 0.4	72.8 ± 6.0	86.2	87.5
α1-t-α2-loop-α3-t-α4	70		9.4 ± 0.1	1.2 ± 0.2	71.8 ± 4.8	84.7	80.4
α3-t-α4-loop-α5-t-α6	70		9.7 ± 0.1	4.4 ± 0.2	20.8 ± 3.0	59.1	62.6
α5-t-α6-loop-α7-t-α8	70		10.0 ± 0.1	5.1 ± 0.1	28.2 ± 3.4	70.2	69.2
α7-t-α8-loop-α9-t-α10	70		9.2 ± 0.1	4.2 ± 0.1	55.4 ± 3.6	65.1	78.4
Cluster of four helix-loop-helix motifs	50		16.3 ± 0.2	6.3 ± 0.4	36.8 ± 2.4	68.9	73.9
Cluster of ten helices	50		23.2 ± 4.3	16.3 ± 4.3	13.0 ± 1.6	34.7	32.6
Cluster of ten helices connected via loops	50		16.0 ± 0.1	3.9 ± 0.2	55.5 ± 2.0	62.9	70.8
Cluster of five helix-turn-helix motifs	50		15.7 ± 0.1	2.3 ± 0.1	62.6 ± 2.7	71.1	75.2
The entire p18	50		16.4 ± 0.1 (16.0)[g]	2.1 ± 0.2	72.8 ± 2.2	96.8	96.7

[a] t_{dis} is the time at which the dissociation of α-helices occurs.

[b] R_g-MD is the mean radius of gyration ± S.D. (standard deviation) of backbone atoms calculated from the last 5 ns of each simulation.

[c] RMSD is the root-mean-square-deviation of backbone atoms compared with the initial structure calculated as the mean ± S.D. from the last 5 ns of each MD simulation.

[d] The mean number of native contacts ± S.D. calculated from the last 5 ns of each MD simulation.

[g] R_g calculated for the X-ray structure.

Pairs of Helices Capped by C-terminus Helix-turn-helix Motifs: No pairs of α-helices capped by their neighboring C-terminal helix-turn-helix motif (Figure 2J) are stable during the 50 ns simulation. The short α3 unfolds quickly (after several picoseconds) in the α1+α2+α3-t-α4 and α3+α4+α5-t-α6 fragments, while the α5 and α7 helices in the α3+α4+α5-t-α6 and α7+α8+α9-t-α10 fragments, respectively, unfold later, at ∼11.5 ns and ∼14 ns, respectively. The other α-helices melt in the nanoseconds-timescale. The α3+α4+α5-t-α6 and α1+α2+α3-t-α4 fragments exhibit the highest and lowest α-helicity (49.1 and 33.5%, respectively; Table II). Generally, α-helices connected by a turn refold more often than α-helices that are not connected with their neighboring α-helices via a tight turn (data not shown). In all of these fragments, the α-helices readily adopt a non-native arrangement. In the α1+α2+α3-t-α4 simulation, the native-like fold is only maintained until ∼700 ps. Then α2 separates from the rest of the fragment, while α1 approaches α4. In the simulation of the α3+α4+α5-t-α6 fragment at ∼1.5 ns the α5-t-α6 segment moves away from the α3+α4 pair and slightly tilts (Supplemental Figure S5). The α6 helix of the α5+α6+α7-t-α8 fragment separates at ∼1 ns and both of its termini concomitantly unwind. The α8 helix moves away from the rest of the α7+α8+α9-t-α10 fragment within the

first several hundred picoseconds, accompanied by the disruption of both of its termini. In these simulations the shifts and reorientation of α-helices lead to a loss of stabilizing tertiary contacts, which is associated with their disruption and melting.

Ankyrin Repeats: Superimposition of all simulated ARs (Figure 2E) shows that the structures of all p18 AR motifs are, as expected, almost identical (Figure 3A and Supplemental Figure S6A). In addition, superimposition of the ARs' average structures during the last 5 ns of the simulations (Figure 3B and Supplemental Figure S6B), together with low numbers of native contacts (Table II), show that all of the simulated ARs lose their native structure within the 50 ns timescale. Further, secondary element analysis (Supplemental Figure S7) shows that all α-helices (except for α5 located in AR III) unfold during the 50 ns simulations. The unfolding events occur within timescales ranging from several hundred picoseconds (α3, α6, α9) to several nanoseconds (α1, α2, α4, α7, α8, α10), but numerous refolding events occur during the AR simulations (Supplemental Figure S7).
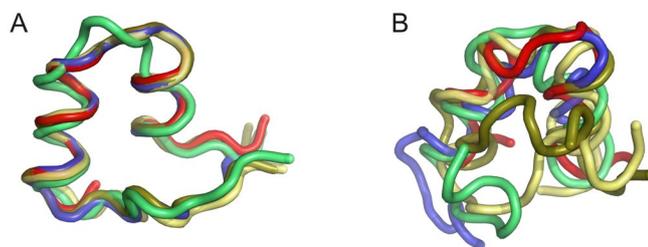
Figure 3: (A) Superimposition of all p18 ARs over C_α atoms, showing that the X-ray structures of individual repeats are highly similar. ARs I-V are colored blue, green, brown, yellow and red, respectively. (B) Superimposition of average structures of ARs calculated from the last 5 ns showing melting of the repeats (same colors as in (A)).

ARs generally have higher α-helicity than the helix-turn-helix motifs, but the percentage of native contacts does not significantly differ between them. All of the α-helices of the helix-turn-helix fragments, apart from α9, unfold in 1 ns timescales, while some α-helices of ARs (α1, α4 and α8) melt in 10+ ns timescales. Clearly, the α-helices unfold more slowly in ARs than in helix-turn-helix fragments, showing that the presence of the N-terminal β-strand, loop and C-terminal β-strand retards their unfolding.

Three-helical Fragments: The α5-t-α6-loop-α7 fragment (Figure 2F) is stable in the 50 ns timescale. This fragment has a high α-helical content (82.0%) and relatively high percentage of native contacts (51.4%). The RMSD curve reaches a plateau at $\sim$7.5 ns, after which it fluctuates below 3.0 Å (Supplemental Figure S8A). The motif is stable, probably due to a shift of α6 towards α7, which occurs at $\sim$0.5 ns. After the shift, the hydrophobic core becomes well packed and protected from the solvent (Supplemental Figure S8B). This structural arrangement remains stable until the end of the simulation. In contrast, the α7-t-α8-loop-α9 fragment has a substantial degree of instability during the 50 ns simulation, as indicated by the low α-helical content and percentage of native contacts (34.6% and 10.7%, respectively) (28).

Judging from its low α-helicity (53.8%) and percentage of native contacts (26.5%), the AR III-loop-α7 fragment (Figure 2G) is not stable in the 50 ns timescale. The α6 helix readily unfolds at $\sim$1 ns, while α5 and α7 unfold later, at $\sim$23.5 ns and $\sim$9.5 ns, respectively. The α5 and α7 helices have significantly higher tendencies to refold than α6 (Supplemental Figure S9). Notably, except for the time between 23.5 ns and 25 ns, the α-helical structure of α5 is retained well during the entire simulation run (Supplemental Figure S9). As shown in our previous study (28), the AR IV-loop-α9 fragment is also unstable in the 50 ns timescale, at the end of which its α-helicity and number of native contacts are 54.0% and 31.6%, respectively.

Finally, the mfu-7-8-9 fragment is stable during the 50 ns simulation. This fragment is topologically identical to the minimal folding module of ARPs proposed by Ferreiro and co-workers (33). It comprises the consensus 1 AR IV linked with the loop to α9 (the first helix of the following C-terminal AR V; Figure 2H). The α-helicity of the fragment is high (81.5%), but its percentage of native contacts is perhaps surprisingly low (40.3%; Table II). However, most (57.1%) native contacts among the α-helices are preserved, and the low overall value (40.3%) can be attributed to the rearrangement/reconformation of the loops during the simulation. The RMSD is convergent, associated and sufficiently low (3.3 Å) to indicate stability. The radius of gyration decreases (at $\sim$3 ns) by almost 1 Å, due to a shift of α8 towards α9 and distortion of the N-terminal turn (Supplemental Figure S10). None of the α-helices unfolds during the simulation. Four interactions stabilize the overall fold of the mfu-7-8-9 fragment (see Supplemental Figure S11).

Pairs of Helix-turn-helix Motifs: Our recently published study, mentioned above (28), included a 50 ns simulation of the α7-t-α8+α9-t-α10 fragment, which proved to have considerable stability during the entire simulation, with high α-helicity (89.1%) and percentage of native contacts (75.3%). In this study, the simulation of this fragment was extended by 20 ns (to 70 ns) and three other pairs of helix-turn-helix motifs (Figure 2K) were also simulated at the 70 ns timescale.

The α1-t-α2+α3-t-α4 and α7-t-α8+α9-t-α10 fragments are stable, while the α3-t-α4+α5-t-α6 and α5-t-α6+α7-t-α8 fragments are unstable in the 70 ns timescale. The α7-t-α8+α9-t-α10 is the most stable of these motifs, as evidenced by its high α-helicity (87.5%) and number of native contacts (72.8%). We have not observed any helix melting or significant helical structure disruption of this fragment, and the α-helical bundle remains in its native-like conformation during the entire simulation (Figure 4D). The α1-t-α2+α3-t-α4 system is also quite stable, with α-helicity and native contact percentages of 69.7 and 56.4%, respectively (Table II), but less stable than the α7-t-α8+α9-t-α10 fragment, probably because of the presence of a short α3, which quickly melts and prevents the tighter packing of hydrophobic residues (Figure 4A). The α2 helix of the α1-t-α2+α3-t-α4 fragment unfolds at $\sim$24 ns and refolds at $\sim$28 ns (and the refolded helix is maintained until the end of the simulation, data not shown), while α4 has flexible C- and N-termini and only the middle turn remains stable during the entire simulation. Notably, at $\sim$40 ns α4 slightly separates from the bundle and the R_g and RMSD values subsequently increase while the α-helicity and native contacts decrease (data not shown).

The fragments consisting of internal helix-turn-helix moieties (α3-t-α4+α5-t-α6 and α5-t-α6+α7-t-α8) do not maintain their native-like structures in the 70 ns timescale. In the simulation of
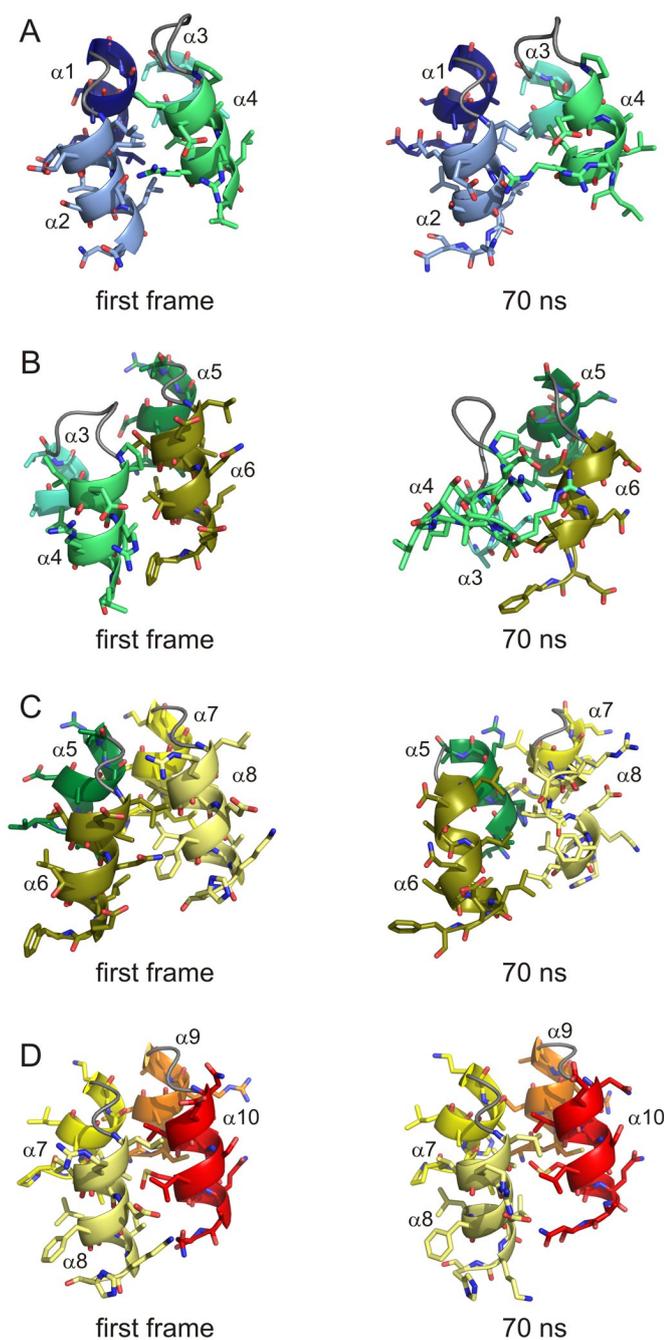
Figure 4: Initial and final structures from the MD simulations of pairs of helix-turn-helix motifs. (A) the $\alpha1$-t-$\alpha2$+$\alpha3$-t-$\alpha4$ fragment ($\alpha1$ in dark blue, turns in gray, $\alpha2$ in light blue, $\alpha3$ in cyan and $\alpha4$ in light green). (B) the $\alpha3$-t-$\alpha4$+$\alpha5$-t-$\alpha6$ fragment ($\alpha3$ in cyan, turns in gray, $\alpha4$ in light green, $\alpha5$ in dark green, and $\alpha6$ in dark olive). (C) the $\alpha5$-t-$\alpha6$+$\alpha7$-t-$\alpha8$ fragment ($\alpha5$ in dark green, turns in gray, $\alpha6$ in dark olive, $\alpha7$ in yellow, and $\alpha8$ in pale yellow). (D) the $\alpha7$-t-$\alpha8$+$\alpha9$-t-$\alpha10$ fragment ($\alpha7$ in yellow, turns in gray, $\alpha8$ in pale yellow, $\alpha9$ in orange, and $\alpha10$ in red).

$\alpha3$-t-$\alpha4$+$\alpha5$-t-$\alpha6$, $\alpha5$ separates from the remaining bundle quickly, after several hundred picoseconds. Following this separation only a few interactions mediate contacts between the helix-turn-helix

moieties (Figure 4B). The α6 of the α5-t-α6+α7-t-α8 fragment unfolds at ∼8 ns and becomes highly flexible, moving away from the rest of the fragment and back again several times. In contrast to the high flexibility of α6, the remaining α-helices (α5, α7, and α8) remain in a close-to-native assembly until ∼40 ns. After that, the fragment slowly adopts a structure in which only a small portion of each helix-turn-helix moiety mediates contacts between them. The final structure (Figure 4C) has residual α-helical content of 37.6% and low percentage of native contacts (23.8%).

Helix-turn-helix-loop-helix-turn-helix Motifs: These fragments coincide with the minimal folding module inferred from experiments by Zhang and Peng (27) (Figure 2L). The α7-t-α8-loop-α9-t-α10 fragment showed high stability throughout an entire 50 ns simulation in our recent study (with 83.7% α-helicity and 53.5% percentage of native contacts) (28). Here, we present simulations of three other helix-turn-helix-loop-helix-turn-helix fragments of the p18 protein in the 70 ns timescale and report a prolonged simulation of the α7-t-α8-loop-α9-t-α10 from 50 ns to 70 ns.

Generally, pairs composed of N- or C-terminal repeats stacked with an internal repeat (α1-t-α2-loop-α3-t-α4 and α7-t-α8-loop-α9-t-α10) have slightly higher stability than pairs consisting of two internal repeats (α3-t-α4-loop-α5-t-α6 and α5-t-α6-loop-α7-t-α8). The α1-t-α2-loop-α3-t-α4 fragment has considerable stability during the 70 ns simulation (with high α-helicity and percentage of native contacts, 80.4 and 71.8%, respectively; Table II) and maintains its native-like fold well (Supplemental Figure S12A). In addition, the mean RMSD of its backbone atoms over the last 5 ns is also lower (1.2 Å) than that of any other simulated helix-turn-helix-loop-helix-turn-helix fragments. The α7-t-α8-loop-α9-t-α10 fragment is also stable (Supplemental Figure S12D), having high α-helical content (78.4%) and percentage of preserved native contacts (55.4%). Its RMSD is higher than that of the α1-t-α2-loop-α3-t-α4 fragment (4.2 Å), due to a rearrangement of the loop, which is not maintained in a native-like conformation (i.e., perpendicular to the α-helix bundle). As a result of the loop rearrangement, the R_g slightly decreases during the simulation (data not shown). The interaction of loop residues (Asn129, Val130, Gly131 and His132) with the C-terminal moiety of α10 most likely initiates the disruption of the α-helix, which concludes with its melting at ∼48 ns.

The systems consisting of pairs of internal repeats (α3-t-α4-loop-α5-t-α6 and α5-t-α6-loop-α7-t-α8) are significantly less stable (Supplemental Figure S12B and S12C), as shown by the low percentages of native contacts at the end of the simulations of α3-t-α4-loop-α5-t-α6 and α5-t-α6-loop-α7-t-α8, (20.8% and 28.2%, respectively), although their α-helicity is somewhat higher (62.6% and 69.2%, respectively) (Table II). The flexible α3 of the α3-t-α4-loop-α5-t-α6 fragment separates from the rest of the fragment at ∼22 ns and native contacts abruptly decrease by ∼20% (data not shown), the hydrophobic core of the fragment then becomes poorly packed (Supplemental Figure S12B), and the loop connecting α4 with α5 rearranges at ∼36.5 ns.

The RMSD steadily increases while the percentage of native contacts steadily decreases during the entire simulation of the α5-t-α6-loop-α7-t-α8 fragment (Supplemental Figure S13A), due to slow shifts in the positions of the helix-turn-helix moieties relative to each other, resulting in a structure in which only α5 and α8 mediate contacts between individual repeats (Supplemental

Figure S13B).

The Cluster of Four helix-loop-helix Motifs: The native-like fold of the four helix-loop-helix fragment (Figure 2O) is maintained until ∼10 ns, then α9 flips and loop IV moves away (Supplemental Figure S14A). At ∼11 ns, the first two loops also move away from each other, and after ∼500 ps (at ∼11.5 ns) on the opposite side of the system α9 swings away from α7 (Supplemental Figure S14B). Then, α2 flips and simultaneously its N-terminus unwinds at ∼14 ns, while loops II and III adopt a splayed arrangement at ∼15.5 ns. At ∼33 ns, the α2-loop-α3 moiety starts to move slowly away from the rest of the system (Supplemental Figure S14C). Notably, the bundle of α-helices α4, α5, α6, α7, and α8 maintains a native-like conformation during the entire MD simulation.

Except for α4, α7 and α9, all other α-helices are stable at the 50 ns timescale (data not shown), resulting in high (73.9%) α-helicity (Table II). The α9 helix melts within the first few nanoseconds, while α4 and α7 unfold significantly later, at ∼34 ns and ∼43 ns, respectively. A non-native α-helix (residues 63–66) forms at ∼42 ns and remains stable until the end of the simulation. As a result of the large structural rearrangement of the α2-loop-α3 and α8-loop-α9 fragments, the system has a low percentage of native contacts (36.8%) and high RMSD of backbone atoms (6.3 Å) at the end of the simulation.

The Cluster of Ten Helices: The ten-helical cluster (Figure 2M and Supplemental Figure S15A) was simulated to check whether the α-helices are able to maintain their native-like arrangement without the presence of loops and short turns in the 50 ns timescale. The cluster is not stable during the simulation and the α-helices only remain in their native-like assembly until ∼1.5 ns. Then, the bundle of α-helices α1-α4 slightly rotates and moves away from the remaining α-helices. After that, at ∼2.5 ns, α5 and α7 tilt toward the C-terminus and contacts among α-helices in the α5-α10 bundle become less tight. After several hundred picoseconds (at ∼3 ns), the α7-α10 bundle of α-helices collapses, while the α1-α6 bundle remains in a close-to-native assembly until ∼8.5 ns. Then, the remaining α-helices (α1-α6) collapse and a globule composed of all α-helices slowly forms, and its formation is complete at ∼29 ns (Supplemental Figure S15B). Finally, at ∼42 ns, α10 separates from the globule and consequently the RMSD and R_g of the entire system significantly increase (Table II; Supplemental Figure S16A). Otherwise the globule remains packed until the end of the simulation. Notably, in the compact α-helical globule both the number of native contacts (13.0%) and the α-helix content (32.6%; Table II) are very low. Moreover, the α-helical content and number of native contacts decrease simultaneously during the simulation (Supplemental Figure S16B). All α-helices unfold during the simulation, but some of them refold back to their native α-helical structure and retain it for a fairly long simulation time. The internal α-helices (α4, α5, α6) are more stable than the other helices (data not shown).

Cluster of Ten Helices Connected Via Loops: This system (Figure 2P and Supplemental Figure S17A), which has two additional α-helices (α1 and α10) compared to the cluster of four helix-loop-helix motifs (Figure 2O), is reasonably stable during the 50 ns MD simulation (Supplemental Figure S17). The mean RMSD is quite low (3.9 Å) and exhibits the typically small fluctuations

of stable molecules (Table II). More than 50% of native contacts are preserved at the end of the simulation (55.5%) and the α-helicity equals 70.8%, showing that the structure of α-helices is well maintained during the simulation, although in general the N-terminal α-helices are less stable than those of the C-terminal (Supplemental Figure S17B). The $\alpha 1$ and $\alpha 2$ helices melt at ~ 1 and ~ 3 ns, respectively. The $\alpha 3$ helix unfolds immediately at the beginning of the simulation and its residues then adopt dynamically equilibrating 3_{10}-helix, turn and α-helix conformations. At ~ 40 ns, $\alpha 5$ unfolds and its α-helical structure does not reform during the following simulation time. The other α-helices are stable during the simulation (Supplemental Figure S18).

The Cluster of Five Helix-turn-helix Motifs: Unlike the cluster of ten α-helices, the system with five stacked helix-turn-helix motifs (Figure 2N and Supplemental Figure S19A) remains stable during the entire 50 ns simulation time, consequently the RMSD between its initial and final structures is low (2.3 Å) while its final α-helicity and percentage of native contacts are high (75.2 and 62.6%, respectively; Table II). More specifically, the $\alpha 1$, $\alpha 4$, $\alpha 5$, $\alpha 6$, $\alpha 7$ and $\alpha 8$ helices are highly stable, but the $\alpha 2$ N-terminus is molten between ~ 26 ns and ~ 43 ns and $\alpha 3$ readily melts and does not refold. The $\alpha 9$ and $\alpha 10$ helices both melt at ~ 11 ns and their α-helical structures repeatedly unfold and refold in the following simulation time (data not shown). The $\alpha 9$-t-$\alpha 10$ pair is surprisingly very flexible. Within ~ 27.5 ns $\alpha 9$ moves away from the rest of the cluster, and its position is replaced by $\alpha 10$ (Supplemental Figure S19B). This arrangement of the $\alpha 9$-t-$\alpha 10$ moiety is maintained until ~ 28.5 ns, when the interface between $\alpha 7$-t-$\alpha 8$ and $\alpha 9$-t-$\alpha 10$ regains a native-like arrangement that is maintained until the end of the simulation (Supplemental Figure S19C).

Stability and Flexibility of the Entire p18 Domain

The native structure of p18 (Figure 1 and 2A) is well maintained during the entire 50 ns simulation. Early in the simulation the short $\alpha 3$ unfolds, but then repeatedly refolds (for at most hundreds of picoseconds) and unfolds during the rest of the simulation. The C-terminus of $\alpha 10$ melts at ~ 25 ns, while the other α-helices of p18 are stable throughout the simulation, as demonstrated by the very high average α-helicity, 96.7% (Table II), low RMSD of backbone atoms (2.1 Å) and high number of native contacts (72.8%). Loops and the C-terminal turn of $\alpha 10$ are the most flexible parts of p18, as indicated by their high B-factors (Figure 5A and Supplemental Figure S20A). In all loops, the most flexible residues are located at their tips, and the most dynamic loop is loop II.

An essential dynamics analysis to extract the correlated atomic motions, which may be biologically relevant (50, 51), was also performed (Figure 5B and Supplemental Figure S20B). The first essential motion is localized to loop I and spreads over $\alpha 2$ and the C-terminal turn of $\alpha 10$. The motions corresponding to the first four eigenvectors are similar, but cover different parts of the p18 protein, causing it to bend towards the C-shape that is typical for long ARPs (15, 52).

Baumgartner et al., (53) suggested that loops of the INK4 proteins are stabilized by H-bonds between sidechains of conserved Asn/Asp residues and backbone amide nitrogens of neighboring residues. In the p18 protein there is a strictly conserved residue in each loop that is important for the loop's stability: Asn35, Asp67, Asp100 and Asn134 in loops I-IV, respectively (Supplemental Figure S21). The native H-bond of the Asn35 sidechain oxygen with the backbone nitrogen of Arg39
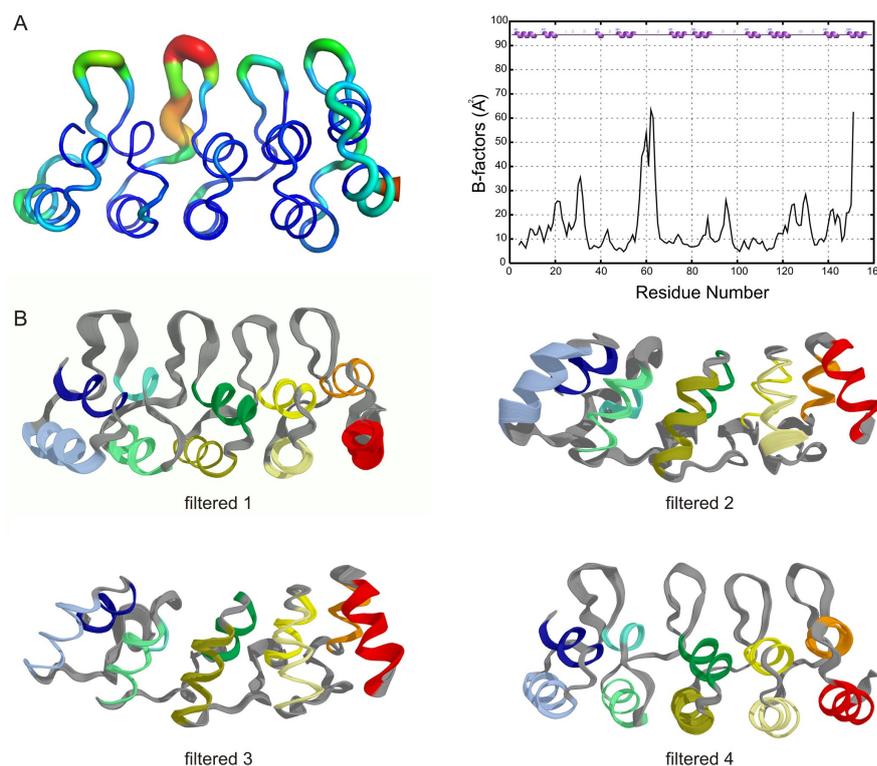
Figure 5: (A) On the left, there is a plot of the atomic temperature B-factors. The highest values (in Å^2) are colored red (the thick regions) while the lowest are blue (the thin regions). On the right, there is a plotting of residue temperature B-factors calculated from the last 5 ns of the simulation. (B) Depiction of the first four essential motions extracted from the complete p18 simulation by the essential dynamics analysis.

(loop I) remains stable during the entire simulation. The H-bond between the Asn134 sidechain oxygen (loop IV) and the backbone nitrogen of Asp138 is also stable in the 50 ns timescale. On the other hand, the H-bond between the Asp67 side-chain and the Phe71 backbone nitrogen (located in loop II) melts within $\sim$250 ps, but repeatedly reestablishes during the following simulation time. Similar patterns are observed in the H-bonds stabilizing the loop III, in that the first sidechain oxygen of Asp100 H-bonds to the Asn104 backbone nitrogen, while the second oxygen bounds to the sidechain nitrogen of Asn104. These H-bonds disrupt at $\sim$7 ns due to the Asp100 sidechain flipping, but the native-like H-bonds of Asp100 to Asn104 are frequently reestablished during the subsequent course of simulation.

Thermally Induced Denaturation of the p18

In addition to the room-temperature simulations, four independent 20 ns simulations of the entire p18 protein at 498 K were run, using a well-established protocol for temperature-induced protein unfolding (47), to obtain complementary information on the mechanism of p18 unfolding at elevated temperature and to pinpoint regions of enhanced stability. The plot of RMSD vs. R_g (Figure 6 and Supplemental Figure S22) shows two densely populated regions corresponding to the native

state (Figure 6, green box) and intermediate structures (Figure 6, red box). The intermediate structures have two first ARs melted and ARs III-V folded. The plot also indicates that structures with RMSD above ~12 Åhave very variable R_g which is an indicator of a completely melted p18 protein (cf. Figure 6 and Supplemental Figure S23).
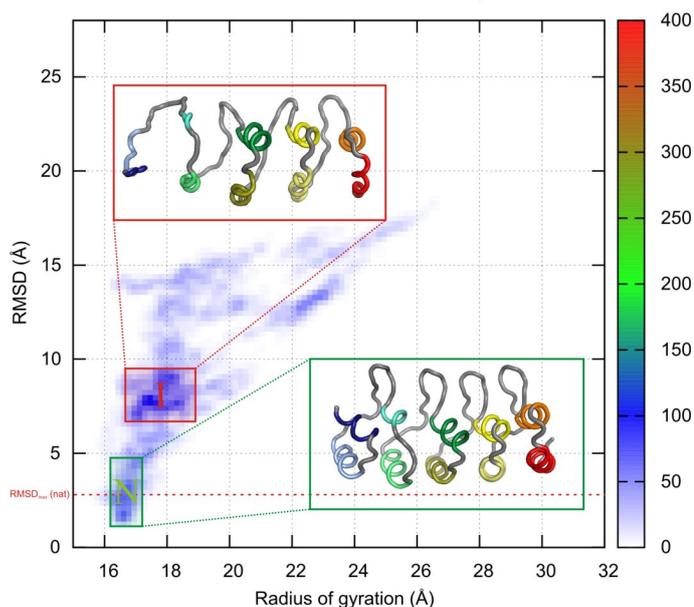
Figure 6: The plot of RMSD vs. radius of gyration calculated from the first 5 ns of four p18 unfolding simulations. The plot is visualized as a density plot giving information about the relative populations of structures in respective 2D bins (number of bins is 100×100). In the lower part of the plot is shown the average structure calculated from structures inside the green region (N stands for native) while in the upper part is depicted the average structure of intermediates (I stands for intermediate) lying in the red-bordered region. The red dashed horizontal bar indicates the maximal RMSD reached in the p18 room-temperature (298 K) simulation. The color coding of helical main-chain atoms (cartoon representation) is consistent with Figure 1.

The p18 has a very fast unfolding kinetics at 498 K. The entire system completely unfolds within ~4 ns (Supplemental Figure S23). The α-helices of ARs I and II remain structured for only several hundred picoseconds while the ARs III-V keep, generally, their α-helical structure for several nanoseconds (Supplemental Figure S24). The fact that ARs I and II melt prior to the ARs III-V is also well documented by number of native contacts preserved between neighboring helix-turn-helix motifs. The percentage of native contacts between α1-t-α2 and α3-t-α4 and between the α3-t-α4 and α5-t-α6 moieties falls below 25% at 237 ps and 137 ps, respectively (Table SII). However, the numbers of native contacts between the α5-t-α6 and α7-t-α8 and between the α7-t-α8 and α9-t-α10 moieties fall below the 25% threshold at 1093 and 999 ps, respectively (Table SII).

Discussion

Native tertiary contacts stabilize α-helices of p18

The MD results strongly indicate that none of the single α-helices of p18 have sufficient stability or propensity to retain their α-helical structure to maintain it when taken out of their native structural context. The α-helix pairs and four-helical bundles also seem to lack the capacity to retain their structures, since their secondary structures melt immediately after dissociation of the helices. Processes such as rearrangements of α-helical positions, leading to non-native arrangements of α-helices in fragments, are also frequently followed by α-helix unfolding. Furthermore, the simulation of the ten-helical bundle (p18 without turns and loops) shows that non-native tertiary contacts, which occur in the collapsed state, do not prevent melting of the α-helices. Generally, as soon as the number of native tertiary interactions between p18 α-helices decreases (due to dissociation/rearrangement), the α-helical structure melts (i.e., the α-helical content decreases). Taken together, our results strongly indicate that the intrinsic α-helical propensity of p18 helices is low, and highlight the importance of native tertiary contacts in the stabilization of the protein's α-helices. The importance of structural context for the stability of secondary structure elements has been previously demonstrated for several globular protein domains (54-59), and the data presented here show that it is also important for ankyrin repeat proteins, i.e., non-globular proteins consisting of repeating motifs.

Pairs of Ankyrin Repeats are Stable, but not Single Motifs

It is generally believed that the single AR motif is partly stabilized by intra-repeat H-bonds between β-strands (10, 53) and hydrophobic contacts between antiparallel α-helices that are not tightly packed (60), but more important contributors to AR stability are the inter-repeat hydrophobic contacts between adjacent AR motifs, which are stronger than the intra-repeat contacts (60, 61). These contributions are difficult to measure, but the theoretical study by Otyepka et al., (32) provides rough estimates (which can be validly used for comparisons, although they are somewhat stronger than true values because they do not account for zero-point vibrations and entropy factors). The values obtained in the cited study indicate that the mean intra-repeat interaction in water is equal to –17 kcal/mol, while the mean inter-repeat interaction is at least –43 kcal/mol; the calculated average ignoring diagonal interactions (Table III). The interactions between the long even-numbered helices are also clearly the major contributors to the inter-repeat interaction. Altogether, the inter-repeat interaction is more than $\sim$2.5 times stronger than the intra-repeat interaction.

Table III Intra-repeat and inter-repeat (between odd- and even-numbered α-helices) interaction energies between pairs of α-helices of p18, from study (32), clustered into three groups.

Interaction type	Pair of α-helices	ΔE_g^a (kcal/mol)	ΔE_{sol}^b (kcal/mol)
	$\alpha1, \alpha2$	−19.9	−16.6
	$\alpha3, \alpha4$	−11.3	−8.4
	$\alpha5, \alpha6$	−32.0	−30.3
Intra-repeat (ARs I-V)	$\alpha7, \alpha8$	−7.2	−15.4
	$\alpha9, \alpha10$	−20.3	−16.3
	average	−18.1	−17.4
	$\alpha1, \alpha3$	−6.4	−9.5
	$\alpha3, \alpha5$	−0.5	−4.1
Inter-repeat (odd numbered α-helices)	$\alpha5, \alpha7$	−5.0	−21.2
	$\alpha7, \alpha9$	−20.4	−12.1
	average	−8.1	−11.7
	$\alpha2, \alpha4$	−64.8	−30.5
	$\alpha4, \alpha6$	−95.6	−29.4
Inter-repeat (even numbered α-helices)	$\alpha6, \alpha8$	−87.2	−40.8
	$\alpha8, \alpha10$	−64.9	−23.5
	average	−78.1	−31.1

[a] ΔE_g is the gas interaction energy between α-helices calculated by DFT-D method (72).

[b] ΔE_{sol} is the interaction energy of α-helices in solvent determined as a sum of gas interaction energy (ΔE_g) and the empirical solvation free energy ($\Delta\Delta G_{sol}$).

The MD simulations presented here reflect the low intra-repeat and higher inter-repeat interaction energies, since the single ARs are not stable, while the fragments formed by at least two ARs are stable. The observation that the single ARs are not structured in water is consistent with the results of previous experimental studies (27, 30), and published MD simulations (19, 28). The finding that two ARs form a stable structural motif is also consistent with the experimental work by Zhang and Peng (27), in which they defined the minimal stable unit of ARPs as a moiety with at least two ARs. This excellent agreement with all of the cited experimental works also illustrates the predictive power of the all-atomic MD simulations at 50 ns timescale applied here, which can be routinely used to check the stability of various ARPs, including designed ARPs (DARPins).

The Minimal Stable unit of the p18 Protein

As previously mentioned, a single AR does not fold independently and does not retain its typical structure. Further, all known natural ARPs contain at least four repeats (62-64), and the hydrophobic interactions among neighboring α-helices (53, 65) and H-bonds between loops (10, 53, 60) are essential for their stability. Both of these kinds of interactions are short-range contacts, typical for "one-dimensional" structures, enabling stable, truncated ARP variants to be generated (22, 27). The reason why smaller ARPs do not occur naturally is difficult to determine, because we do not know what evolutionary pressures have acted on ARPs, but they are probably related to the proteins' function rather than their stability *per se*. However, regardless of the reason,

determination of the minimal autonomous ARP folding unit could greatly assist attempts to design ARPs for specific applications (19). Such structural units can be also considered nucleation centers for ARP folding. Zhang and Peng identified two C-terminal ARs of the p16 ARP as the smallest AR module that was able to fold independently and represented the minimal structural unit of ARPs (27). Later, Ferrerio et al., proposed a three-helical fragment to be the minimal folding unit of ARPs (33). However, the proposed three-helical model of the minimal folding unit has been recently questioned by Sklenovský et al., (28). In the present study, we have significantly extended the set of three- and four-helical fragments and can conclude that pairs of ARs retain their native structure well and can undoubtedly be considered to be at least very close to minimal stable folding and structural units. On the other hand, some fragments containing three α-helices (one complete AR with the consecutive C-terminal α-helix) are unstable while others are stable, showing that three-helical motifs lie at the "stability edge", i.e., three-helical motifs with sufficient stability for various purposes could be designed or found, but they would be rare.

The Functional Importance of p18 Turns and Loops

The reported MD simulations also provide information about the roles of turns and loops. The tight turns connecting antiparallel α-helices significantly reduce the conformational space of the motif, thereby enhancing stability and likely also accelerating the folding of the AR. The difference in behavior between the ten- helical cluster, which collapses, and the complex of five helix-turn-helix motifs, which is stable, shows that the inter-helix interaction itself does not provide sufficient stabilization for the p18 system, and the conformational restrictions imposed by the turns are required to stabilize the structure. On the other hand, the turns must be sufficiently malleable to accommodate the sharp bend in the helix-turn-helix motif of the AR (cf. Figure 3A). This structural malleability is encoded in the primary sequence of the turn, which bears a highly conserved Gly residue at its tip (3, 5, 39). In summary, the tight turn plays a significant structural role in the stability and conformation of the AR.

The loops connecting ARs also restrict the conformational space of the system, as illustrated by the simulation of the cluster of ten helices connected via loops. The complex is stable and its native-like assembly is well maintained during the simulation. Comparison of the simulations of the five helix-turn-helix complex and the cluster of ten helices connected via loops provides indications of the relative importance of turns and loops for the stabilization of the ten-helical bundle. The mean RMSD of backbone atoms calculated from the initial structure over the last 5 ns amounts to 2.3 Åfor the five helix-turn-helix complex and 3.9 Åfor the cluster of ten helices connected via loops. The latter system also has slightly lower α-helicity and percentage of native contacts at the end of the MD simulation in comparison with the five helix-turn-helix complex (Table II; cf. also Supplemental Figure S17 and S19). Taken together, differences in structural parameters of both systems are small, but significant, and turns appear to contribute more to the stability of p18 than loops.

The analysis of flexibility of the complete p18 protein shows that the helix-turn-helix motifs are rigid, while the loops are flexible (Figure 5 and Supplemental Figure S20). Loop II, which contacts the β-sheet of the CDK6 N-terminal lobe (66) and has been found to be important for binding

to CDK4 (67), shows the highest flexibility. Consequently, the loops are not only important for p18 stability, but also play roles in CDK4/6 recognition, enabling through their intrinsic flexibility a better fit of the p18 to the contacted protein. In contrast, the rigidity of the helix-turn-helix motifs suggests that they supply a rigid scaffold. Collectively the available information suggests that the helix-turn-helix motifs form the structural scaffold of the p18 protein while the loops are responsible for the recognition processes (functionality in the context of the INK4 family) as well as helping to stabilize the structure.

p18 Terminal Repeats Stabilize the Internal Repeats

The terminal repeats of ARP proteins are often called capping repeats, because one side is in contact with the adjacent internal repeat while the other side interacts with the solvent. Hence, in capping repeats there are hydrophilic residues at positions occupied by hydrophobic residues in internal repeats, making interactions of the capping repeats with the solvent more favorable and preventing aggregation of ARPs (25, 26, 68, 69). Provision of suitable capping repeats is also essential for constructing foldable, stable DARPins (25, 26).

The MD simulations show that the addition of $\alpha 1$ and $\alpha 10$ to the complex of four helix-loop-helix motifs (systems depicted in Figure 2O and 2P) improves the stability of all of the helix-loop-helix moieties and the maintenance of their native-like structures (cf. Supplemental Figure S14 and S17). MD simulations also show that pairs of internal and terminal ARs (internal repeats capped by terminal repeats) are more stable in the 70 ns timescale than pairs of internal repeats (cf. Figure 4 and Supplemental Figure S12). These findings show that capping repeats have a strong stabilizing effect in the naturally occurring protein p18, in accordance with results obtained by Interlandi et al., from simulations and experiments carried out for DARPins. The cited authors suggested that the C-terminal capping repeat of the protein E3_5 and all NIxC systems (where x denotes the number of identical internal repeats) significantly affects the stability of DARPins (19). Moreover, the MD simulations of DARPins E3_5 and E3_19 showed there were differences in stability between them, mainly due to the difference in stability of each C-terminal capping repeat (70).

The Unfolding Mechanism of p18

As described in the Methods section below, we explored the unfolding mechanism of p18 in four independent ~20 ns simulations (to enhance the sampling of the conformational space) of the thermally induced denaturation (498 K) of the protein. In all of these simulations, melting of the p18 protein from the N-terminus can be observed. As shown by the time courses of changes in native contacts between adjacent helix-turn-helix motifs and the secondary element plots, ARs I and II melt in the timescale of hundreds of picoseconds, while ARs III-V remain structured and retain their native-like assembly for significantly longer simulation times (Supplemental Figure S23). At ~4 ns, in all simulations, the entire p18 becomes unstructured with residual α-helical content (Supplemental Figure S24) and melted tertiary structure (Supplemental Figure S23). At the beginning of the p18 simulations structures having ARs I and II unstructured while the remaining ARs folded occur (Figure 6). The topology of the plot (Figure 6) clearly indicates that these structures represent

an intermediate between folded and unfolded states. As a result, the mechanism of p18 unfolding during the thermally induced denaturation can be considered as three-state (sequential) because one reaction intermediate populates on the unfolding pathway. The sequential N-terminal $\rightarrow$ C-terminal unfolding of p18, which our simulations exhibit, is consistent with the experimental measurements of Löw and co-workers (22) on the homolog p19 (71). The cited authors suggested that there is an on-pathway intermediate of p19 with structured ARs III-V and unstructured ARs I and II at equilibrium, unfolding and refolding conditions.

Conclusions

- The fragment simulations indicate that tertiary native interactions are important contributors to the stability of p18 α-helices, because in the absence of their native context the α-helices readily melt.

- Simulations of p18 fragments show that neither ARs nor their building blocks are stable in the 50 ns timescale. The intra-repeat interactions that contribute to stabilizing the ARs' structure are not sufficiently strong to maintain the stability of ARs separated from the intact system (i.e., of single ARs in solution).

- Two adjacent ARs in complex are stable in the 70 ns timescale, showing that inter-repeat interactions significantly contribute to the AR stability. In addition, a pair of ARs represents a minimal folding module of the p18, in accordance with previously published results (27-29).

- Fragment simulations, together with simulation of the complete p18, provide valuable information that would be extremely difficult to obtain from experimental studies, using currently available methodology at least, regarding the functional importance of loops and turns. The turns play an important role in p18 stability, mediating covalent linkage of antiparallel helices, while the loops also stabilize the p18 fold to a degree, and play a major role in the recognition process.

- The simulations of the thermally induced denaturation of the p18 indicate that the protein unfolds from the N-terminus to the C-terminus, via a three-state, sequential process with one intermediate populated on the reaction pathway.

- All-atomic MD simulation in explicit water at 50 ns timescales has proved to be a valuable tool for elucidating the roles of various fragments and interactions in the stability of ARP p18. This approach could also be useful in structural and functional analyses of similar proteins, for instance to compare the stability of various DARPins and their building blocks.

Supplementary Material

Supplementary material containing Figures S1 to S24 and Tables SI and SII is available free of charge online at `http://fch.upol.cz/jbsd/jbsd_sklenovsky.pdf` or from Adenine Press for a cost of US $50.00

Acknowledgments

References and Footnotes

1. J. N. Li, A. Mahajan, and M. D. Tsai. 1. Biochemistry 45, 15168-15178 (2006).

2. A. K. Bjorklund and D. Ekman, A. Elofsson. 2. PLoS Comput Biol 2, 959-970 (2006).

3. S. G. Sedgwick and S. J. Smerdon. 3. Trends Biochem Sci 24, 311-316 (1999).

4. E. R. G. Main, A. R. Lowe, S. G. J. Mochrie, S. E. Jackson, and L. Regan. 4. Curr Opin Struct Biol 15, 464-471 (2005).

5. L. K. Mosavi, T. J. Cammett, D. C. Desrosiers, and Z. Y. Peng. 5. Protein Sci 13, 1435-1448 (2004).

6. A. Hryniewicz-Jankowska, A. Czogalla, E. Bok, and A. F. Sikorski. 6. Folia Histochem Cytobiol 40, 239-249 (2002).

7. M. A. Andrade, C. Perez-Iratxeta, and C. P. Ponting. 7. J Struct Biol 134, 117-131 (2001).

8. P. Bork. 8. Proteins 17, 363-374 (1993).

9. G. Lee, K. Abdi, Y. Jiang, P. Michaely, V. Bennett, and P. E. Marszalek. 9. Nature 440, 246-249 (2006).

10. M. Sotomayor, D. P. Corey, and K. Schulten. 10. Structure 13, 669-682 (2005).

11. L. W. Li, S. Wetzel, A. Pluckthun, and J. M. Fernandez. 11. Biophys J 90, L30-L32 (2006).

12. M. T. Stumpp, H. K. Binz, and P. Amstutz. 12. Drug Discov Today 13, 695-701 (2008).

13. D. Steiner, P. Forrer, and A. Pluckthun. 13. J Mol Biol 382, 1211-1227 (2008).

14. D. U. Ferreiro, A. M. Walczak, E. A. Komives, and P. G. Wolynes. 14. PLoS Comput Biol 4, e1000070 (2008).

15. E. R. G. Main, S. E. Jackson, and L. Regan. 15. Curr Opin Struct Biol 13, 482-489 (2003).

16. E. Kloss, N. Courtemanche, and D. Barrick. 16. Arch Biochem Biophys 469, 83-99 (2008).

17. D. Barrick, D. U. Ferreiro, and E. A. Komives. 17. Curr Opin Struct Biol 18, 27-34 (2008).

18. G. Interlandi, G. Settanni, and A. Caflisch. 18. Proteins 64, 178-192 (2006).

19. G. Interlandi, S. K. Wetzel, G. Settanni, A. Pluckthun, and A. Caflisch. 19. J Mol Biol 375, 837-854 (2008).

20. K. W. Tripp and D. Barrick. 20. J Mol Biol 344, 169-178 (2004).

21. K. W. Tripp and D. Barrick. 21. J Mol Biol 365, 1187-1200 (2007).

22. C. Low, U. Weininger, M. Zeeb, W. Zhang, E. D. Laue, F. X. Schmid, and J. Balbach. 22. J Mol Biol 373, 219-231 (2007).

23. K. W. Tripp and D. Barrick. 23. J Am Chem Soc 130, 5681-5688 (2008).

24. T. O. Street and D. Barrick. 24. Protein Sci 18, 58-68 (2009).

25. P. Forrer, M. T. Stumpp, H. K. Binz, and A. Pluckthun. 25. FEBS Lett 539, 2-6 (2003).

26. H. K. Binz, M. T. Stumpp, P. Forrer, P. Amstutz, and A. Pluckthun. 26. J Mol Biol 332, 489-503 (2003).

27. B. Zhang and Z. Y. Peng. 27. J Mol Biol 299, 1121-1132 (2000).

28. P. Sklenovsky, P. Banas, and M. Otyepka. 28. J Mol Model 14, 747-759 (2008).

29. L. K. Mosavi and Z. Y. Peng. 29. Protein Eng 16, 739-745 (2003).

30. L. K. Mosavi, D. L. Minor, and Z. Y. Peng. 30. Proc Natl Acad Sci USA 99, 16029-16034 (2002).

31. C. C. Mello and D. Barrick. 31. Proc Natl Acad Sci USA 101, 14102-14107 (2004).

32. M. Otyepka, P. Sklenovsky, D. Horinek, T. Kubar, and P. Hobza. 32. J Phys Chem B 110, 4423-4429 (2006).

33. D. U. Ferreiro, S. S. Cho, E. A. Komives, and P. G. Wolynes. 33. J Mol Biol 354, 679-692 (2005).

34. M. F. Roussel. 34. Oncogene 18, 5311-5317 (1999).

35. A. Carnero and G. J. Hannon. 35. Curr Top Microbiol Immunol 227, 43-55 (1998).

36. D. O. Morgan. 36. Annu Rev Cell Dev Biol 13, 261-291 (1997).

37. S. Ortega, M. Malumbres, and M. Barbacid. 37. Biochim Biophys Acta-Rev Cancer 1602, 73-87 (2002).

38. N. P. Pavletich. 38. J Mol Biol 287, 821-828 (1999).

39. R. Venkataramani, K. Swaminathan, and R. Marmorstein. 39. Nat Struct Biol 5, 74-81 (1998).

40. D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman. Comput Phys Commun 91, 1-41 (1995).

41. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. J Am Chem Soc 117, 5179-5197 (1995).

42. J. M. Wang, P. Cieplak, and P. A. Kollman. 42. J Comput Chem 21, 1049-1074 (2000).

43. I. Bartova, M. Otyepka, Z. Kriz, and J. Koca. 43. Protein Sci 13, 1449-1457 (2004).

44. I. Bartova, M. Otyepka, Z. Kriz, and J. Koca. 44. Protein Sci 14, 445-451 (2005).

45. M. Otyepka, I. Bartova, Z. Kriz, and J. Koca. 45. J Biol Chem 281, 7271-7281 (2006).

46. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 46. J Chem Phys 79, 926-935 (1983).

47. V. Daggett. 47. Chem Rev 106, 1898-1916 (2006).

48. W. Kabsch and C. Sander. 48. Biopolymers 22, 2577-2637 (1983).

49. D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. J Comput Chem 26, 1701-1718 (2005).

50. A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. 50. Proteins 17, 412-425 (1993).

51. I. Bartova, J. Koca, and M. Otyepka. 51. Protein Sci 17, 22-33 (2008).

52. P. Michaely, D. R. Tomchick, M. Machius, and R. G. W. Anderson. 52. Embo J 21, 6387-6396 (2002).

53. R. Baumgartner, C. Fernandez-Catalan, A. Winoto, R. Huber, R. A. Engh, and T. A. Holak. Structure 6, 1279-1290 (1998).

54. J. L. Neira, L. S. Itzhaki, A. G. Ladurner, B. Davis, G. D. Gay, and A. R. Fersht. 54. J Mol Biol 268, 185-197 (1997).

55. M. L. DeMarco, D. O. V. Alonso, and V. Daggett. 55. J Mol Biol 341, 1109-1124 (2004).

56. J. D. Hirst and C. L. Brooks. 56. Biochemistry 34, 7614-7621 (1995).

57. K. A. Scott, D. O. V. Alonso, Y. P. Pan, and V. Daggett. 57. Biochemistry 45, 4153-4163 (2006).

58. L. S. Itzhaki, D. E. Otzen, and A. R. Fersht. 58. J Mol Biol 254, 260-288 (1995).

59. V. Daggett and A. Fersht. 59. Nat Rev Mol Cell Biol 4, 497-502 (2003).

60. A. Kohl, H. K. Binz, P. Forrer, M. T. Stumpp, A. Pluckthun, and M. G. Grutter. 60. Proc Natl Acad Sci USA 100, 1700-1705 (2003).

61. V. S. Devi, H. K. Binz, M. T. Stumpp, A. Pluckthun, H. R. Bosshard, and I. Jelesarov. Protein Sci 13, 2864-2870 (2004).

62. Y. W. Yang, S. Nanduri, S. Sen, and J. Qin. 62. Structure 6, 619-626 (1998).

63. M. Serrano, G. J. Hannon, and D. Beach. 63. Nature 366, 704-707 (1993).

64. P. Michaely and V. Bennett. 64. J Biol Chem 268, 22703-22709 (1993).

65. K. S. Tang, B. J. Guralnick, W. K. Wang, A. R. Fersht, and L. S. Itzhaki. 65. J Mol Biol 285, 1869-1886 (1999).

66. P. D. Jeffrey, L. Tong, and N. P. Pavletich. 66. Genes Dev 14, 3115-3125 (2000).

67. J. N. Li, I. J. L. Byeon, K. Ericson, M. J. Poi, P. O'Maille, T. Selby, and M. D. Tsai. Biochemistry 38, 2930-2940 (1999).

68. M. E. Zweifel and D. Barrick. 68. Biochemistry 40, 14344-14356 (2001).

69. M. E. Zweifel and D. Barrick. 69. Biochemistry 40, 14357-14367 (2001).

70. H. Yu, A. Kohl, H. K. Binz, A. Pluckthun, M. G. Grutter, and W. F. van Gunsteren. 70. Proteins 65, 285-295 (2006).

71. C. H. Yuan, J. A. Li, T. E. Selby, I. J. L. Byeon, and M. D. Tsai. 71. J Mol Biol 294, 201-211 (1999).

72. P. Jurecka, J. Cerny, P. Hobza, and D. R. Salahub. 72. J Comput Chem 28, 555-569 (2007).

10.3 Appendix C

ATP and magnesium drive conformational changes of the Na$^+$/K$^+$-ATPase cytoplasmic headpiece

Lenka Gryčová • **Petr Sklenovský** • Zdeněk Lánský • Marika Janovská • Michal Otyepka • Evžen Amler • Jan Teisinger • Martin Kubala

Abstract

Conformational changes of the Na$^+$/K$^+$-ATPase isolated large cytoplasmic segment connecting transmembrane helices M4 and M5 (C45) induced by the interaction with enzyme ligands (i.e. Mg^{2+} and/or ATP) were investigated by means of the intrinsic tryptophan fluorescence measurement and molecular dynamic simulations. Our data revealed that this model system consisting of only two domains retained the ability to adopt open or closed conformation, i.e. behavior, which is expected from the crystal structures of relative Ca^{2+}-ATPase from sarco(endo)plasmic reticulum for the corresponding part of the entire enzyme. Our data revealed that the C45 is found in the closed conformation in the absence of any ligand, in the presence of Mg^{2+} only, or in the simultaneous presence of Mg^{2+} and ATP. Binding of the ATP alone (i.e. in the absence of Mg^{2+}) induced open conformation of the C45. The fact that the transmembrane part of the enzyme was absent in our experiments suggested that the observed conformational changes are consequences only of the interaction with ATP or Mg^{2+} and may not be related to the transported cations binding/release, as generally believed. Our data are consistent with the model, where ATP binding to the low-affinity site induces conformational change of the cytoplasmic part of the enzyme, traditionally attributed to E2 $\rightarrow$ E1 transition, and subsequent Mg^{2+} binding to the enzyme–ATP complex induces in turn conformational change traditionally attributed to E1 $\rightarrow$ E2 transition.

Introduction

The Na$^+$/K$^+$-ATPase (sodium pump, EC 3.6.3.9) translocates sodium and potassium ions across the plasma membrane against their concentration gradients using the energy from ATP hydrolysis. This is crucial for maintaining both the membrane potential and low cytoplasmic sodium concentration. As the gradient of sodium is also used by several secondary transporters of various solutes, it is not surprising that malfunction of the Na$^+$/K$^+$-ATPase is involved in pathology of several dozens of diseases, like e.g. ischemia or diabetes.

Na$^+$/K$^+$-ATPase belongs to the P-type ATPases superfamily [1]. Its designation is derived from the finding that the enzyme is autophosphorylated during the catalytic cycle. The Albers–Post model postulates that the enzyme adopts two conformations during the catalytic cycle, which are traditionally designated as E1 and E2 [2,3]. The enzyme in the E1 conformation has a high affinity

to both ATP and sodium ions, while in the E2 conformation it has a low affinity to ATP and a high affinity to potassium. The cations binding sites, which are located within the transmembrane domain, are open toward cytoplasm in the E1 conformation and toward extracellular milieu in the E2 conformation. Notably, already early experiment of Skou [4] revealed the important role of the Mg^{2+} cations for the proper function of the enzyme. Although Mg^{2+} is not transported across the plasma membrane, it must be present in the cytoplasm as an essential cofactor.

Na^+/K^+-ATPase and H^+/K^+-ATPase are the only members of the P-type ATPases superfamily characterized by the presence of two subunits. The catalytic α-subunit contains the conserved sequences characteristic for the P-type ATPase superfamily [5] and it is responsible for the cations translocation as well as ATP hydrolysis. The β-subunit is a glycoprotein with single transmembrane helix and it is essential for the proper maturation and targeting of the enzyme into the plasma membrane, though it is very likely that it interacts with the α-subunit also during the catalytic cycle [6].

Successful crystallization of another P-type ATPase family member, the Ca^{2+}-ATPase from sarco(endo)plasmic reticulum (SERCA, EC 3.6.3.8), provided a very valuable information about the structure of the catalytic subunit [7]. Moreover, SERCA was later crystallized in various conformations that were assigned to the E1 or E2 states of the enzyme [8–16]. There are numerous spectroscopic and biochemical experiments suggesting that the topologies of the Na^+/K^+-ATPase α-subunit and SERCA are very similar [for review, see [17]]. They were recently supported by the publication of the Na^+/K^+-ATPase crystal structure at 3.5 Å resolution [18], which confirmed the close structural relationship of these two enzymes. The α-subunit has ten transmembrane helices (M1–M10) and two longer cytoplasmic segments that are organized into three well-separated domains. The one, designated as A, is formed by the C23 (cytoplasmic segment between the M2 and M3), the other two, designated as N and P are formed by C45. Designation of the P-domain has origin in the fact that it contains the phosphorylated aspartyl residue, the N-domain contains the nucleotide- binding site and the A-domain (actuator or anchor) assists to the conformational changes that are necessary for the cation transport. The cation-binding sites are located within the transmembrane region; the cations are coordinated by the residues of M4, M5, M6 and M8 [7].

It was shown that the large cytoplasmic segment C45 of the α-subunit containing both the nucleotide binding- and phosphorylation sites can be overexpressed in E. coli and purified without the rest of the enzyme [19–23]. It was demonstrated that this isolated C45 retains its functional properties, such as ATP- or TNP-ATP-binding, suggesting that the 3D-structure is preserved [19–25]. This is also supported by the finding that the structure of the isolated N-domain independently solved by both NMR and crystallography fitted well to the N-domain from the crystals of SERCA [26]. This artificial system became popular, because (i) it uncouples the enzyme/nucleotide interaction from the cation transport, which facilitates the data interpretation, and (ii) solubility of the isolated C45 greatly facilitates the experimental work.

Intrinsic tryptophan fluorescence is very sensitive to the changes in the environment of the tryptophanyl residue, and thus, it could be a useful marker of the conformational changes. However, in order to localize the molecular events, the spectroscopic data can be easily interpreted only for a protein that contains no more than a single Trp residue. Therefore, we have created a set of

single-tryptophan mutants of the C45, and monitored changes induced on the C45 upon interaction with the Mg^{2+}, Na_2ATP or MgATP ligands. Our data indicate that ATP and magnesium act in a coordinated manner in order to open or close the C45. As the abovementioned crystallographic studies provide only the static structural information, our experiments monitoring the C45 dynamic well complement these data.

Materials and Methods

Site-directed mutagenesis of the Na^+/K^+-ATPase C45

DNA sequence of the C45 of the Na^+/K^+-ATPase α-subunit was prepared by the polymerase chain reaction (PCR) from the 1260 base DNA segment encoding Lys354–Lys774 α1-subunit of the mouse brain Na^+/K^+-ATPase used previously in our laboratory [21–23,27]. This sequence (L354 – I777) was amplified using the primers allowing the introduction of NheI and HindIII restriction sites at the 5′ and 3′ termini, double digested with NheI, HindIII restriction enzymes, and ligated into the expression vector pET28b. DNA sequencing of the resulting construct was performed on ABI Prism automated sequencer (facility of the Academy of Sciences of the Czech Republic) in order to confirm the in-frame insertion of the C45 DNA into the expression vector. The wild-type sequence of the C45 contains two native tryptophan residues W411 and W385. These two amino acid residues were subsequently substituted by phenylalanine residues, yielding two single-tryptophan mutants WT-W385F and WT-W411F and the tryptophanless (TL) double mutant WT-W411F-W385F, which was thereafter used as a template for other point mutations. Following single-tryptophan mutants were constructed: TL-F404W, TL-F426W, TL-F571W, TL-I627W, TL-F683W, and TL-S732W. Site directed mutagenesis PCRs were performed using PfuUltra High-fidelity DNA Polymerase (Stratagene). The upstream primers for performed mutations were (altered nucleotides are in italic):

WT-W385F 5′ggatgacagtggctcacatgt*t*ctttgacaatcaaatcc3′

TL-F404W 5′gagaatcagagtggggtctcct*gg*gacaagacgtcagccacc3′

WT-W411F 5′gacaagacgtcagccacct*t*cttcgctctgtccagaattgc3′

TL-F426W 5′tgtaacagggcagtgt*gg*caggctaaccaagaaaacc3′

TL-F571W 5′gacactgatgaagtcaatt*gg*cccgtggataacctctgc3′

TL-I627W 5′gggggtgggcattt*gg*tcagaaggtaacg3′

TL-F683W 5′gcggtaccacacggagattgtct*gg*gctaggacctctcctc3′

TL-S732W 5′gccatggggattgttggct*g*ggatgtgtccaagc3′

Downstream primers were reverse complementary. All constructs were verified by DNA sequencing.

Fusion protein expression and purification

The C45 of the Na^+/K^+-ATPase DNA and all its mutations were transformed into BL21 *Escherichia coli* bacteria (Promega) and were cultured at 37 ℃ to OD(600 nm) of 0.6. Induction was carried

out with 0.2 mM IPTG (isopropyl β-D-thiogalactoside) overnight at 17 ℃. Cells were centrifuged, resuspended in 20 mM Tris–HCl, 140 mM NaCl, pH 7.4 containing protein inhibitors (2 μg/ml leupeptin, 2 μg/ml pepstatin and 1 mM Phenylmethylsulfonyl fluoride (PMSF)), disrupted by sonication and the homogenate was again centrifuged. All constructs were expressed as a $(His)_6$–tag fusion protein; the $(His)_6$–tag was attached to the N-terminus. Purification affinity chromatography was performed according to standard TALON Metal Affinity Resin (Clontech) manufacturer protocol. The protein samples were eluted with 0.5 M Imidazol and were dialyzed against 1 liter of 50 mM Tris–HCl, 140 mM NaCl, 2 mM DTT, pH 7.5 overnight at 4 ℃. The purity of protein samples was verified using 12% SDS-PAGE. Concentrations were estimated using the Bradford assay [28].

Tryptophan fluorescence quenching

The steady-state tryptophan fluorescence emission was quenched using 0–300 mM acrylamide as a quencher. Data were collected using an excitation and emission wavelengths of 295 nm and 360 nm, respectively, slits were set to 5 nm and 10 nm for the excitation and emission channel, respectively, the integration time was 3 s for recording of each point, the measurements were performed at 22 ℃. Protein samples were diluted to the final 3 μM concentration into 20 mM Tris–HCl, 140 mM NaCl, pH 7.5 buffer, where oxygen was removed by argon, and were titrated by aliquots of the quencher. The sample was gently stirred and the fluorescence spectrum was recorded.

The efficiency of quenching was evaluated by the non-linear leastsquares analysis using the Stern–Volmer formula:

$$F = \frac{F_0}{1 + K_{SV}\,[Q]},$$

where F_0 and F are the fluorescence intensities in the absence or in the presence of the quencher, respectively, K_{SV} is the Stern–Volmer quenching constant and $[Q]$ is the concentration of the quencher. Both the F_0 and K_{SV} were the fitted parameters. Notably, attempts to use more complicated models (e.g. two populations of fluorophors with different quencher accessibility) did not improve the quality of the fit. Therefore, we used the formula given above.

Quenching experiments were performed in four different setups i.e. without any ligand, with 15 mM $MgCl_2$ (Lach-Ner), 15 mM Na_2ATP (Sigma) or 15 mM MgATP (Sigma). The experiments in the absence of magnesium (i.e. free C45, or in the presence of Na_2ATP) were carried out in the presence of 5 mM EDTA. In order to identify the influence of individual ligand on the C45 conformation, quenching constants from at least three independent measurements were compared to the ligand-free experiments quenching constants.

Time-resolved fluorescence measurements

Time-resolved fluorescence data were obtained by the time-correlated single photon counting (TCSPC) method on the spectrometer PicoHarp 300 (PicoQuant), using the pulsed LED centered at 298 nm (PLS300, PicoQuant) with the pulse frequency 10 MHz and emission monochromator set to 350 nm. Detected photons were collected into a histogram covering the time scale of 100 ns and

with a 32 ps/channel resolution. All the experiments were carried out at 20 ℃ (bath controlled). In the experiments for fluorescence lifetime determination, data were sampled for 20 min under the magic-angle conditions, and 2000–11000 counts in the leading channel were collected in the histogram. In the anisotropy decay experiments with W385 mutant, each I_{VV} and I_{VH} were collected for 10 min and about 8000 counts in the leading channel for I_{VV} was obtained. The G-factor was determined in the separate experiment. The instrument–response function was estimated using the colloid–silica (Ludox) as a scatterer. Fluorescence decays were fitted using the FLUOFIT 4.2.1 software (PicoQuant) as a sum of exponentials:

$$I\left(t\right) = \sum_i \alpha_i \exp\left(-\frac{t}{\tau_i}\right).$$

The intensity-weighted mean fluorescence lifetime was calculated as

$$\tau_M = \frac{\sum \alpha_i \tau_i{}^2}{\sum \alpha_i \tau_i}.$$

Fluorescence anisotropy decays were fitted by the Pulse5Q software using the maximum entropy method (MEDC, Ltd.). Fluorescence anisotropy decays were analyzed as a sum of exponentials:

$$r\left(t\right) = \sum_i \beta_i \exp\left(-\frac{t}{\phi_i}\right),$$

where the set of the amplitudes β_i represents a distribution of the correlation times ϕ_i. The β_i are related to the initial anisotropy r_0 by the formula:

$$\sum_i \beta_i = r_0.$$

100 correlation times ϕ_i equidistantly spaced at the logarithmic scale ranging from 50 ps to 100 ns were used.

Molecular modeling

At the beginning of the study, no experimentally determined structure of the entire Na^+/K^+-ATPase C45 was known. Therefore we created the C45 model in the closed conformation (see also Discussion) based on the homology with SERCA and it was used as a starting structure in the molecular dynamic simulations described below. However, during the preparation of the manuscript, the crystal structure of the Na^+/K^+-ATPase at 3.5 Å resolution has been published revealing the open C45 conformation. In order to check the validity of our simulations, we excised the C45 from the crystal structure and the subsequent molecular dynamic simulation revealed that the loop relaxed into the closed conformation, thus, justifying our homology model based molecular dynamic simulations (see below).

Homology model construction   The homology model of the C45 of the Na^+/K^+-ATPase was created using the MOE [29] software package. The C45 sequence was aligned to the sequence of the rabbit SERCA1a and the alignment was manually refined with respect to the known Na^+/K^+-

ATPase structural data, namely to the NMR-solved N-domain structure [26] (Suppl. Mat., Fig. S1). Our previous data suggested that the isolated C45 adopted the E2-like conformation; therefore the model was built using the high-resolution structure of SERCA in E2-conformation (PDB entry 2C8L) as a template. Using the default settings and the force-field of the *parm99* [30,31], 10 structures were generated using the fine minimization and the average structure was taken as the initial state for the molecular dynamics simulations. The C-terminal loop of the C45 homology model was truncated by 13 residues in order to reduce computational demands.

Molecular dynamics simulations   Molecular dynamics (MD) simulations have been performed with a homology model of the C45 with sodium or magnesium counterions in the absence of ATP (hereafter C45-Na and C45-Mg), as well as with ATP bound to the active site of C45 (hereafter C45-Na$_2$ATP and C45-MgATP). The ligand was immersed to the previously relaxed structure of C45-Na using the 1MO8 structure [26] as a template. The ATP phosphates were oriented toward the phosphorylation site of P-domain at D369 residue. The relaxed structure was taken at the time of $\sim$2000 ps from the C45-Na MD simulation. To compensate the negative charge of C45-Na$_2$ATP and C45-MgATP, 22 Na$^+$ or 11 Mg^{2+} counterions were added, respectively. To check the quality of C45 homology model simulations, a control MD run at 10 ns time scale with the crystal structure of the Na$^+$/K$^+$ ATPase (PDB ID 3B8E[18]) was performed. The MD simulation of unliganded Na$^+$/K$^+$ ATPase (denoted as C45-Xray) was carried out with Na$^+$ counterions to neutralize the system. All simulations were carried out using the SANDER module of AMBER 9.0 [32] suite of programs with the *parm99* force field. The simulation protocol was used as follows: at first all protonation states of histidine residues were checked by visual inspection to maximize H-bond contacts. Then, all hydrogens and individual counterions were added using LEaP program of AMBER 9.0 package. Each system was immersed in a rectangular water box with a minimum distance between the molecule and the box wall of 9 Å. Then, each system was minimized prior to the production part of the molecular dynamics run in the following way. The protein was constrained and the solvent molecules with counterions were allowed to move during a 1000 step minimization followed by 10 ps long molecular dynamics run under $[NpT]$ conditions ($p = 1$ atm, $T = 298.15$ K). The next step covers relaxation of side chains by several consequent minimizations with decreasing force constants applied to the backbone atoms. After the relaxation, the entire system was heated. The thermalization protocol was set up as follows: each system was heated from 10 K to 50 K for 20 ps, then from 50 K to 250 K for 20 ps and then from 250 K to 298.15 K for 30 ps. The particle-mesh Ewald (PME) methods for treating electrostatic interactions were used. All simulations were performed under periodic boundary conditions in the $[NpT]$ ensemble at 298.15 K and 1 atm using the 2 fs integration step. The SHAKE algorithm with a tolerance of 10^{-5} Å was used to fix positions of all hydrogens. The cutoff of 9.0 Å was applied to treat nonbonding interactions. Coordinates were stored every picosecond. The total duration of the production phases for C45-Na, C45-Mg, C45-Na$_2$ATP and C45-MgATP is equal to 7 ns, 3.5 ns, 4 ns and 4.5 ns, respectively. The C45-Na system contained in total 60,219 atoms (411 residues, 18 Na$^+$ counterions and 17,992 water molecules). The C45-Mg system contained 58,983 atoms (411 residues, 9 Mg^{2+} atoms and 17,583 water molecules). The C45-Na$_2$ATP system contained 57,431 atoms (411

residues, 22 Na$^+$ counterions, one ATP molecule and 17,047 water molecules). The C45-MgATP contained 57,315 atoms (411 residues, 11 Mg^{2+} counterions, one ATP molecule and 17,012 water molecules). The C45-Xray structure contained 52,710 atoms (411 residues, 18 Na$^+$ counterions and 15,489 water molecules). The trajectory stability was checked by the root-mean-square-deviation (RMSD) of backbone atoms from the initial structure and radius of gyration (R_g). The RMSD, R_g and secondary structure analysis were calculated by PTRAJ module of AMBER 9.0 package. The radial distribution function (RDF), describes how the density of surrounding matter varies as a function of the distance from a distinguished point. The RDF values of water molecules oxygens at 1.0 nm distance to center of mass of selected residues have been used to assess a spatial accessibility of the residues from solvent.

Results

Expression and purification of the single-tryptophan C45 mutants

The native C45 sequence contains two tryptophanyl residues (W385 and W411). They were subsequently replaced by phenylalanines to obtain a tryptophanless (TL) construct. Further, artificial reporter tryptophans were inserted into this TL construct to yield a set of single-tryptophan mutants. Based on the homology model of the C45 (see below), these reporter tryptophans were placed on the various positions on the C45 surface (Fig.1). Wherever it was possible, we attempted to perform conserved mutations (i.e. the residue to be replaced by Trp was chosen as aromatic or hydrophobic one), in order to minimize the effect of the mutation on the protein folding. Indeed, expression of all mutants resulted in a high yield of soluble protein (typically > 5 mg of purified protein from 0.2 L of the culture), suggesting that the protein tertiary structure was not substantially altered. The purification of the (His)$_6$–tagged protein on the Co^{2+}-affinity column yielded a single band on the SDS-electrophoresis gel (not shown).

Fluorescence decays

Time-correlated single photon counting (TCSPC) method was used to characterize fluorescence decays of the single-tryptophan mutants at various conditions (without any ligand, in the presence of Mg^{2+}, MgATP or Na$_2$ATP). All the decays were satisfactorily fitted to the three-exponential decay scheme (Suppl. Mat., Table S1) with reduced χ^2 approaching 1.00, random distribution of the first residuals and the autocorrelation function in all cases (not shown). The only exceptions were all decays of the TL-S732W mutant, decays for TL-V648W mutant in the absence of any ligand and in the presence of Mg^{2+}, which were satisfactorily described by the 2-exponential decay model. The main goal of these experiments was to estimate the mean excited-state lifetime (the penultimate column of the Table S1 in Suppl. Mat.) that characterizes the time, during which the acrylamide can effectively quench the fluorescence.

Table 1

Protein	Location	No substrate	Mg^{2+}	MgATP	Na_2ATP
WT-W411F	N-domain	0.86 ± 0.02	0.78 ± 0.06	0.65 ± 0.02	0.68 ± 0.03
TL-F404W	N-domain	1.79 ± 0.23	1.37 ± 0.17	1.30 ± 0.33	0.98 ± 0.17
WT-W385F	N-domain	1.16 ± 0.08	1.21 ± 0.19	1.13 ± 0.22	1.14 ± 0.22
TL-F426W	N-domain	1.66 ± 0.03	1.46 ± 0.09	1.53 ± 0.09	1.48 ± 0.12
TL-F571W	N-domain	1.77 ± 0.18	1.00 ± 0.12	1.34 ± 0.09	1.28 ± 0.08
TL-I627W	P-domain	0.97 ± 0.07	1.07 ± 0.07	1.05 ± 0.22	0.89 ± 0.18
TL-V648W	P-domain	1.42 ± 0.33	1.36 ± 0.26	1.37 ± 0.19	1.34 ± 0.23
TL-F683W	P-domain	1.33 ± 0.08	1.39 ± 0.26	1.18 ± 0.11	0.92 ± 0.07
TL-S732W	P-domain	1.36 ± 0.07	1.26 ± 0.02	1.25 ± 0.17	1.22 ± 0.09

Bimolecular quenching constants for quenching of the Trp fluorescence by acrylamide given in 10^9 M^{-1} s^{-1} are calculated as the ratio of the Stern–Volmer quenching constants given in Table S2, and the corresponding mean fluorescence lifetime in Table S1. The native sequence of C45 contains two Trp residues on the positions 385 and 411. Thus, WT-W411F or WT-W385F designates mutants where only Trp385 or Trp411 was present, respectively, the other Trp residue was mutated to Phe. For all the other mutants, both the Trp385 and Trp411 were mutated to phenylalanines, and the designation of the protein indicates the position of the reporter Trp residue that was inserted into the tryptophanless mutant.
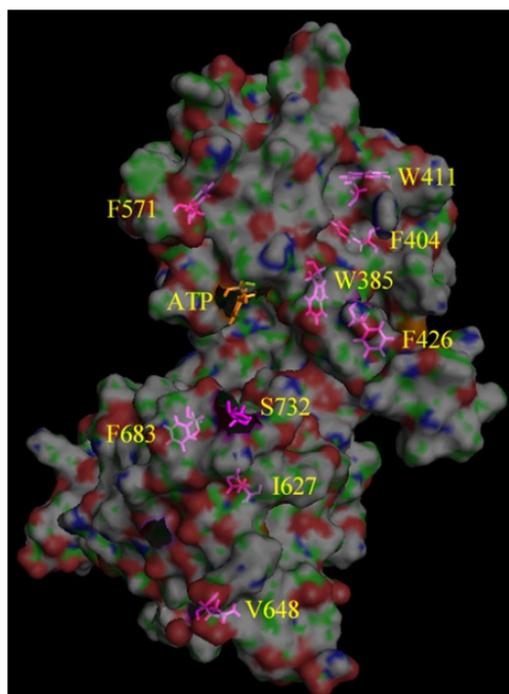

Figure 1: Reporter positions on the C45. The homology C45 model is displayed with the Connolly type semitransparent surface, the residues that where mutated to the reporter tryptophanyl residue and subsequently evaluated in the spectroscopic experiments are in the stick representation and colored purple.

Fluorescence quenching by acrylamide

Fluorescence quenching experiments give a unique opportunity to monitor changes of the protein shape. The Stern–Volmer quenching constant (K_{SV}) reflects the effectivity of the quenching (Suppl. Mat., Table S2) and is dependent mainly on the spatial accessibility of the fluorophore and its mean excited-state lifetime (τ_M, Table S1, Suppl. Mat.). Dividing the K_{SV} by τ_M yields the bimolecular quenching constant (k_q, Table 1), which already characterizes the spatial accessibility of the fluorophor. Therefore, even in the experiments without any ligand the k_q varied between 1.79×10^9 M^{-1}s^{-1} for the best accessible tryptophan in the TL-F404W mutant, down to 0.86×10^9 M^{-1}s^{-1} for the worst accessible native W385 residue (WT-W411F mutant), within our set of single-tryptophan mutants (Table 1). Notably, after binding of ligand (Mg^{2+}, MgATP or Na$_2$ATP) we observed significant k_q changes for the quenching of some tryptophans, as discussed below.

Mg^{2+} binding

It was expected that binding of this small ligand would induce only subtle changes in the C45 shape that will be hardly detectable. Indeed, quenching of most tryptophans deviated only within the experimental error from the experiments performed in the absence of any ligand. The only exception was the quenching of the TL-F571W mutant (this residue is located on the N-domain), where we observed almost 2-fold decrease in the k_q (from 1.77×10^9 to 1.00×10^9 M^{-1}s^{-1}), reflecting substantial protection of this residue against quenching ($p<0.01$). Notably, the k_q value significantly differs also from the value obtained in the MgATP presence ($p<0.03$, see below), suggesting that the observed effect is a result of free Mg^{2+} binding.

MgATP binding

We expected that binding of ATP in the presence of magnesium would induce substantial changes in the C45 conformation. Surprisingly, the observed changes differed only very slightly from the experiments performed with Mg^{2+} only. Again, quenching of most Trp residues deviated only within the range of experimental error, when compared to the experiments in the absence of any ligand, and the only mutant exhibiting substantial changes was the TL-F571W mutant where we observed decrease in k_q from 1.77×10^9 M^{-1}s^{-1} to 1.34×10^9 M^{-1}s^{-1} ($p<0.06$). This value was significantly different also from that obtained in the presence of Mg^{2+} only ($p<0.03$), but not from that obtained in the Na$_2$ATP presence ($p<0.51$, see below), suggesting that it is an effect of the ATP binding to nucleotide-binding site on the N-domain. Further, for the native W385 residue (WTW411F mutant), we observed decrease in k_q from 0.86×10^9 to 0.65×10^9 M^{-1}s^{-1} ($p<0.01$). However, this value does significantly differ neither from that obtained in the Mg^{2+} presence ($p<0.70$) nor in the Na$_2$ATP presence ($p<0.95$, see below), and it seems that this half-buried native tryptophan is sensitive to binding of all ligands.

Table 2

Ligand	β_1	β_2	ϕ_1 ns	ϕ_2 ns	χ_R^2
None	0.062	0.145	0.053	39.8	1.09
Mg^{2+}	0.060	0.144	0.069	41.5	1.07
MgATP	0.038	0.143	0.069	39.6	1.01
Na_2ATP	0.039	0.133	0.104	31.5	1.05

Anisotropy decay parameters for the native W385 residuum (WT-W411F mutant), β_i are the fractional anisotropies, ϕ_i are the corresponding rotational correlation times, reduced χ^2 reflects the goodness of fit.

Na_2ATP binding

The most complex changes were observed when the C45 mutants were incubated with the Na_2ATP. The k_q decreased significantly (when compared to the quenching in the absence of any ligand) in the quenching of the native W385 residue (WT-W411F mutant) from 0.86×10^9 to 0.68×10^9 $M^{-1}s^{-1}$ ($p<0.01$), for the TL-F404W mutant from 1.79×10^9 to 0.98×10^9 $M^{-1}s^{-1}$ ($p<0.02$), for the TL-F571W mutant from 1.77×10^9 to 1.28×10^9 $M^{-1}s^{-1}$ ($p<0.05$) and for the TL-F683W mutant from 1.33×10^9 to 0.92×10^9 $M^{-1}s^{-1}$ ($p<0.01$), revealing that also these residues are partially protected against quenching in the Na_2ATP presence. Notably, the changes concerning the N-domain residues (i.e. WT-W411F, TL-F404W and TL-F571W mutants) go in the same direction as in the MgATP presence (mutual comparison of results in the MgATP or Na_2ATP presence yields values of $p<0.38$, $p<0.31$ or $p<0.51$, respectively), but are much more distinct. However, for the TL-F683W mutant, the result significantly differs also in the comparison to the MgATP ($p<0.06$). The quenching of the Trp residues on the other mutants did not significantly differ from the experiments in the absence of any ligand.

Anisotropy decay

Fluorescence anisotropy contains the information about the mobility of the fluorophor. Typically, time-resolved measurement of the anisotropy decay enables to distinguish three components for the tryptophanyl residue on the protein. The shortest rotation correlation time (typically about 70 ps) corresponds to the movement of Trp sidechain. The middle one (typically in ns time-range) reflects segmental motion, while the longest one reflects rotation of the whole protein, and is therefore dependent on its size (typically tens of ns) [33].

We selected the mutant containing the native W385 residue (WT-W411F mutant) located on the N-domain for the anisotropy decay experiments because according to our model this residue is halfburied, and we expected that it should not exhibit the segmental motion, which simplifies the data analysis. Indeed, the anisotropy decay could be described as a sum of two components (Table 2) in all cases. The $\sim$70 ps correlation time fits well to the values expected for the Trp side-chain movement, and the $\sim$40 ns correlation time observed for the protein without any ligand, in the presence of Mg^{2+} or MgATP could be also expected for the 48 kDa protein. Interestingly, in the presence of ATP alone, this longer component has been shortened to $\sim$30 ns only, reflecting that the N-domain is only weakly linked to the P-domain in this case. In principle, we should

observe three components in the anisotropy decay in this situation. As the N-domain constitutes roughly one half of the C45, the longer correlation time should be split to the ∼20 ns component reflecting the N-domain "segmental" movement and the ∼40 ns component characterizing the C45 rotation. However, resolution the two correlation times differing only by the factor of 2 in this complex system (3-exponential fluorescence decay and 3-exponential anisotropy decay) is beyond the limits of this method, and in fact, we observe only some average value. Note that we cannot safely interpret the increase of the correlation time in the shorter component of the anisotropy decay in the presence of the ATP, as this component is poorly resolved in the experimental setup with 32 ps/channel resolution (we can just detect that it is present).

Molecular dynamics simulations

The stability of studied systems   All trajectories are stable during the production part of the MD simulations and reliable for further analyses. The trajectory stability confirms RMSD of backbone atoms from the initial structure, R_g and secondary structure analysis. The curve of RMSD against time calculated for the C45-Na system reaches its plateau at ∼1.5 ns. The RMSD of C45-Mg system points out that the system equilibrates at ∼1 ns. The C45-Na$_2$ATP system equilibrates within 2 ns and the C45-MgATP system within 1 ns. The time evolution of RMSD of the C45-Xray reaches its plateau at ∼2 ns. The R_g values of C45-Na, C45-Mg and C45-MgATP systems fluctuate around the mean values of 24.9 Å, 25.0 Å and 24.9 Å, respectively. However, the R_g of C45-Na$_2$ATP starts to increase from 25.1 Å after 2 ns reaching a new plateau ($R_g = 25.7$ Å) at ∼2.4 ns (Fig. 2). The visual inspection of the C45-Na$_2$ATP MD simulation unravels that the increase of R_g correlates with opening of both domains (Fig. 3). The domain opening motion can be described by the angle among the N-domain center of mass, center of mass of loops connecting both domains and P-domain center of mass (hereafter angle-m-m-m). The mean value ± SEM of angle-m-m-m calculated from the last nanosecond of C45-Na$_2$ATP simulation is equal to 129.9° ± 2.1° and is significantly larger than the mean value 121.6° ± 2.9° from the first nanosecond. The angle 129.9° between both domains (C45-Na$_2$ATP) is also significantly larger than the same angle 117.5° ± 2.1° calculated for C45-MgATP system.

Structure, dynamics and ATP binding   Generally, the systems containing magnesium ions are less flexible than the systems with sodium ions. The secondary structure is well preserved during all simulations and the highly flexible regions correspond to the loops between secondary structure elements. The N-terminal loop (residues 386–412) is the most flexible segment in all investigated systems. Another flexible C45 loop is formed by residues 422–467. Two loops (residues 625–667 and 721–733) are the most flexible regions of the P-domain.

The ATP molecule is housed by a narrow active site formed by antiparallel β-sheet (residues 481–487 and 496–502), loop of residues 474–480 and loop of residues 540–545 in our models. The ATP purine ring stacks to the F475 benzene ring and ATP N^{6}H$_2$ group may form H-bond to D443. The ATP ribose moiety (O2' and O3') makes two H-bonds to R544 and D612, respectively. A salt bridge among the α- and β-phosphate moieties and K480 also occurs during simulations. These structural characteristics correspond well to previously published spectroscopic experiments [21–23,26,34].
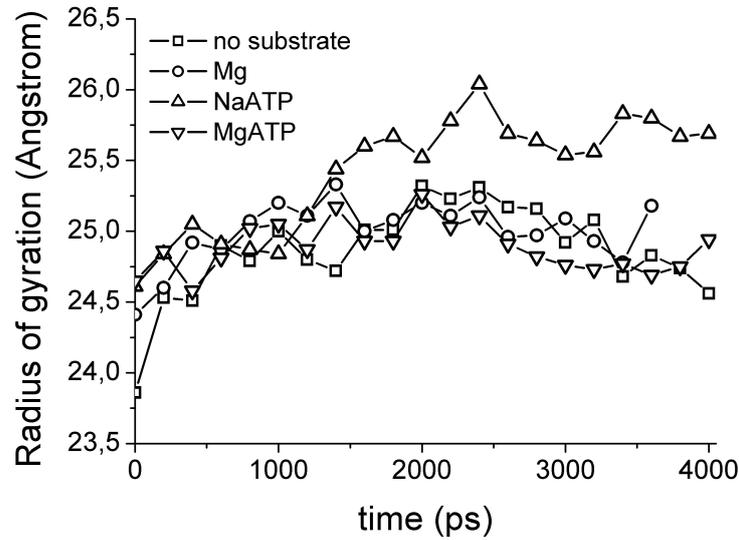
Figure 2: Overall C45 shape in the MD simulation was evaluated using the radius of gyration. Although this parameter was evaluated every 1 ps of the MD simulation, only the values after every 200 ps are displayed in this graph for clarity. The C45 conformation in the Na_2ATP presence clearly differed from the conformations seen in all the other simulations.
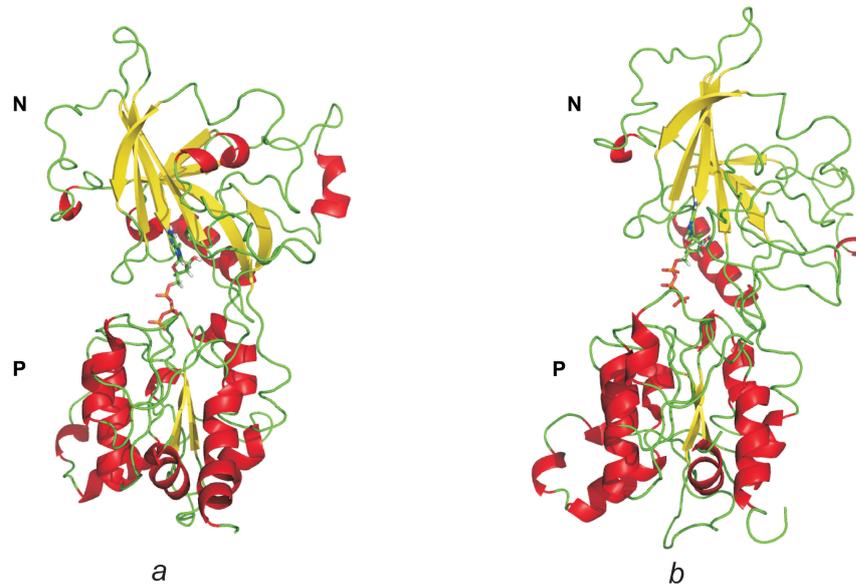
Figure 3: Two snapshots taken from MD simulation of C45-Na_2ATP system at the beginning and at the end. On the left-hand side there is the equilibrated system of C45-Na_2ATP (700 ps). The angle-m-m-m between N- and P-domains is equal to 117.5°. The figure on the right-hand side shows the snapshot taken from 2250 ps of the corresponding MD simulation. The inter-domain angle is equal to 133.4° and documents opening of both domains.

Two Mg^{2+} counterions are coordinated to α-, β- and γ-phosphates of ATP in the C45-MgATP system, similarly as revealed in SERCA crystallographic studies [16]. One Mg^{2+} ion is housed

among D369, V609 backbone oxygen and the ATP γ-phosphate moiety. The other one Mg^{2+} ion resides between ATP α- and β-phosphate group oxygens. Both Mg^{2+} counterions remain in their positions and maintain all abovementioned interactions during the entire MD simulation.

On the other hand, three sodium counterions are tightly bound to ATP phosphate moiety in the C45-Na_2ATP system. In the closed state at the beginning of the simulation, one sodium ion mediates the interaction between D369 and ATP γ-phosphate moiety. As soon as both domains open (interdomain angle increases by $\sim$20°), the interaction between D369 and ATP disrupts. In the open state, two sodium ions mediate the interaction of D612 to ATP β- and γ-phosphates. We assume that the loss of the D369 interaction to ATP starts the domain opening (cf. Fig. 3).

Correlation with experimental data   Our homology model corresponds well to the known experimentally determined structural data. Comparison of the C_α's positions of the NMR-solved N-domain and the corresponding part of our C45 yielded RMSD = 1.7 Å, and comparison of the entire C45 structure to the corresponding part of the SERCA (2c8l) yielded RMSD = 1.6 Å. Both these values are lower than a resolution of any available crystallographic data on P-type ATPases. The overall fold and secondary structure elements of the C45 homology model are consistent with the Na^+/K^+ ATPase crystal structure (at 3.5 Å resolution), which has been published during preparation of this manuscript [18]. The crystal structure represents the C45 in the open conformation while the homology model is in the closed state. This fact prevents a direct comparison of entire structures of both systems. To compare both structures we have superimposed N- and P-domains separately (Suppl. Mat. Fig. S2). There was a good agreement in the overall fold for each domain, larger differences were observed only for the loops that were flexible during our molecular dynamic simulations. The active site of the relaxed homology model (after 2 ns MD simulation of C45-Na) well preserves all main structural features of the active site revealed by the crystal structure (Suppl. Mat. Fig. S3). Only small changes in the side-chains conformations occur for residues Arg544 and Asp443 and can be explained by the flexibility of these residues. Hence, we can conclude that the homology model saves main structural and functional features of the C45 system.

Notably, there is also a good correlation ($p<0.0001$) between spectroscopic and modeling experiments (Fig. 4). The solvent accessibility of a residue in the model could be evaluated by RDF that estimates number of water molecules within the sphere of given radius. The value could be directly compared to the Stern–Volmer quenching constant from the acrylamide-quenching experiments, which is also proportional to the steric angle around the residue accessible from the solvent. Further, the anisotropy decay experiments performed in this study (Table 2) reflected a weak linking of the N- and P-domains in the case of the C45-Na_2ATP system while a tighter one in the case of C45-MgATP system. This fact well coincides with the number of H-bonds between both domains calculated as a mean from the last 2 ns of MD simulations. The numbers of interdomain H-bonds are equal to 4.0 ± 1.9 and 9.5 ± 2.3 for the C45-Na_2ATP and C45-MgATP, respectively, where ATP is considered as an integral part of the N-domain. The introduced terminology "open" and "closed" conformation of the enzyme is attributed the state of the cytoplasmic headpiece. In the open conformation, the N-domain is loosely connected to the other cytoplasmic domains,

while in the closed conformation it is closely connected to the P- and A-domains and displays multiple interdomain hydrogen bonds. In this sense, we use the terminology analogously, despite the A-domain is missing in our case.
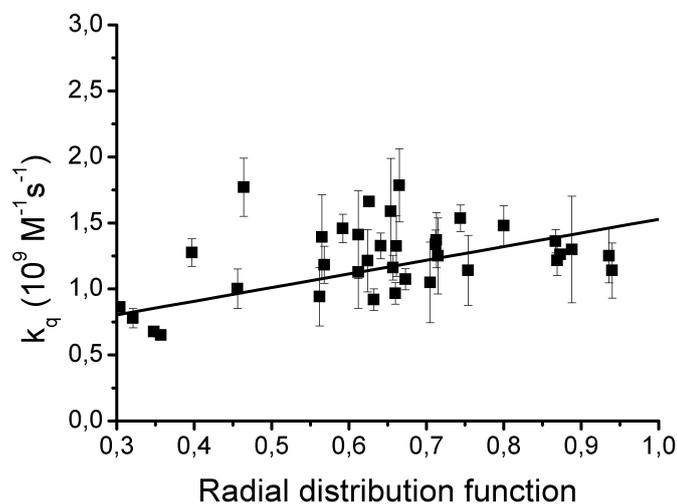

Figure 4: Correlation between the spectroscopic and modeling data. The accessibility of individual residues in the models was evaluated using the radial distribution function for the 1.0 nm radius around the center of mass of the monitored residues. These values could be directly compared to the bimolecular quenching constants from the acrylamide-quenching experiments. See also text for detailed comments.

Discussion

Several crystallographic structures of SERCA in various conformations were published [7–16], and recently, also the first crystal structure of Na^+/K^+-ATPase appeared [18]. There are numerous indications that SERCA is a good representative of all P-type ATPases (or at least Type II ones) and that the displayed conformations are highly relevant also for Na^+/K^+-ATPase [17,35]. Nevertheless, the crystallographic approach has some substantial drawbacks, as thoroughly discussed in [35]. First, it provides only static picture. Moreover, spectroscopic experiments performed under physiological conditions revealed that the protein crystals obtained under very extreme conditions can inevitably influence substantial details in the resolved structure [36]. Finally, all the crystallographic structures displayed enzymes complexed only with molecules that are incompatible with enzyme function, e.g. nucleotide analogs functioning as inhibitors. Thus, in order to observe the protein dynamics under physiological conditions and with native ligands, the conventional spectroscopic techniques and molecular dynamic simulations are the preferred methods of choice [37]. Notably, experimental monitoring of the fine dynamics of the entire transmembrane enzyme encounters serious practical difficulties and presents still a challenge for the future work. Fortunately, the isolated C45 retains the same structural and dynamic features as when being a part of the entire enzyme, and represents a useful system for the studies of the enzyme dynamics, as discussed below.

Our previous results indicated that the isolated C45 contains the low-affinity ATP-binding

site [21–25], and similar results were reported also from other laboratories for the isolated C45 of various P-type superfamily members [19,20,38,39]. Moreover, the NMR studies revealed that the structure of the isolated N-domain in the absence of any substrate fits better to the N-domain of SERCA crystallized in the E2 conformation [26]. This corresponds well also to our observations in the molecular dynamic experiments. The C45 homology model turned out to be stable during subsequent molecular dynamic simulation only when it was built on the basis of the 2C8L (E2 conformation) template, in contrast to the model based on the 1SU4 template (E1 conformation, not shown). The recently published Na^+/K^+-ATPase crystal structure revealed the state with the open conformation of C45. Nevertheless, when we excised the C45 from this structure, it rapidly relaxed into the closed conformation during the C45-Xray MD simulation (data not shown). Thus, we can conclude that isolated C45 adopts in the absence of any substrate the closed conformation characteristic for the E2 state.

It was expected that Mg^{2+} itself could hardly dramatically influence the C45 structure. Indeed, most observed experimental parameters varied only slightly and also molecular dynamic simulation revealed that the C45 remained in the closed conformation with one Mg^{2+} cation coordinated by residues D369 and D710 that were already identified in previous studies [40]. Experimental data revealed changes in the environment of the F571 and the molecular model suggested that it may be a consequence of the coordination of one Mg^{2+} cation by the main-chain oxygens of V569 and P572 (Fig. 5a). It should be noted that this site is rather distal to all known important binding sites on Na^+/K^+-ATPase, and thus, it is not clear whether this second Mg^{2+}-binding site could have any physiological importance.

Also in the case when ATP was bound to the C45 in the presence of magnesium, the changes in the C45 conformation were rather small and it preserved the closed conformation. It could be concluded that the MgATP stabilized the closed conformation of the C45. Notably, there were two magnesium cations coordinated by the nucleotide triphosphate chain, similarly as revealed in the crystal structures of SERCA [16]. One of the cations was coordinated by the oxygens of the α- and β-phosphates, the other one was positioned between the γ-phosphate and phosphorylation site at D369, with the assistance of the main-chain oxygen of V609 (Fig. 5b).

The most dramatic changes in all measured parameters were observed in both the spectroscopic experiments and molecular dynamic simulations when ATP was bound to the C45 in the absence of magnesium (Na_2ATP). The fluorescence anisotropy decay experiments revealed that the contact between N- and P-domains was weakened after Na_2ATP binding, which is in good agreement with our molecular dynamic simulations, where we observed rapid opening of the C45, resembling the open conformation observed for the SERCA crystal in the E1-2Ca conformation [7]. Obviously, the huge global conformational change was reflected also by the local changes within various C45 parts. This was reflected also in our spectroscopic experiments, where numerous residues on both the N- and P-domains exhibited changes in its accessibility (see Results).

Usability and limits of the isolated C45

The obvious question arises, how much are the experiments with the isolated C45 relevant also for the entire enzyme. As discussed above, the isolated C45 seems to support its tertiary structure, and

a

b
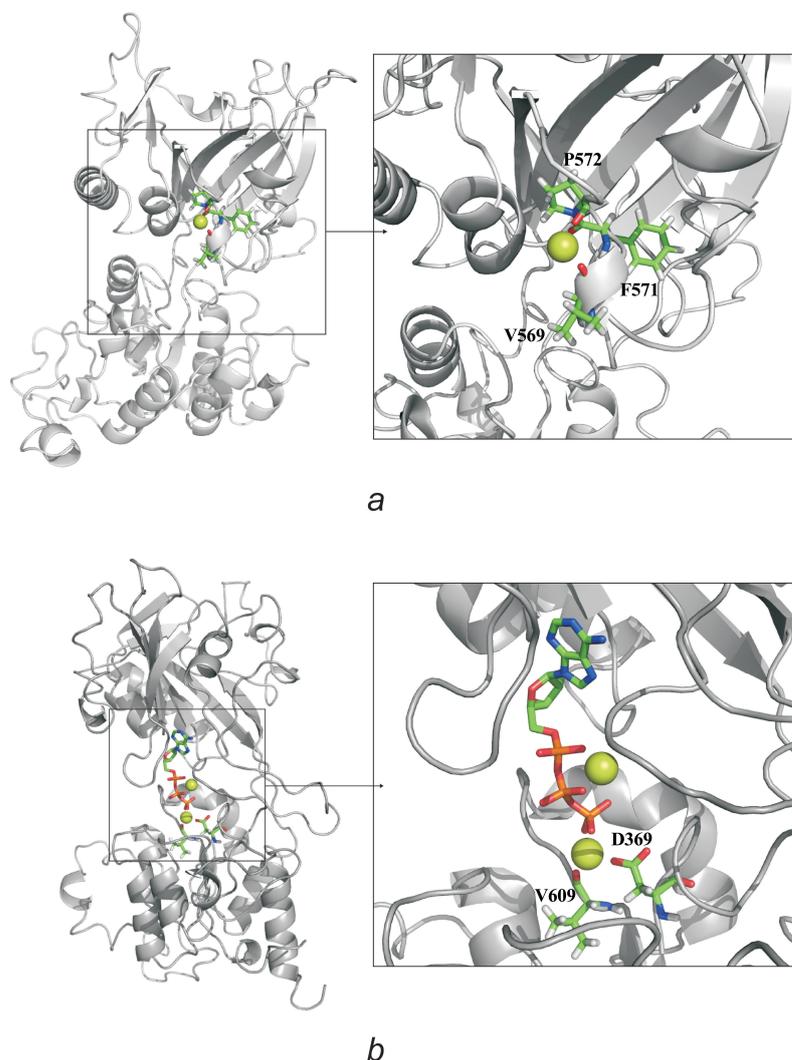
Figure 5: Mg^{2+}-binding sites. (a) Coordination of the Mg^{2+} in the proximity of Phe571. (b) Coordination of the Mg^{2+} in the proximity of the triphosphate chain. ATP and residues coordinating Mg^{2+} cations are depicted in stick representation, magnesium cations are depicted as green spheres. In the left-hand side pictures, N-domain is on the top.

our data suggest that the C45 can adopt the open and close conformations, i.e. dynamic behavior expected for this loop when being part of the entire enzyme. Finally, it is necessary to analyze some logical consequences to the Albers–Post cycle, which is usually used for description of the P-type ATPases working mechanism. It postulates that the enzyme adopts two major conformations during the catalytic cycle. One of them, traditionally designated as E1, has high affinity to both ATP and sodium and low affinity to potassium ions. Contrary, the E2 conformation has low affinity to both ATP and sodium and high affinity to potassium ions. Already early experiments of Skou revealed that magnesium must be present in the cytoplasm as an essential cofactor that is not transported to the other side of the membrane. The fact that recent experiments located the Mg^{2+}-binding site in the proximity of the phosphorylation site lead to the consideration of

MgATP as the only true P-type ATPases ligand [41,42] and it is generally agreed that magnesium is essential for the phosphorylation of the conserved aspartyl residue (D369 on Na^+/K^+-ATPase). Additionally, it is known that ATP can bind also to the enzyme in the E2 conformation, although with lower affinity ($K_M \sim 200$ μM).

As discussed above, the isolated C45 without any substrate adopts the closed conformation that was attributed to the E2 state. Our experiments and molecular dynamic simulation revealed that binding of Mg^{2+} or MgATP does not substantially alter the C45 conformation. In contrast, binding of the Na_2ATP resulted in the rapid mutual movement of both N- and P-domains yielding the open conformation, which is characteristic for the E1 state. Moreover, subsequent addition of Mg^{2+} induced the C45 closure again (Fig. 6). However, it should be noted that this conformation (after subsequent addition of ATP and Mg) turned out to be unstable in molecular dynamic simulation on longer time scales, and we observed some kind of oscillations. Comparison with the closed conformation where the Mg and ATP were added at once (and which was very stable on the long time scales) revealed that the difference consists in $\sim 20°$ mutual rotation of the N- and P-domains with respect to the axis perpendicular to the membrane (Fig. 7) and the γ-phosphate cannot properly reach the phosphorylation site, which is in agreement with previous findings that the isolated C45 cannot efficiently hydrolyze the ATP [43]. Apparently, here we encountered the limits of the isolated C45 system. Based on the crystal structures of SERCA, where the A-domain was in the contact with both the N- and P-domains, we suggest that the flanking stroke of the A-domain into C45 is necessary to correct the mutual orientation of the N- and P-domains. Unfortunately, this could not have been observed in our studies with the isolated C45.

Our experiments suggested that the state of the cytoplasmic headpiece may not be related to the binding of the transported cations to the transmembrane domain. It correlates with the observation that the cytoplasmic headpiece in the crystal structures of SERCA obtained in the presence of Ca^{2+} ions can be found in the open (PDB ID: 1SU4) or closed (PDB ID: 1T5S) conformation. Our experiments suggest that C45 conformation depends on the presence or absence of the nucleotide and magnesium; the open conformation is found only for Na_2ATP bound to the C45, in all the other cases (i.e. Mg^{2+}, Mg + ATP or an empty loop) the C45 is closed.

Based on these observations, it seems reasonable to describe the conformation of the cytoplasmic- and transmembrane domains within the Albers–Post cycle independently, and instead of traditional E1 and E2 to use $^{o/c}E_{Na/K}$, where the superscript "o" or "c" denotes the open or closed conformation of the cytoplasmic headpiece, and the subscript "Na" or "K" reflects the high affinity of the sodium or potassium ions to the transmembrane domain (Fig. 8). In contrast to the previously published models, binding of ATP and Mg^{2+} occurs at different steps of the catalytic cycle. ATP binding induces the $^cE_K \rightarrow {}^oE_K$ conformational change. This conformational change is transmitted to the transmembrane domain yielding the $^oE_K \rightarrow {}^oE_{Na}$ and the cations can be exchanged on the cytoplasmic side. Only then, binding of the Mg^{2+} causes the $^oE_{Na} \rightarrow {}^cE_{Na}$ transition, enabling the ATP hydrolysis and the enzyme phosphorylation (notably, the ADP bound form is probably only a transient state). This is the impulse for the $^cE_{Na} \rightarrow {}^cE_K$ transition and the cation exchange on the extracellular side can take place. Potassium binding induces phosphate and Mg^{2+} release and the enzyme is ready to bind another ATP molecule – the cycle is
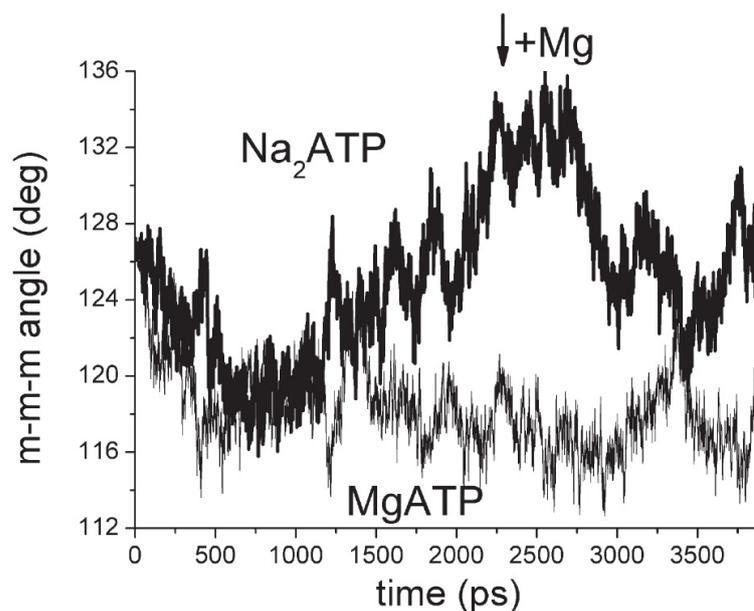
Figure 6: Opening and closing of the C45. The angle between centers of mass of the N- and P-domain and the hinge between these two domains was monitored during the MD simulations. Addition of the pure ATP the closed C45 conformation induced rapid repulsion of the domains, while subsequent addition of magnesium enabled closure of the domains. Notably, addition of the MgATP left the loop in the closed conformation.
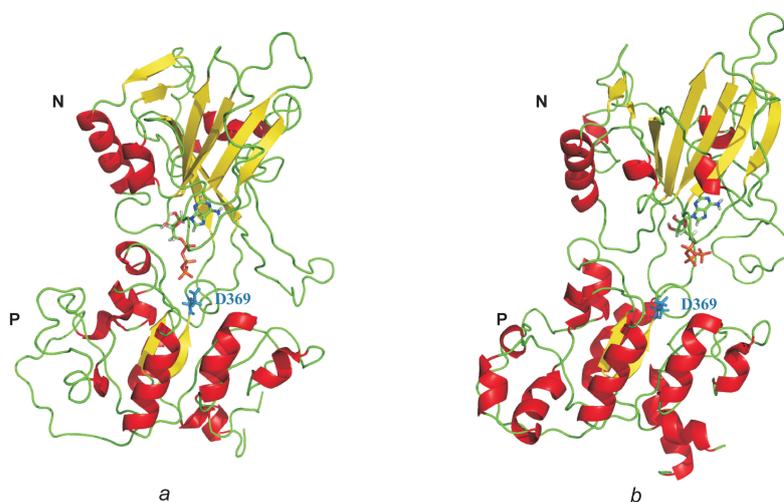
Figure 7: Comparison of the C45 structures when the MgATP was bound at once and when ATP and Mg^{2+} were bound subsequently (last frames from the experiments described in Fig. 6).

completed. Logically, the alternative conformational changes in the cytoplasmic headpiece (CH) and transmembrane domain (TD), i.e. CH $\rightarrow$ TD $\rightarrow$ CH $\rightarrow$ TD $\rightarrow$ etc., seem to be an essential feature that supports the vectoriality of the cation transport.

This scheme fits well to the generally known fact that Mg^{2+} is essential for the enzyme phosphorylation. Moreover, it seems to provide the structural explanation for the results of kinetic

OUT
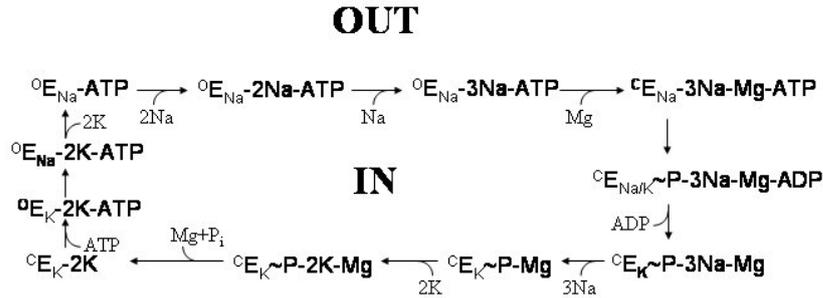
Figure 8: Proposed modification of the Albers–Post model based on our experiments. See text for detailed comments.

experiments performed by Plesner and Plesner [44] on the entire enzyme. They demonstrated that the cytoplasmic ligands can bind subsequently to the enzyme also in the dissociated form, ATP being bound to the enzyme as the first ligand, and Mg^{2+} as the second one. It should be stressed that the estimated affinity of the ATP to the so-called low-affinity binding site ($\sim 200\ \mu M$) should be fully sufficient for the proper enzyme work under physiological cytoplasmic free ATP concentrations ($\sim 600\ \mu M$, assuming 3 mM total ATP concentration, and considering the Mg + ATP $\leftrightarrow$ MgATP equilibrium; under cytoplasmic ionic strength, approx. 20% of the nucleotide should be in the free ATP form). In fact, such an affinity enables much finer regulation than eventual modulation of the high-affinity binding site ($K_M \sim 0.2\ \mu M$). With subtle differences, similar hypothesis was recently proposed by Jensen et al. [13] based on the experiments with SERCA. They postulated that ATP binds in fact to the low-affinity (or modulatory, in Jensen's terminology) site, which is later converted to the high-affinity (or catalytic, in Jensen's terminology) site. Finally, the proposed scheme is also in agreement with experiments, where the conformational change was observed as a consequence of sodium/potassium exchange [45]. They can be attributed to the $^oE_K \rightarrow {}^oE_{Na}$ or $^cE_{Na} \rightarrow {}^cE_K$ transitions in our scheme.

Conclusion

Series of successful crystallographic experiments on SERCA provided good structural basis for understanding of the P-type ATPases working mechanism. Our experiments examining the dynamics of the enzyme complement these data. The work with the isolated C45 enabled efficient separation of the events occurring within the transmembrane- and cytoplasmic parts of the Na^+/K^+-ATPase during the pumping cycle. Combination of the intrinsic tryptophan fluorescence experiments and molecular dynamics simulation suggested fine coordination of the cytoplasmic ligands (ATP and Mg^{2+}) binding in order to drive the opening and closure of the cytoplasmic domains. Based on our results and in agreement with numerous previously published experiments performed with the entire enzyme, we propose that free ATP is bound to the low-affinity binding site inducing opening of the cytoplasmic headpiece (traditionally denoted as E2 $\rightarrow$ E1 transition). Magnesiumis bound later and induces closure of the cytoplasmic headpiece, thus (likely with the A-domain assistance) enabling autophosphorylation (traditionally denoted as E1 $\rightarrow$ E2 transition). However, we cannot

exclude that the catalytic cycle is more complex, and our mechanism of sequential binding of ATP and Mg^{2+} represents only part of this scheme and that MgATP can be bound in some other part of the catalytic cycle. This possibility could be very attractive namely in models assuming oligomeric cooperation of two or more subunits [46–48]. Obviously, although our conclusions fit well to the recent state of knowledge about the pumping mechanism, it will be necessary to prove our proposals also in the experiments with the entire enzyme, which still presents a challenging task.

Acknowledgements

Supplementary data

Supplementary data associated with this article can be found, in the online version, at `doi: 10.1016/j.bbamem.2009.02.004`.

References

1. K.B. Axelsen, M.G. Palmgren, Evolution of substrate specificities in the P-type ATPase superfamily, J. Mol. Evol. 46 (1998) 84–101.

2. R.W. Albers, Biochemical aspects of active transport, Annu. Rev. Biochem. 36 (1967) 727–756.

3. R.L. Post, S. Kume, C. Hegyvary, Activation by adenosine-triphosphate in phosphorylation kinetics of sodium and potassium ion transport adenosinetriphosphatase, J. Biol. Chem. 247 (1972) 6530–6540.

4. J.C. Skou, Further investigations on a Mg++ + Na+-activated adenosintriphosphatase, possibly related to the active, linked transport of Na+ and K+ across the nerve membrane, Biochim. Biophys. Acta 42 (1960) 6–23.

5. J.V. Moller, B. Juul, M. leMaire, Structural organization, ion transport, and energy transduction of P-type ATPases, Biochim. Biophys. Acta-Rev. Biomembr. 1286 (1996) 1–51.

6. K. Geering, The functional role of beta subunits in oligomeric P-type ATPases, J. Bioenerg. Biomembranes 33 (2001) 425–438.

7. C. Toyoshima, M. Nakasako, H. Nomura, H. Ogawa, Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 angstrom resolution, Nature 405 (2000) 647–655.

8. M. Takahashi, Y. Kondou, C. Toyoshima, Interdomain communication in calcium pump as revealed in the crystal structures with transmembrane inhibitors, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 5800–5805.

9. K. Obara, N. Miyashita, C. Xu, L. Toyoshima, Y. Sugita, G. Inesi, C. Toyoshima, Structural role of countertransport revealed in Ca2+ pump crystal structure in the absence of Ca2+, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 14489–14496.

10. C. Toyoshima, T. Mizutani, Crystal structure of the calcium pump with a bound ATP analogue, Nature 430 (2004) 529–535.

11. C. Toyoshima, H. Nomura, Structural changes in the calcium pump accompanying the dissociation of calcium, Nature 418 (2002) 605–611.

12. T.L.M. Sorensen, C. Olesen, A.M.L. Jensen, J.V. Moller, P. Nissen, Crystals of sarcoplasmic reticulum Ca2+-ATPase, J. Biotechnol. 124 (2006) 704–716.

13. A.M.L. Jensen, T.L.M. Sorensen, C. Olesen, J.V. Moller, P. Nissen, Modulatory and catalytic modes of ATP binding by the calcium pump, EMBO J. 25 (2006) 2305–2314.

14. J.V. Moller, C. Olesen, A.M.L. Jensen, P. Nissen, The structural basis for coupling of Ca2+ transport to ATP hydrolysis by the sarcoplasmic reticulum Ca2+-ATPase, J. Bioenerg. Biomembranes 37 (2005) 359–364.

15. C. Olesen, T.L.M. Sorensen, R.C. Nielsen, J.V. Moller, P. Nissen, Dephosphorylation of the calcium pump coupled to counterion occlusion, Science 306 (2004) 2251–2255.

16. T.L.M. Sorensen, J.V. Moller, P. Nissen, Phosphoryl transfer and calcium ion occlusion in the calcium pump, Science 304 (2004) 1672–1675.

17. K.J. Sweadner, C. Donnet, Structural similarities of Na,K-ATPase and SERCA, the Ca2+-ATPase of the sarcoplasmic reticulum, Biochem. J. 356 (2001) 685–704.

18. J.P. Morth, B.P. Pedersen, M.S. Toustrup-Jensen, T.L.M. Sorensen, J. Petersen, J.P. Andersen, B. Vilsen, P. Nissen, Crystal structure of the sodium–potassium pump, Nature 450 (2007) 1043–U1046.

19. C.M. Tran, R.A. Farley, Catalytic activity of an isolated domain of Na,K-ATPase expressed in Escherichia coli, Biophys. J. 77 (1999) 258–266.

20. C. Gatto, A.X. Wang, J.H. Kaplan, The M4M5 cytoplasmic loop of the Na,K-ATPase, overexpressed in Escherichia coli, binds nucleoside triphosphates with the same selectivity as the intact native protein, J. Biol. Chem. 273 (1998) 10578–10585.

21. M. Kubala, K. Hofbauerova, R. Ettrich, V. Kopecky, R. Krumscheid, J. Plasek, J. Teisinger, W. Schoner, E. Amler, Phe(475) and Glu(446) but not Ser(445) participate in ATP-binding to the alpha-subunit of Na+/K+-ATPase, Biochem. Biophys. Res. Commun. 297 (2002) 154–159.

22. M. Kubala, J. Teisinger, R. Ettrich, K. Hofbauerova, V. Kopecky, V. Baumruk, R. Krumscheid, J. Plasek, W. Schoner, E. Amler, Eight amino acids form the ATP recognition site of Na+/K+-ATPase, Biochemistry 42 (2003) 6446–6452.

23. Z. Lansky, M. Kubala, R. Ettrich, M. Kuty, J. Plasek, J. Teisinger, W. Schoner, E.E. Amler, The hydrogen bonds between Arg(423) and Glu(472) and other key residues, Asp(443), Ser(477), and Pro(489), are responsible for the formation and a different positioning of TNP-ATP and ATP within the nucleotide-binding site of Na+/K+-ATPase, Biochemistry 43 (2004) 8303–8311.

24. M. Kubala, J. Plasek, E. Amler, Limitations in linearized analyses of binding equilibria: binding of TNP-ATP to the H-4-H-5 loop of Na/K-ATPase, Eur. Biophys. J. Biophys. Lett. 32 (2003) 363–369.

25. M. Kubala, J. Plasek, E. Amler, Fluorescence competition assay for the assessment of ATP binding to an isolated domain of Na+,K+-ATPase, Physiol. Res. 53 (2004) 109–113.

26. M. Hilge, G. Siegal, G.W. Vuister, P. Guntert, S.M. Gloor, J.P. Abrahams, ATP-induced conformational changes of the nucleotide-binding domain of Na,K-ATPase, Nat. Struct. Biol. 10 (2003) 468–474.

27. T. Obsil, F. Merola, A. Lewit-Bentley, E. Amler, The isolated H-4–H-5 cytoplasmic loop of Na,K-ATPase overexpressed in Escherichia coli retains its ability to bind ATP, FEBS Lett. 426 (1998) 297–300.

28. M.M. Bradford, Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.

29. C.C.G. Inc., Molecular Operating Environment (MOE), Canada, 2006.

30. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.

31. J.M. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21 (2000) 1049–1074.

32. D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules, Comput. Phys. Commun. 91 (1995) 1–41.

33. J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Kluwer/Plenum, NewYork, 1999, pp. 496–499.

34. H.R. Hinz, T.L. Kirley, Lysine-480 is an essential residue in the putative ATP site of lamb kidney (Na,K)-Atpase – identification of the pyridoxal 5'-diphospho-5'-adenosine and pyridoxal-phosphate reactive residue, J. Biol. Chem. 265 (1990) 10260–10265.

35. M. Kubala, ATP-binding to P-type ATPases as revealed by biochemical, spectroscopic, and crystallographic experiments, Proteins-Struct. Funct. Bioinformatics 64 (2006) 1–12.

36. M. Picard, C. Toyoshima, P. Champeil, The average conformation at micromolar [Ca2+] of Ca2+-ATPase with bound nucleotide differs from that adopted with the transition state analog ADP center dot AlFx or with AMPPCP under crystallization conditions at millimolar [Ca2+], J. Biol. Chem. 280 (2005) 18745–18754.

37. M. Kubala, T. Obsil, V. Obsilova, Z. Lansky, E. Amler, Protein modeling combined with spectroscopic techniques: an attractive quick alternative to obtain structural information, Physiol. Res. 53 (2004) S187–S197.

38. M. Abu-Abed, T.K. Mal, M. Kainosho, D.H. MacLennan, M. Ikura, Characterization of the ATP-binding domain of the sarco(endo)plasmic reticulum Ca2+-ATPase: probing nucleotide binding by multidimensional NMR, Biochemistry 41 (2002) 1156–1164.

39. E. Capieaux, C. Rapin, D. Thines, Y. Dupont, A. Goffeau, Overexpression in Escherichia-coli and purification of an ATP-binding peptide from the yeast plasma-membrane H+-Atpase, J. Biol. Chem. 268 (1993) 21895–21900.

40. G. Patchornik, R. Goldshleger, S.J.D. Karlish, The complex ATP-Fe2+ serves as a specific affinity cleavage reagent in ATP-Mg2+ sites of a Na,K-ATPase: altered ligation of Fe2+ (Mg2+) ions accompanies the E1P →E2P conformational change, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 11954–11959.

41. P.A. Pedersen, J.R. Jorgensen, P.L. Jorgensen, Importance of conserved alpha-subunit segment (709)GDGVND for Mg2+ binding, phosphorylation, and energy transduction in Na,K-ATPase, J. Biol. Chem. 275 (2000) 37588–37595.

42. J.E. Mahaney, D.D. Thomas, J.P. Froehlich, The time-dependent distribution of phosphorylated intermediates in native sarcoplasmic reticulum Ca2+-ATPase from skeletal muscle is not compatible with a linear kinetic model, Biochemistry 43 (2004) 4400–4416.

43. R. Krumscheid, K. Suankova, R. Ettrich, J. Teisinger, E. Amler, W. Schoner, Localization of catalytic active sites in the large cytoplasmic domain of Na+/K+-ATPase, Na,K-Atpase and Related Cation Pumps, vol. 986, New York Acad Sciences, New York, 2003, pp. 242–244.

44. L. Plesner, I.W. Plesner, The steady-state kinetic mechanism of ATP hydrolysis catalyzed by membrane-bound (Na+ + K+)-ATPase from ox brain. 1. Substrate identity, Biochim. Biophys. Acta 643 (1981) 449–462.

45. S.J.D. Karlish, D.W. Yates, Tryptophan fluorescence of (Na+ + K+)-ATPase as a tool for study of enzyme mechanism, Biochim. Biophys. Acta 527 (1978) 115–130.

46. D. Thoenges, E. Amler, T. Eckert, W. Schoner, Tight binding of bulky fluorescent derivatives of adenosine to the low affinity E(2)ATP site leads to inhibition of Na+/K+-ATPase — Analysis of structural requirements of fluorescent ATP derivatives with a Koshland–Nemethy–Filmer model of two interacting ATP sites, J. Biol. Chem. 274 (1999) 1971–1978.

47. K. Tanoue, S. Kaya, Y. Hayashi, K. Abe, T. Imagawa, K. Taniguchi, K. Sakaguchil, New evidence for ATP binding induced catalytic subunit interactions in pig kidney Na/K-ATPase, J. Biochem. 140 (2006) 599–607.

48. R.J. Clarke, D.J. Kane, Two gears of pumping by the sodium pump, Biophys. J. 93 (2007) 4187–4196.

10.4   Appendix D

Changes in Electrostatic Surface Potential of Na$^+$/K$^+$-ATPase Cytoplasmic Headpiece Induced by Cytoplasmic Ligand(s) Binding

Martin Kubala • Lenka Gryčová • Zdeněk Lánský • **Petr Sklenovský** • Marika Janovská • Michal Otyepka • Jan Teisinger

Abstract

A set of single-tryptophan mutants of the Na$^+$/K$^+$-ATPase isolated, large cytoplasmic loop connecting transmembrane helices M4 and M5 (C45) was prepared to monitor effects of the natural cytoplasmic ligands (i.e., Mg^{2+} and/or ATP) binding. We introduced a novel method for the monitoring of the changes in the electrostatic surface potential (ESP) induced by ligand binding, using the quenching of the intrinsic tryptophan fluorescence by acrylamide or iodide. This approach opens a new way to understanding the interactions within the proteins. Our experiments revealed that the C45 conformation in the presence of the ATP (without magnesium) substantially differed from the conformation in the presence of Mg^{2+} or MgATP or in the absence of any ligand not only in the sense of geometry but also in the sense of the ESP. Notably, the set of ESP-sensitive residues was different from the set of geometry-sensitive residues. Moreover, our data indicate that the effect of the ligand binding is not restricted only to the close environment of the binding site and that the information is in fact transmitted also to the distal parts of the molecule. This property could be important for the communication between the cytoplasmic headpiece and the cation binding sites located within the transmembrane domain.

Introduction

Most biochemical reactions are facilitated by the assistance of enzymes. The secret of the enzymes consists in their ability to adopt various conformations, and thus, to provide catalytic residues for chemical reactions, to open or close the pathways for solutes or to change affinities for ligands (1,2). However, description of the enzyme conformation is usually limited to the description of the geometrical features (3). Although the geometry analysis can provide lot of useful information, usually the information about molecular forces causing the movements or molecular interactions is missing. This study describes how we can monitor local changes in the electrostatic surface potential. As the electrostatic forces play an important role in the molecular interactions, this approach can deepen our understanding to the causes and consequences of the enzyme conformational changes. We describe the changes induced by the ligand binding to the cytoplasmic part of the Na$^+$/K$^+$-ATPase.

The Na^+/K^+-ATPase (sodium pump, EC 3.6.3.9) is the main cytoplasmic sodium regulator within all animal cells. During one catalytic cycle, the pump translocates three sodium ions across the plasma membrane out of the cell, and two potassium ions in the reverse direction; the transport is energized by the ATP hydrolysis (4). The steep gradient of sodium ions concentration created by the Na^+/K^+-ATPase substantially contributes to the plasma membrane potential and could be used for the secondary transport of other solutes. Hence, it is not surprising that inhibition or malfunction of Na^+/K^+-ATPase can result in variety of serious diseases, e.g., ischemia, diabetes, etc. (5,6).

Na^+/K^+-ATPase is a member of the superfamily of cation translocating membrane enzymes, designated as P-type ATPases (7). This designation has origin in the finding that the enzyme is transiently autophosphorylated on the conserved aspartyl residue during the catalytic cycle. Another common feature of these pumps is that they need the Mg^{2+} ions as a nontransported cofactor to their proper function (8). The first high-resolution structural information about the P-type ATPases was obtained in 2000 in the crystallographic studies of the Ca^{2+}-ATPase from sarco(endo)plasmic reticulum (SERCA) (9). The SERCA was later crystallized under various conditions revealing the different conformations of this enzyme (10–16). A crystallographic structure of Na^+/K^+-ATPase at 3.5 Å resolution has been published recently (17). It confirms the close relationship of these two enzymes.

Two cation-binding sites (common to both sodium and potassium) were localized in the transmembrane domain. They are formed by the residues of the transmembrane helices M4, M5, M6, and M8 (17). The third binding site for the sodium ion has not been identified so far, but recent data revealed that truncation of the C-terminus greatly reduced the affinity to Na^+ (17). Three large domains could be identified on the cytoplasmic side of the membrane. Domain A (actuator) is formed by the N-terminus and the loop between the second and third transmembrane helices (C23). The other two domains are formed by C45. The nucleotide-binding site was localized on the N-domain, whereas the conserved phosphorylation site Asp369 is localized on the central P-domain. The first crystal structure of SERCA showed that the three cytoplasmic domains were well separated each from other, and this structure was assigned to the conformation with high affinity to ATP (9). Later, crystallographic structures (including those of Na^+/ K^+-ATPase) obtained in the presence of various ATP or phosphate analogs, showed several slightly different conformations, where these three domains interacted in the closed complex (10,11,13–15,17). Analysis of these structures revealed that N-domain can rotate by $\sim 90°$ around the hinge connecting the N- and P-domains, thus, mutually approaching the nucleotide binding and phosphorylation sites. The A-domain displays two kinds of motion during various conformational transitions. Combining the $\sim 110°$ rotation around the axis perpendicular to the membrane or $\sim 30°$ rotation around axis parallel with membrane, various A-domain residues can participate to the interactions on the cytoplasmic domains interfaces. Previous kinetic studies suggested that the enzyme conformational change could be the rate limiting step of the catalytic cycle (18).

Cytoplasmic C45 containing the N- and P-domains constitutes $\sim 40\%$ of the α-subunit mass, and experiments showed that it could be isolated from the rest of the enzyme (19–21) retaining its structure, as well as dynamic and some of its functional (e.g., TNP-ATP- or low-affinity

ATP-binding) properties (22–25). Recently, we have described the conformational changes of C45 induced by the cytoplasmic ligand(s) binding (from the point of view of geometry), using the intrinsic tryptophan fluorescence quenching by acrylamide, fluorescence anisotropy decay and molecular dynamic (MD) simulations. We showed that in the absence of any ligand, in the presence of Mg^{2+} alone or MgATP, the C45 adopted closed conformation, whereas addition of the ATP in the absence of Mg^{2+} induced the C45 opening (26), i.e., dynamic behavior that is expected for the corresponding part of the entire enzyme. In this study, modification of the quenching experiment (quenching by iodide) enabled us to analyze the conformational changes from the point of view of the electrostatic surface potential. Our data revealed that the set of the sensitive residues was different, and thus, this kind of information can be considered as complementary to the traditional geometrical approach. As the electrostatic forces play an important role in the inter- or intramolecular interactions (27,28), our results indicate that changes in the electrostatic surface potential could mediate the propagation of information about ATP binding to the distal parts of the molecule.

Materials and Methods

Site-directed mutagenesis and protein expression

The native sequence of the C45 from the mouse brain Na^+/K^+-ATPase contains two Trp residues on the position 385 and 411. The cloning of the isolated C45, subsequent replacement of the two native tryptophans by phenylalanines (WT-W385F, WT-W411F, and WT-W385FW411F = TL; tryptophanless), insertion of the mutations TL-F404W, TL-F426W, TL-F571W, TL-I627W, TL-F683W, and TL-S732W into the tryptophanless construct to yield the set of single tryptophan mutants (Fig. 1), the expression and purification of the $(His)_6$–tagged proteins was described in detail previously (26). Additionally, the mutation TL-V648W was carried out using the upstream primer (altered nucleotides are in italic): TL-V648W 5'ggcatgtctgggggttccac*tg*gttcactgg3'; the downstream primer was reverse complement.

Tryptophan fluorescence quenching

Steady-state tryptophan fluorescence emission was quenched using acrylamide or KI as a quencher under conditions described previously (26). The excitation wavelength was set to 295 nm to avoid excitation of Tyr residues and the inner filter effect caused by ATP or iodide (in the cuvette with the optical path of 3 mm, both in excitation and emission, the absorbance of 15 mM or 250 mM KI is < 0.01, and the inner filter effect is therefore negligible). Emission was set to 350 nm because i), the water-exposed Trp residues have typically emission spectra with maxima at wavelengths > 340 nm; and ii), we aimed to avoid contamination of the signal by the Raman scattering of water, which appears near 330 nm. Quenching experiments were carried out in four different setups, i.e., without any ligand, with 15 mM $MgCl_2$ (Lach-Ner, Brno, Czech Republic), 15 mM Na_2ATP (Sigma, St. Louis, MO), or 15 mM MgATP (Sigma). The experiments in the absence of magnesium (i.e., free C45, or in the presence of Na_2ATP) were carried out in the presence of 5 mM EDTA.

The efficiency of quenching was evaluated by the nonlinear least-squares analysis using the
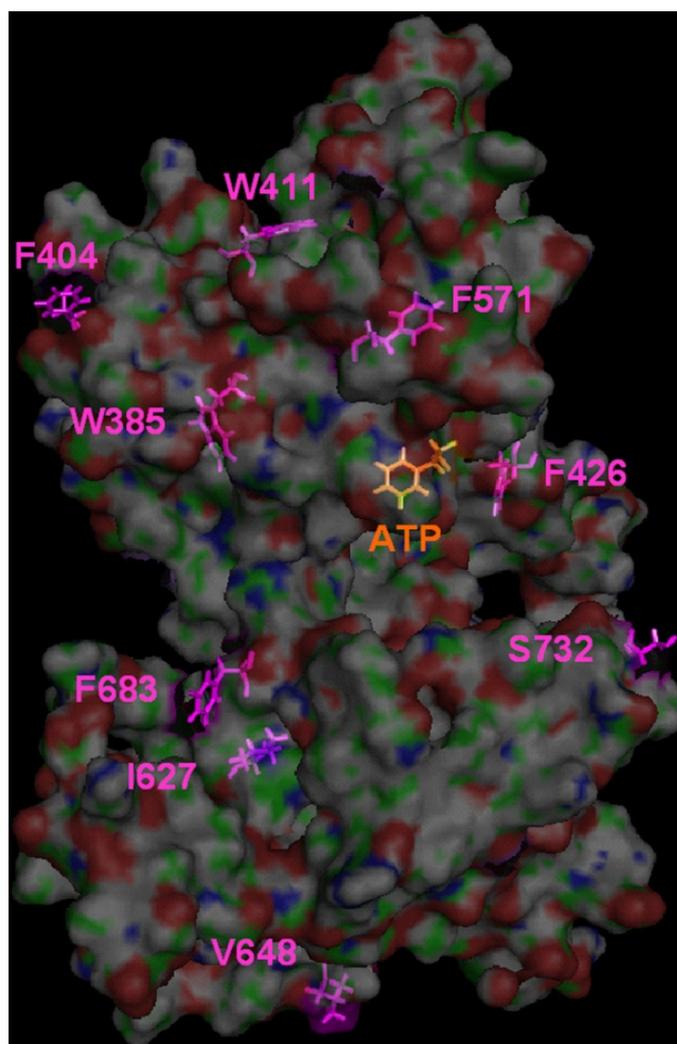
Figure 1: Locations of the Trp residues within our set of single-tryptophan mutants. The ATP binding site is represented by the residuum F475.

Stern-Volmer formula (29):

$$F = \frac{F_0}{1 + K_{SV}\,[\mathrm{Q}]},$$

where F_0 and F are the fluorescence intensities in the absence or in the presence of the quencher, respectively, K_{SV} is the Stern-Volmer quenching constant and [Q] is the concentration of the quencher. Both the F_0 and K_{SV} were the fitted parameters, and the mean $\pm$ SD from three to five independent measurements was calculated. The bimolecular quenching constant was calculated as ratio of the K_{SV} and the mean fluorescence lifetime.

The Stern-Volmer quenching constants obtained for the quenching by acrylamide or iodide can substantially differ. In addition to the fluorophor accessibility and fluorescence lifetime (that are the same in both cases), the quenching efficiency is influenced also by the diffusion coefficients of quenchers (diffusion coefficient of Trp on a bulky protein is negligibly small), and in the case of the

negatively charged iodide (in contrast to the neutral acrylamide), also by electrostatic repulsion or attraction between the quencher and protein. The information about the electrostatic surface potential in the environment of the quenched tryptophan can be extracted by defining the charge parameter (CP) as:

$$\text{CP} = \frac{K_{\text{SV}}(\text{iodide})}{K_{\text{SV}}(\text{acrylamide})}.$$

The higher is this value, the more is the iodide attracted to the tryptophan by positively charged residues in its environment, and vice versa. Contribution of all the other factors vanish as they are common to both experiments, only the ratio of diffusion coefficients $D(\text{iodide})/D(\text{acrylamide})$ remains as a multiplicative parameter, which is constant within the whole data set, and thus, does not affect the evaluation of the relative changes.

Note, that these assumptions hold true for the tryptophans located on surface of the protein. However, it was shown that the acrylamide is able to diffuse into the protein interior, in contrast to iodide, which is hydrated, and its ability to penetrate under the surface is very limited (30). Therefore, this approach should be used carefully, namely in the case of tryptophanyl residues buried under the protein surface. Fortunately, according to our model (26), all the examined tryptophans are located on the surface of the protein (except for native W385, which is partly buried), and therefore similarly accessible for both the acrylamide and iodide.

The SD of the CP (ΔCP) was calculated according to error propagation theory (31) as

$$\Delta\text{CP} = \sqrt{\frac{1}{K_{\text{AA}}^2}\Delta K_{\text{KI}}^2 + \frac{K_{\text{KI}}^2}{K_{\text{AA}}^4}\Delta K_{\text{AA}}^2},$$

where K_{KI} and K_{AA} represent the Stern-Volmer quenching constants for the iodide and acrylamide quenching, respectively, and ΔK_{KI} and ΔK_{AA} are the corresponding standard deviations. The significance of the differences in CP was evaluated using Student's t-test (MICROSOFT EXCEL utility).

Fluorescence decay measurement

Fluorescence decays of the mutant TL-V648W were measured using the time-correlated single photon counting (TCSPC) method on the Pico- Harp300 instrument (PicoQuant, Berlin, Germany). The source of excitation pulses was pulsed LED centered at 298 nm (PLS300; PicoQuant) operating at 10 MHz frequency. The emission was sampled under magic-angle conditions through the monochromator set to 350 nm. The data were collected into the histogram spanning the 100 ns time-range with the 32 ps/channel resolution. The experiment proceeded until 10,000 counts in the peak channel were sampled. The instrument response function was measured using the colloid silica (Ludox, Sigma) as a scatterer. The least-squares fitting was carried out by the Levenberg-Marquardt method using the software FLUOFIT 4.2.1 (PicoQuant). The model used was the sum of exponentials with instrument response function (IRF) deconvolution:

$$I(t) = IRF \otimes \sum_i \alpha_i \times \exp\left(-\frac{t}{\tau_i}\right),$$

and the intensity-weighted mean fluorescence lifetime was calculated as

$$\tau_M = \frac{\sum \alpha_i \tau_i^2}{\sum \alpha_i \tau_i}.$$

All the experiments were carried out at 20 ℃ (bath-controlled).

Electrostatic potential calculations

The electrostatic potential isocontour maps were calculated for the last snapshots of liganded and unliganded C45 MD simulations (C45-Na, C45-Na$_2$ATP, C45-Mg, and C45-MgATP) reported in Grycova et al. (26) using the program APBS (32) as implemented in the PMV software (33). The dielectric constants of the protein and the solvent were set to 2 and 78.5, respectively. The partial atomic charges for amino acid residues were assigned by AMBER force field of the PMV built-in PDB2PQR converter.

Results

The intrinsic tryptophan fluorescence presents a unique probe for examination of molecular processes on proteins. However, the interpretation of the spectroscopic data is straightforward only for the proteins containing no more than a single tryptophanyl residue. Therefore, we have prepared the set of single-tryptophan mutants, where the Trp residues were placed as reporters on the various locations on the C45 surface. The expression levels of all mutants were high, suggesting that the carried out mutations did not influence the correct protein folding.

Fluorescence decay of the TL-V648W mutant

Fluorescence decays of the TL-V648W mutant were measured under various conditions (i.e., without any ligand, in the presence of Mg^{2+}, MgATP, or Na$_2$ATP). In all cases, the satisfactory fit with the reduced χ^2 approaching 1.00, random distribution of the first residuals and autocorrelation function, was obtained for the 3-exponential fit (Table 1) and the decay scheme was very similar in all cases.

TABLE 1 Estimated parameters for the fluorescence decays and acrylamide quenching of TL-V648W mutant

Ligand	A_1	A_2	A_3	τ_1(ns)	τ_2(ns)	τ_3(ns)	τ_M(ns)	χ_R^2	K_{SV} (M^{-1})	k_q (10^9 M^{-1}s^{-1})
None	44.21	43.08	12.71	0.72	2.75	7.04	4.083	1.027	5.78 ± 1.35	1.42 ± 0.33
Mg	42.81	43.35	13.84	0.74	2.69	6.88	4.074	1.033	5.55 ± 1.07	1.36 ± 0.26
MgATP	45.68	43.48	10.84	0.75	2.87	7.45	4.108	0.966	5.64 ± 0.80	1.37 ± 0.19
Na$_2$ATP	47.76	42.07	10.18	0.97	3.16	7.82	4.203	1.015	5.63 ± 0.97	1.34 ± 0.23

The τ_i represent calculated fluorescence lifetimes, A_i their normalized preexponential factors, τ_M is the intensity-weighted mean fluorescence lifetime, χ_R^2 characterizes the goodness of fit. K_{SV} represents the Stern-Volmer quenching constant for the acrylamide quenching, and k_q the bimolecular quenching constant calculated as K_{SV}/τ_M.

Acrylamide quenching of TL-V648W mutant

Acrylamide quenching can reveal the information about the steric accessibility of the fluorophor. In our previous study (26), we described how the ligand binding changes the protein conformation. In the absence of any ligand, in the presence of Mg^{2+}, MgATP, or Na_2ATP we observed the Stern-Volmer quenching constants (K_{SV}) 5.78 ± 1.35 M^{-1}, 5.55 ± 1.07 M^{-1}, 5.64 ± 0.80 M^{-1}, or 5.63 ± 0.97 M^{-1}, respectively. These constants must be corrected for the possible changes of the excited-state lifetime (i.e., the time, during which the quencher can effectively quench the fluorescence). The bimolecular quenching constant (see Table 1) calculated as the ratio of the K_{SV} and the mean fluorescence lifetime describes already the steric accessibility of the fluorophor. However, for the mutant TL-V648W no significant changes were observed on ligand binding. The acrylamide-quenching experiments for the other mutants were presented in our previous study (26). Briefly, in comparison to the unliganded C45, binding of Mg^2 resulted in decreased accessibility of Trp residue in the TL-F571W mutant only, binding of MgATP resulted in decreased accessibility of tryptophan in TL-F571W and WT-W411F mutants, and the most complex changes were observed on Na_2ATP binding, where we observed decreased accessibility of tryptophan in WT-W411F, TL-F404W, TL-F571W, and TL-F683W mutants (26).

Iodide quenching

In comparison to the acrylamide quenching (26), the experiments using quenching by iodide ions (Table 2) carry also interesting information about the local electrostatic surface potential in the fluorophor microenvironment (Fig. 2). Within our set of mutants, the charge parameter (see Materials and Methods) ranged between values 1.35 and 3.50, showing that in the absence of any ligand, the F404 have the most negatively charged environment and the I627 and V648 are surrounded by the positively charged residues (Table 3).

TABLE 2 Iodide quenching

Protein	No substrate	Mg^{2+}	MgATP	Na_2ATP
WT-W411F	8.44 ± 0.40	9.19 ± 0.94	9.10 ± 0.50	8.44 ± 0.18
TL-F404W	9.96 ± 0.50	11.92 ± 0.75	9.47 ± 0.68	9.36 ± 1.02
WT-W385F	10.77 ± 0.93	12.21 ± 0.70	10.79 ± 0.20	10.36 ± 0.51
TL-F426W	12.78 ± 0.56	13.09 ± 1.00	12.53 ± 0.60	11.23 ± 1.05
TL-F571W	14.43 ± 1.23	12.55 ± 0.63	11.22 ± 0.13	9.65 ± 0.27
TL-I627W	14.27 ± 0.68	14.49 ± 0.73	15.25 ± 0.82	15.93 ± 0.84
TL-V648W	20.22 ± 2.38	18.49 ± 0.41	15.80 ± 1.80	10.42 ± 3.07
TL-F683W	10.32 ± 1.00	8.83 ± 0.34	10.21 ± 2.14	12.18 ± 1.12
TL-S732W	12.99 ± 0.75	11.93 ± 0.78	10.97 ± 0.66	10.84 ± 0.52

Stern-Volmer quenching constants for quenching of the Trp fluorescence by iodide given in M^{-1} are expressed as mean $\pm$ SD from three to five independent titrations. The native sequence of C45 contains two Trp residues on the positions 385 and 411. Thus, WT-W411F or WT-W385F designate mutants where only Trp^{385} or Trp^{411} was present, respectively, the other Trp residue was mutated to Phe. For all the other mutants, both the Trp^{385} and Trp^{411} were mutated to phenylalanines, and the designation of the protein indicates the position of the reporter Trp residue that was inserted into the tryptophanless (TL) mutant.
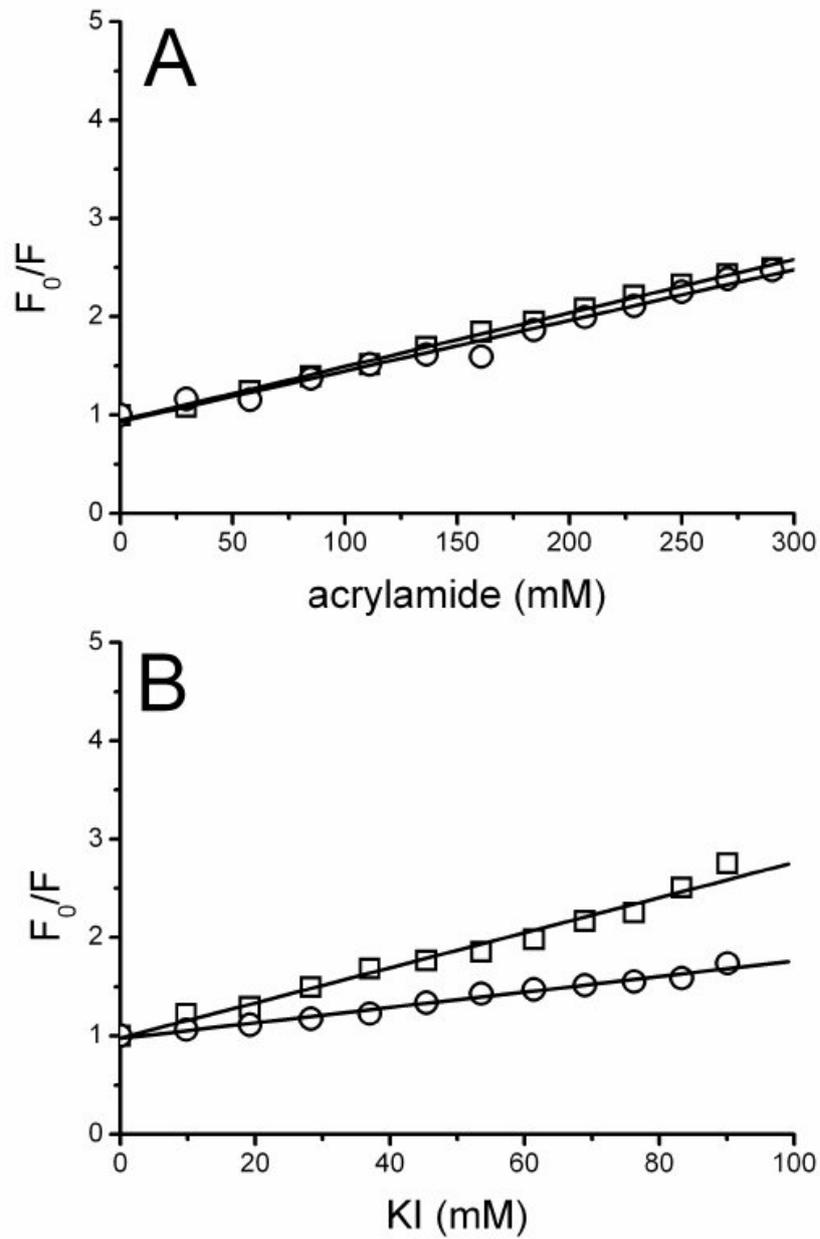
Figure 2: Example of the data from the quenching experiments. Quenching of TL-V648W mutant by (A) acrylamide or (B) iodide in the absence of any ligand (squares) or in the Na$_2$ATP presence (*circles*).

TABLE 3 Charge parameters

Protein	No substrate	Mg^{2+}	MgATP	Na$_2$ATP
WT-W411F	1.98 ± 0.10	2.39 ± 0.30	2.83 ± 0.18	2.55 ± 0.12
TL-F404W	1.35 ± 0.18	2.09 ± 0.29	1.77 ± 0.47	2.29 ± 0.47
WT-W385F	2.04 ± 0.22	2.19 ± 0.37	2.12 ± 0.42	2.02 ± 0.40
TL-F426W	1.77 ± 0.08	1.98 ± 0.19	1.82 ± 0.13	1.72 ± 0.22
TL-F571W	1.80 ± 0.24	2.73 ± 0.36	1.86 ± 0.13	1.74 ± 0.13
TL-I627W	3.35 ± 0.28	3.07 ± 0.24	3.37 ± 0.72	4.13 ± 0.87
TL-V648W	3.50 ± 0.91	3.33 ± 0.64	2.80 ± 0.51	1.85 ± 0.63
TL-F683W	1.80 ± 0.21	1.45 ± 0.28	2.07 ± 0.48	3.07 ± 0.36
TL-S732W	2.40 ± 0.19	2.33 ± 0.16	2.29 ± 0.34	2.28 ± 0.21

Charge parameters ± SD calculated as a ratio of the Stern-Volmer constants for the iodide quenching (Table 2) and for acrylamide quenching (26). Description of the proteins is identical to that used in Table 2.

Mg^{2+} binding

The binding of magnesium caused changes in the charge distribution in the proximity of two residues located on the N-domain. When compared to the experiments in the absence of any substrate, the charge parameter increased from 1.35 ± 0.18 to 2.09 ± 0.29 ($p < 0.05$) or from 1.80 ± 0.24 to 2.73 ± 0.36 ($p < 0.05$) for the TL-F404W and TL-F571W mutants, respectively, suggesting a shift of the charge toward positive values. Contrary, for the residues located on the P-domain, no significant differences were observed.

MgATP binding

Similarly as in acrylamide-quenching experiments described in our previous study (26), binding of MgATP resulted only in little changes in the observed parameters for most of the mutants. Significant shift of the charge parameter from 1.98 ± 0.10 to 2.83 ± 0.18 ($p < 0.01$) was observed only in the experiments with the mutant containing the native W385 (WT-W411F mutant). It shows that the charge of the residue environment is shifted toward positive values. Experiments with the other residues showed no significant changes.

Na$_2$ATP binding

Like the acrylamide quenching described in our previous study (26), the most complex changes were observed in the experiments carried out in the Na$_2$ATP presence. On the N-domain, the charge parameter for the mutants WT-W411F or TL-F404W increased from 1.98 ± 0.10 to 2.55 ± 0.12 ($p < 0.01$) or from 1.35 ± 0.18 to 2.28 ± 0.47 ($p < 0.02$, respectively. For the TL-F571W mutant we observed change in K_{SV} but not in the charge parameter, indicating only steric but not charge changes in the residue environment. For the other mutants (WT-W385F and TL-F426W) remained both K_{SV} and charge parameter essentially unaltered. On the P-domain, the charge in the environment of the residue F683 is shifted toward positive values, as shown by the shift of the charge parameter from 1.80 ± 0.21 to 3.07 ± 0.36 ($p < 0.02$). Contrary, the charge in the V648 environment is shifted toward negative values, as shown by the charge parameter change from

3.50 ± 0.91 to 1.85 ± 0.63 ($p < 0.02$). For the mutants TL-I627W and TL-S732W, both the K_{SV} and charge parameter varied only within the experimental error.

Electrostatic potential maps from the MD simulations

We attempted to correlate the above-mentioned spectroscopic results to the information from the MD simulations reported previously (26) and we calculated the electrostatic potential isocontour maps from the last snapshots of the MD simulations (Fig. 3). We observed a good agreement between the spectroscopic- and MD data for all the N-domain sensitive residues.
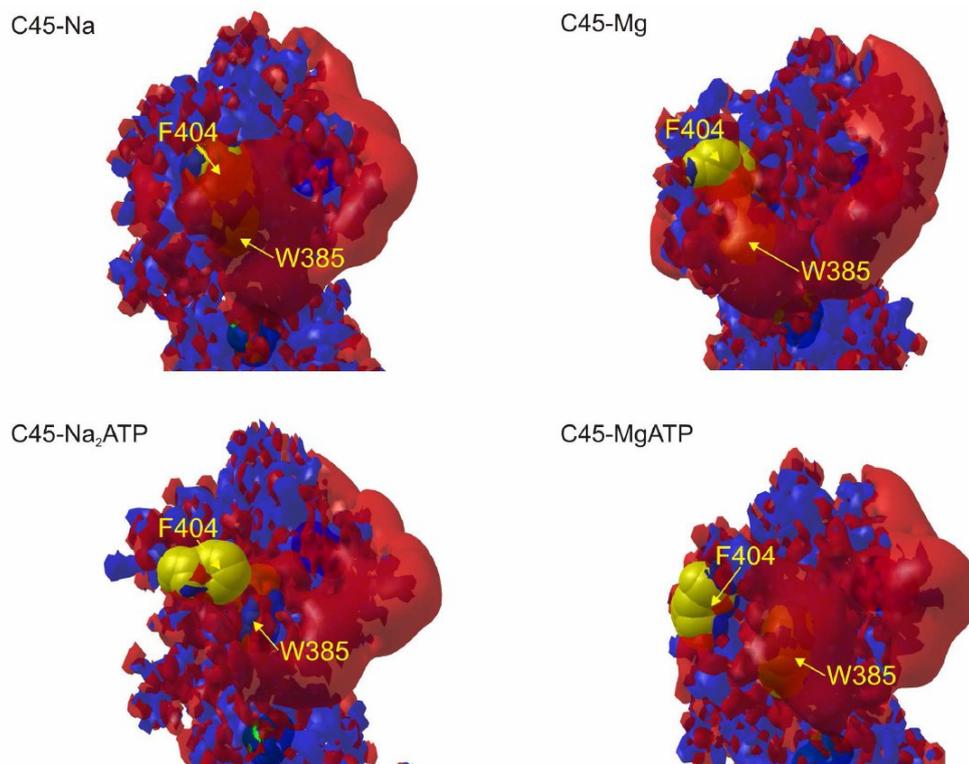
Figure 3: Semitransparent electrostatic potential maps from the final snapshots of the MD simulations of the C45 in the absence of any ligand (C45-Na), in the presence of Mg^{2+} only (C45-Mg), in the presence of ATP only (C45-Na$_2$ATP), and in the simultaneous presence of Mg^{2+} and ATP (C45-MgATP). Contour levels were set to -4.0 kT/e (*red*) and $+4.0$ kT/e (*blue*). The picture was generated using the PMV software (33). Residues W385 and F404 are displayed as yellow balls, the residue F571 as well as the nucleotide binding site are located on the rear side of the molecule.

The N-domain seems to be rather rigid (except for the region 393–410; see the Supporting Material), and we observe the positively charged environment of the nucleotide-binding pocket, and a large negatively charged cloud on the opposite site of the N-domain, which fully covers the residue W385, and F404 and F571 are located on its edge. This cloud is shifted on the Na$_2$ATP or MgATP binding, and the W385 gets to the boarder of the cloud, which corresponds to the experimental observations that the electrostatic surface potential shifted toward positive values in these cases. The residue F404 seems to be displaced from the negatively charged cloud in all

simulations with ligands. However, evaluation of the results for the F404 residue must be done with caution, as it is located within a very flexible region, and (e.g., in the case of MgATP binding) the difference in the experimentally determined charge parameter was not statistically significant on the $p < 0.05$ level. The previously suggested binding of the Mg^{2+} cation in the proximity of F571 (see Fig. 5A in Grycova et al. (26)) could explain the observation of the TL-F571W charge parameter change in the presence of Mg^{2+}.

In contrast, we were not able to reliably link the spectroscopic and MD data for the P-domain. First, numerous P-domain regions became very flexible on Na_2ATP binding (see the Supporting Material) and almost every snapshot offered substantially different isocontour map. Second, as reported previously, the last 13 C-terminal residues were truncated in our MD simulation to reduce computational demands. Although this truncation should not substantially influence the C45 geometry during the simulation, it can substantially influence the electrostatic potential maps. Hence, the comparison of MD and spectroscopic data requires further investigation in this case.

Discussion

Conformational changes of the proteins are frequently described only as changes in the molecule geometry. For the P-type ATPases, we have recently obtained detailed information from the crystallographic experiments (9–15,17). However, crystallography data suffer from the facts that buffer composition may be substantially different from the physiological conditions or that the natural ligands usually cannot be used (34). Therefore, these data should be complemented by the conventional spectroscopy experiments, which do not have these limitations (26,35,36). Furthermore, the geometrical approach usually describes only the initial and final states, however, providing only indirect idea about forces causing the movement. In this study, we describe an experimental approach allowing us to monitor the changes in the electrostatic surface potential, which can support the theoretical approaches reported previously (37,38).

The experiments describing the dynamic of transmembrane proteins present still a challenge due to numerous practical difficulties. Fortunately, the central cytoplasmic segment (C45) of Na^+/K^+-ATPase can be isolated as a soluble protein without the rest of the enzyme, retaining its structural and functional properties (19–25), which greatly facilitates the experimental work. In our previous study (26), we described that the C45 has also the dynamic properties that are expected for this part of the molecule within the entire enzyme. We found that the C45 adopted a closed conformation in the absence of any ligand, in the presence of Mg^{2+} or in the simultaneous presence of Mg^{2+} and ATP. Presence of ATP in the absence of Mg^{2+} induced the C45 opening. Data presented in this study confirmed that the C45 conformations are indeed similar in the absence of any ligand, in the presence of Mg^{2+} and in the presence of MgATP, whereas the conformation in the presence of ATP (without Mg^{2+}) is substantially different. However, the set of sensitive residues is not the same, which was best demonstrated by the mutant TL-V648W that exhibits little changes from the point of view of the geometry, but large changes in the electrostatic surface potential. Other mutants, however (e.g., WT-W385F or TL-F571W), were sensitive to the changes in geometry (26), but showed no significant change in the electrostatic surface potential. Hence, the potential approach can be considered as complementary to the geometry approach. Moreover,

the electrostatic interactions are long-range ones, and we propose that local changes in electrostatic surface potential could be important for the communication between distal parts of the molecule.

Communication between the C45 and other parts of the enzyme

The cation-binding sites are formed by the residues located on the helices M4, M5, M6, and M8 (9,17). Crystallographic experiments with SERCA showed completely different arrangement of the transmembrane helices in the E1 or E2 conformations, respectively. The complex changes were described in detail previously (10), and they concern all 10 of the transmembrane helices. In contrast, interaction partner for the external cofactors (ATP and Mg^{2+}) on the cytoplasmic side is only the C45. Thus, it is interesting to see how the information about the cytoplasmic events can be propagated to the various parts of the transmembrane domain. The influence of the M4 and M5 by the processes occurring on the C45 is easily comprehensible, and it was shown that M4 moves toward cytoplasm in the E2 conformation and M5 inclines on nucleotide binding (10). The N-terminal transmembrane helices (M1–M3) are apparently linked to the C45 through the cytoplasmic A-domain, which is formed by the cytoplasmic N-terminus and C23. The crystalline structures showed that A-domain can form a compact assembly with the N- and P-domains (formed by C45) on nucleotide binding. It seems that the presence of the ATP molecule itself can be a sufficient signal for the A-domain to approach, because the C23- and C45-contact sites are located near the nucleotide-binding- or phosphorylation sites (10). In contrast, it is not straightforward to understand how the information about the nucleotide binding is transmitted to the C-terminal transmembrane helices (M6–M10), because they are located on the side that is reverse to the location of the ATP-binding site on the C45. Obviously, the transmembrane helices are tightly packed and could interact each with other. Nevertheless, numerous authors have already speculated that the short cytoplasmic loop C67 can mediate the information between C45 and C-terminal transmembrane helices (4,39–43). Comparison of our C45 MD simulations (26) with the crystal structure of the complete Na^+/K^+-ATPase suggests that the interaction partner on the C45 could be the flexible Asn^{642}-Asn^{649} loop (Fig. 4 and Fig. S1). Indeed, when our reporter tryptophan was located on the position 648, we observed that the ESP in its environment shifted toward more negative values when the C45 adopted the open conformation (induced by Na_2ATP binding) as compared to the closed conformation. We propose that this change in the local surface charge could serve as a signal for the charged residues on C67. This hypotheses is supported by mutagenesis experiments, where alterations of charged amino acids within C67 resulted in the decrease of Na^+/K^+-ATPase activity (40). However, further mutagenesis experiments are required to reveal details of this interaction.

Further, we have estimated that the surface charge of four residues located on both N- and P-domains shifted toward positive values. Interestingly, they are all located on the side of the C45 that is reverse to the nucleotide-binding site, on the C-terminal transmembrane helices M7-M10 (Fig. 5). Therefore, it is unlikely that this change could serve as a signal in the communication between the C45 and the A-domain. One may speculate that this could be a signal in the entry/exit pathway of the transported cations, for the cytoplasmic part of the γ-subunit or any other regulatory molecule. However, little structural information is available to support either of these hypotheses.
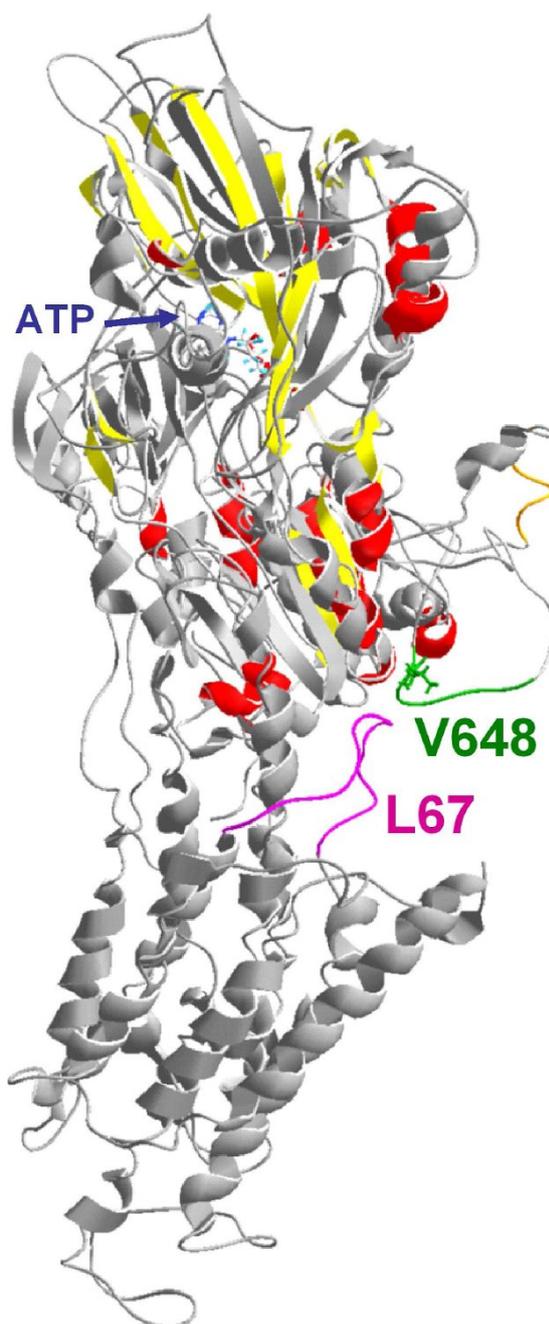
Figure 4: C45 to C67 signaling. The information about the ATP binding to the N-domain of the C45 (the top domain of the colored structure) could be mediated to distal parts of the molecule. Our experiments showed changes in the ESP when the reporter Trp residue was placed on the position 648 within the flexible Asn^{642}-Asn^{649} turn (*green* in the final snapshot of the MD simulation of the C45 in the ATP presence described in Grycova et al. (26), or *orange* in the crystal structure of Na^+/K^+-ATPase). Comparison with the Na^+/K^+-ATPase crystal structure (*gray*; Protein Data Bank entry 3b8e) suggests that this turn could be in the proximity of the C67 (*magenta*).
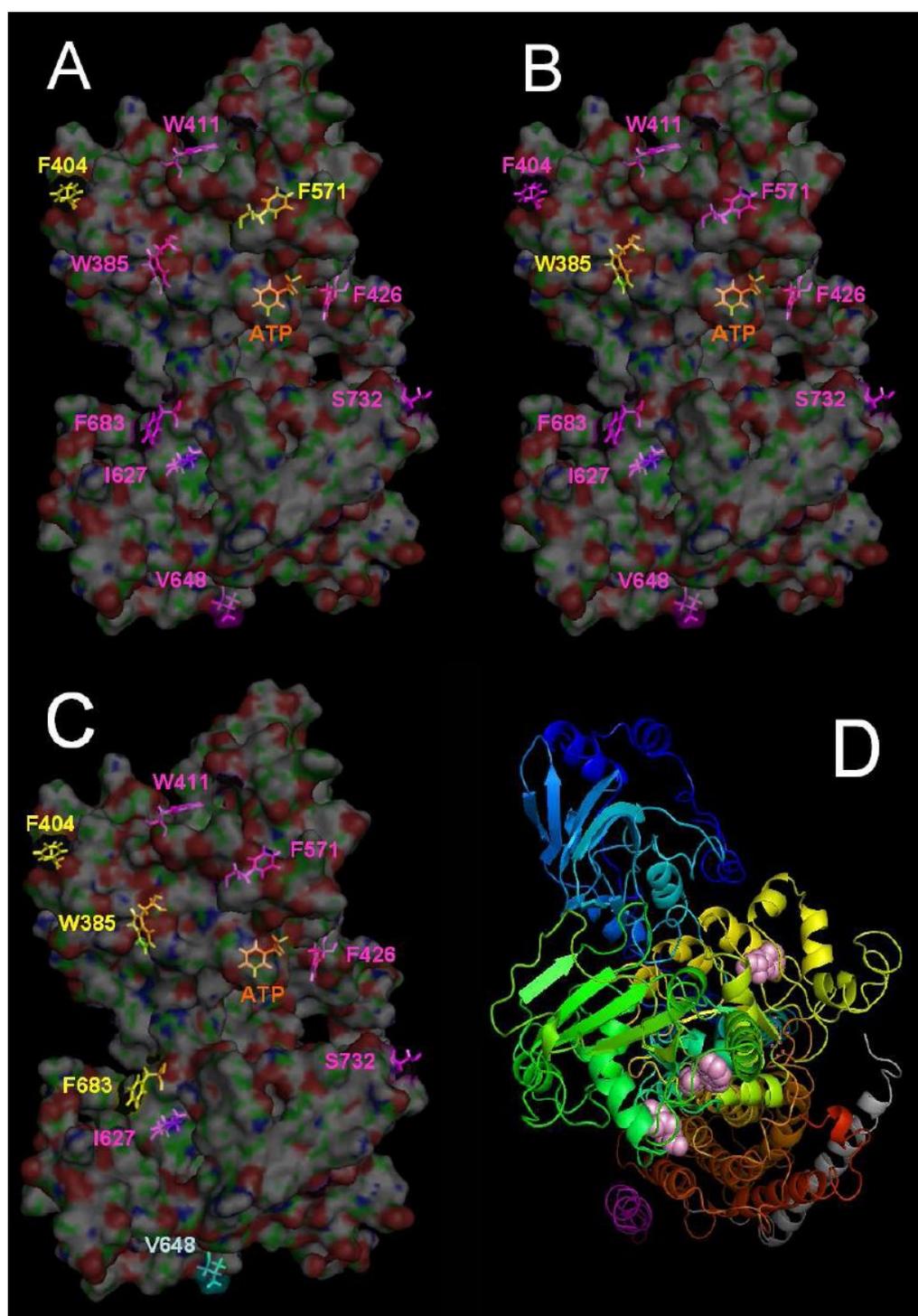
Figure 5: Local changes in the ESP on ligand binding. The color code for the local surface charge changes on (*A*) Mg^{2+}, (*B*) MgATP, or (*C*) Na_2ATP binding: *yellow* are the residues with a ESP shift toward positive values, *magenta* for the residues without change, and *cyan* for the residue with a ESP shift toward negative values. The *orange* residue F475 is located within the ATP binding site. (*D*) View roughly perpendicular to the membrane. The structure of the entire Na^+/K^+-ATPase is rainbow-colored from the N-terminus (*blue*) to C-terminus (*red*), the transmembrane domain of the β-subunit is in *gray*, transmembrane helix of the γ-subunit is in *magenta*. *Pink* spheres represent residues with the positive ESP shift on Na_2ATP binding (*yellow* in *C*; left to right, positions 404, 385, 683).

Conclusion

This study introduced a novel concept for the evaluation of the enzyme conformational changes. It is based on the monitoring of the local changes of the electrostatic surface potential using the intrinsic tryptophan fluorescence quenching. We found that the effect of the nucleotide binding to the C45 is not restricted only to the close environment of the binding site and that the information is in fact transmitted also to the distal parts of the molecule. We detected changes on the nucleotide binding within the set of residues that are located on the side that is reverse to the nucleotide-binding site. At the moment, however, we are not able to reliably link this observation to any step within the enzyme catalytic cycle. Further, our experiments showed ESP changes in the surrounding of the residue V648 on ATP binding, and we propose that this change could be a signal in the interaction of the Asn^{642}-Asn^{649} loop and C67. This could be one of the important links in the communication between the cytoplasmic headpiece and the cation binding sites located within the transmembrane domain.

Supporting Material

A figure is available at `http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01209-0`.

References

1. Benkovic, S. J., and S. Hammes-Schiffer. 2003. A perspective on enzyme catalysis. Science. 301:1196–1202.

2. Kantarci-Carsibasi, N., T. Haliloglu, and P. Doruker. 2008. Conformational transition pathways explored by Monte Carlo simulation integrated with collective modes. Biophys. J. 95:5862–5873.

3. Wen, P. C., and E. Tajkhorshid. 2008. Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis. Biophys. J. 95:5100–5110.

4. Jorgensen, P. L., K. O. Hakansson, and S. J. D. Karlish. 2003. Structure and mechanism of Na,K-ATPase: Functional sites and their interactions. Annu. Rev. Physiol. 65:817–849.

5. Therien, A. G., and C.M.Deber. 2002. Oligomerization of a peptide derived from the transmembrane region of the sodium pump gamma subunit: effect of the pathological mutation G41R. J. Mol. Biol. 322:583–590.

6. Yatime, L., M. J. Buch-Pedersen, M. Musgaard, J. P. Morth, A. M. L. Winther, et al. 2009. P-type ATPases as drug targets: tools for medicine and science. Biochim. Biophys. Acta. 1787:207–220.

7. Moller, J. V., B. Juul, and M. leMaire. 1996. Structural organization, ion transport, and energy transduction of P-type ATPases. Biochim. Biophys. Acta. 1286:1–51.

8. Skou, J. C. 1960. Further investigations on a Mg^{++} + Na^+-activated adenosinetriphosphatase, possibly related to the active, linked transport of Na^+ and K^+ across the nerve membrane. Biochim. Biophys. Acta. 42:6–23.

9. Toyoshima, C., M. Nakasako, H. Nomura, and H. Ogawa. 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 angstrom resolution. Nature. 405:647–655.

10. Toyoshima, C., and H. Nomura. 2002. Structural changes in the calcium pump accompanying the dissociation of calcium. Nature. 418:605–611.

11. Toyoshima, C., H. Nomura, and T. Tsuda. 2004. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. Nature. 432:361–368.

12. Toyoshima, C., and T. Mizutani. 2004. Crystal structure of the calcium pump with a bound ATP analogue. Nature. 430:529–535.

13. Sorensen, T. L. M., J. V. Moller, and P. Nissen. 2004. Phosphoryl transfer and calcium ion occlusion in the calcium pump. Science. 304:1672–1675.

14. Sorensen, T. L.M., C. Olesen, A.M.L. Jensen, J. V.Moller, and P. Nissen. 2006. Crystals of sarcoplasmic reticulum Ca^{2+}-ATPase. J. Biotechnol. 124:704–716.

15. Olesen, C., T. L. M. Sorensen, R. C. Nielsen, J. V. Moller, and P. Nissen. 2004. Dephosphorylation of the calcium pump coupled to counterion occlusion. Science. 306:2251–2255.

16. Jensen, A.M.L., T. L.M.Sorensen, C.Olesen, J. V.Moller, and P. Nissen. 2006. Modulatory and catalytic modes of ATP binding by the calcium pump. EMBO J. 25:2305–2314.

17. Morth, J. P., B. P. Pedersen, M. S. Toustrup-Jensen, T. L. M. Sorensen, J. Petersen, et al. 2007. Crystal structure of the sodium-potassium pump. Nature. 450:1043–1046.

18. Lupfert, C., E. Grell, V. Pintschovius, H. J. Apell, F. Cornelius, et al. 2001. Rate limitation of the Na^+,K^+-ATPase pump cycle. Biophys. J. 81:2069–2081.

19. Capieaux, E., C. Rapin, D. Thines, Y. Dupont, and A. Goffeau. 1993. Overexpression in *Escherichia coli* and purification of an ATP-binding peptide from the yeast plasma-membrane H^+-ATPase. J. Biol. Chem. 268:21895–21900.

20. Gatto, C., A. X. Wang, and J. H. Kaplan. 1998. The M4M5 cytoplasmic loop of the Na,K-ATPase, overexpressed in *Escherichia coli*, binds nucleoside triphosphates with the same selectivity as the intact native protein. J. Biol. Chem. 273:10578–10585.

21. Kubala, M., K. Hofbauerova, R. Ettrich, V. Kopecky, R. Krumscheid, et al. 2002. Phe(475) and Glu(446) but not Ser(445) participate in ATP-binding to the alpha-subunit of Na^+/K^+-ATPase. Biochem. Biophys. Res. Commun. 297:154–159.

22. Kubala, M., J. Plasek, and E. Amler. 2003. Limitations in linearized analyses of binding equilibria: binding of TNP-ATP to the H-4-H-5 loop of Na/K-ATPase. Eur. Biophys. J. Biophys. Lett. 32:363–369.

23. Kubala, M., J. Plasek, and E. Amler. 2004. Fluorescence competition assay for the assessment of ATP binding to an isolated domain of Na^+,K^+-ATPase. Physiol. Res. 53:109–113.

24. Kubala, M., J. Teisinger, R. Ettrich, K. Hofbauerova, V. Kopecky, et al. 2003. Eight amino acids form the ATP recognition site of Na^+/K^+-ATPase. Biochemistry. 42:6446–6452.

25. Lansky, Z., M. Kubala, R. Ettrich, M. Kuty, J. Plasek, et al. 2004. The hydrogen bonds between Arg(423) and Glu(472) and other key residues, Asp(443), Ser(477), and Pro(489), are responsible for the formation and a different positioning of TNP-ATP and ATP within the nucleotide-binding site of Na^+/K^+-ATPase. Biochemistry. 43:8303–8311.

26. Grycova, L., P. Sklenovsky, Z. Lansky, M. Janovska, M. Otyepka, et al. 2009. ATP and magnesium drive conformational changes of the Na^+/K^+-ATPase cytoplasmic headpiece. Biochim. Biophys. Acta. 1788:1081–1091.

27. Garciamoreno, B., L. X. Chen, K. L. March, R. S. Gurd, and F. R. N. Gurd. 1985. Electrostatic interactions in sperm whale myoglobin—site specificity, roles in structural elements, and external electrostatic potential distributions. J. Biol. Chem. 260:4070–4082.

28. Warshel, A., P. K. Sharma, M. Kato, and W. W. Parson. 2006. Modeling electrostatic effects in proteins. Biochem. Biophys. Acta. 1764:1647–1676.

29. Lakowicz, J. R. 1999. Principles of Fluorescence Spectroscopy. Kluwer/Plenum, New York.

30. Albani, J. R. 2004. Structure and Dynamics of Macromolecules: Absorption and Fluorescence Studies. Elsevier, Amsterdam, The Netherlands.

31. Arras, K. O. 1998. An introduction to error propagation: derivation, meaning and examples of equation $C_y = F_x C_x F_x{}^T$. *In* Technical Report of the Autonomous Systems Lab. Swiss Federal Institute of Technology Lausanne, EPFL-ASL-TR-98–01 R3.

32. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA. 98:10037–10041.

33. Sanner, M. F. 1999. Python: a programming language for software integration and development. J. Mol. Graph. 17:57–61.

34. Kubala, M. 2006. ATP-binding to P-type ATPases as revealed by biochemical, spectroscopic, and crystallographic experiments. Proteins. 64:1–12.

35. Pratap, P. R., O. Dediu, and G. U. Nienhaus. 2003. FTIR study of ATP-induced changes in Na^+/K^+-ATPase from duck supraorbital glands. Biophys. J. 85:3707–3717.

36. Picard, M., C. Toyoshima, and P. Champeil. 2005. The average conformation at micromolar $[Ca^{2+}]$ of Ca^{2+}-ATPase with bound nucleotide differs from that adopted with the transition state analog ADP center dot AlFx or with AMPPCP under crystallization conditions at millimolar. J. Biol. Chem. 280:18745–18754.

37. Mead-Savery, F. C., R. Wang, B. Tanna-Topan, S. R. W. Chen, W. Welch, et al. 2009. Changes in negative charge at the luminal mouth of the pore alter ion handling and gating in the cardiac ryanodinereceptor. Biophys. J. 96:1374–1387.

38. Ledvina, P. S., N. H. Yao, A. Choudhary, and F. A. Quiocho. 1996. Negative electrostatic surface potential of protein sites specific for anionic ligands. Proc. Natl. Acad. Sci. USA. 93:6786–6791.

39. Clausen, J. D., and J. P. Andersen. 2004. Functional consequences of alterations to Thr(247), Pro(248), Glu(340), Asp(813), Arg(819), and Arg(822) at the interfaces between domain P, M3, and L6–7 of sarcoplasmic reticulum Ca^{2+}-ATPase—roles in Ca^{2+} interaction and phosphoenzyme processing. J. Biol. Chem. 279:54426–54437.

40. Xu, G. Y., D. J. Kane, L. D. Faller, and R. A. Farley. 2004. The role of loop 6/7 in folding and functional performance of Na,K-ATPase. J. Biol. Chem. 279:45594–45602.

41. Zhang, Z. S., D. Lewis, C. Sumbilla, G. Inesi, and C. Toyoshima. 2001. The role of the M6–M7 loop (L67) in stabilization of the phosphorylation and Ca^{2+} binding domains of the sarcoplasmic reticulum Ca^{2+}-ATPase (SERCA). J. Biol. Chem. 276:15232–15239.

42. Lenoir, G., M. Picard, J. V. Moller, M. le Maire, P. Champeil, et al. 2004. Involvement of the L6–7 loop in SERCA1a Ca^{2+}-ATPase activation by Ca^{2+} (or Sr^{2+}) and ATP. J. Biol. Chem. 279:32125–32133.

43. Corre, F., C. Jaxel, J. Fuentes, T. Menguy, P. Falson, et al. 2002. Involvement of the cytoplasmic loop L6–7 in the entry mechanism for transport of Ca^{2+} through the sarcoplasmic reticulum Ca^{2+}-ATPase. J. Biol. Chem. 277:13016–13028.

10.5   Appendix E

Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact

Petra Florová • **Petr Sklenovský** • Pavel Banáš • Michal Otyepka

J Chem Theory Comput (2010) 6:3569-3579

Abstract

Conventional molecular dynamics simulations on 50 ns to 1 μs time scales were used to study the effects of explicit solvent models on the conformational behavior and solvation of two oligopeptide solutes: α-helical EK-peptide (14 amino acids) and a β-hairpin chignolin (10 amino acids). The widely used AMBER force fields (ff99, ff99SB, and ff03) were combined with four of the most commonly used explicit solvent models (TIP3P, TIP4P, TIP5P, and SPC/E). Significant differences in the specific solvation of chignolin among the studied water models were identified. Chignolin was highly solvated in TIP5P, whereas reduced specific solvation was found in the TIP4P, SPC/E, and TIP3P models for kinetic, thermodynamic, and both kinetic and thermodynamic reasons, respectively. The differences in specific solvation did not influence the dynamics of structured parts of the folded peptide. However, substantial differences between TIP5P and the other models were observed in the dynamics of unfolded chignolin, stability of salt bridges, and specific solvation of the backbone carbonyls of EK-peptide. Thus, we conclude that the choice of water model may affect the dynamics of flexible parts of proteins that are solvent-exposed. On the other hand, all water models should perform similarly for well-structured folded protein regions. The merits of the TIP3P model include its high and overestimated mobility, which accelerates simulation processes and thus effectively increases sampling.

Introduction

Molecular dynamic (MD) simulations of biomacromolecules are based on empirical force fields, which relate potential energy and molecular structure. Several in-depth overviews of current trends in the field of empirical potentials have been published.[1-5] Among all of the available empirical potentials, the AMBER,[6,7] GROMOS,[8] CHARMM,[9] and OPLS[10] families of force fields have performed well for biomacromolecules. However, in this study we deal only with the AMBER family of force fields, which have been thoroughly tested on a variety of biomolecular systems. The AMBER family of force fields perform well for a range of systems including proteins,[11] RNA,[12] and DNA[13,14] and their homo- and heterocomplexes.[15]

One of the AMBER force fields is ff99 (also referred to as parm99),[16] which is a second-generation force field based on the Cornell et al. set of effective parameters[17] employing the pair-additive potential used for condensed-phase allatomic simulations. Simmerling et al. have shown

that the ff99 force field does not accurately represent glycine behavior in protein simulations and that it is prone to other inaccuracies, including overstabilization of α-helical peptide conformations and underestimation of β-bend propensity.[18] Simmerling et al. also suggested a reparametrization of backbone torsion terms, yielding the modified force field ff99SB,[18] which improved the balance in secondary structure propensities. A third-generation AMBER force field, ff03, was introduced by Duan et al., who revised all ψ/φ torsion parameters and recalculated atomic partial charges.[19] Despite such efforts, none of these recent force fields are perfect; for example, Hummer et al. showed that ff03 overestimated and ff99SB underestimated α-helical propensities.[20,21]

The quest for a sufficiently accurate force field has mainly focused on solute behavior, with only marginal consideration of the water environment. However, the water environment plays an essential role in biomolecular processes, and some important effects, such as the hydrophobic effect and Coulomb interaction screening, vanish if the water molecules do not surround a biomolecule.[22-25] An explicit consideration of water molecules is therefore essential for a reasonably accurate description of, at least, solute-solvent interactions.[26] A large number of explicit water models have been developed in attempts to accommodate all of the physicochemical properties of water.[27-37] Among them, Jorgensen's TIPnP[28,29,38] and Berendsen's SPC/E[39] models (Figure 1, Table 1) are the most widely used for biomolecular simulations. It should be noted, however, that although these explicit solvent models have been derived to represent the physico-chemical properties of bulk water well, i.e., their solvent-solvent interactions, the balance of solute-solvent interactions in these models remains questionable and requires further study.

All of the popular TIPnP and SPC/E water models agree well with bulk water characteristics at ambient temperatures. The three-site TIP3P model (in which point charges are centered on each of the three atoms) is the most commonly used model in AMBER simulations. TIP3P reproduces well the key features of bulk water at 25 °C and 1 atm (i.e., a density of 0.997 g/cm^3 and heat of vaporization of 10.53 kcal/mol), but it underestimates the height of the second (tetrahedral) peak in the O—O radial distribution function and overestimates the diffusion constant (Table 1). Both the four-site TIP4P and five-site TIP5P models give better fits to the experimental O—O radial distribution function and also behave better than the TIP3P model in many other respects. However, improving the representation of water properties by adding additional extra point charges increases computational costs considerably, with TIP4P being approximately 1.5 times and TIP5P 2.5 times as expensive as TIP3P in terms of the simulation time (Table 1). The original three-site SPC water model has been superseded by the SPC/E model, which includes corrections for self-polarization and improved structural and diffusion properties. The computational demands of the SPC/E model are comparable to those of the TIP3P model.

Despite the development of sufficiently representative force fields and explicit water models, less attention has been paid to the effects of explicit water models on solute behavior, namely, on the structure, dynamics, and kinetics of solute molecules. Moreover, explicit water models differ in ways that are likely to generate differences in solute-solvent interactions, which might then propagate to differences in solute behavior in various explicit solvent models. In response to the call for studies on the effects of explicit water models on solute behavior reported in the literature,[1,40-43] Nutt and Smith presented CHARMM simulations of N-methylacetamide, other
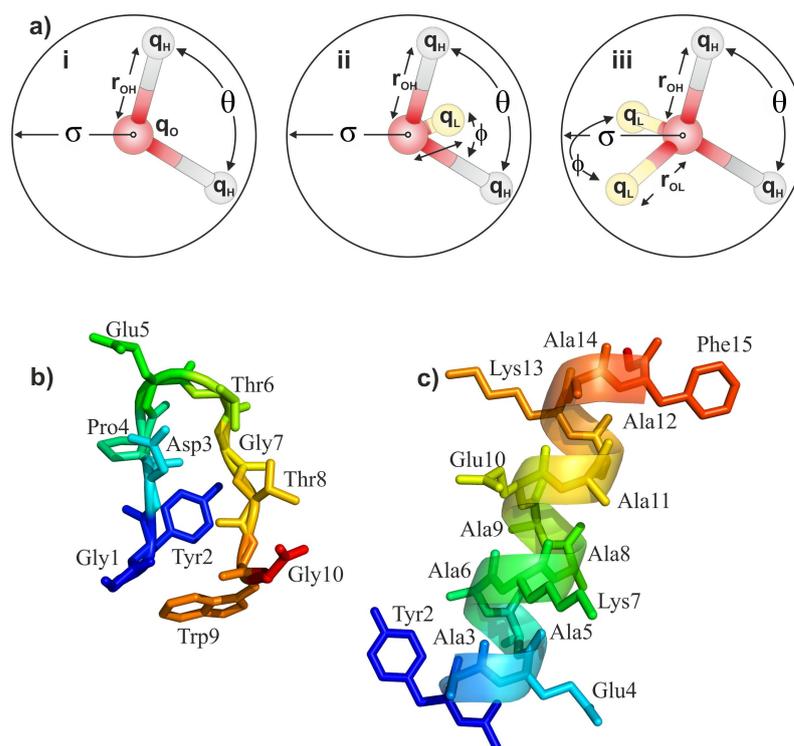
Figure 1: (a) Three-dimensional representations of three-site TIP3P and SPC/E (left), four-site TIP4P (middle), and five-site TIP5P (right) water models. The labels correspond to the parameters listed in Table 1. (b) Chignolin and (c) EK-peptide structures in stick representations.

small solute molecules, and a small protein (crambin) in various solvent models, concluding that although TIP3P, TIP4P, and TIP5P differed in solute-solvent interactions, they provided similar overall descriptions of solvation.[40] However, their results on biomacromolecular solute behavior were, unfortunately, limited by their use of only a 2-3 ns time scale and of an alternative modified TIP3P solvent model (mTIP3P; bearing modified van der Waals parameters for hydrogen, σ_H = 0.449 Å and ε_H = 0.046 kcal/mol).[44] In a recent study, Vymětal and Vondrášek studied the effects of different explicit water models on the free energy φ—ψ profiles of alanine dipeptide using metadynamics. They concluded that the choice of solvent model had no significant effect on the conformational preferences of alanine dipeptide.[45] Molecular dynamics simulations of small proteins carried out by Wong and Case in ff99SB showed that protein diffusion occurs rapidly due to the high self-diffusion constant of TIP3P, whereas the SPC/E and TIP4P water models with more realistic self-diffusion constants have too large a protein rotation diffusion constant.[42] Shirts et al. calculated hydration free energies of the amino acid side chain analogues in several water models and concluded that modified TIP3P (TIP3P-MOD; having different σ_O = 3.12171 Å and ε_O = 0.190 kcal/mol parameters compared to the standard TIP3P parameters; see Table 1)[46] gave the closest match to the experimental data. On the other hand, the ability of explicit solvent models to accurately represent pure water properties did not necessarily determine the ability to correctly predict solute/solvent behavior.[41,47]

Table 1 Parameters and Physicochemical Properties of the Explicit Water Models Studied, As Used in the AMBER Package[a]

param (units)	TIP3P	TIP4P	TIP5P	SPC/E	exptl
q_H^b (e)	0.417	0.520	0.241	0.424	
q_O^b (e)	-0.834	0.000	0.000	-0.848	
q_L^b (e)		-1.040	-0.241		
r_{OH}^c (Å)	0.9572	0.9572	0.9572	1.0000	0.9572[76]
r_{OL}^c (Å)		0.1500	0.7000		
θ_{HOH}^d (deg)	104.520	104.520	104.520	109.470	104.474[76]
ϕ_{LOL}^d (deg)			109.470		
ε^e (kcal/mol)	0.1521	0.1550	0.1600	0.1554	
σ_O^e (Å)	3.1506	3.1537	3.1200	3.1656	
cost[f]	100	140	240	100	
dipole moment (D)	2.35[28]	2.18[30]	2.29[28]	2.35[30]	2.95[77]
dielectric constant	82.0[30]	53.0[30]	81.5[28]	71.0[30]	78.4[78]
density(298 K, 1 atm) (g/cm^3)	0.986 ± 0.010^g	0.994 ± 0.010^g	0.985 ± 0.010^g	0.999 ± 0.010^g	0.997
self-diffusion (298 K, 1 atm) (10^{-5} cm^2/s)	5.5[g]	3.5[g]	2.7[g]	2.5[g]	2.3[50]
density maximum (K)	$\sim 182.15^{31}$	$\sim 248.15^{28}$	$\sim 277.15^{28}$	$\sim 235.15^{32}$	277.134[79]
melting temperature (K)	145.55[32]	232.45[32]	273.95[32]	214.95[32]	273.15

[a] See Figure 1. [b] q_H, q_O, and q_L are the partial charges of hydrogen and oxygen and that on the lone pair, respectively. [c] r_{OH} and r_{OL} are the oxygen–hydrogen and oxygen–lone pair distances, respectively. [d] θ_{HOH} and ϕ_{LOL}^d are the hydrogen–oxygen–hydrogen and lone pair–oxygen–lone pair angles, respectively. [a] e and σ are the well depth and van der Waals radius Lennard-Jones parameters, respectively. [f] Computer cost in percent with respect to TIP3P. [g] Data from this study.

Table 2 List of Simulations Performed for Each Studied System

		no. of simulations run in tested force fields		
system		ff99	ff99SB	ff03
chignolin	TIP3P	6×100 ns; 1×1 μs	6×50 ns; 1×1 μs	6×50 ns; 1×1 μs
	TIP4P	6×50 ns	6×50 ns	6×50 ns
	TIP5P	6×50 ns	6×50 ns	6×50 ns
	SPC/E	6×100 ns; 1×1 μs	6×50 ns; 1×1 μs	6×50 ns; 1×1 μs
EK-peptide	TIP3P	6×100 ns; 1×1 μs	6×50 ns; 1×1 μs	6×100 ns; 1×1 μs
	TIP4P	6×50 ns	6×50 ns	6×50 ns
	TIP5P	6×50 ns	6×50 ns	6×50 ns
	SPC/E	6×100 ns; 1×1 μs	6×50 ns; 1×1 μs	6×50 ns; 1×1 μs

In the present study we examine the structure and dynamics of the small peptide solutes chignolin and EK-peptide (Figure 1) in MD simulations on 50 ns to 1 μs time scales (20.7 μs in total). Chignolin is an artificial peptide consisting of 10 residues (GYDPETGTWG) which adopts a β-hairpin conformation in solution under in vitro conditions (Figure 1b). The peptide is stabilized by H-bonds between atoms Asp3(O) and Gly7(N) (the strongest), Asp3(N) and Thr8(O), and Asp3(O$^\delta$) and Glu5(N). Gly7 plays a further key role in chignolin structure stability because its left-handed α-helical conformation enables propagation of the C-terminal strand. It has been suggested that the side chain interaction between Tyr2 and Trp9 also stabilizes chignolin. The

experimental data concerning chignolin stability suggest that the peptide has a ratio of folded to unfolded states equal to ~60:40 at 300 K.[48] EK-peptide is an artificial 14 residue long α-helix (YAEAAKAAEAAKAF).[49] An α-helicity of 40% at 273 K was measured for EK-peptide using circular dichroism (CD) spectroscopy. Ghosh and Dill found that the α-helicity of EK-peptides generally decreases with increasing temperature and amounts to ~20% at 300 K for the studied EK-peptide.[50] In the present study, four explicit solvent models (TIP3P, TIP4P, TIP5P, and SPC/E) combined with three AMBER family force fields (ff99, ff99SB, and ff03) are analyzed to elucidate the role of explicit solvent models on the behavior of both oligopeptides.

Methods

Studied Systems.   A designed β-hairpin peptide, chignolin, and an artificial α-helical EK-peptide were chosen as test systems for the MD simulations. The chignolin NMR structure (PDB ID 1UAO) was used as a starting structure in MD simulations, and both chignolin termini were charged (N-terminus positively and C-terminus negatively) in all MD simulations. The starting structure of EK-peptide was modeled as an α-helix in accordance with CD spectra.[49] Although, to our best knowledge, no structural data such as X-ray or NMR spectra are available for EK-peptide, the data from CD spectra are sufficient to provide relevant information about its structure. The N-terminus of EK-peptide was capped by an acetyl group and the C-terminus by N-methylamide for MD simulations.

Molecular Dynamics Simulations.   All MD simulations were carried out using the AMBER suite of programs with the all-atomic force fields ff99,[16] ff99SB,[18] and ff03.[19] The simulation protocol, which has been repeatedly shown to perform well for proteins,[51-55] was set up as follows. The hydrogen atoms were added by the LEaP module of AMBER. Systems were then neutralized by adding counterions (Na$^+$ or Cl$^-$, according to the solute charge) and immersed into a rectangular box of explicit water molecules. Each system was solvated by four explicit water models—TIP3P, TIP4P, TIP5P, and SPC/E (Table 1). The initial coordinates of solute, as well as solvent, atoms were identical for a given peptide system in all water models and all force fields. Consequently, we examined the systems' dynamics in all 12 possible combinations of solute force field and solvent type (Table 2). In each simulation, the minimal distance between the solute and the box wall was set to 10 Å. Prior to the production phase of the MD run, each system was minimized by first optimizing the positions of the hydrogen atoms while the heavy atoms remained constrained; then all protein atoms were constrained, and the solvent molecules with counterions were allowed to move during a 1000-step minimization, followed by 10 ps long MD runs under $[NpT]$ conditions (p = 1 atm, T = 298.16 K). After this, the side chains were relaxed by several minimization runs, with decreasing force constants applied to the backbone atoms. After the relaxation, each system was heated from 10 to 298.16 K for 100 ps. The particle-mesh Ewald (PME) method was used for treating electrostatic interactions, and all simulations were performed under periodic boundary conditions in the $[NpT]$ ensemble at 298.16 K and 1 atm using a 2 fs integration step. The SHAKE algorithm, with a tolerance of 10^{-5} Å, was used to fix the positions of all hydrogen atoms, and a 9.0 Åcutoff was applied to nonbonding interactions. The Berendsen thermostat was used.[56]

All systems studied are listed in Table 2. Generally, the dynamics of each system in respective combinations of solute-solvent description were modeled via six MD simulations on 50-100 ns time scales, in which the coordinates were stored every picosecond. The number of parallel MD simulations in each run was chosen according to Day and Daggett,[57] who suggested that, for capturing average properties of simulated systems, 5-10 simulations in multiple MD runs are sufficient. Moreover, additional extensive 1 μs long MD runs with coordinates stored every 10 ps were performed for all force fields in combinations with the TIP3P and SPC/E models. The total simulation time in the present study (the sum of the simulation times of all systems) reached 20.7 μs.

In addition to the peptide simulations, we carried out water box simulations (i.e., only water molecules without solute) for the TIP3P, TIP4P, TIP5P, and SPC/E models under $[NpT]$ conditions ($p = 1$ atm, $T = 298.16$ K) and using an 8.0 Å cutoff for nonbonding interactions on 10 ns time scales. A periodic rectangular cubic box with dimensions of $20 \times 20 \times 20$ Å filled with 375 explicit water molecules was used in each water box simulation.

Analyses of Trajectories.   The α-helix dynamics were monitored using the secondary structure analysis (implemented in ptraj from the AMBER package), the time evolution of the root-mean-square deviations (rmsd's) of the backbone atoms (C, N, C$_\alpha$) with respect to the initial structure, and the time evolution of the distances between atoms forming salt bridges (Glu3(C$_\delta$)-Lys6(N$_\zeta$), Lys6(N$_\zeta$)-Glu9(C$_\delta$), and Glu9(C$_\delta$)-Lys12(N$_\zeta$)). The numbers of water molecules in the first solvation shell (<3.4 Å) around the carbonyl oxygen of each residue were calculated for ff03 EK-peptide simulations using ptraj. The α-helix was said to be unfolded if there were fewer than four neighboring helical residues for more than 100 ps (in accordance with Daggett[58]). The following structural analyses for chignolin simulations were performed: the time evolution of the rmsd's of the backbone atoms (C, N, C$_\alpha$) from the initial structure and the distances between C$_\gamma$ atoms of Tyr2 and Trp9. In this study, all structures with an rmsd of main chain atoms up to 1.7 Å were considered to be native-like, because all chignolin structures from the NMR ensemble fitted into this interval. In addition, structures from the MD ensemble with rmsd below 1.7 Å displayed a native H-bond network and native distance between Tyr2 and Trp9 (C$_\gamma$ atoms). We considered the chignolin molecule as unfolded if the rmsd of the backbone atoms was above 1.7 Å for more than 100 ps.

The effect of water models on the unfolding or refolding rates in ff99 chignolin simulations were analyzed by the following statistical model. Unfolding and refolding are stochastic processes having Bernoulli distributions with associated probabilities p_{unfold} and p_{refold}. The p_{unfold} probability can be estimated from the number of unfolding events (within a 1 ps time frame) divided by the number of snapshots where the system was folded. Similarly, the estimate of p_{refold} equals the number of refolding events divided by the number of snapshots where the system is unfolded. These estimates have a binomial distribution, which can be approximated by a normal distribution, and thus, the confidence intervals documenting statistical relevance of these estimated probabilities can be expressed by the Wilson score interval:[52]

$$p_{\text{min,max}} = \frac{p + \frac{1}{2n}z_{1-\frac{\alpha}{2}}^2 \pm \sqrt{\frac{p(1-p)}{n} + \frac{z_{1-\frac{\alpha}{2}}^2}{4n^2}}}{1 + \frac{1}{n}z_{1-\frac{\alpha}{2}}^2},$$

where p is the estimated probability, p_{min} and p_{max} are the lower and upper limits of the Wilson score interval of the estimated probability, n is the number of realizations (i.e., total number of snapshots) where the system is folded/unfolded, and $z_{1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ percentile of the normal distribution (the α value used was 5%). Finally, the corresponding kinetic constants ($k_{\text{unfold/refold}}$) were derived from the estimated probabilities of unfolding and refolding ($p_{\text{unfold/refold}}$) using the first-order kinetic equation

$$p_{\text{unfold/refold}} = \exp\left(-k_{\text{unfold/refold}}\Delta t\right),$$

where Δt denotes the time interval between two consecutive snapshots. Subsequently, these kinetic constants were transformed to the corresponding potential free energy barriers ($\Delta G^{\ddagger}_{\text{unfold/refold}}$) using the Eyring equation

$$k_{\text{unfold/refold}} = \left(\frac{k_{\text{B}}T}{h}\right)\exp\left(-\frac{\Delta G^{\ddagger}_{\text{unfold/refold}}}{RT}\right),$$

in which T is the absolute temperature and k_{B}, R, and h are the Boltzmann, universal gas, and Planck constants, respectively.

The population and evolution of chignolin conformations in various water models in the ff99 simulations were monitored by the Bayesian clustering algorithm[59] implemented in ptraj, with all variables set to the default (i.e., without the critical distance metric). The optimal cluster number was found iteratively with visual inspection of indices measuring the clustering performance.

The Ramachandran plots of peptide residues were depicted as scatter plots and density plots using an in-house script. The names of canonical regions and their positions in the Ramachandran plot were assigned according to Schlick.[60]

The following analyses of specific solvation in various solvents were performed only for ff99SB chignolin simulations, given the sufficient stability of chignolin in ff99SB; density maps of water models were calculated using ptraj (grid analysis). Prior to the analysis, chignolin was image-centered and rms-fitted, and then finally the density of water molecules was calculated using cubic grids spread over the entire box volume. The output file was visualized using VMD. Further, the numbers of water molecules in the first and second solvation shells around the chignolin molecule were calculated using ptraj. The distance of the chignolin first hydration shell was set to 3.4 Å and the second to 5.0 Å. The analysis of water residence times on the chignolin surface was based on a survival probability correlation function as implemented in ReTiNal (Resident Time Analyzer v1.0, Petr Kulhanek, NCBR Brno, `http://troll.chemi.muni.cz/whitezone/development/root/`) software (see ref 61 for details). Water molecules interacting with oxygen or nitrogen atoms of chignolin with residence time over 0.5 ns were further analyzed.

Self-diffusion coefficients (Table 1) were calculated from the last 5 ns of 10 ns long water box simulations using the Einstein equation

$$D = \frac{1}{6tN} \lim_{t \to \infty} \left\langle \sum_{i=1}^{N} [r_i(t) - r_i(0)]^2 \right\rangle,$$

where D is the self-diffusion coefficient, t is time, N is the total number of atoms, and $r_i(t)$ is the displacement vector of the ith atom at time t. The radial distribution functions of each water model (Figure S1, Supporting Information) were calculated over the entire time scale of the water box simulations by ptraj.

Results

Chignolin

ff99 Force Field. The simulations showed that chignolin did not maintain its native-like structure in the ff99 force field and melted readily in almost all simulations with different solvent models (Table 3). The differences among estimated probabilities and corresponding free energy barriers of unfolding/refolding processes in different solvation models were not statistically significant ($\alpha = 0.05$). Nonetheless, the free energy barriers of unfolding and refolding differed significantly in different force fields, as discussed below (Table 4). The agreement between free energy barriers of unfolding estimated from a series of 100 ns simulations, and from the 1 μs long simulation in the TIP3P and SPC/E water models, shows that the simulations sufficiently converged on a 100 ns time scale, in terms of estimation of unfolding probability. On the other hand, this does not apply for the probability of refolding, because, on longer simulation time scales, higher free energy barriers of refolding were in most cases estimated. This could have been a consequence of further structural relaxation of the unfolded state toward some energetically deeper minima.

Table 3 Populations (%) of the Native-like Chignolin Structures and α-Helicity (%) of EK-Peptide in the Respective MD Simulations

| force field | water model | native-like[a] | | α-helicity[b] | |
		multiple[c]	long[d]	multiple[c]	long[d]
ff99	TIP3P	18 ± 26	3	16 ± 20	15 ± 20
	TIP4P	24 ± 26		14 ± 19	
	TIP5P	53 ± 36		20 ± 23	
	SPC/E	36 ± 35	11	21 ± 23	17 ± 21
ff99SB	TIP3P	95 ± 7	74	21 ± 26	4 ± 12
	TIP4P	94 ± 7		17 ± 24	
	TIP5P	94 ± 6		35 ± 30	
	SPC/E	96 ± 6	83	20 ± 29	4 ± 12
ff03	TIP3P	96 ± 2	46	70 ± 22	68 ± 23
	TIP4P	80 ± 22		74 ± 17	
	TIP5P	96 ± 2		73 ± 22	
	SPC/E	78 ± 19	79	79 ± 15	70 ± 22

[a] The mean populations of the native chignolin structure with their standard deviations were calculated from multiple simulations. [b] The α-helicities are mean values with standard deviations, calculated from all multiple MD simulations or from one 1 μs run, respectively. [c] Six independent simulations. [d] A 1 μs MD simulation.

After chignolin lost its native-like structure, it adopted one of five misfolded conformations (designated clusters 1–5; Figure S2, Supporting Information), which were subsequently identified by a cluster analysis (Table 5). We did not observe any significant difference in populations of unfolded clusters between simulations with the TIP3P and TIP4P models. However, the TIP5P model produced simulations significantly different from those of both TIP3P and TIP4P. The cluster 1 conformation, which was the dominant unfolding state in the TIP3P and TIP4P simulations, was hardly populated in the TIP5P simulations. Instead, the cluster 2 conformation was significantly preferred in the TIP5P simulations, but negligible in populations from the TIP3P and TIP4P simulations (Table 5). Note that the cluster 2 conformation, despite being unfolded, retained a left-handed α-helical conformation of Gly7, which seems to be the critical residue for proper chignolin folding.[48] Both clusters 1 and 2 were occupied in the series of 100 ns simulations with SPC/E, and their populations did not significantly differ from those of any of the TIPnP simulations. When we extended sampling of simulations with TIP3P and SPC/E, the most widely used models, to a 1 μs time scale, we found the cluster populations in TIP3P 100 ns simulations were more or less converged, while further relaxation was observed with the SPC/E model. Specifically, cluster 1 became the most populated unfolded state while cluster 2 was no longer observed in 1 μs long SPC/E simulations. Thus, chignolin in the SPC/E water model on microsecond time scales occupied the same unfolded clusters as in the TIP3P and TIP4P models, which were, however, significantly different from those clusters populated in the TIP5P model. It is worth noting that the transitions between unfolded clusters happened more rapidly in TIP3P than in the SPC/E model, which seems to be a consequence of the higher self-diffusion coefficient of TIP3P (Table 1) and could be the reason for better convergence of cluster populations in TIP3P 100 ns simulations compared to SPC/E simulations (see Figure S3, Supporting Information).

Ramachandran density and scatter plots revealed that some residues moved apart from the native regions (defined from the NMR structure;[48] Figure S2, Supporting Information) and also populated the non-native ones (Figure S4, Supporting Information). All simulations showed that the residues Thr6 and Glu5 could be found not only in the native right-handed α-helical region, but also in a region of noncanonical artificial structures $\varphi = \sim -145°$ and $\psi = \sim 0°$, which is a known artifact of ff99.[18] The most flexible Gly7 readily lost its native left-handed α-helical conformation in TIP3P, TIP4P, and (eventually) SPC/E simulations and shifted to the artificial region around $\varphi = \sim -165°$ and $\psi = \sim 35°$ or to the glycine β_S region.[62] The population of the non-native and artificial region ($\varphi = \sim -145°$ and $\psi = \sim 0°$) was significantly reduced in TIP5P simulations, due to either slightly higher stability of chignolin in the TIP5P model or its favoring of cluster 2, causing less dense population of this artificial region.

Table 4 Estimated Free Energy Barriers (kcal/mol) of Unfolding ($\Delta G^{\ddagger}_{\text{unfold}}$) and Refolding ($\Delta G^{\ddagger}_{\text{refold}}$) and Corresponding Free Energy Differences (kcal/mol) between Unfolded and Folded States ($\Delta G = G_{\text{unfold}} - G_{\text{fold}}$) of Chignolin in Various Solvent Models and Force Fields (298 K, 1 atm)[a]

force field	water model	$\Delta G^{\ddagger}_{\text{unfold}}$	$\Delta G^{\ddagger}_{\text{refold}}$	ΔG
ff99	TIP3P	6.8 ± 0.4 (6.3 ± 0.5)	8.5 ± 0.8 (8.3 ± 0.5)	$-1.7 \pm 0.9 (-2.1 \pm 0.7)$
	TIP4P	6.5 ± 0.4	7.8 ± 0.6	-1.3 ± 0.7
	TIP5P	7.1 ± 0.4	7.3 ± 0.6	-0.2 ± 0.7
	SPC/E	6.7 ± 0.3 (6.9 ± 0.4)	7.2 ± 0.3 (8.2 ± 0.5)	$-0.5 \pm 0.4 (-1.3 \pm 0.6)$
ff99SB	TIP3P	7.5 ± 0.5 (7.3 ± 0.3)	5.7 ± 0.5 (6.8 ± 0.3)	1.9 ± 0.7 (0.6 ± 0.4)
	TIP4P	7.4 ± 0.4	5.4 ± 0.4	1.9 ± 0.6
	TIP5P	7.5 ± 0.5	5.5 ± 0.5	2.0 ± 0.7
	SPC/E	8.2 ± 0.8 (7.8 ± 0.3)	6.2 ± 0.8 (6.9 ± 0.4)	2.0 ± 1.1 (0.9 ± 0.5)
ff03	TIP3P	7.8 ± 0.6 (7.2 ± 0.3)	4.5 ± 0.6 (7.3 ± 0.3)	3.3 ± 0.8 (-0.1 ± 0.4)
	TIP4P	7.1 ± 0.4	6.3 ± 0.4	0.9 ± 0.5
	TIP5P	8.2 ± 0.8	5.6 ± 1.0	2.6 ± 1.3
	SPC/E	7.0 ± 0.3 (7.0 ± 0.2)	6.5 ± 0.4 (6.1 ± 0.2)	0.6 ± 0.5 (0.8 ± 0.3)

[a] The estimates in parentheses relate to 1 μs long simulations.

Table 5 Population (%) of the Native-like Chignolin Structure and Relative Populations (%) of Unfolded Chignolin Clusters in Multiple ff99 Simulations with the TIP3P, TIP4P, TIP5P, and SPC/E Models[a]

cluster number	TIP3P[b]	TIP4P[b]	TIP5P[b]	SPC/E[b]	TIP3P[c]	SPC/E[c]
native-like	18 ± 26	24 ± 26	53 ± 36	36 ± 35	3	11
1	67 ± 19	44 ± 31	5 ± 4	19 ± 32	72	71
2	8 ± 5	9 ± 15	50 ± 29	30 ± 35	0	10
3	1 ± 1	4 ± 5	8 ± 10	22 ± 37	1	1
4	19 ± 13	32 ± 39	24 ± 9	12 ± 13	12	6
5	9 ± 12	14 ± 21	17 ± 16	22 ± 36	13	12

[a] Note that the population of unfolded clusters is normalized, so it represents the portion of unfolded states. [b] Six independent simulations; the percentages of native-like structures and misfolded chignolin clusters of particular simulation runs are listed in Table S1 (Supporting Information). [c] A 1 μs MD simulation.

ff99SB Force Field. In contrast to the ff99 simulations, chignolin was highly stable in all ff99SB simulations (Table 3). In most simulations, no unfolding event occurred, and even if the structure melted, it adopted near-native structures that swiftly refolded to the native chignolin. Both increased unfolding and decreased refolding free energy barriers contributed to significantly higher stability of the chignolin native structure in ff99SB simulations. On the other hand, neither stability nor estimated free energy barriers of unfolding and refolding differed in the set of four explicit solvent models (Table 4). The artificial population of the non-native region ($\varphi = \sim -145°$ and ψ

= ∼ 0°) of the Ramachandran plots vanished in all ff99SB simulations, and backbone torsions of chignolin residues fluctuated around the NMR native values (Figure S5, Supporting Information).

The finding that chignolin was highly stable in all ff99SB simulations allowed us to monitor the behavior of water molecules in the vicinity of chignolin and to consider the effects of various solvent models on the specific solvation of chignolin. Both thermodynamic (the number and position of binding sites of long-residency water molecules) and kinetic (the mobility of these long-residency water molecules) perspectives on the preferential hydration of the chignolin surface were taken into account.[63] Similar local densities of water molecules were observed for simulations with the TIP4P and TIP5P models, while identifying large numbers of specific hydration sites. The three-site TIP3P and SPC/E water models also yielded similar density maps, but with significantly fewer specific hydration sites and with lower densities, in comparison with the TIP4P and TIP5P models (Figure 2). In addition, the analysis of long-residency water molecules (here, those with residence times longer than 0.5 ns) identified significant differences among solvation models, such that the number of long-residency water molecules decreased in the order TIP5P ≫ SPC/E ≈ TIP4P > TIP3P (Table 6).
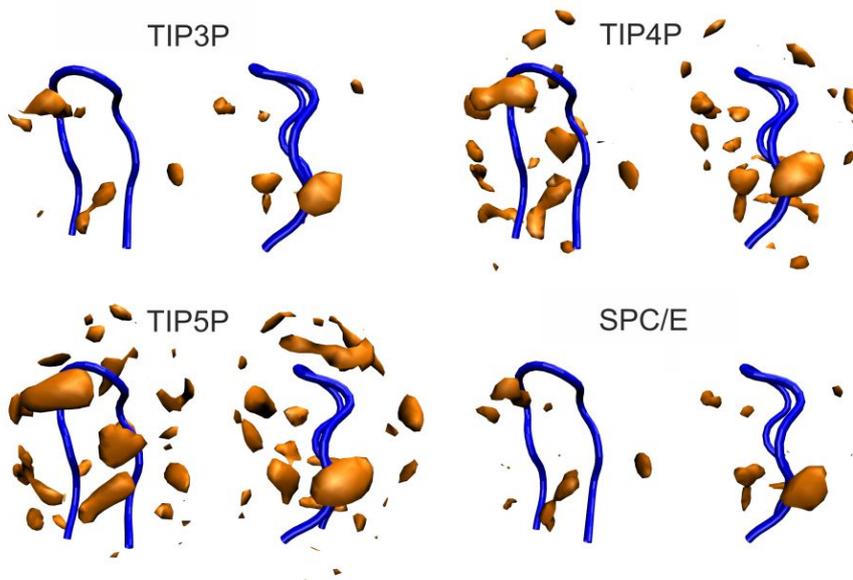

Figure 2: Water density maps (orange) around chignolin (the blue tube represents the chignolin backbone) for ff99SB simulations.

Table 6 Average Number of Water Molecules in the First Two Solvation Shells of Chignolin and Average Number of Long-Residency Water Molecules (Residence Times > 0.5 ns) in ff99SB Chignolin Simulations with Corresponding Standard Deviations

	TIP3P	TIP4P	TIP5P	SPC/E	TIP3P$^{1\mu s}$	SPC/E$^{1\mu s}$
first shell, < 3.4 Å	93 ± 8	93 ± 6	100 ± 8	94 ± 7	90 ± 6	95 ± 8
second shell, < 5.0 Å	185 ± 14	182 ± 12	196 ± 15	186 ± 13	179 ± 11	189 ± 15
long-residency water molecules	0.0 ± 0.0	2.0 ± 1.3	12.8 ± 4.5	3.7 ± 2.7		

These findings show that the specific hydration of chignolin is determined by the mobility of a given water model in terms of its self-diffusion coefficient (Table 1) and its propensity to bind to specific hydration sites. Thus, the highest specific solvation in TIP5P simulations was caused by a large number of specific hydration sites and the low (and most realistic) mobility of TIP5P water molecules. On the other hand, no long-residency water molecules were identified for the TIP3P model, due to the small number of preferred hydration sites and rapid exchange of TIP3P water molecules at these sites. Interestingly, although the SPC/E model has a self-diffusion coefficient similar to that of TIP5P, the significantly smaller quantity of long-residency SPC/E water molecules was caused by a low number of specific hydration sites, most likely stemming from solute–solvent interactions and thermodynamic factors. Similarly, although the water density map of TIP4P was similar to that of TIP5P, with the same number of specific hydration sites, the higher self-diffusion coefficient of TIP4P resulted in significantly more rapid dynamics at these hydration sites, and thus, the specific hydration was reduced for kinetic reasons.

ff03 Force Field. As in ff99SB simulations, the chignolin native structure was highly populated in ff03 simulations. The rare unfolding events were frequently quickly followed by chignolin refolding in almost all cases, except in some SPC/E and TIP4P simulations and the 1 μs TIP3P simulation, where we observed minor but apparent populations of unfolded chignolin with Gly7 shifted from the left-handed α-helical region to the β_{PR} region with $\varphi = \sim 80°$ and $\psi = \sim -150°$ (Table 3; Figure S6, Supporting Information).[62] These non-native chignolin conformations closely match the β-hairpin topology, but lack the native hydrogen bond network and contacts between Tyr2 and Trp9 residues. On the other hand, the shift of Gly7 to the β_{PR} region was always fully reversible, and refolding was sooner or later observed. It seems that the shift of Gly7 does not depend on the solvation model and can be considered as an ff03 force field effect. The difference in the balance between left-handed α-helical and β_{PR} conformations of Gly7 in the ff03 and ff99SB force fields can be explained by differences in the free energy landscapes of the φ and ψ dihedrals in ff03 and ff99SB.[45]

EK-Peptide

ff99 Force Field. The EK-peptide readily lost its α-helical structure in all ff99 simulations (Table 3). Melting that propagated from the termini occurred within a time scale of hundreds of picoseconds to several nanoseconds. Melting times were progressively longer in the TIP5P and SPC/E water models, whose self-diffusion coefficients were smaller (and more realistic) than those of TIP3P, which had the highest self-diffusion coefficient of the explicit models in this study (Figures S13 and S14, Supporting Information). In the unfolded state, the peptide was highly dynamic, often switching among the 3_{10}-helical structure, turn, random coil, and α-helical conformations. Some residues also populated the artificial region around $\varphi = \sim -145°$ and $\psi = \sim 0°$, which is a known artifact of the ff99 force field.[18] Notably, internal residues of the EK-peptide had a propensity to refold for up to $\sim$10 ns in the TIP5P and SPC/E models.

Table 7 Salt Bridge Probabilities (Distance up to 4.5 Å) Calculated from EK-Peptide Simulations with ff03[a]

water model	Glu3···Lys6	Lys6···Glu9	Glu9···Lys12
TIP3P	0.36 ± 0.04 (0.35)	0.06 ± 0.02 (0.06)	0.31 ± 0.02 (0.33)
TIP4P	0.33 ± 0.07	0.03 ± 0.01	0.36 ± 0.04
TIP5P	0.49 ± 0.20	0.08 ± 0.10	0.51 ± 0.15
SPC/E	0.33 ± 0.07 (0.29)	0.06 ± 0.06 (0.06)	0.36 ± 0.07 (0.29)

[a] The mean probabilities and corresponding standard deviations were calculated from multiple simulations having helicity above 70%. The estimates in parentheses relate to 1 μs simulations.

ff99SB Force Field.   As in the previous case, the native α-helical fold was unstable throughout the time scale of the conducted simulations (Figures S15 and S16, Supporting Information). Melting of the helical structure occurred early, within a time scale of several nanoseconds, in simulations with the most mobile models, TIP3P and TIP4P. In contrast, in simulations using the water models with lower (and more realistic) mobility, TIP5P and SPC/E, the system melted later, generally on a time scale of tens of nanoseconds. The residues of the unfolded EK-peptide populated the left- and right-handed helical regions, the region of noncanonical structures around $\varphi = \sim -145°$ and $\psi = \sim 0°$, and the antiparallel β-sheet and triple-stranded collagen helix regions of the Ramachandran plot (Figures S8 and S11, Supporting Information). Nonetheless, the 1 μs TIP3P and SPC/E MD runs characterized the behavior of the system after melting, revealing that the antiparallel β-sheet, triple-stranded collagen helix, and left-handed regions were significantly populated in this state.

ff03 Force Field.   In all simulations, the EK-peptide displayed a significant level of intrinsic α-helicity lying in the interval from $\sim$70% to $\sim$80% (Table 3; Figures S17 and S18, Supporting Information). The terminal residues had less helical propensity than internal residues and in the nonhelical state populated the triple-stranded collagen helix, antiparallel β-sheet, left-handed α-helix, and $\varphi = \sim -160°$ and $\psi = \sim -25°$ regions (Figures S9 and S12, Supporting Information).

The high stability of α-helical content in ff03 simulations of EK-peptide led us to monitor the stability of salt bridges between Lys and Glu side chains (Table 7) and the specific hydration of backbone carbonyl oxygen atoms (Figure 3). The formation of salt bridges and shielding of backbone H-bonds forming group (–C=O, –N–H) from water molecules were suggested to contribute to the stability of alanine-based α-helical peptides (see refs 64 and 65 and references therein). The TIP3P, TIP4P, and SPC/E models behave similarly; i.e., the stability of salt bridges and hydration of backbone carbonyl oxygen atoms do not significantly differ among these three solvent models. On the other hand, MD simulations with TIP5P show higher stabilities of salt bridges (Table 7) and more extensive hydration of backbone carbonyl oxygen atoms (Figure 3).

Discussion

In the present study we aimed to elucidate the effects of four popular explicit solvent models (TIP3P, TIP4P, TIP5P, and SPC/E) on the behavior of two oligopeptides: a β-hairpin chignolin and an
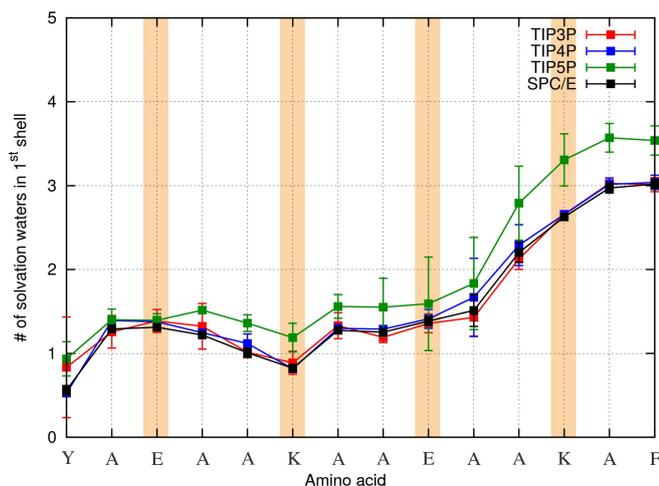
Figure 3: Average numbers of water molecules within the first hydration shell (<3.4 Å) of backbone carbonyl oxygen atoms calculated from ff03 simulations of EK-peptide.

α-helical EK-peptide. We combined these solvent models with three recent protein force fields (ff99, ff99SB, and ff03) from the AMBER family. Our results suggest that the choice of solvent model does not significantly affect the stability of the studied peptides, which was completely governed by the force field employed. Furthermore, even the conformational behavior of the peptides in their native folded state was not influenced by the choice of solvent model.

We found that the ff99 force field destabilized the native fold of both peptides to an extent that significantly underestimated peptide stabilities in comparison with experimentally observed values.[48,50] In addition, in ff99 simulations we observed a significant population of noncanonical conformations in the form of φ—ψ torsions ($\varphi = \sim -145°$ and $\psi = \sim 0°$), which are a known ff99 force field artifact.[18] Elimination of this artifact motivated the development of a reparametrized force field named ff99SB.[18] The ff99SB force field overstabilized the chignolin structure, while it did not improve the stability of the EK-peptide, destabilizing helical EK-peptide in favor of β-structured states. This indicates that although the ff99SB force field slightly improves the behavior of β-structured chignolin, it yields unbalanced force fields biased toward β-structures. Finally, the ff03 force field seemed the best choice among the tested force fields, as the chignolin stability estimated from all ff03 simulations was $72 \pm 21\%$, in good agreement with the experimentally observed chignolin stability of 60%. Moreover, the ff03 simulations did not cause any force field artifacts. They displayed folding/unfolding processes in chignolin connected with shifts of the flexible Gly7 between the left-handed α-helix and β_{PR} conformations, which represents the tolerated conformational variability of this Gly residue. Nonetheless, the ff03 force field still overestimated the stability of α-helical EK-peptide, as the α-helicity observed in the ff03 simulations amounted to ~70-80%, while the experimental helical content equals $\sim20\%$ at 300 K. Thus, the ff03 force field seems the best, but still not perfect, choice from the studied set.

While we did not observe any effect of explicit solvent models on the stability of the studied peptides, we found substantial differences in specific solvation of the backbone carbonyls of the

EK-peptide α-helix and of the chignolin native structure in the set of studied water models. The TIP5P model showed more extensive hydration of backbone carbonyl oxygen atoms in comparison with TIP3P, TIP4P, and SPC/E, which gave almost the same results (Figure 3). The higher hydration of H-bonds between the backbone amide and carbonyl groups of the α-helix was suggested to destabilize the α-helicity of alanine-based peptides.[64,65] Thus, the lower α-helicity of EK-peptide in the TIP5P simulations in comparison with other solvation models was anticipated as an effect of the more extensive backbone hydration. However, the α-helicity of EK-peptide did not significantly differ among various solvent models (Table 3). This finding can be most likely explained by the higher probabilities of salt bridges in TIP5P water (Table 7) and compensation between the specific solvation of backbone carbonyl oxygens and salt bridge stabilities.

Chignolin was highly solvated in the TIP5P simulations, while significantly lower solvation was observed in other water model simulations. We analyzed the reasons for decreased specific solvation in the remaining three models and found that specific solvation by TIP4P was reduced by rapid exchange of water molecules in the solvation binding sites, i.e., for kinetic reasons. This finding can be explained by the larger self-diffusion coefficient of TIP4P in comparison with the TIP5P model. The specific solvation by SPC/E was reduced due to the small number of specific solvent binding sites compared to those in the TIP5P and TIP4P models, i.e., for thermodynamic reasons. Finally, the reduced specific solvation in TIP3P simulations was a consequence of both the aforementioned kinetic and thermodynamic factors (Figure 4). Consequently, assuming that the studied solvation models differ significantly in the specific solvation of solute and do not concurrently affect the stability of the studied peptide solutes, the question arises of whether these differences in specific solvation can influence the solute and, if so, how.

As the differences in specific solvation of the studied models did not affect the behavior of the studied peptides in their folded states, we focused rather on their behavior in unfolded states in ff99 simulations. We found that the conformational behavior of chignolin in unfolded states is similar in all three water models with lower specific solvation, i.e., in TIP3P, TIP4P, and SPC/E, while a different conformational behavior of unfolded chignolin was observed in simulations with TIP5P, demonstrating a high degree of specific solvation. Specifically, we observed different preferences in the population of clusters of unfolded structures. The TIP3P, TIP4P, and SPC/E simulations resulted in preferred occupancy of the unfolded cluster with an artificial conformation of Gly7 ($\varphi = \sim -165°$ and $\psi = \sim 35°$), while in TIP5P simulations an alternative cluster with a left-handed α-helical conformation of Gly7 was preferred. This result implies that the solvation models will most likely not significantly affect the structure, stability, and flexibility of structured segments of folded proteins. However, the differences in specific solvation might contribute to conformational variability and flexibility of solvent-exposed flexible parts, such as flexible glycine-rich loops. Notably, these most flexible parts of proteins are frequently connected with specific biological functions and are at the focus of attention of many MD simulations.[51,52,66–72] Moreover, specific solvation is also the subject of many other studies concerning, e.g., drug design.[67,73–75] Therefore, in these cases, the correct description of specific solvation might be critical for correct results.

Unfortunately, we are not able to ultimately conclude which model provides the most realistic description of specific solvation. On the other hand, our data suggest that the TIP5P model
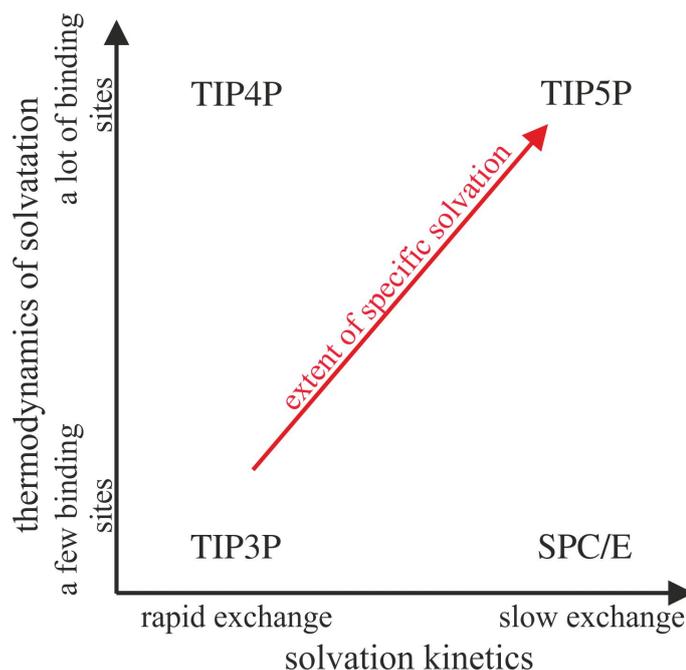
Figure 4: Schematic relationship among the solvation kinetics, thermodynamics of specific solvation (expressed by the number of specific binding sites), and extent of specific solvation of chignolin peptide in four studied water models.

provides a significantly different specific solvation that could propagate to solute behavior via stability of salt bridges and modification of conformational preferences in unfolded states. As the TIP5P model has the most realistic solvent parameters (among studied solvent models), we suggest that TIP5P is the most promising candidate for the realistic description of solute–solvent interactions and specific solvation. Nonetheless, this implication is not straightforward and might be questionable, so direct comparison with reference experimental data of the specific solvation or of the solute–solvent interaction is needed to determine which model performs the best.

In addition, we observed a relationship between the mobility of a water model, represented by its self-diffusion coefficient, and the speed of processes in our simulations, namely, the kinetics of conversion between the unfolded clusters and the structural relaxation of unfolded states in ff99 simulations. Thus, we conclude that the mobility of solvent models can accelerate the kinetics of processes in solutes.

Taken together, in sharp contrast to the choice of force field, choosing the most reliable water model is not straightforward. In cases where it can be assumed that specific solvation will not play an important role (i.e., in wells-tructured proteins lacking highly flexible loops), all the model performances may be similar in quality. Furthermore, in such cases, simulations can benefit from the high (overestimated) mobility of the TIP3P model, because it does not significantly affect the thermodynamic properties of the solutes, such as the structure of thermodynamically stable conformers, but it significantly reduces the barriers between these states and thus accelerates the kinetics and effectively increases sampling in the simulations. On the other hand, if the solute contains flexible and solvent-exposed loops, it would be better to use the more realistic description

of solvation by the TIP5P model, even though this would entail an increased simulation time due to the relatively slower processes and lower sampling.

References

1. Mackerell, A. D. J. Comput. Chem. 2004, 25, 1584.

2. Ponder, J. W.; Case, D. A. AdV. Protein Chem. 2003, 66, 27.

3. van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J. J. R. Soc. Interface 2008, 5 (3), S173.

4. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. Angew. Chem., Int. Ed. 2006, 45, 4064.

5. Sponer, J.; Spackova, N. Methods 2007, 43, 278.

6. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Comput. Phys. Commun. 1995, 91, 1.

7. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. J. Comput. Chem. 2005, 26, 1668.

8. Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. Comput. Phys. Commun. 1995, 91, 43.

9. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J. Comput. Chem. 1983, 4, 187.

10. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. J. Phys. Chem. B 2001, 105, 6474.

11. Guvench, O.; MacKerell, A. D., Jr. Methods Mol. Biol. 2008, 443, 63.

12. Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. Acc. Chem. Res. 2010, 43, 40.

13. Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. Nucleic Acids Res. 2008, 36, 2379.

14. Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., III; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. Nucleic Acids Res. 2010, 38, 299.

15. MacKerell, A. D.; Nilsson, L. Curr. Opin. Struct. Biol. 2008, 18, 194.

16. Wang, J. M.; Cieplak, P.; Kollman, P. A. J. Comput. Chem. 2000, 21, 1049.

17. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. J. Comput. Chem. 1995, 16, 1357.

18. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Proteins 2006, 65, 712.

19. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. J. Comput. Chem. 2003, 24, 1999.

20. Best, R. B.; Buchete, N. V.; Hummer, G. Biophys. J. 2008, 95, 4494.

21. Best, R. B.; Hummer, G. J. Phys. Chem. B 2009, 113, 9004.

22. Chandler, D. Nature 2005, 437, 640.

23. Roca, M.; Messer, B.; Warshel, A. FEBS Lett. 2007, 581, 2065.

24. Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Biochim. Biophys. Acta 2006, 1764, 1647.

25. Leach, A. R. Molecular Modelling: Principles and Applications, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 2001; p 768.

26. Orozco, M.; Luque, F. J. Chem. ReV. 2000, 100, 4187.

27. Guillot, B. J. Mol. Liq. 2002, 101, 219.

28. Mahoney, M. W.; Jorgensen, W. L. J. Chem. Phys. 2000, 112, 8910.

29. Mahoney, M. W.; Jorgensen, W. L. J. Chem. Phys. 2001, 114, 363.

30. Kusalik, P. G.; Svishchev, I. M. Science 1994, 265, 1219.

31. Vega, C.; Abascal, J. L. F. J. Chem. Phys. 2005, 123, 144504.

32. Vega, C.; Sanz, E.; Abascal, J. L. F. J. Chem. Phys. 2005, 122, 114507.

33. Baez, L. A.; Clancy, P. J. Chem. Phys. 1994, 101, 9837.

34. Vega, C.; McBride, C.; Sanz, E.; Abascal, J. L. F. Phys. Chem. Chem. Phys. 2005, 7, 1450.

35. Mark, P.; Nilsson, L. J. Phys. Chem. A 2001, 105, 9954.

36. Zielkiewicz, J. J. Chem. Phys. 2005, 123, 104501.

37. Hess, B.; van der Vegt, N. F. J. Phys. Chem. B 2006, 110, 17616.

38. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J. Chem. Phys. 1983, 79, 926.

39. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. J. Phys. Chem. 1987, 91, 6269.

40. Nutt, D. R.; Smith, J. C. J. Chem. Theory Comput. 2007, 3, 1550.

41. Shirts, M. R.; Pande, V. S. J. Chem. Phys. 2005, 122, 1.

42. Wong, V.; Case, D. A. J. Phys. Chem. B 2008, 112, 6013.

43. Glass, D. C.; Krishnan, M.; Nutt, D. R.; Smith, J. C. J. Chem. Theory Comput. 2010, 6, 1390.

44. Price, D. J.; Brooks, C. L. J. Chem. Phys. 2004, 121, 10096.

45. Vymetal, J.; Vondrasek, J. J. Phys. Chem. B 2010, 114, 5632.

46. Sun, Y. X.; Kollman, P. A. J. Comput. Chem. 1995, 16, 1164.

47. Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. J. Chem. Phys. 2003, 119, 5740.

48. Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. Structure 2004, 12, 1507.

49. Scholtz, J. M.; Barrick, D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. Proc. Natl. Acad. Sci. U.S.A. 1995, 92, 185.

50. Ghosh, K.; Dill, K. A. J. Am. Chem. Soc. 2009, 131, 2306.

51. Bartova, I.; Koca, J.; Otyepka, M. J. Mol. Model. 2008, 14, 761.

52. Wilson, E. B. J. Am. Stat. Assoc. 1927, 22, 209.

53. Sklenovsky, P.; Banas, P.; Otyepka, M. J. Mol. Model. 2008, 14, 747.

54. Skopalik, J.; Anzenbacher, P.; Otyepka, M. J. Phys. Chem. B 2008, 112, 8165.

55. Sklenovsky, P.; Otyepka, M. J. Biomol. Struct. Dyn. 2010, 27, 521.

56. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. J. Chem. Phys. 1984, 81, 3684.

57. Day, R.; Daggett, V. Proc. Natl. Acad. Sci. U.S.A. 2005, 102, 13445.

58. Scott, K. A.; Alonso, D. O. V.; Pan, Y. P.; Daggett, V. Biochemistry 2006, 45, 4153.

59. Shao, J. Y.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. J. Chem. Theory Comput. 2007, 3, 2312.

60. Schlick, T. Molecular Modeling and Simulation: An Interdisciplinary Guide; Springer Science + Business Media LLC: New York, 2002; Vol. 21.

61. Kriz, Z.; Otyepka, M.; Bartova, I.; Koca, J. Proteins 2004, 55, 258.

62. Ho, B. K.; Brasseur, R. BMC Struct. Biol. 2005, 5, 14.

63. Priya, M. H.; Shah, J. K.; Asthagiri, D.; Paulaitis, M. E. Biophys. J. 2008, 95, 2219.

64. Garcia, A. E.; Sanbonmatsu, K. Y. Proc. Natl. Acad. Sci. U.S.A. 2002, 99, 2782.

65. Ghosh, T.; Garde, S.; Garcia, A. E. Biophys. J. 2003, 85, 3187.

66. Dvorakova-Hola, K.; Matuskova, A.; Kubala, M.; Otyepka, M.; Kucera, T.; Vecer, J.; Herman, P.; Parkhomenko, N.; Kutejova, E.; Janata, J. J. Mol. Biol. 2010, 396, 1197.

67. Bartova, I.; Otyepka, M.; Kriz, Z.; Koca, J. Protein Sci. 2005, 14, 445.

68. Bartova, I.; Otyepka, M.; Kriz, Z.; Koca, J. Protein Sci. 2004, 13, 1449.

69. Wong, L.; Jennings, P. A.; Adams, J. A. Acc. Chem. Res. 2004, 37, 304.

70. Narayanan, A.; Jacobson, M. P. Curr. Opin. Struct. Biol. 2009, 19, 156.

71. Prudent, R.; Sautel, C. F.; Cochet, C. BBAsProteins Proteomics 2010, 1804, 493.

72. Otyepka, M.; Bartova, I.; Kriz, Z.; Koca, J. J. Biol. Chem. 2006, 281, 7271.

73. de Beer, S. B.; Vermeulen, N. P.; Oostenbrink, C. Curr. Top. Med. Chem. 2010, 10, 55.

74. Yan, A.; Grant, G. H.; Richards, W. G. J. R. Soc. Interface 2008, 5 (3), S199.

75. Zhang, B.; Tan, V. B.; Lim, K. M.; Tay, T. E. J. Chem. Inf. Model. 2007, 47, 1877.

76. Hasted, J. B. Liquid Water: Dielectric Properties. Water: A ComprehensiVe Treatise; Plenum Press: New York, 1972; Vol. 1.

77. Gubskaya, A. V.; Kusalik, P. G. J. Chem. Phys. 2002, 117, 5290.

78. Fernandez, D. P.; Mulev, Y.; Goodwin, A. R. H.; Sengers, J. M. H. L. J. Phys. Chem. Ref. Data 1995, 24, 33.

79. Jancso, G.; van Hook, W. A. Chem. Rev. 1974, 74, 689.

Acknowledgment

Supporting Information Available

Figures S1–S21 and Tables S1–S4 are available free of charge via the Internet at `http://pubs.acs.org`.