doi:10.1016/j.jmb.2004.06.063 *J. Mol. Biol.* (2004) **xx**, 1–8

JMB

Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

The Protein Folding Network

Francesco Rao and Amedeo Caflisch*

Department of Biochemistry
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich, Switzerland

*Corresponding author

The conformation space of a 20 residue antiparallel β-sheet peptide, sampled by molecular dynamics simulations, is mapped to a network. Snapshots saved along the trajectory are grouped according to secondary structure into nodes of the network and the transitions between them are links. The conformation space network describes the significant free energy minima and their dynamic connectivity without requiring arbitrarily chosen reaction coordinates. As previously found for the Internet and the World-Wide Web as well as for social and biological networks, the conformation space network is scale-free and contains highly connected hubs like the native state which is the most populated free energy basin. Furthermore, the native basin exhibits a hierarchical organization, which is not found for a random heteropolymer lacking a predominant free-energy minimum. The network topology is used to identify conformations in the folding transition state (TS) ensemble, and provides a basis for understanding the heterogeneity of the TS and denatured state ensemble as well as the existence of multiple pathways.

Keywords: complex networks; protein folding; energy landscape; transition state; denatured state ensemble

Proteins are complex macromolecules with many degrees of freedom. To fulfil their function they have to fold to a unique three-dimensional structure (native state). Protein folding is a complex process governed by non-covalent interactions involving the entire molecule. Spontaneous folding in a time-range of microseconds to seconds[1] can be reconciled with the large amount of conformers by using energy landscape analysis.[2–4] The main difficulty of this analysis is that the free energy has to be projected on arbitrarily chosen reaction coordinates (or order parameters). In many cases, a simplified representation of the free-energy landscape is obtained where important information on the non-native conformation ensemble and the folding TS ensemble are hidden. Moreover, the possible transitions between free-energy minima cannot be displayed in such projections, which hinders the study of pathways and folding intermediates. The characterization of the free-energy minima and the connectivity among them, i.e. possible transitions between minima, for peptides and proteins is still a challenging problem despite the fact that several elegant approaches have been proposed.[5–7]

In the last five years, many complex systems, like the World-Wide Web, metabolic pathways, and protein structures have been modeled as networks.[8–11] Intriguingly, common topological properties have emerged from their organization.[12] The conformation space of a short two-dimensional lattice polymer chain has been mapped to a network where a link between two nodes indicates the interconversion in a single Monte Carlo move of the chain.[13] A description of the potential energy landscape without the use of any projection has been given in terms of networks for a Lennard–Jones cluster of atoms.[14]

Here, we use complex network analysis[12] to study the conformation space and folding of beta3s, a designed 20 residue sequence whose solution conformation has been investigated by NMR spectroscopy.[15] The NMR data indicate that beta3s in aqueous solution forms a monomeric (up to more than 1 mM concentration) triple-stranded antiparallel β-sheet (Figure 1, bottom), in equilibrium with the denatured state.[15] We have previously shown that in implicit solvent[16] molecular dynamics simulations beta3s folds reversibly to the NMR solution conformation, irrespective of the starting conformation.[17,18] We consider

Abbreviations used: RMSD, root-mean-square deviations; TS, transition state; TR, trap; FS, folded state.

E-mail address of the corresponding author: caflisch@bioc.unizh.ch

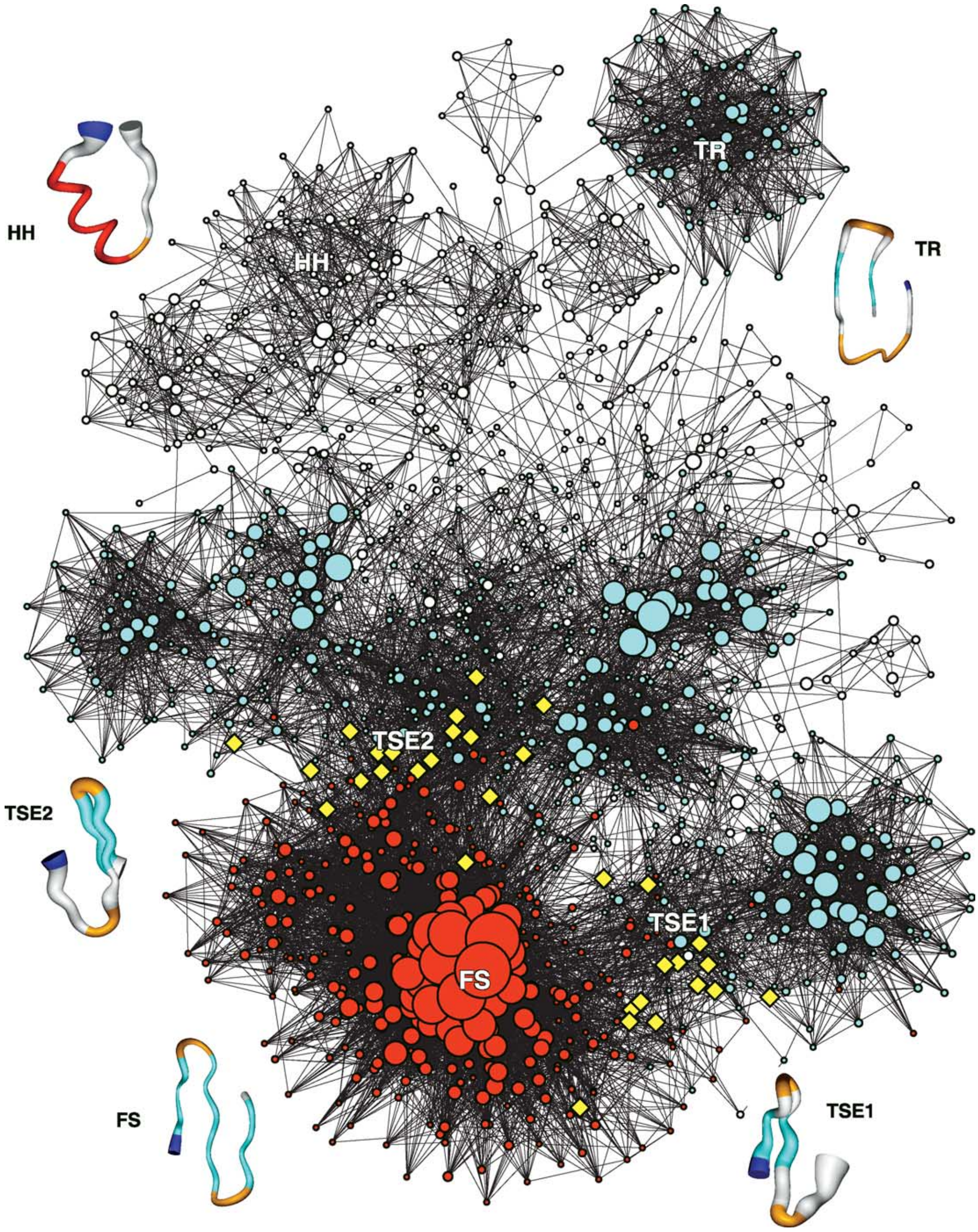

Figure 1. The beta3s conformation space network. The size and color coding of the nodes reflect the statistical weight w and average neighbor connectivity k_{nn} respectively. White, cyan, and red nodes have $k_{nn} < 30$, $30 \le k_{nn} \le 70$, and $k_{nn} > 70$, respectively. Representative conformations are shown by a pipe colored according to secondary structure: white stands for coil, red for α-helix, orange for bend, cyan for strand and the N terminus is in blue. The variable radius of the pipe reflects structural variability within snapshots in a conformation. The yellow diamonds are folding TS conformations (TSE1, TSE2, see the text for details) characterized by a connectivity/weight ratio $k/2\bar{w} > 0.3$, a clustering coefficient $C < 0.3$, and $60 < k_{nn} < 80$. This Figure was made using visone (www.visone.de) and MOLMOL[40] visualization tools.

conformations sampled by molecular dynamics simulations and the transitions between them as the network nodes and links, respectively. The network analysis allows us to identify the topological properties that are common to both beta3s, which folds to a unique three-dimensional structure,[15,19] and a random heteropolymer which lacks a single preferential conformation like the native state despite the fact that it has the same residue composition as beta3s. These properties include the presence of several free-energy minima and highly connected conformations (hubs). On the other hand, a hierarchical modularity[20] in the proximity of the native state is peculiar of a folding sequence.

Model and Methods

Molecular dynamics simulations

The simulations and part of the analysis of the trajectories were performed with the program CHARMM.[21] beta3s was modeled by explicitly considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field.[21]) A mean field approximation based on the solvent-accessible surface was used to describe the main effects of the aqueous solvent on the solute.[16] The two parameters of the solvation model were optimized without using beta3s. The same force field and implicit solvent model have been used recently in molecular dynamics simulations of the early steps of ordered aggregation,[22] and folding of structured peptides (α-helices and β-sheets) ranging in size from 15 to 31 residues,[16,17,23] as well as small proteins of about 60 residues.[24,25] Despite the absence of collisions with water molecules, in the simulations with implicit solvent the separation of time-scales is comparable with that observed experimentally. Helices fold in about 1 ns,[26] β-hairpins in about 10 ns[26] and triple-stranded β-sheets in about 100 ns,[18] while the experimental values are $\sim$0.1 μs,[27] $\sim$1 μs[27] and $\sim$10 μs,[15] respectively. Recently, four molecular dynamics simulations of beta3s were performed at 330 K for a total simulation time of 12.6 μs.[19] There are 72 folding events and 73 unfolding events, and the average time required to go from the denatured state to the folded conformation is 83 ns. The 12.6 μs of simulation length is about two orders of magnitude longer than the average folding or unfolding time, which are similar because at 330 K the native and denatured states are almost equally populated.[19] For the network analysis the first 0.65 μs of each of the four simulations were neglected so that along the 10 μs of simulations there are a total of 5×10^5 snapshots because coordinates were saved every 20 ps. The sequence of the random heteropolymer is a randomly scrambled version of the beta3s sequence with the same residue composition. It was simulated for 2 μs and 10^5 snapshots were saved. The conditions for the molecular dynamics simulations, i.e. force field, solvation model, temperature, and time interval between saved snapshots were the same for both peptides.

Construction of the protein folding network

To define the nodes and links of the network the secondary structure was calculated[28] for each snapshot (Cartesian coordinates of the atomic nuclei) saved along the molecular dynamics trajectory. A “conformation” is a single string of secondary structure,[28] e.g., the most populated conformation for beta3s (FS in Figure 1) is:
`-EEEESSEEEEEES SEEEE-`
There are eight possible “letters” in the secondary structure “alphabet”: “H”, “G”, “I”, “E” “B”, “T”, “S”, and “-”, standing for α-helix, 3_{10} helix, π-helix, extended, isolated β-bridge, hydrogen bonded turn, bend, and unstructured, respectively. Since the N and C-terminal residues are always assigned an “-”[28] a 20 residue peptide can, in principle, assume $8^{18}\approx10^{16}$ conformations. Conformations are nodes of the network and the transitions between them are links. A weight $\bar{w}$ is assigned to each node to take into account the free-energy of each conformation and is equal to the number of snapshots with a given secondary structure string. The statistical weight w of a node is equal to $w=\bar{w}/N$, where N is the total number of snapshots in the simulation (N is equal to 5×10^5 and 10^5 for beta3s and the random heteropolymer, respectively). Considering all the conformations visited during a microsecond-scale simulation can yield to a computationally intractable network size. For this reason we used for the network analysis the 1287 conformations of beta3s with significant weight ($\bar{w}\geq20$ per conformation). Two nodes are connected by an undirected link (and called neighbors) if they either include a pair of snapshots that are visited within 20 ps or they are separated by one or more conformations with less than 20 snapshots each. For the 2 μs of the random heteropolymer, a threshold of $\bar{w}\geq4$ was used, so that $w\geq4\times10^{-5}$ as in the beta3s network. The choice of a threshold value is somewhat arbitrary but the network properties are robust for a large range of threshold values (see Supplementary Material).

The properties of the network are robust also with respect to the length of the simulation time and the definition of the nodes. The topological properties are independent from simulation lengths if one considers more than 2 μs. The correlation between statistical weight and connectivity, as well as power-law behavior of the connectivity distribution, and $1/k$ behavior of the clustering coefficient distribution (see below) are essentially identical after 2 μs, 4 μs, and 10 μs. As an example, the exponent of the power-law is 2.0 for the beta3s networks based on 2 μs, 4 μs and 10 μs of simulation time. Defining nodes by grouping snapshots according to root-mean-square deviations (RMSD) in coordinates of C^{α}–C^{β} atoms yields the same overall properties, i.e. power-law distribution of the links (with a scaling factor γ of 2.2) and $1/k$ tail of the clustering distribution. Grouping snapshots according to secondary structure motifs does not require the use of an arbitrarily chosen RMSD cutoff, and is able to capture the fluctuations of partially structured conformations.[28]

Evaluation of P_{fold}

The TS ensemble can be defined as the set of structures which have the same probability of folding (P_{fold}) or unfolding in trajectories started with varying initial conditions.[29] For each putative TS conformation, the probability to fold before unfolding was calculated by 100 very short

Table 1. Energetic comparison of folded and denatured state

	$\langle E\rangle$[a]	$\langle\Delta\mathcal{F}\rangle$[b]
Folded state (FS)		
-EEEESSEEEEEEESSEEEE-	−7.6	0
-EEE-STTEEEEEESEEEE-	−8.6	0.1
-EEEESSEEEEE-STTEEE-	−8.4	0.5
-EEE-STTEEEE-STTEEE-	−9.2	0.7
Helical conformations (HH)		
---HHHHHHHHHHS------	0.9	3.1
-HHHHHHHHHHHHS------	−1.9	3.3
---HHHHHHHHHHTT-----	0.7	3.5
---HHHHHHHHH-------	0.5	3.7
-HHHHHHHHHHHHTT-----	−0.8	3.7
--TT--HHHHHHHHHHHH-	−0.8	3.8
Curl-like trap (TR)		
---SSGGG-EEE-STTTEE-	−7.8	3.4
---SSSS--EEE-STTTEE-	−7.0	3.5
---S-GGG-EEE-STTTEE-	−9.3	3.7
---SSGGG-EEE-SGGGEE-	−9.6	3.7
---SSTTT-EEE-STTTEE-	−8.4	3.7

The free-energy of conformation *i* is $\mathcal{F}_i=-k_\mathrm{B}T\ 1og(w_i)$, where w_i is the probability along the trajectory to find the peptide in the conformation *i*.

[a] Average effective energy.

[b] Free-energy relative to the most populated conformation. All values are in kcal/mol. The conformational entropy of the peptide is equal to $(\langle\mathcal{E}\rangle-\mathcal{F})/T$. Note that the curl-like traps are entropically penalized with respect to the native state.

trajectories at 330 K started from ten snapshots within a node. The only difference between the ten runs was the seed for the random number generator used for the initial assignment of the atomic velocities. A trajectory was considered to lead to folding (unfolding) if it visits first structures with a fraction of native contacts $Q>22/26$ ($Q<4/26$).[17] The 33,381 snapshots with $Q>22/26$ have a distribution of the pairwise C^α RMSD peaked at 1.1 Å (see Supplementary Material).

Results and Discussion

To study the conformation space network of polypeptides we concentrate on the analysis of topology, i.e. on the study of the connectivity between different conformations, leaving for a later study the analysis of transition rates. We have investigated the network topologies of several peptides but, here, we focus on beta3s and the random scrambled version of it. Additional details can be found in the Supplementary Material, where the network properties of another structured peptide and a glycine homopolymer are presented.

Conformation space network of a structured peptide

The conformation space network and relevant structures of beta3s are shown in Figure 1. The group of nodes at the bottom of Figure 1 (red nodes) represents the native state basin (FS). The native basin is connected to a wide region of nodes with significant native content (cyan circles in the middle of Figure 1). Although many heterogeneous routes can be taken to reach the folded state (in agreement with lattice simulations),[30,31] most of the folding events have common structural features that define two average folding pathways. The less frequented average pathway[18] (see the density of transitions in Figure 1, bottom right) consists of conformations that have the N-terminal hairpin formed while the C-terminal strand is mostly unstructured with non-native hydrogen bonds at the turn (TSE1 in Figure 1). The second and most frequented average pathway includes conformations with a well formed C-terminal hairpin while the N-terminal strand is disordered (TSE2 in Figure 1), namely it can be out-of-register or mostly unstructured. It is interesting to note that the same two folding pathways were observed experimentally for a 24 residue peptide with the same folded state as beta3s.[32] Furthermore, multiple folding pathways have recently been detected by kinetic analysis of a β-sandwich protein.[33]

The denatured state ensemble is very heterogeneous and includes high-enthalpy, high-entropy conformations (e.g. the partially helical conformations, denoted HH in Figure 1) but also low-enthalpy, low-entropy conformations (e.g., the curl-like trap, TR). The former are loosely linked clusters of conformations with similar secondary structure (see Table 1) which are characterized by an unfavorable effective energy (sum of peptide potential energy and solvation energy) and fluctuating unstructured residues (e.g. the terminal of the helix shown on top left of Figure 1). On the contrary, low-enthalpy, low-entropy traps form tightly linked clusters with almost identical secondary and tertiary structure, favorable effective energy (similar to the one of the native structure, see Table 1) and no fluctuating residues (e.g. Figure 1, top right). Taken together, these results indicate that FS is entropically favored over low-enthalpy conformations like TR, i.e. FS has more flexibility than TR. A possible

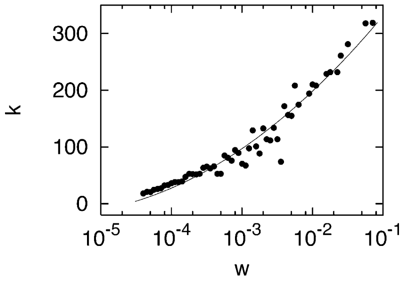

Figure 2. Correlation between the statistical weight w and the connectivity k for beta3s. The connectivity can be fitted to $\log^2(w)$ (with a correlation coefficient of 0.88, continuous line) indicating a deviation from a purely diffusive dynamics where $k \sim w$. The correlation and the fit are calculated over all nodes of the network but in the Figure logarithmic binning is applied to reduce noise.

explanation is that the C-terminal carboxy group is involved in four hydrogen bonds in TR (with the backbone NH groups of residues 4–7), whereas both termini undergo rather large fluctuations in FS. In addition, a more favorable van der Waals energy in TR is consistent with a denser packing in TR than in FS. Entropically favored structures (like FS) are destabilized by lowering the temperature. Hence, there should be a temperature (not accessible to conventional MD simulations) where the system becomes frustrated and a glass-like scenario emerges.

Note that the network description of non-native conformations is more detailed than the one obtained by projecting the free energy surface on progress variables (e.g. based on fraction of native contacts). In such projections, for low values of the fraction of native contacts structures as diverse as helices and the curl-like conformations mentioned above are not distinguished. Even the ensemble with half of the native contacts is heterogeneous and hard to classify. Using as reaction coordinate the RMSD (with respect to a given structure) or the radius of gyration is even less selective. Only when a clever combination of variables is used is it possible to have a more detailed description of the free-energy landscape. The network description of the conformation space gives a synthetic and systematic view of all the possible conformations accessed by the system and their transitions. By considering the statistical weight of the nodes a thermodynamical description of the system is obtained.

The high correlation between the statistical weight of a node and its number of links (Figure 2) shows that the most connected nodes are also low-lying minima on the free-energy landscape. This indicates that the conformation space network describes the significant free energy minima and their dynamic connectivity, without projection, where highly populated nodes are minima of free-energy and the set of nodes densely connected to them make up the basins of such minima. The connectivity can be fitted to $\log^2(w)$, which indicates that the dynamics is not diffusive (see Figure 2).

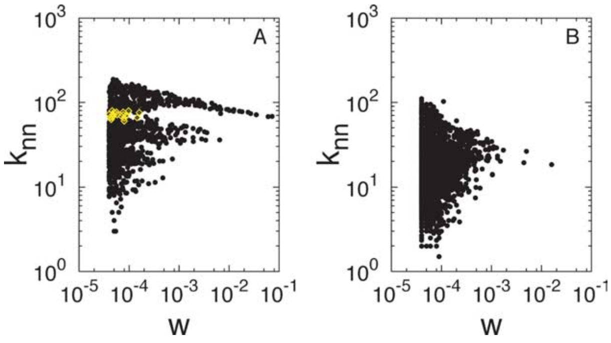

Figure 3. Average neighbor connectivity k_{nn} plotted as a function of the statistical weight for the 1287 nodes of beta3s (A) and for the 2658 nodes of the random heteropolymer (B). k_{nn} of node i is the average number of links of the neighbors of node i. The yellow diamonds are folding TS conformations (see also Figure 1 and the text) characterized by a connectivity/weight ratio $k/2\bar{w} > 0.3$, a clustering coefficient $C < 0.3$, and $60 < k_{nn} < 80$.

Folding and network topology

The average neighbor connectivity k_{nn} of beta3s (Figure 3A), i.e. the average number of links of the neighbors of a given node, is rather heterogeneous, highlighting the presence of different connection rules in different regions of the network. This is not the case for the random heteropolymer (Figure 3B), whose basins have organization and statistical

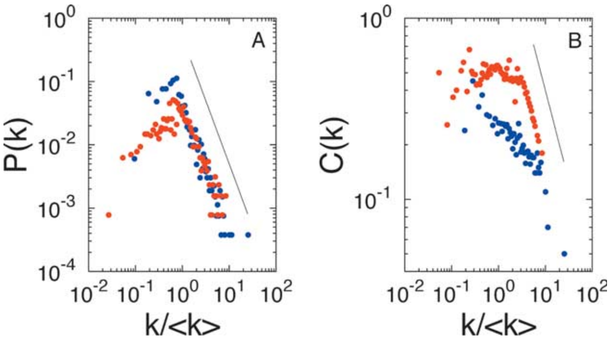

Figure 4. Topological properties of conformation space networks. Red and blue data points are plotted for beta3s and a random heteropolymer, respectively. For a direct comparison, the connectivity k is normalized by the average connectivity $\langle k \rangle$ of each network. Logarithmic binning is applied to reduce noise. A, The connectivity distribution $P(k)$ is the probability that a node (conformation) has k links (neighbor conformations). The straight line corresponds to a power-law fit $y = x^{-\gamma}$ on the tail of the distribution with $\gamma = 2.0$. B, The clustering coefficient C describes the cliques of a node. For node i it is defined as $C_i = 2n_i/k_i(k_i - 1)$, where k_i is the number of neighbors of node i and n_i is the total number of connections between them. Values of C are averaged over the nodes with k links. The straight line corresponds to a power-law fit $y = x^{-1}$ on the tail of the distribution of beta3s.

weight similar among each other as previously found for most homopolymers.[10] Note that for beta3s the native state is well discriminated by k_{nn} (red nodes in Figure 1 and top band in Figure 3A).

The connectivity distribution of conformation space networks shows a well pronounced power-law tail $P(k)\sim k^{-y}$ with $\gamma=2.0$ for both beta3s and the random heteropolymer (Figure 4A) as well as another structured peptide[34] and homoglycine, i.e. $(Gly)_{20}$ (see Supplementary Material). The power-law is due to the presence of a few largely connected "hubs" while the majority of the nodes have a relatively small number of links.[35] This behavior has been previously observed for several biological,[8] social[36] and technological networks,[9] which in the literature take the name of scale-free networks. In terms of free energy this means that only a few low lying minima are present but they act as "hubs" with a large number of routes to access them.

The average clustering coefficient C is a measure of the probability that any two neighbors of a node are connected. beta3s and the heteropolymer have C values of 0.49 and 0.28, respectively. These values are one order of magnitude larger than random realizations of the two networks with the same amount of nodes and links. The native basin of beta3s includes the nodes with the largest number of links of the network. These nodes give rise to the $1/k$ tail of the clustering distribution (Figure 4B), i.e. an inherently hierarchical organization[20] of the conformations in the native basin of beta3s. Such organization is not apparent for the non-native region of beta3s and the random heteropolymer. Note that the power-law scaling of the connectivity distribution can be considered as a general property of free-energy landscapes of polypeptides, whereas a hierarchical organization of the nodes reflects a pronounced free-energy basin of attraction (like the native state).

Transition state ensemble

As mentioned above, folding is a complex process with many degrees of freedom involved and it is difficult (or even not possible) to define a single reaction coordinate to monitor folding events.[37,38] Hence, it is very difficult to isolate transition state (TS) conformations from equilibrium sampling. The TS conformations are saddle points, i.e. local maxima with respect to the reaction coordinate for folding and local minima with respect to all other coordinates. For this reason, we identified the nodes with a high connectivity/weight ratio $k_i/2\bar{w}>0.3$ and low clustering coefficient value C_i as putative TS conformations. The former criterion guarantees that these nodes are accessed and exited, most of the time, by a different route, i.e. they can be directly reached from different conformations of the network space. The low clustering coefficient value guarantees that the neighbors of these conformations are likely to be disconnected. These two conditions are necessary but not sufficient because they do not distinguish folding TS conformations from saddle points between unfolded conformations. Since the folding TS conformations are linked to both nodes in the native state (having large number of links) and in the denatured state (small/intermediate number of links), we speculated that folding TS conformations should have values of the average neighbor connectivity k_{nn} within a certain range. For nodes with high connectivity/weight ratio and low clustering coefficient, a remarkable correlation of 0.89 was found between the average neighbor connectivity k_{nn} and P_{fold} (Figure 5), which is the probability of a given conformation to fold before unfolding.[29] A P_{fold} value close to 0.5 is expected for conformations on top of the folding TS barrier[25] and the correlation suggests that network properties can be used to predict folding TS conformations. These are shown in Figures 1 and 3A with yellow diamonds. As discussed above, two main average folding pathways are observed. The less frequent one is characterized by a TS ensemble of conformations with the first hairpin in a native form (residues 1–13) and a bend corresponding to the second native turn (residues 14 and 15). The C-terminal residues form a straight structure with almost no contacts, either native or non-native. The second average pathway shows a TS with the second native hairpin formed (residues 7–20) and a bend corresponding to the first native turn (residues 5 and 6). Such a symmetrical behavior is presumably due to the simplicity and symmetry of the native conformation as well as the symmetry in the sequence (sequence identity of 67% between the two hairpins). The folding TS conformations of beta3s form a heterogeneous ensemble with C^{α} RMSD within contributing structures between 3 Å and 6 Å. In contrast to previous molecular dynamics studies in which progress variables based on fraction of native contacts were used to describe TS conformations,[17,39] the network properties yield a description of the folding TS ensemble (Figure 1) which does not depend on the choice of reaction coordinates. Interestingly, the folding TS conformations of beta3s have about one-half of the

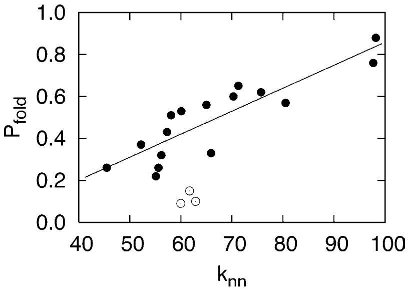

Figure 5. Correlation between P_{fold} and average neighbor connectivity k_{nn}. Three nodes used as a negative control (low connectivity/weight ratio and/or high clustering coefficient but similar fraction of native contacts) are shown with open circles.

native contacts formed but this is not a sufficient criterion (Table S1 in Supplementary Material). Moreover, there is no correlation between the fraction of native contacts and the probability of folding. As a control, P_{fold} values smaller than 0.15 were obtained for five nodes with an average fraction of native contacts similar to the folding TS conformations but low connectivity/weight ratio and/or high clustering coefficient.

Conclusions

Complex network theory was used to analyze the conformation space of a structured peptide and that of a random heteropolymer of the same residue composition. Four main results have emerged. First, as it was already observed for a variety of networks as diverse as the World-Wide Web and the protein interactions in a cell, the conformation space network of polypeptide chains is a scale-free network (power-law behavior of the degree distribution). Second, the native basin of the structured peptide shows a hierarchical organization of conformations. This organization is not observed for the random heteropolymer which lacks a native state. Third, free energy minima and their connectivity emerge from the network analysis without requiring projections into arbitrarily chosen reaction coordinates. As a consequence, it is found that the denatured state ensemble is very heterogeneous and includes high-entropy, high-enthalpy conformations as well as low-entropy, low-enthalpy traps. Fourth, the network properties were used to identify TS conformations and two main average folding pathways. It was found that the average neighbor connectivity k_{nn} correlates with P_{fold}, the probability of folding. P_{fold} is computationally very expensive to evaluate. Hence, it will be important to generalize this result by analyzing other structured peptides, which is work in progress in our research group. In conclusion, the network analysis seems to be particularly useful to study the conformation space and folding of structured peptides including the otherwise elusive TS ensemble.

Acknowledgements

We thank M. Cecchini, Professor P. De Los Rios, E. Guarnera, Professor M. Karplus, Dr E. Paci, Dr M. Seeber and Dr G. Settanni for interesting discussions. The molecular dynamics simulations were performed on the Matterhorn Beowulf cluster at the Informatikdienste of the University of Zurich. We thank C. Bollinger and Dr A. Godknecht for their help in setting up and maintaining the cluster, and the Canton of Zurich for generous hardware support. This work was supported by the Swiss National Science Foundation.

Supplementary data

Supplementary data associated with this article can be found on doi:10.1016/j.jmb.2004.06.063.

References

1. Daggett, V. & Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18–25.
2. Bryngelson, J. & Wolynes, P. (1989). Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915.
3. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl Acad. Sci. USA*, **89**, 8721–8725.
4. Karplus, M. (1997). The Levinthal paradox: yesterday and today. *Fold. Des.* **2**, S69–S75.
5. Becker, O. M. & Karplus, M. (1997). The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. *J. Chem. Phys.* **106**, 1495–1517.
6. Wales, D., Doye, J., Miller, M., Mortenson, P. & Walsh, T. (2000). Energy landscapes: from clusters to biomolecules. *Advan. Chem. Phys.* **115**, 1–111.
7. Krivov, S. V. & Karplus, M. (2002). Free energy disconnectivity graphs: application to peptide models. *J. Chem. Phys.* **117**, 10894–10903.
8. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.
9. Albert, R., Jeong, H. & Barabási, A.-L. (1999). Diameter of the World-Wide Web. *Nature*, **401**, 130–131.
10. Vendruscolo, M., Dokholyan, N. V., Paci, E. & Karplus, M. (2002). Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. ser. E*, **65**, 061910.1–061910.4.
11. Greene, L. H. & Higman, V. A. (2003). Uncovering network systems within protein structures. *J. Mol. Biol.* **334**, 781–791.
12. Newman, M. (2003). The structure and function of complex networks. *SIAM REV.* **45**, 167–256.
13. Scala, A., Amaral, L. A. N. & Barthelemy, M. (2001). Small-world networks and the conformation space of a short lattice polymer chain. *Europhys. Letters*, **55**, 594–600.
14. Doye, J. (2002). Network topology of a potential energy landscape: a static scale-free network. *Phys. Rev. Letters*, **88**, 238701.
15. De Alba, E., Santoro, J., Rico, M. & Jiménez, M. A. (1999). De novo, design of a monomeric three-stranded antiparallel β-sheet. *Protein Sci.* **8**, 854–865.
16. Ferrara, P., Apostolakis, J. & Caflisch, A. (2002). Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Struct., Funct. Genet.* **46**, 24–33.
17. Ferrara, P. & Caflisch, A. (2000). Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl Acad. Sci. USA*, **97**, 10780–10785.
18. Cavalli, A., Ferrara, P. & Caflisch, A. (2002). Weak temperature dependence of the free energy surface and folding pathways of structured peptides. *Proteins: Struct., Funct. Genet.* **47**, 305–314.

19. Cavalli, A., Haberthür, U., Paci, E. & Caflisch, A. (2003). Fast protein folding on downhill energy landscape. *Protein Sci.* **12**, 1801–1803.
20. Ravasz, E. & Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Phys. Rev. ser. E*, **67**, 026112.
21. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
22. Gsponer, J., Haberthür, U. & Caflisch, A. (2003). The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid-forming peptide from the yeast prion sup35. *Proc. Natl Acad. Sci. USA*, **100**, 5154–5159.
23. Hiltpold, A., Ferrara, P., Gsponer, J. & Caflisch, A. (2000). Free energy surface of the helical peptide Y(MEARA)$_6$. *J. Phys. Chem. ser. B*, **104**, 10080–10086.
24. Gsponer, J. & Caflisch, A. (2001). Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *J. Mol. Biol.* **309**, 285–298.
25. Gsponer, J. & Caflisch, A. (2002). Molecular dynamics simulations of protein folding from the transition state. *Proc. Natl Acad. Sci. USA*, **99**, 6719–6724.
26. Ferrara, P., Apostolakis, J. & Caflisch, A. (2000). Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. ser. B*, **104**, 5000–5010.
27. Eaton, W. A., Munoz, V., Hagen, S., G, S., Jas, L. J., Lapidus, E. R. & Henry, J. (2000). Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359.
28. Andersen, C. A. F., Palmer, A. G., Brunak, S. & Rost, B. (2002). Continuum secondary structure captures protein flexibility. *Structure*, **10**, 174–184.
29. Du, R., Pande, V., Grosberg, A., Tanaka, T. & Shakhnovich, E. (1998). On the transition coordinate for protein folding. *J. Chem. Phys.* **108**, 334–350.
30. Onuchic, J., Socci, N., Luthey-Schulten, Z. & Wolynes, P. (1996). Protein folding funnels: the nature of the transition state ensemble. *Fold. Des.* **1**, 441–450.
31. Schonbrun, J. & Dill, K. A. (2003). Fast protein folding kinetics. *Proc. Natl Acad. Sci. USA*, **100**, 12678–12682.
32. Griffiths-Jones, S. R. & Searle, M. S. (2000). Structure, folding, and energetics of cooperative interactions between the β-strands of a *de novo* designed three-stranded antiparallel β-sheet peptide. *J. Am. Chem. Soc.* **122**, 8350–8356.
33. Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. (2003). Parallel protein-unfolding pathways revealed and mapped. *Nature Struct. Biol.* **10**, 658–662.
34. Demarest, S. J., Hua, Y. X. & Raleigh, D. P. (1999). Local interactions drive the formation of nonnative structure in the denatured state of human alpha-lactalbumin: a high resolution structural characterization of a peptide model in aqueous solution. *Biochemistry*, **38**, 7380–7387.
35. Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, **286**, 509–512.
36. Newman, M. (2001). The structure of scientific collaboration networks. *Proc. Natl Acad. Sci. USA*, **98**, 404–409.
37. Chan, H. S. & Dill, K. A. (1998). Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins: Struct., Funct. Genet.* **30**, 2–33.
38. Karplus, M. (2000). Aspects of protein reaction dynamics: deviations from simple behaivor. *J. Phys. Chem. ser. B*, **104**, 11–27.
39. Lazaridis, T. & Karplus, M. (1997). "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science*, **278**, 1928–1931.
40. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

Edited by J. Thornton

(Received 23 March 2004; received in revised form 10 June 2004; accepted 15 June 2004)