Mem. S.A.It. Suppl. Vol. 4, 24

Memorie della
Supplementi

Folding and mis–folding of peptides and proteins: insights from molecular simulations

G. Colombo[1,2]

[1] Istituto di Chimica del Riconoscimento Molecolare, CNR via Mario Bianco 9, 20131 Milano, Italy
[2] Centre for biomolecular Interdisciplinary Studies and Industrial Applications, University of Milan, Italy

Abstract. In this paper, the main achievements and problems of the application of all–atom molecular simulations, with particular attention for Molecular Dynamics (MD), will be critically reviewed. Starting from unfolding simulations, through biased simulations, which require a knowledge of the native state conformation, to folding studies based on the simple knowledge of the protein (or peptide) sequence, the strengths and weaknesses of theoretical approaches to the study of folding and their matching with experimental observations will discussed. Finally, we will give a critical outlook on the possible development of this field in the near future.

1. Introduction

Protein folding, mis–folding and unfolding are fundamental events in living organisms that have been difficult to characterize in detail, even in vitro. However, the past decade has witnessed a revolution in experimental and theoretical methods that can describe folding processes at high resolution. This exploding interest in the study of the correlations between the linear sequence and the three dimensional organization of a protein has been boosted by the ever increasing number of protein sequences available from the genome projects and the realization that mis–folding of proteins can lead to diseases. Molecular illnesses such as Alzheimer's and Parkinson's disease, Kreutzfeld-Jacob syndrome, Bovine Spongiform Encephalopathy (BSE, or mad cow disease) share a common cause represented by the aggregation of mis–folded proteins or peptides (Kelly 1997; Booth 1997; Ma & Nussinov 2002). Understanding the molecular reasons of folding and mis–folding assumes thus a paramount importance in the development of treatments or drugs targeting these conditions.

From the practical point of view the use of single point mutations as targeted probes can define the kinetic and thermodynamic influence of particular residues on the pathways leading to the native state (Fersht 1995). Temperature jump relaxation and fluorescence spectroscopy experiments yield information about the global content of secondary structure in the early stages of folding (Snow & Nguyen & Pande & Gruebele 2002). At the present moment, the only two experimental approaches providing atomic level details on molecular structures are X-ray crystallography and NMR spectroscopy. The former can only give a static picture of the protein under exam in the confined environment of the crystal lattice. NMR methods provide a picture of the molecule in solution, and have also proved to pin down the unfolded states (Shortle & Ackermnan 2001). The conformations obtained are based on the application of NOE restraints which in general contain a highly diverse ensemble of fast interconverting structures on the NMR timescale; the ensemble of NOE compatible conformations is in general averaged and minimized to give a set of representative structures of the protein under exam. However, despite this high level of sophistication and detail, none of these approaches has been able so far to unravel all of the atomic determinants of the folding reaction.

In order to study the self-organization of biomolecules and fully characterize the initial stages of folding at residue or atomic resolution one cannot but turn to simulation techniques. Simulations, and Molecular Dynamics (MD) in particular, can provide the ultimate detail concerning individual particle motions as a function of time. Thus, they can be used to address specific questions about the properties of a model system, which could be inaccessible experimentally. The use of these methods in close collaboration with more sophisticated experiments has led to a general understanding of the mechanism of protein folding. However, much remains to be learned about the folding of individual proteins, including the complete prediction of the structure of a protein from its sequence. Because of the difficulty of folding a protein by brute force techniques (the fastest protein folding reactions require 10 microseconds, which is at the limit of accessible simulation times), simplified models – for example, lattice models and Cα off lattice models – have been used to obtain insights concerning the mechanism of protein folding (Settanni & Hoang & Micheletti & Maritan 2002; Dill & Chan 1997; Chan & Dill 1997). Recently, it has become possible to follow the folding and unfolding of model peptides in solution (vanGunsteren & Burgi & Peter & Daura 2001). Such simulations are likely to be extended to study the folding of actual proteins before long. In this paper, we will review the main developments in the applications of Molecular Dynamics (MD) simulations in the last few years, and discuss their relative importance as an aid to medicinal chemistry.

2. All Atom MD simulations of peptides and proteins

2.1. Understanding folding from unfolding simulations

Detailed atomic models, e.g. all–atom force fields with explicit solvent, play a fundamental role in the study of the folding process as they were shown to provide specific microscopic information about the conformational characteristics of specific sequences of peptides, or of specific mutations within proteins (Colombo & Merz 1999; vanGunsteren & Burgi & Peter & Daura 2001).

Due to the experimental time scales required for folding, of the order of a few microseconds in the case of small proteins, and to the complexity of the systems under exam, realistic MD simulations of folding of ”real proteins” are still not currently possible. As a consequence, MD simulations have been widely applied to probe unfolding and limited refolding events, with the assumption that the principle of microscopic reversibility dictates that folding and unfolding processes are the same under the same conditions. In this framework, the group of Daggett employed strongly denaturing conditions, such as extremely high temperatures to break the interactions stabilizing secondary and tertiary structures, forcing the protein to unfold (Brooks 1993; Dagget & Levitt 1994; Dagget 2000; Brooks 1998).

The analysis of multiple unfolding trajectories for model proteins, like CI2, Engrailed homeodomains, (Mayor 2003) barnase etc., allowed the authors to identify the structures of unfolding intermediates and Transition States (TS) (Fig. 1) (Dagget & Levitt 1993; Caflisch & Karplus 1994).

The TS of protein folding/unfolding is the state of highest free energy, and unfortunately, free energies cannot be reliably calculated from MD unfolding trajectories with the purpose of mining for conformational states. So, Daggett and coworkers devised a method to identify transition state structures employing

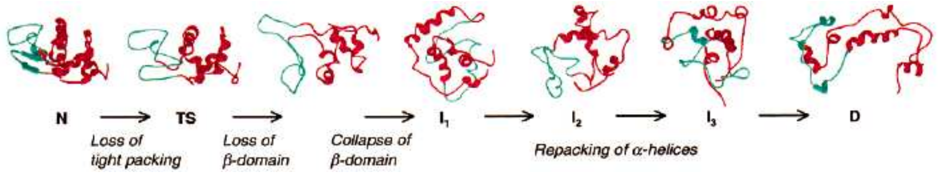

Fig. 1. Thermal Unfolding trajectory to determine the foldign TS.

a conformational clustering procedure (Li & Dagggett 1996). This method assumes that the state of highest free energy relates to leaving the first (most populated) cluster near the native state. This corresponds to a state in which packing interactions are disrupted but the protein has not progressed far enough along the unfolding pathway to allow for a substantial increase in entropy and therefore to a decrease in free energy. The results were compared to experimentally derived measures of the degree of structuring of residues in the transition state based on Fersht's φ–value analysis. The φ–value analysis approach maps precisely particular regions of a protein transition or intermediate state, and relates the destabilization of the TS upon mutation to the effect on stability, i.e.

$$\varphi = \Delta\Delta G_{TS-D}/\Delta\Delta G_{N-D} \quad (1)$$

where φ–values typically range from 0 (denatured like structure for transition state) to 1 (native like structure for transition state). In the case of the proteins used by Daggett et al., hydrophobic deletion mutations were performed throughout the core. Using a geometrical approach based on the relative number of contacts of one particular residue in the native and in the transition state ensembles, Daggett and coworkers were able to generate computationally a set of theoretical φ–values in good agreement with the experimentally determined ones. While novel insights can, and have been gained from this type of molecular simulations, this approach suffers from the potential problem of yielding only anedoctal characterizations of the process of interest, reflecting (and being heavily biased from) the initial conditions sampled in the study (Shea & Brooks 2001).

2.2. Determining Free Energy folding pathways from the knowledge of native state conformation

Brooks and coworkers used a different approach based on the application of biased or umbrella sampling as a framework for calculations of thermodynamic properties for folding projected onto important (and predetermined) progress coordinates, (Shea & Brooks 2001) knowing the structure of the folded state and using it as a target for MD simulations. The free energy surfaces obtained with this procedure, form the basis for exploring the time scale (kinetic aspects) and mechanism of folding for peptides and proteins as well as their thermodynamic properties. The application of biased MD simulations has the primary objective of the calculation of changes in the free energy occurring along a small and specified set of variables describing the transformation of a physical system from one well defined thermodynamic state into another. In the context of folding, these generally refer to thermodynamic states well described by the nature of their configurational distributions. In folding of complex polypeptides such as proteins, such coordinates would distinguish the folded ensemble from the manifold of unfolded configurations. Typical variables might be the radius of gyration, or the number of native contacts. What is really worth noting is that these methods yield free energy curves (or surfaces in higher dimensions) that provide thermodynamic information on populations of conformational states, mechanistic interpretations of folding pathways and approximate kinetic time scales for folding processes form detailed atomic level descriptions of both peptide and solvent. After the initial studies on the mechanism of formation of simple secondary structure motifs in model peptides, in which simple one dimensional progress coordinates could be identified and used to map the free energy change with umbrella sampling along that very coordinate, (Tobias & Sneddon & Brooks 1990; Brooks & Boczko 1995) the attention was shifted to the description of complete folding free energy landscapes. Critical issues were the distribution of conformational states sampled by a folding protein as it approached its native conformational basin and the nature of interactions near the protein folding TS. This approach utilized the following protocol: sampling different regions of conformational space between the folded and the putative unfolding states using methods of high temperature or conformational pulling (Brooks & Boczko 1995); assessing and partitioning the space over which the sampling was run by clustering methods (Brooks & Boczko 1995; Sheinerman & Brooks 1998); employing newly developed umbrella potentials in more collective folding coordinates, such as the radius of gyration, or the number of native contacts, to affect sampling in the partitioned regions of conformational space. The sampling from each of these regions is then combined in total via weighted histogram techniques to give thermodynamic properties, including the folding free energy landscape, as a function of the progress coordinates for folding and under the thermodynamic conditions of the sampling simulations (generally 298K). Using this approach, Brooks and coworkers were able to reconstruct the distribution of distinct conformations sampled during the folding process for fragment B of staphylococcal protein A, a three-helical bundle fragment (Fig. 2). Measuring the progress in folding using the radius of gyration of the protein, the degree of dissimilarity and the free energy for each conformation, it was possible to build a three dimensional surface representing the possible pathways to the native structure. Importantly, while prior to the folding TS the protein samples a very diverse set of conformations almost uniformly, as folding progresses through the transition state, the conformational distribution narrows significantly. The mechanistic implication for this protein is that there are very many different pathways leading to the TS. After passing beyond this region, the number of possible paths to the native basin is drastically reduced. The analysis of the "shape" of the free energy folding surface for proteins with different topologies, show sensitive differences: for the fragment B1 of streptococcal protein G (shortly called herein protein G), and for cold shock protein A (CspA), representative of the α β and all β protein families, the surfaces are much more L–shaped, suggesting a nonspecific collapse preceding the formation of the final ordered interactions. These observations suggest the possibility of kinetic differences between the folding of proteins adopting native states comprised of helical structures and those dominated by longer range interactions, with the latter being slower because of the need to escape local compact conformational traps. In passing, we note that with this approach Brooks and coworkers could analyze in detail the role of water molecules in the folding process, something which is often neglected due to the use of implicit or simplified models of solvation. Water molecules exist in the protein core late in folding and they can act as lubricant, mediating and facilitating the interchange between mis–registered hydrogen bonds and mis–folded helices. Water, thus appears to facilitate the intramolecular search for correct hydrogen bonding partners or side chain pairing in the near native conformations of the protein (Kim & Fisher & Baker & 2000).

2.3. *All–Atom* ab initio *folding simulations*

The methods illustrated so far have had a reasonable degree of success in shedding light on different issues connected with protein folding, but they all rely on a knowledge of the native state, which is used either as a starting point for unfolding, or as the target for biased MD simulations. The first major effort to obtain the native conformation of a reasonably sized protein (Villin Headpiece subdomain, 36 residues) starting from a completely extended conformation and the only knowledge of the sequence is due to Duan and Kollman (Duan &

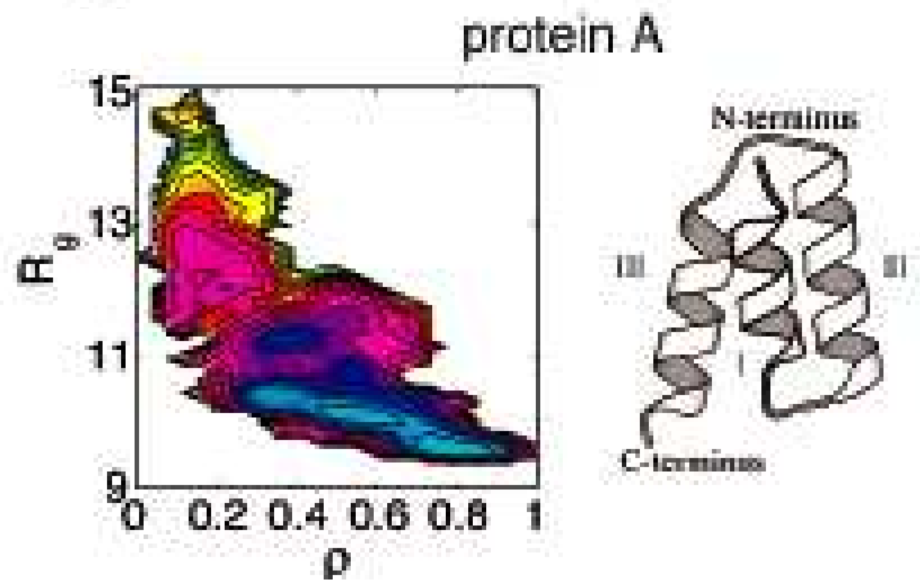

Fig. 2. Folding Free Energy Surface of Protein A.

Kollman 1998). They used an implementation of classical molecular dynamics on parallel computers of increased efficiency for a simulation of protein folding with explicit representation of water for 1 microsecond, about two orders of magnitude longer than the longest simulation reported at that time and still among the longest MD simulations on real systems. Starting with a completely unfolded state of villin headpiece (HP 36), Kollman and Duan observed hydrophobic collapse and helix formation in an initial phase, followed by conformational rearrangements. Most importantly, a marginally stable state was found with a lifetime of 150 ns, a favorable solvation free energy and significant resemblance to the native state structure. During the whole trajectory the protein showed a tendency to principally populate compact states as proved by the values of the radius of gyration, very close to the native one. The main chain RMSD (root mean square deviation) of all residues varied between 1.2 and 0.45 nm, that of the middle portion (residues 9 to 32) fluctuated between 0.88 and 0.3 nm. Up to 80% of the native helical content and up to 62% of the native contacts were formed. Interestingly, the solvation component of the free energy of solvation (SFE) also reached levels comparable to those of the native structure. In terms of mechanism, the folding began with a burst phase, characterized by a steady rise in native helical content and in native contacts and the decrease of the SFE, which lasted from the beginning of the simulation to about 60 ns. The analysis of solvation energy terms and the solvent accessible area indicated that the initial phase was driven by the burial of hydrophobic surface. Therefore, the initial phase closely represents the so called hydrophobic collapse. This occurs on the same time scale as formation of secondary structure. According to Duan and Kollman, this makes physical sense in that a protein, as it buries it hydrophobic groups, tries to avoid burying its hydrogen bonding functions, and secondary structure formation provides a way to do so. A clustering procedure was applied to characterize the most populated conformational families and the analysis of the transitions among different clusters showed that there are two distinct possible pathways for the folding to occur. Moreover, the simulations indicated that tertiary contacts were less likely to form and be maintained in the early stages of folding. Therefore, the formation of tertiary contacts is likely to be the bottleneck of the folding process. These results are consistent with kinetic measurements on the contact order introduced by Baker and coworkers. The work by Duan and Kollman has been of fundamental importance as it proved that the folding of a protein to a set of conformations resemblant of the native state by detailed computer simulations is not so impossible as many workers in the field believed at the moment (and still believe).

2.4. Validation of MD results with experimental data and small peptide folding

The two main problems with MD simulations of real systems with high levels of detail are first of all that existing computers cannot sample enough conformations in a reasonable time to come up with the thermodynamically stable native structure, and, second of all, that available force field descriptions might not be accurate enough to yield the reliable free energy of a conformation. These two aspects are evidenced in the paper by Duan and Kollman, where folding to the stable native state does not occur and the simulation does not seem to contain relevant statistics on the process. The real protein should fold and unfold multiple times, eventually stumbling into the stable conformation with the lowest free energy. An approach to the simulation of multiple folding/unfolding events in polypeptides would be to choose a simpler system on which sufficient statistics can actually be obtained. This is what Daura et al. have done. The method and the attention dedicated to the validation of simulation results by comparison with NMR-NOE derived restraints has made the work by Daura et al. of central importance in the study of biomolecular systems by the application of all–atom MD simulations. These authors could study several short, 3 to 10 residues, β peptides designed to fold into either helical or β-hairpin conformations (Fig. 3)(see van Gunsteren, Angew. Chem. Int. Ed. 2001, 352–355 and refs. therein) (vanGunsteren & Burgi & Peter & Daura 2001). The results of MD simulations of each short peptide at different temperatures and in various solvents showed that the peptides could actually fold and unfold multiple times, also starting from completely extended conformations. The use of multiple trajectories with different temperatures, and the application of a suitable conformational clustering technique allowed to identify the family of most populated native-like conformations and to identify the transitions leading to this cluster from other non-native families. Thanks to the observation of multiple folding–unfolding events, the free energy differences between the folded and unfolded ensembles could be computed and even the melting temperature for the peptides in the force could be estimated.

The validation of these MD simulations was performed by comparing calculated interatomic distances with experimental NOE restraints. This approach proved that even at 340K, in highly denaturing conditions, most of the NOE restraints were not violated in the simulations, a fact that demonstrates on the one hand the high conformational variability of short peptide segments, and, on the other hand, that the denatured states for those peptides comprises vastly less conformations than one would think based only on the number of conformational degrees of freedom present. For peptides with about 20 rotatable bonds, the denatured state can be characterized by about 10^2–10^3 conformers. The small size of the denatures state explains why these peptides fold on a nanosecond time scale (vanGunsteren & Burgi & Peter & Daura 2001; Daura & vanGunsteren & Mark 1999). The finding that the denatured state of small peptides is not completely random has the important consequence of explaining why these peptides fold experimentally on a nanosecond time scale. Extending these concepts to "real" proteins which fold on a millisecond time scale, their denatured state will extend to approximately 10^9 conformers, vastly smaller than the 10^{90} hypothesized earlier. The increase in computational power and, most importantly, the improvement in our force field description of the unfolded state should make the computation of

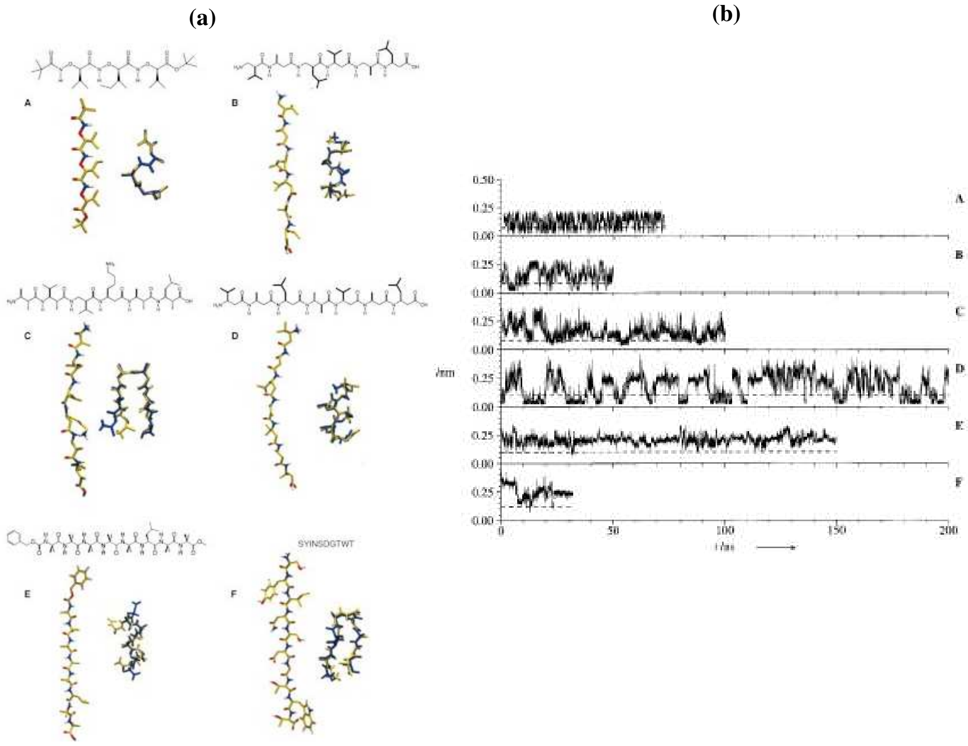

Fig. 3. The peptides (a) and their time evolution in MD (b) studied by Van Gunsteren and coworkers.

folding processes for bigger systems amenable in a limited number of years.

An interesting thing to note about the study of Daura et al. is that evident folding–unfolding behavior is observed only in the case of helical forming peptides: hairpin forming sequences tend to compact conformations without exploring completely extended states. This is another evidence of the fact that the denatured state for peptides is not completely random, and that hairpin forming sequences fold by a first collapse to the compact state followed by conformational rearrangements and hydrogen bond register optimization mediated by the solvent.

2.5. The study of β-sheet forming structures

The study of β-hairpin folding is at the moment of particular interest as the formation of β-sheets seems to be the main cause of the conformational changes involved in several pathological disorders, like Alzheimer's disease and spongiform encephalopathies, that are characterized by the formation of amyloid fibrils (Kelly 1997; Booth 1997; Ma & Nussinov 2002). In this framework, Colombo and coworkers have studied several systems designed to form either stable β-sheets or stable β-hairpins in aqueous solutions. In a first paper, (Colombo & Roccatano & Mark 2002) these authors studied The dynamics of the three-stranded β-sheet peptide Betanova at four different temperatures (280, 300, 350, and 450 K) by long time scale molecular dynamics simulations, in explicit water. Two 20-ns simulations at 280 K indicate that the peptide remains very flexible under "folding" conditions sampling a range of conformations that together satisfy the nuclear magnetic resonance (NMR)-derived experimental constraints. Two simulations at 300 K (above the experimental folding temperature) of 20 ns each show partial formation of ”native”-like structure, which also satisfies most of the NOE constraints at 280 K. At higher temperature, the presence of compact states, in which a series of hydrophobic contacts remain present, are observed. This is consistent with experimental observations regarding the role of hydrophobic contacts in determining the peptide’s stability and in initiating the formation of turns and loops. A set of different structures was shown to satisfy NMR-derived distance restraints and a possible mechanism for the folding of the peptide into the NMR-determined structure is proposed. This work underlined the importance of hydrophobic core formation in imposing the necessary conformational constraints on the peptide to drive it to the native structural basin, and the importance of using an ensemble of multiple conformations to satisfy all the NOE restraints, proving that a single molecule representation of such flexible systems is in general not accurate enough.

The relative importance of the hydrophobic interactions and of the turn sequence in determining the formation of ordered β-hairpin structures was the subject of another paper by Colombo and coworkers. (Colombo & DeMori & Roccatano 2003) The authors investigated the structural determinants of the stability of a designed beta-hairpin containing a natural hydrophobic cluster from the protein GB1 and a D-Pro-Gly turn forming sequence, a sequence known to be a good β-turn inducer. The results of the simulations could shed light on the factors leading to an ordered secondary structure in a model peptide: in particular, the importance of the so-called diagonal interactions in forming a stable hydrophobic nucleus in the beta-hairpin, together with the more obvious lateral interactions, was examined. With the use of long timescale (200 ns) MD simulations in explicit water, the authors showed the role of diagonal interactions in driving the peptide to the correct folded structure (formation of the hydrophobic core with Trp 2, Tyr 4, and Phe 9 in the first stages of refolding) and in keeping it in the ensemble of folded conformations. The special turn sequence favors not only the attainment of the compact structure, but also the correct alignment of the two strands in the hairpin in order to obtain the right hydrogen bonding registry. The combination of the stabilizing effects of the D-Pro-Gly turn sequence and of the hydrophobic nucleus formation thus favors thus an ordered secondary structure compatible with the one determined experimentally. Moreover, the data from these simulations underline the importance of the juxtapositions of the side chains of amino acids not directly facing each other in the three-dimensional structure (diagonal interactions). The combination of these interactions forces the peptide to sample a nonrandom portion of the conformational space, as can be seen in the rapid collapse to an ordered structure in the refolding simulation, and shows that the unfolded state can be closely correlated to the folded ensemble of structures, at least in the case of small model peptides. Given these observations, this paper represents a further evidence of the limited amount of conformational space sampled by peptides in the unfolded or denatured state (Fig. 4).

The study of peptides able to form ordered secondary structure is extremely important also in terms of a more general examination of the folding reaction. The discovery of short sequences derived from bigger proteins with the capability to fold in isolation into the same structure as they have in the original protein, (Espinosa & Munoz & Gellman 2001) helped confirm the theory about the hierarchical nature of protein folding. In other words, if simple isolated sequences isolated in solution can fold into the same structure they display when embedded in a longer sequences in proteins, then we can think of protein folding as a process in which simple elements of secondary structure (Local Elementary Structures, LES) are formed first, and then assembled into higher order architectures stabilized by tertiary contacts. The process of formation and assembly of LES occurs *in vivo* in the hydrophobic internal regions of chaperonins or in the hydrophobic environment represented by the interior of proteins. Thus, a model to mimic these conditions is needed in order to pinpoint all the factors leading to a cor-

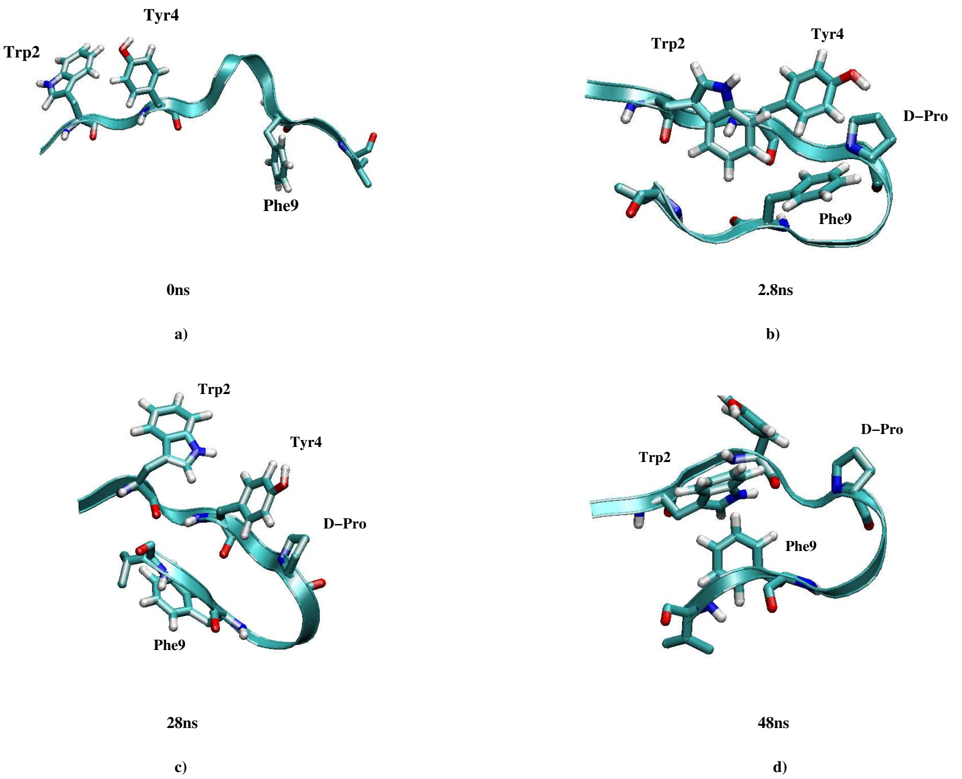

Fig. 4. Spontaneous, fast refolding of a beta-hairpin containing the DPro-Gly turn motif.

rect fold. Experimentally 2,2,2,trifluoroethanol (TFE) has been widely used for its solubilizing and structure inducing properties. From the theoretical point of view, a good model of this membrane environment mimicking solvent has been missing for a long time. Fioroni and coworkers (Fioroni & Burger & Mark & Roccatano 2000) were however able to implement a new model of TFE correctly reproducing all of the thermodynamic characteristics of the bulk solvent. After developing the suitable solvent model, Roccatano and coworkers (Roccatano & Colombo & Fioroni & Mark 2002) used MD simulation techniques to investigate the effect of 2,2,2-trifluoroethanol (TFE) as a cosolvent on the stability of three different secondary structure-forming peptides: the alpha-helix from Melittin, the three-stranded beta-sheet peptide Betanova, and the beta-hairpin 41-56 from the B1 domain of protein G. The peptides were studied in pure water and 30% (vol/vol) TFE/water mixtures at 300 K. The simulations suggest that the stabilizing effect of TFE is induced by the preferential aggregation of TFE molecules around the peptides. This coating displaces water, thereby removing alternative hydrogen-bonding partners and providing a low dielectric environment that favors the formation of intrapeptide hydrogen bonds. Because TFE interacts only weakly with nonpolar residues, hydrophobic interactions within the peptides are not disrupted. As a consequence, TFE promotes sta-

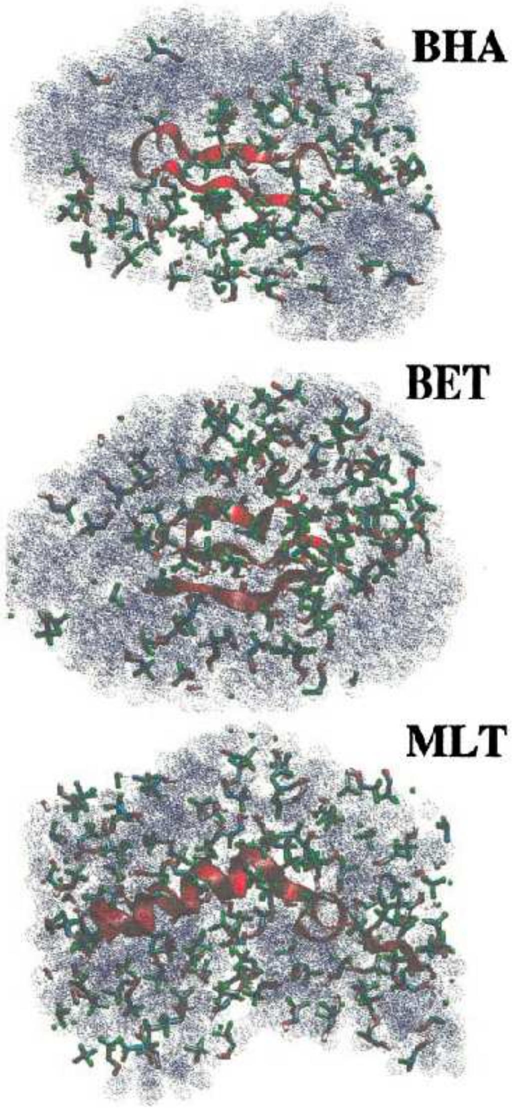

Fig. 5. TFE covering of peptide BHA, Betanova (Bet) and Melittin (MLT).

bility rather than inducing denaturation (Fig. 5).

2.6. A short summary of alternative theoretical approaches to protein folding

All of the examples reported, from the seminal work of Kollman to peptide simulations highlighting the role of the environment on folding, clearly show the importance of examining the full atomic detail and the time evolution of the systems under examination. Important insights can be obtained this way on both the folding process and the stabilizing factors determining the three dimensional structures of peptides and proteins. The drawback of these approaches, is that the number of particles in the systems becomes overwhelmingly big as attention is shifted to bigger proteins. Two approaches have been proposed to overcome this problem. In the first we examine, the solvent degrees of freedom are neglected through the use of implicit solvation models based either on a rescaling of force field interatomic interaction parameters or on the Generalized Born continuum model. The former solvation model was first applied by Lazaridis and Karplus in the study of 24 unfolding trajectories of CI2. (Lazaridis & Karplus 1997) With the use of these multiple unfolding trajectories, the authors were able to reconstruct the folding energy landscape of CI2 reporting the first extensive application of the statistical interpretation of protein folding. In the same framework, Ferrara and Caflish (Ferrara & Caflisch 2000) could build the whole free energy surface of the designed β-sheet forming peptide beta3, using a combination of high and low temperature simulations. This study showed that the use of such simplified models for solvation can overestimate the stability of the folded structure. The use of the GB–SA solvent model seems to yield better representations of peptide systems, while allowing a sensitive increase in the speed of the simulation, thanks to the neglect of solvent degrees of freedom. Simmerling and coworkers were able to refold the 20 residue designed peptide Trp-Cage to within 0.1 nm of main chain rmsd. (Simmerling & Stockbine & Roitberg 2002) The most impressive thing about their work, is that both of these groups were also able to reproduce the relative orientation and packing of the hydrophobic side chains defining the hydrophobic nucleus of the mini-protein.

The second, impressive, approach to overcome the performance problem associated with computers, is represented by the use of a network of thousands of computers scattered all-over the world attempted by the Pande group. (Snow & Nguyen & Pande & Gruebele 2002) The purpose of this approach was to show that it is possible to overcome the current incompatibilities between the time scales of protein folding and molecular dynamics simulation by using a large number of short simulations of only tens of nanoseconds (distributed computing). Pande and coworkers were actually able to obtain an atomic representation of the unfolded state of HP-36 and to study some of its transitions to the native state. This simulation approach is based on the assumption that folding is a first order kinetic process, so that a sufficiently large number of short simulations will include a small number o long time scale events going to completion. But protein folding is not an elementary kinetic process: folding has a series of early conformational steps that lead to lag phases at the beginning of the kinetics. The presence of these lag phases can bias short simulations toward selecting minor pathways that have fewer or faster lag phases, thus missing the major folding pathways. Nevertheless, the procedure of using parallel independent simulations is perfectly valid and quite feasible, once the time scale of simulation proceeds past the lag phases into a single exponential region.

3. Conclusions and Outlook

In this minireview, we have tried to summarize the most important achievements in the field of MD studies of protein and peptide folding, showing the impressive accuracy and level of details these methods can yield, once the appropriate protocols have been put in place and validated against experimental data. Starting from studies in which the knowledge of the folded conformation is necessary, such as the approaches by Daggett and Brooks, the methodology has evolved to the extent that it almost allows the study of complete folding only from sequence. The steady pace improvement of force fields, simulation methodologies and theory combined with the continuous increase in computer performances is allowing all–atom simulations to reach timescales which are closer to the experimental ones. On the other hand, experimental procedures are heading toward levels of detail which were unexpected up to a few years ago, allowing the observation of phenomena on timescales of nanoseconds. The two realms of experiment and theory are thus very likely to be merging in the next few years to yield a unified view of the folding process. This will clarify at atomic level the correlations between sequence and 3D structures of proteins, which is one of the ultimate goals of present day genomic revolution.

Based on these observations, we think that the main future developments and contributions coming from the field of all atom simulations lay in the identification of the effect of mutations on the very early stages of folding, opening the way to the design of highly specific leads and drugs targeted to unfolded or mis–folded states of proteins, with the potential of inhibiting the undesired effects of negative mutations. In this context, a fully detailed atomistic view of the folding process can be highly influential also in understanding the causes of many genetic diseases, for which the mutations on genes have been identified, but no correlation with the gene products are yet available. All–atom theoretical approaches, and MD simulations in particular, are going to be essential in addressing the key questions of atomic interactions in the formation of denatured states and their evolution to the folded ensembles. Many challenges are thus open for the future development of the studies in this exciting field.

References

Booth D., 1997, Nature, 385

Brooks III C.L., 1993, Curr. Op. Struct. Biol. 3, 92

Brooks III C.L., Boczko E.M., 1995, Science, 269, 393

Brooks III C.L., 1998, Curr. Op. Struct. Biol., 8, 222

Caflish A., Karplus M., 1994, Proc. Natl. Acad. Sci. USA, 91, 1746

Chan H.S., Dill K., 1997, Proteins: Struct. Funct. and Genet., 8, 2

Colombo G., Merz Jr. K.M., 1999, J. Am. Chem. Soc., 121, 6895

Colombo G., Roccatano D., Mark A.E., 2002, Proteins: Struct., Funct. and Genet., 46, 380

Colombo G., deMori G.M.S., Roccatano D., 2003, Prot. Sci., 12, 538

Daggett V., Levitt M., 1994, Curr. Opin. in Struct. Biol.4, 291

Daggett V., Levitt M., 1993, J. Mol. Biol., 232, 600

Daggett V., 2000, Curr. Op. Struct. Biol., 10, 160

Dill K.A., Chan H.S., 1997, Nat. Struct. Biol., 4, 10

Daura X., van Gunsteren W.F., Mark A.E., 1999, Proteins, 34, 269

Duan Y., Kollman P.A., 1998, Science, 282, 740

Espinosa J.F., Munoz V., Gellman S.H., 2001, J.Mol.Biol., 306, 397

Ferrara P., Caflisch A., 2000, Proc. Natl. Acad. Sci. USA, 97, 10780

Fersht A.R., 1995, Curr.Opin.Struct.Biol., 5, 79

Fioroni M., Burger K., Mark A.E., Roccatano D., 2000, J. Phys. Chem. B, 104, 12347

Kelly J., 1997, Structure, 5, 595

Kim D.E., Fisher C., Baker D., 2000, J. Mol. Biol., 298, 971

Lazaridis T., Karplus M., 1997, Science, 278, 1928

Li A., Daggett V., 1996, J. Mol. Biol., 277, 412

Ma B.Y., Nussinov R., 2002, Proc. Natl. Acad. Sci. USA, 99, 14126

Mayor U., 2003, Nature, 421, 863

Roccatano D., Colombo G., Fioroni M., Mark A.E., 2002, Proc. Natl. Acad. Sci. USA, 99, 12179

Settanni G., Hoang T.X., Micheletti C., Maritan A., 2002, Biophysical J., 83, 3533

Shea J.E., Brooks III C.L., 2001, Annu. Rev. Phys. Chem., 52, 499

Sheinerman F.B., Brooks III C.L., 1998, J. Mol. Biol., 278, 439

Shortle D., Ackerman M.S., 2001, Science, 293, 487

Simmerling C., Stockbine B., Roitberg A.E., 2002, J. Am. Chem. Soc, 124, 11258

Snow C.D., Nguyen H., Pande V.S., Gruebele M., 2002, Nature, 420, 33

Tobias D.J., Sneddon S.F., Brooks III C.L., 1990, J. Mol. Biol., 216, 783

van Gunsteren W.F., Burgi R., Peter C., Daura X., 2001, Angew. Chem. Int. Ed., 40, 352