PROTEINS: Structure, Function, and Bioinformatics 54:513–525 (2004)

Comparison of the Transition State Ensembles for Folding of Im7 and Im9 Determined Using All-Atom Molecular Dynamics Simulations With ϕ Value Restraints

Emanuele Paci,[1] Claire T. Friel,[2] Kresten Lindorff-Larsen,[4] Sheena E. Radford,[2] Martin Karplus,[3] and Michele Vendruscolo[4]*

[1]*Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland*
[2]*School of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom*
[3]*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138*
[4]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

ABSTRACT **Delineation of the structural properties of transition states is key to deriving models for protein folding. Here we describe the structures of the transition states of the bacterial immunity proteins Im7 and Im9 obtained by all-atom molecular dynamics simulations with ϕ value restraints derived from protein engineering experiments. This pair of proteins is of special interest because, at pH 7 and 10 °C, Im7 folds via an intermediate while Im9 folds with a two-state transition. The structures of the transition states for Im7 and Im9, together with their radii of gyration and distances from the native state, are similar. The typical distance between any two members of the transition state ensemble of both proteins is large, with that of Im9 nearly twice that of Im7. Thus, a broad range of structures make up the transition state ensembles of these proteins. The ensembles satisfy the set of rather low ϕ values and yet are consistent with high β_T values (> 0.85 for both proteins). For both Im7 and Im9 the interhelical angles are highly variable in the transition state ensembles, although the native contacts between helices I and IV are well conserved. By measuring the distribution of the accessible surface area for each residue we show that the hydrophobic residues that are buried in the native state remain buried in the transition state, corresponding to a hydrophobic collapse to a relatively ordered globule. The data provide new insights into the structural properties of the transition states of these proteins at an atomic level of detail and show that molecular dynamics simulations with ϕ value restraints can significantly enhance the knowledge of the transition state ensembles (TSE) provided by the experimental ϕ values alone. Proteins 2004;54:513–525.**

INTRODUCTION

The comparison of homologous proteins can provide information about the determinants of the folding process[1–3] beyond that offered by single protein studies. Such comparisons contribute to the understanding of how the native structure is selected in the folding process and how, for similar native structures, changes in the sequence may affect the folding mechanism. For example, the SH3 domains of Src,[4] α-spectrin,[5] and Fyn[6] appear to fold via similar transition states, while the protein Sso7d, which has an SH3 domain-like fold, has a different, but closely related transition state.[7] In the bacterial immunity protein family, Im7 (see Fig. 1) folds via a populated, on-pathway intermediate, while Im9 folds in a two-state manner under identical conditions (pH 7, 10°C).[8] Im9 can be induced to fold in a three-state manner at lower pH, although the structural relationship of the intermediate formed to that populated during Im7 folding is not known.[9] Since the structures of the native states of Im7 and Im9 are very similar,[10,11] it is clear that the differences in the amino acid sequence and the resulting interactions play a significant role in the folding kinetics. Qualitatively, this effect can be described as a perturbation of the free energy landscape induced by the changes in the sequence.[12]

The protein engineering method[13] provides information concerning the role of individual side-chains in stabilizing the rate limiting transition state ensemble (TSE) for folding. The essential element of this technique is the measurement of ϕ values. For residue i, the ϕ_i value is the ratio between the change in stability of the transition state ($\Delta\Delta G_i^{TS}$) and the change in stability of the native state ($\Delta\Delta G_i^{NS}$) caused by the mutation of residue i; the unfolded state is used as the reference state. However, while the experimental values provide an image of the overall properties of the TSE, they do not yield detailed structural information, such as the interactions of different atoms, the distribution of specific conformations, or the orientation of secondary structural elements. To obtain such information, we have developed a method for the determination of the TSE for protein folding using the ϕ values as restraints in a molecular dynamics sampling procedure.[14,15]

*Correspondence to: Michele Vendruscolo, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK E-mail: mv245@cam.ac.uk

Received 28 February 2003; Accepted 7 July 2003

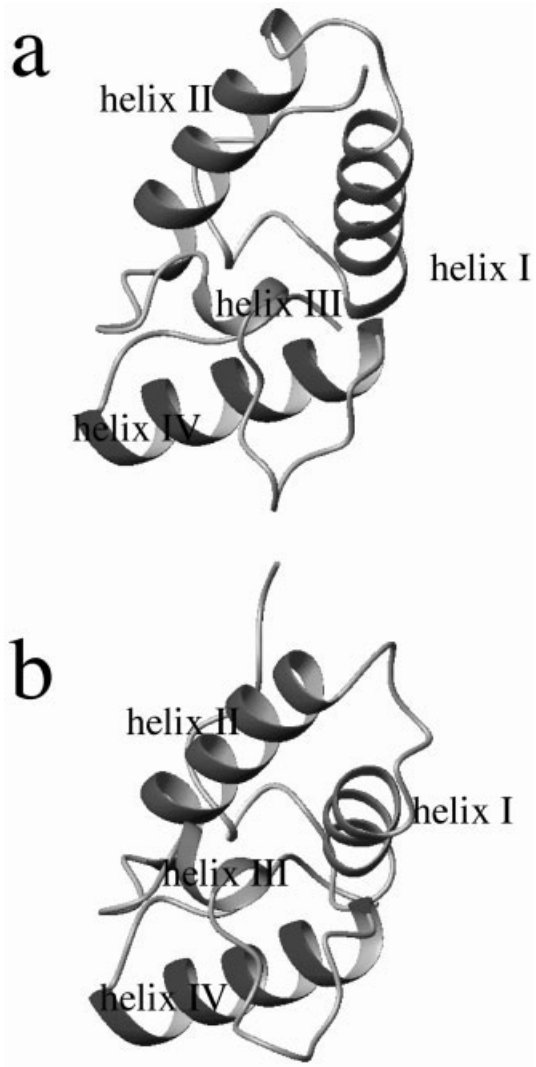

Fig. 1. Native structures of Im7 and of Im9. The secondary structure is represented as ribbon. The Im7 secondary structure elements are: helix I (residues 12–26), helix II (residues 32–45), helix III (residues 52–55), helix IV (residues 65–79). The Im9 secondary structure elements are: helix I (residues 14–23), helix II (residues 30–44), helix III (residues 50–54), helix IV (residues 64–78). The residues involved in the secondary structure, computed using DSSP[43] differ slightly from those given in the PDB files used by Friel et al.[22] The figure was drawn with the program MolMol.[44]

The value of ϕ_i can be related to the fraction of the native interactions present in the transition state[16] when non-native interactions do not make a significant contribution. It has been shown that this condition is satisfied for several proteins.[17,18] More importantly, it was also shown that the non-native interactions that do exist in the TSE tend to involve residues whose neighbors along the sequence form at least some native contacts.[17] Thus, even though non-native interactions are important in the folding intermediate of Im7,[19] their presence is not likely to disrupt the essential properties of the network of native interactions that stabilize the transition state. This assumption was recently tested by use of a method[20] in which the structures of the TSE determined from molecular dynamics sampling[14,15] were used to back-calculate the ϕ values based on estimates of the free energy differences introduced by the mutations. Application of this method to the TSE's of Im7 and Im9 calculated in this paper showed a high correlation (the correlation coefficient was 0.9 for Im7 and 0.7 for Im9) between the sets of ϕ values calculated with the two different definitions (one based on the number of native contacts and the other based on changes in free-energy differences), demonstrating the consistency of the approach. It is important to note that non-native interactions are included in the calculations since we use molecular dynamics simulations with an energy function that combines a standard molecular mechanics force field for the physico-chemical interactions and a pseudo-energy for the ϕ value restraints,[14,21] i.e., the non-native interactions are included, although they are not used in the bias.

In this paper, we determine the ensemble of structures that comprise the TSE for folding of Im7 and Im9 using experimental ϕ values.[22,19] We interpret the experimental ϕ values as the fraction of native contacts present in the structures that comprise the TSE.[14,21] With these restraints, the TSE becomes the most stable state on the potential energy surface, rather than being an unstable region, as it is for the true energy function of the protein. Unlike the native state, which can be well represented by a set of very similar structures, we show that the TSE of both Im7 and Im9 is structurally highly heterogeneous with large RMSD values between its members (see below). Nevertheless, by appropriate clustering techniques, the important structural features of the TSE can be determined and the nature of the ensemble revealed. As the experimental ϕ values are ensemble averages, we extend the Monte Carlo procedure for computing general conformational ensemble averages presented by Davis et al.[23] to show that the TSE's of Im7 and Im9 are correctly described even when the simulations are carried out without ensemble averaging (see Methods). The data that we present here provide detailed insights into the distribution of structures that comprise the TSE. Furthermore, we predict residues thought to contribute to the stability of TSE through non-native interactions that were not analyzed in previous experimental measurements.[22,19]

An important motivation for the present work was to provide a structural characterization of the TSE's of Im7 and Im9 with the unusual property of having a large β_T (0.85) but low ϕ^{exp} values. By determining the structures making up the TSE we show that these two conditions are compatible because, even though the averaged transition state seems to be highly compact, the contributing structures have considerable conformational heterogeneity with the TSE of Im9 being significantly more heterogeneous than that of Im7.

RESULTS AND DISCUSSION

The Transition State Ensembles

We determined the TSE of Im7 and of Im9 by use of restrained molecular dynamics simulations as described previously[14,15] (see Methods). In brief, we performed mo-

lecular dynamics simulations where, for the residues whose experimental ϕ values are available, the fraction of native contacts is restrained to remain close to the experimental ϕ values through an additional term in the energy function. A large number of conformations satisfying the experimental restraints are generated and representative structures are selected by a clustering technique (see Methods). For the selected TSE, the correlation coefficient between the calculated and experimental ϕ values is greater than 0.97 for Im7 and greater than 0.91 for Im9.

The most representative structures of the TSEs of Im7 and Im9 are shown in Figure 2. These structures illustrate the heterogeneity of the TSE of these proteins which is remarkable compared with that for the TSE of acylphosphatase (see Figure 6 of Paci et al.[14]). For example, while several of the most populated clusters have native-like secondary structure in the TSE of Im7 and Im9, other clusters possess little secondary structure (Fig. 2). In addition, there is a certain propensity (20%) to form a small parallel β sheet involving residues 9–11 and 83–85 in the TSE of both Im7 and Im9; this is absent in the native state. The relative arrangement of the α helices is also very variable in the TSE. Nevertheless, the structures that comprise the TSE are highly compact; the radius of gyration and the solvent-accessible surface increase by less than 20% relative to the native proteins, as might be expected from the large β_T values (0.85 and 0.9 for Im7 and Im9, respectively[22,19]). In Table I we report the average properties of the two TSEs. The TSE of Im7 is slightly more native-like than that of Im9: the average ϕ_i^{calc} is larger and the RMSD from the native state is smaller for Im7 than for Im9, in agreement with the higher ϕ^{exp} values of the former (the average $\langle\phi^{\mathrm{exp}}\rangle$ values for Im7 and Im9 are 0.43 and 0.29, respectively,[22,19] when the values larger than 1 are set to 1.0). The fact that ϕ values are larger for Im7 than Im9 is interesting[22] because β_T of Im9 is a slightly larger than that for Im7. These results contrast with those of Ternström et al.,[24] who found that the β_T of U1A increases when the ϕ^{exp} values increase. In our calculations, we found that the TSE of Im7 is as compact and solvent protected than that of Im9 (differences are within the standard deviations; see Table I). Despite this fact, the ensemble of representative structures for Im9 is significantly broader than that of Im7.

Figure 3 shows the probability that two structures of the TSE differ by a given RMSD value. The typical RMSD between structures belonging to the TSE of Im7 is about 7 Å while that of Im9 is nearly twice as large (12 Å). By comparison the breadth of the TSE for acylphosphatase is 6.4 ± 1.0 Å[14] and TNfn3 is 6.4 ± 1.4 Å.[15] Thus while the TSE of Im7 is similar to that of these α/β and all β proteins, the TSE of Im9 is very broad. Another measure of the similarity between two structures is the overlap between contact maps which is defined as the number of heavy atom contacts in common between two contact maps, normalized by the total number of contacts in the contact map with more contacts among the two that are being compared.[25] Thus, two identical contact maps have overlap equal to 1. The average overlap between contact maps corresponding to structures in the TSE of Im7 is 0.72 ± 0.07. For Im9, the average overlap in the TSE is lower (0.66 ± 0.06), indicating that the contact maps in the TSE of Im9 are more different than those in the TSE of Im7. For comparison, the average overlap between contact maps corresponding to different NMR models of the native state of Im9 (PDB code 1IMP[11]) is 0.95 ± 0.01. It is remarkable that the extreme heterogeneity of the transition state of Im9 is compatible with a compactness close to that of the native state, as revealed by the experimental β_T of 0.94, and by the radius of gyration and the solvent-accessible surface estimated from the ensemble of structures (these values increase by 10% and 20%, respectively, relative to the native state; see Table I). Furthermore, the distribution of structures in the TSE is associated with the presence of a specifically-collapsed hydrophobic core (see below).

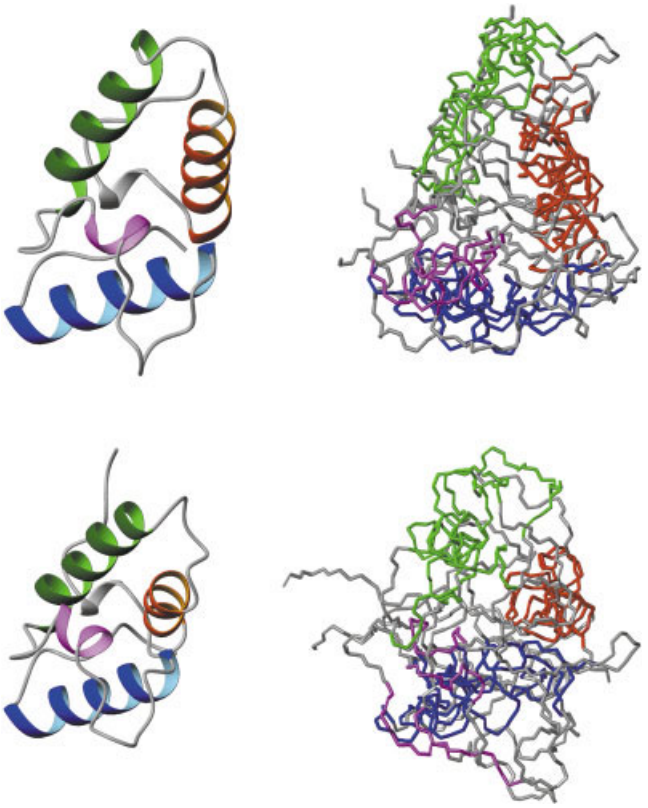

Fig. 2. Four representative structures of the TSE of Im7 (top, right) and Im9 (bottom, right) compared with the native structure (left). Representative structures are chosen as the centers of the most populated clusters; their populations range between 1000 members, for the most native-like clusters, to about 20 for the least native-like ones (right).

As the two sets of mutants for Im7 and Im9 do not coincide completely, we repeated the determination of the respective TSE's using only a subset of 17 mutants at equivalent positions for the two proteins. The results, also shown in Table I, reveal that the overall properties of the two transition states do not change when the more limited set of experimental ϕ values is used in the structure determination. More importantly, as shown in

TABLE I. Properties of the TSE of Im7 and Im9[†]

	$N\langle\phi^{exp}\rangle$	$\langle\phi^{calc}\rangle$	$\langle\phi^{exp}\rangle$	RMSD (Å)	R_g (Å)	ΔR_g (%)	S (Å^2)	ΔS (%)
Im7	22	0.37 ± 0.04	0.43	6.5 ± 0.9	13.9 ± 0.7	10.4	6600 ± 600	20
Im9	20	0.23 ± 0.02	0.29	8.2 ± 1.5	14.3 ± 1.0	11.3	6700 ± 600	20
Im7	17	0.30 ± 0.03	0.45	6.0 ± 2.0	13.7 ± 1.3	10.0	6100 ± 900	25
Im9	17	0.25 ± 0.02	0.30	8.5 ± 1.0	14.5 ± 0.8	14.3	6700 ± 600	20

[†]For Im7 22 ϕ^{exp} values[19] and 20 for Im9[22] were used. Corresponding results obtained using only a subset of 17 ϕ^{exp} in equivalent position in the sequences are also shown. The average $\langle\phi_i^{exp}\rangle$ is computed for the residues for which there is a ϕ^{exp} value while the average $\langle\phi_i^{calc}\rangle$ is computed for all the residues which have a non-zero number of side-chain native contacts (80 residues for both proteins). RMSD indicates the root mean square deviation from the native state, R_g the radius of gyration, S the solvent accessible surface area, and ΔR_g and ΔS are the deviations from the native values. The errors reported represent one standard deviation.

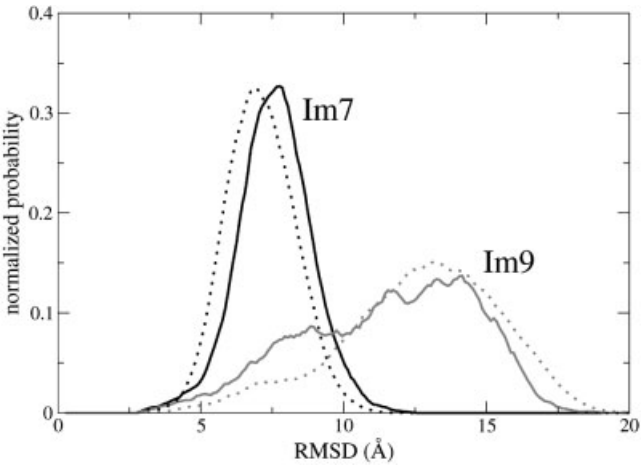

Fig. 3. Probability distribution of the pairwise RMSD for structures in the TSE of Im7 and Im9 determined with the original set of ϕ values compared with those obtained with the set of ϕ^{exp} from mutations at equivalent positions (dashed lines).

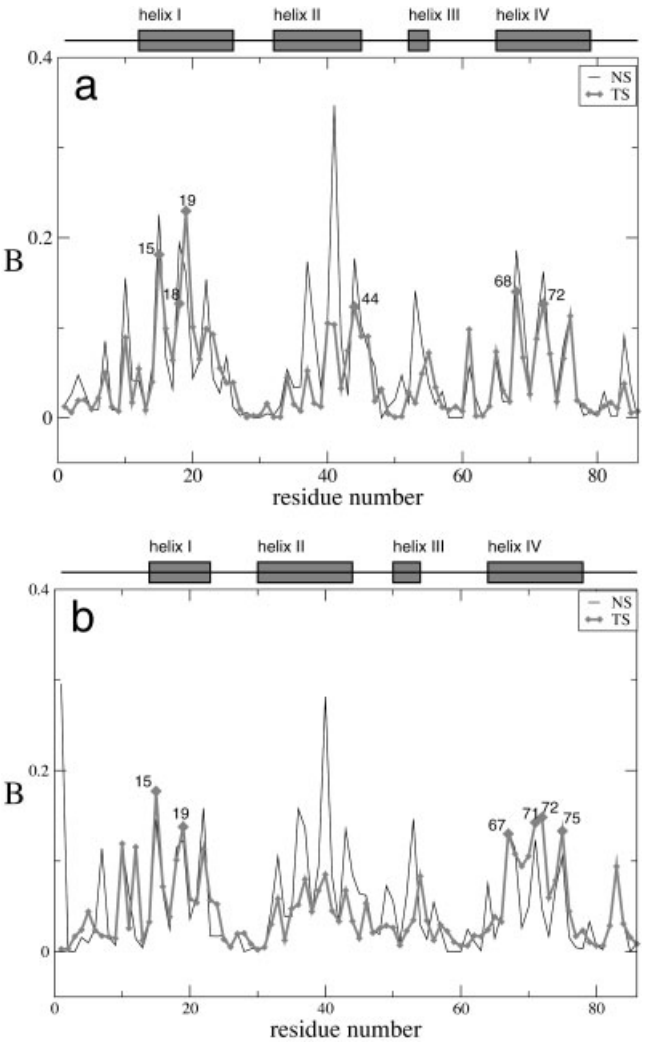

Fig. 4. Profile of the *B* score of Im7 (**a**) and of Im9 (**b**) in the TSE and in the native state (thin line). The *B* score measures the centrality of a residue in the network of interactions in the TSE.[26] NS, native state; TS, transition state.

Figure 3, the widths of the two distributions do not change when only ϕ values measured at equivalent positions in the two proteins are included in the bias.

To identify the parts of the structures in the TSE of Im7 and Im9 that play an important role for folding we measured the *B* score[26] of each residue in the two ensembles (see Fig. 4). To define the *B* score a protein structure is represented as a graph in which nodes correspond to amino acids and links correspond to contacts between them. The resulting graph represents the network of interactions that characterizes a given protein structure. Within such a network two residues can be connected by paths running along the links. A residue has a *B* score of unity if all the shortest paths connecting any two residues pass through it and a *B* score of zero if no shortest path passes through it.[26,27] Thus, the *B* score is a measure of the centrality of residues in the network of interactions that characterize a protein structure. Residues with a *B* score larger than a given threshold (set here at 0.12) were considered as the most connected ones in the TSE of Im7 and Im9. From the analysis we found that residues F15, L19, and I72 are important in determining the network of interactions that stabilize the transition state of Im7 and, correspondingly, F15, V19, and K72 are important for Im9. In addition, I68, L18, and I44 are important in Im7, and V71, R75, and I67 play a significant role in Im9. Interestingly, these residues (apart from I44 in Im7) are in topologically equivalent positions in the native states of the two proteins. They stabilize the packing of helices I and IV. Their high degree of structure formation in the TSE is consistent with their relatively high experimen-

tal ϕ values (> 0.35 in both proteins) and indicates that helices I and IV are not only substantially formed in the rate-limiting transition state, but that their hydrophobic packing is also essentially native-like.

The residues with the largest *B* score in the native state are in helix II in both Im7 and Im9, demonstrating that helix II occupies a central position in the network of interactions that stabilizes the native protein. The helix II residues have a lower, although still significant, *B* score in the TSE of both proteins, indicating that the packing of helix II on helices I and IV is an important element of the TSE.

Prediction of ϕ Values

One important application of the present method for determining the structure of transition states is the prediction of ϕ values for residues that have not been mutated experimentally. We therefore calculated the ϕ values for all the residues in the TSE of Im7 and of Im9 (see Fig. 5). For the residues that have been mutated, the experimental and calculated ϕ values are essentially identical, thus supporting the internal consistency of the experimental date and their interpretation in terms of fraction of native contacts.

To provide a further validation of the TSE's, we used a test recently introduced by Lindorff-Larsen et al.[20] In this procedure, the program Fold-X[28] is used to compute changes in free energy upon mutation for the native and the transition state, in direct correspondence with the protein engineering experiments.[13] From these two changes, within the transition state theory, it is possible to reproduce closely the ϕ value as it is measured experimentally. Using this approach, we obtained a coefficient of correlation of 0.9 for Im7 and of 0.7 for Im9 between the estimated free energies and the experimental values.

The uncertainty in the predictions of the ϕ values of residues that have not been mutated experimentally measures the degree of local structural variability of the TSE. Thus, the present procedure is able to single out those residues whose experimental ϕ value could add valuable information to the picture of the TSE. To this end, we considered residues whose ϕ^{calc} values have a large standard deviation (> 0.2), that are not totally exposed to the solvent and that make a non-negligible number of interatomic contacts in the native state (more than 30). Using these criteria, the data show that the TSE's of Im7 and Im9 are well determined by the available sets of experimental ϕ values. Thus, only two residues in Im7 (E2 and K4) and four residues in Im9 (E14, M43, K72, N78) show significant variability. These residues were not mutated in the original experimental studies[22,19] since their properties do not allow the selective removal of hydrophobic groups, as is preferable in protein engineering studies.[29] Their mutation, e.g., via a leucine as a pseudo wild-type, however, would allow the N-terminal region of Im7, and the loop connecting helices II and III and the C-terminal region of Im9 to be more clearly defined.

Although the number of experimental ϕ values available for Im7 and Im9 is similar, the prediction of the ϕ values for residues which have not been mutated is more precise for Im7. Thus the standard deviation $\sigma(\phi^{calc})$, on average, is larger for Im9 (see Fig. 5), due to the fact that the ϕ^{exp} of Im9 are on average smaller than those for Im7 and its TSE is more heterogeneous. Interestingly, two residues in Im7 are predicted to have large ϕ values (see Fig. 5) and should therefore be structured in the TSE. These residues (D31 with $\phi^{calc} = 0.49 \pm 0.06$, and H40 with $\phi^{calc} = 0.7 \pm 0.1$) are in helix II, and preserve most of their native contacts in the TSE.

Transition State Energetics

The energy maps in Figure 6 show the network of pairwise inter-residue contacts weighted by their energy. They illustrate the changes in the patterns of interactions between the TSE and the native state. The figure shows that while some native interactions persist in the TSE, others are lost. For example residue L53 in Im7 and L52, I53 in Im9, lose most of their native interactions in the TSE (see also Fig. 7). The energy maps also show that non-native interactions occur in the TSEs of both Im7 and of Im9. In most cases, however, non-native interactions do not appear in arbitrary positions, but as a consequence of the partial loss of native order. Such a loss corresponds to a "smearing out" of the native interactions. In other words, most of the non-native interactions are low in energy and between residues that are close in the native structure. This implies that the overall native architecture is to a large extent present in the TSE and that the network of interactions does not undergo large-scale changes as the native state forms from the TSE. Nonetheless, some short- to medium-scale rearrangements are required to reach the native structure.

Non-native interactions can be quite strong in the TSE of Im7. For example the residue pairs L18-T45 and Y55-E46 interact very strongly in the TSE despite being well separated in the native state (the residues are 8.5 and 12 Å apart, respectively, in the native state). Interestingly, in Im7, L18 in helix I forms non-native interactions in the intermediate (since $\phi_I > \phi_{TS}$), and also makes strong non-native interactions with T45 (the C-terminal residue of helix II) in the TSE, as mentioned above. Other residues that may form non-native interactions in the intermediate of Im7 make weaker non-native interactions in the TSE, such as I68 with V16 and Y55, or V42 with W75 and R76. Further analysis, for example, making use of double mutant cycles[30] could be employed to confirm the identity of residue pairs predicted here to form non-native interactions in the TSE.

In several structures of the TSE of Im7, W75 forms a contact with Y55 and Y56 in helix III. Similarly, the interactions of W74 with Y54 with and Y55 are present in the TSE of Im9. This is a particularly interesting result as these non-native contacts have been found to persist in both Im7 and Im9 in the presence of 6M urea (NOEs between these residues, which are not close in the native state, have been observed in the denatured states, C.

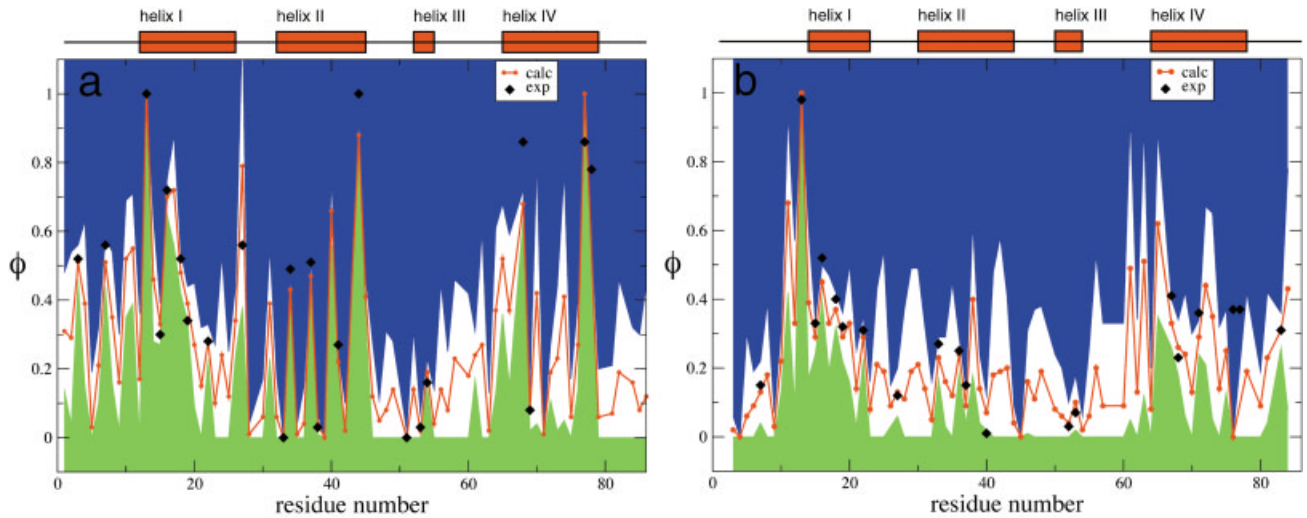

Figure 5.

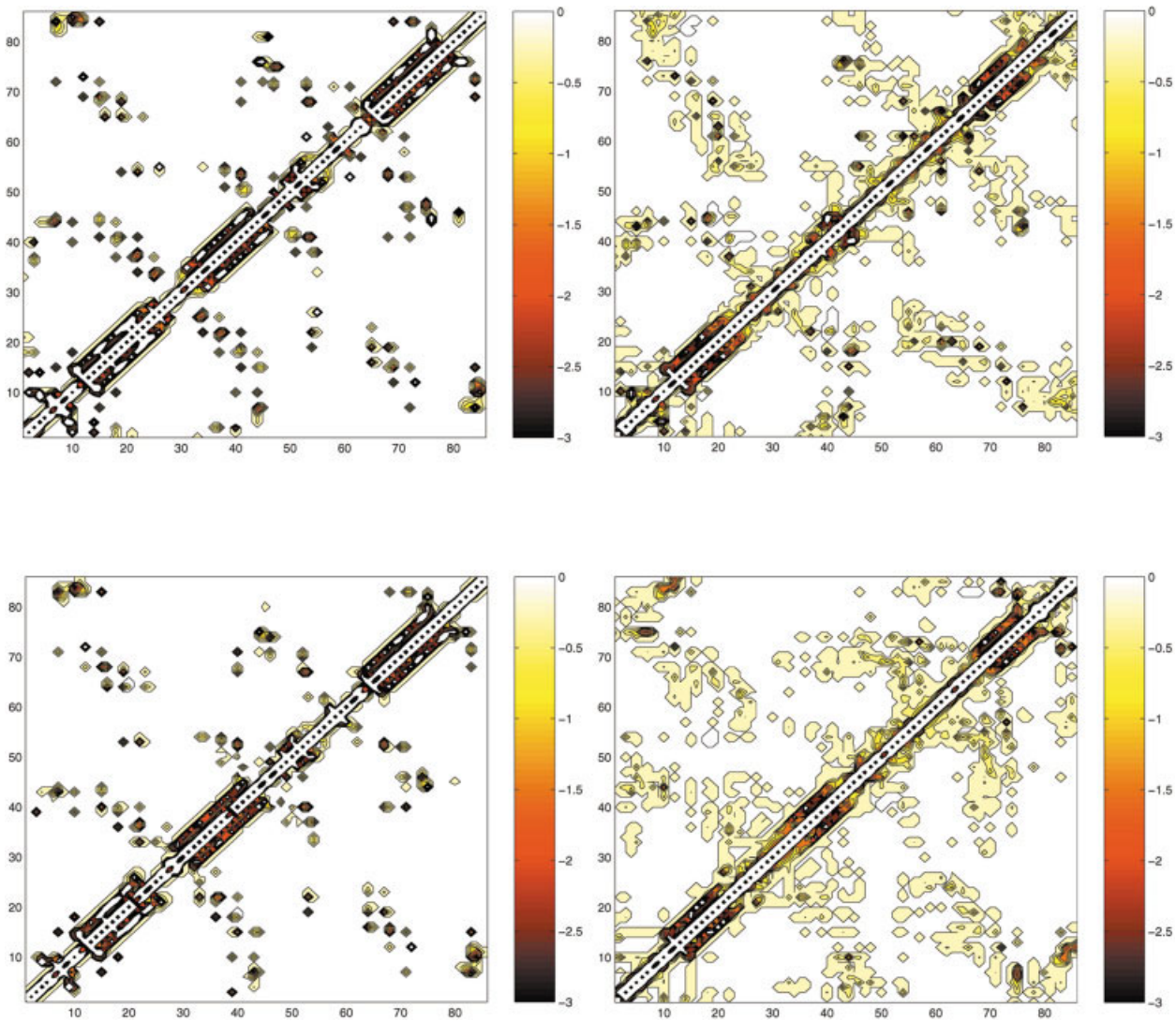

Figure 6.

LeDuff, A. Bijelic, and G. R. Moore, unpublished data). These simulations suggest experiments using site-directed mutagenesis to investigate the role of these tyrosine residues (which are highly solvent exposed in the native proteins) in stabilizing the TSE of Im7 and Im9.

Several charged residues are also predicted to be involved in long-range, non-native interactions in the TSE in both Im7 and Im9. For example, in Im7, the charged residue R61 makes several non-native interactions with residues K20-E23 and E46-H47. Although these residues are not in contact in the native structure (the distance between any pair of heavy-atoms of the side-chain is more that 10 Å), their interactions exist simultaneously in the least native-like members of the TSE (see Fig. 2). Another charged residue, K43, makes strong non-native interactions with N5 and R76. In Im9, residues R75 and S6-I7-S8 interact much more strongly in the TSE than in the native state.

The total effective energy of the TSE is similar for Im7 and Im9 (about −230 kcal/mol considering all pairs of residues more that three positions apart in the chain). Of the total average effective energy stabilizing the TSE, 36% in Im7 and 41% in Im9 arises from non-native interactions. The importance of non-native interactions in folding has been described previously.[17] The transition states of both Im7 and Im9 have the largest fraction of their energy from non-native interactions of all the proteins studied so far with this approach (acylphosphatase,[14] TNfn3,[15] and several other proteins, E.P., M.V., C.M. Dobson, and M.K., unpublished results). While the TSE of Im9 has globally more non-native interactions than Im7, these are in large part local and small in magnitude. In contrast, the TSE of Im7 has fewer but stronger non-native interactions, particularly involving residues which are believed to form non-native interactions that stabilize the intermediate.[19]

Energies for individual residues in the TSE of Im7 and Im9 are shown in Figure 7. The single-residue energy is the sum of all the interactions of a given residue; local contributions less than three residues apart in the sequence are not included. In Im7, R61 is more stable in the TSE than in the native state. Residues Y10, Y55, and R76 in Im7 and Y10, Y54, and R75 in Im9, also have large negative energies in the TSE of both proteins.

Taken together, these data suggest that the TSE's of both Im7 and Im9 contain a significant number of non-native interactions that play an important role in stabilizing this state. Further experiments are needed to test these predictions and to validate the nature of the interacting partners and their role in folding.

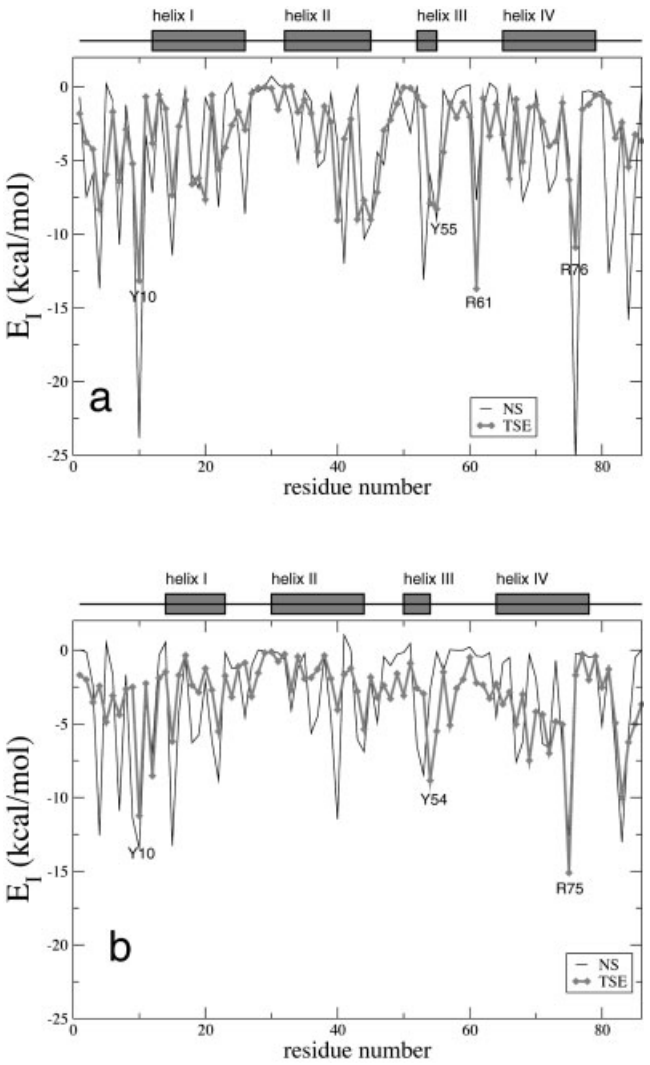

Fig. 7. Single residue energy $E_I = \Sigma_J E_{IJ}$ (see Methods), excluding local contacts (more than three residues apart). **a**: Im7, **b**: Im9.

Helix-Helix Docking

One notable difference between the TSEs of Im7 and Im9 is that in Im7 the interfaces between helices I, II, and IV are more ordered than the corresponding regions in Im9. From the direct analysis of the experimental ϕ values, the interface I–IV, and in particular the point where the N-termini of the two helices meet, has been identified as the most highly structured region in the transition state ensemble of both Im9[22] and Im7.[19] In Figure 8 we show the average experimental ϕ value (blue triangles) for residues belonging to helices I and II and to helices I and IV. From the TSE structures that we determined in this study, the degree of native ordering at the inter-helical interfaces can be measured by the fraction of inter-helical side-chain native-like contacts in the TSE (black squares). The latter values, calculated from the structures that we determined, are systematically lower than the experimental ones. This difference can be attributed to the fact that the experimental values are averaged

Fig. 5. Profiles of the ϕ^{exp} and ϕ^{calc} values for Im7 (**a**) and for Im9 (**b**). The red line is the average over the sample of representative structures of the TSE. The upper and lower bounds, in blue and green respectively, correspond to plus or minus one standard deviation.

Fig. 6. Energy maps of Im7 (top) and Im9 (bottom); the native state is shown on the left while the TSE is shown on the right. The figures show the contour plots of the average residue-residue interaction energy E_{IJ}. The relation of the colors in the maps to the free energies in kcal/mol is shown on the scale on the right.

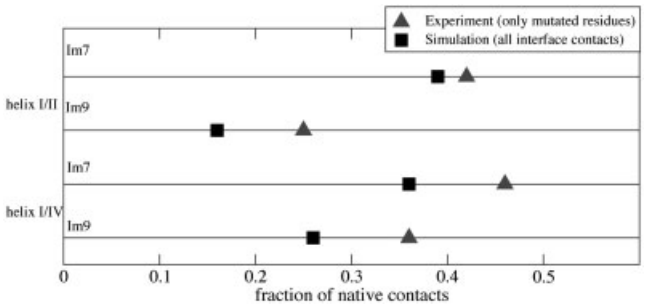

Fig. 8. Experimental ϕ values averaged over residues in helices I and II, and in helices I and IV (triangles); Fraction of native contacts between residues at the interface between helices I and II and between helices I and IV in the TSE determined in this study (squares). Results are shown for Im7 and Im9.

over only a few sites, while the results from the simulations include all the contacts between the helices. Thus, the latter provides a more complete description of the inter-helical interactions. The trends from the simulations and the ϕ values per se correspond in that in Im9 both interfaces are less ordered than those in Im7. In addition, for both proteins, more contacts are preserved in the interface between helices I and IV than between helices I and II.

Capaldi et al.[19] noted that non-native interactions in the intermediate of Im7 involve helices I, II, and IV and that, in particular, helix II forms numerous non-native contacts ($\phi_I > \phi_{TS}$ for five of the seven residues mutated in the helix[19]). The present results indicate that in the transition state such non-native structure is largely lost, but that in Im7 this interface is more ordered than the corresponding one in Im9. These data suggest that a more ordered TSE for Im7 may result from the folding of this protein via a compact intermediate state that is highly populated. Interestingly, Im9 can also be induced to fold via an intermediate, either by reducing the pH or by introducing one or more mutations that increase hydrophobicity,[9] including residues in helix II (C.T.F. and S.E.R., unpublished data). These observations, together with the structures of the TSE presented here, suggest new experiments in which the role of the intermediate in determining the breadth of the TSE, both experimentally and in simulations, could be measured. Such studies are currently under way.

Specific Hydrophobic Collapse

Experiments[22,19] and simulations (see Table I) show that the transition state is highly compact for both Im7 and Im9. Experimentally, specific mutations were designed to probe whether the transition state has a native-like hydrophobic core or if the collapse is non-specific. For both Im9[22] and Im7,[19] the mutation of hydrophobic residues which are solvent-exposed in the native structure has little effect on the folding rate constant and on the stability of the native state, indicating that these residues are solvent exposed in both the native and transition state ensembles. The structures of the TSE of Im7 and Im9 obtained from the simulations allow us to measure directly the solvent accessible surface for every residue, averaged

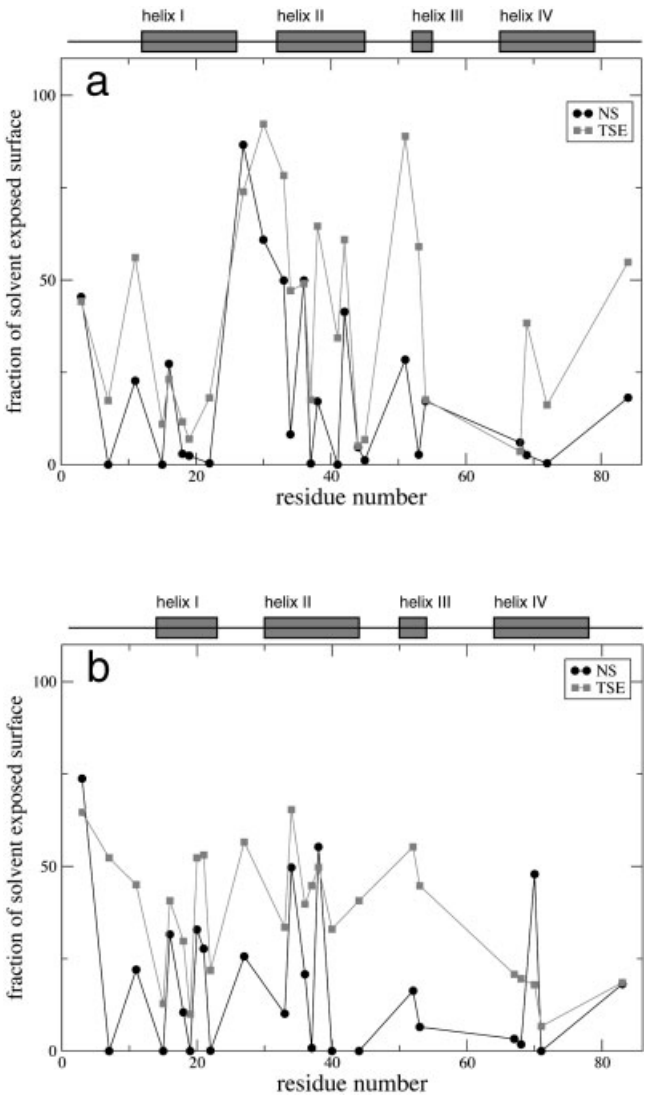

Fig. 9. Fraction (relative to the G-X-G peptide) of solvent-accessible surface area, as computed by NACCESS (S. Hubbard and J. M. Thornton, http: //wolf.bms.umist.ac.uk/naccess/), of hydrophobic residues in the native state and in the TSE of Im7 (**a**) and Im9 (**b**).

over all the conformations in the TSE. The results for all hydrophobic residues (L, I, V, F) and those for T are shown in Figure 9. Overall, the hydrophobic residues are more solvent exposed in the TSE than in the native state. However, those which are buried in the native state tend to remain buried in the TSE, and vice versa, those which are solvent-exposed in the native state tend to remain solvent-exposed in the TSE. The good correlation between the solvent exposure of hydrophobic residues in the TSE and in the native state for both Im7 and Im9 indicates that the hydrophobic core is specific in both proteins.

In both Im7 and Im9, a large increase in the solvent-accessible area, with respect to the native state occurs in the TSE around the residues at positions 51 and 53 in helix III. These residues form part of the hydrophobic core in the native state and make contacts with all other helices. Consistent with the experimental result of Friel et al.[22]

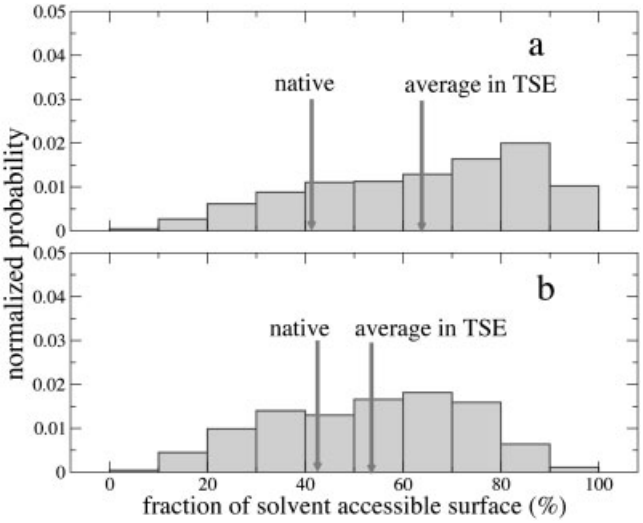

Fig. 10. Histogram of the accessible surface area in the TSE of (**a**) W75 in Im7 and (**b**) W74 in Im9. The native-state values and the average values in the TSE are indicated by arrows.

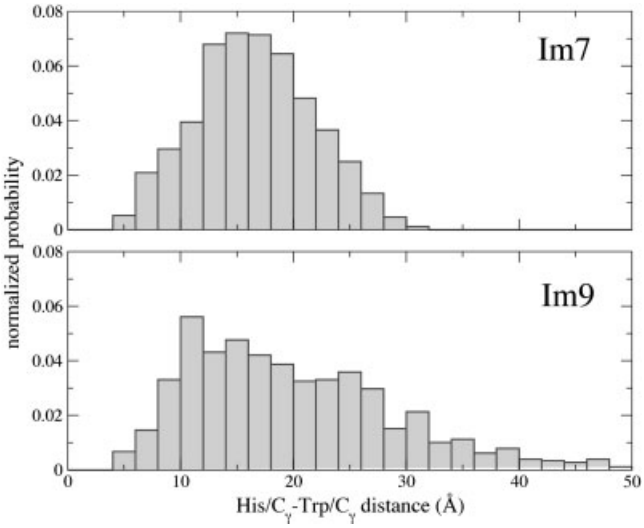

Fig. 11. Distribution of distances between the C_ϕ of W75 in Im7 and W74 in Im9 and the C_γ of H47 in Im7 and H46 in Im9.

and Capaldi et al.[19] these important native interactions are absent in the TSE, confirming that, while the native hydrophobic core is mostly formed at the transition state, residues that contribute to the core from helix III are missing.

In Im7, W75 (in helix IV) has a 40% accessible surface area in the native state (solvent accessibilities are estimated relative to a G-X-G peptide). In the TSE its surface exposure varies considerably, ranging from 10% to 100%, with an average of about 63% (see Fig. 10). The surface area of W74 in Im9 shows a similar variability, in agreement with the suggestion of Friel et al.[22] that the C-terminal part of helix IV is relatively unrestrained in the TSE. It is interesting that, on average, W74 in Im9 is more buried than the W75 in Im7, despite the fact that the ϕ^{exp} values in this region are greater for Im7 (the average ϕ values for residues in helix IV are 0.98 and 0.35 for Im7 and Im9, respectively). The simulation results demonstrate that while the TSE is relatively heterogeneous in this region of Im9, W74 in the TSE remains highly buried in the majority of conformations.

As another example of the results obtained, we consider the distribution in the TSE of the distances between the Trp (W75 in Im7 and W74 in Im9) and the His (H47 in Im7 and H46 in Im9) that quenches its fluorescence in the native state[31] (D. Gill, G. R. Spence, C.T.F., and S.E.R., unpublished results). The results of the simulations show that this distance is highly variable, particularly in Im9, in agreement with the observation that the TSE is broader in Im9 than in Im7 (see Fig. 11). In the experiment, W75 (Im7) and W74 (Im9) were not mutated, as they were necessary for the fluorescence measurements. From the TSE that we determined we obtained $\phi^{calc} = 0.4 \pm 0.3$ (Im7) and $\phi^{calc} = 0.1 \pm 0.2$ (Im9). These results provide further evidence that the TSE of both Im7 and Im9 are highly variable, but that the TSE of the latter protein is more so.

Secondary Structure

The probability of formation of secondary structure in the TSE is shown in Figure 12. The results provide a direct prediction of the location of elements of secondary structure in the TSE of both proteins, that can be compared with the experimental studies of a series of Ala to Gly mutants in Im9 performed by Friel et al.[22] Since these ϕ values were not used as restraints in the simulations, the availability of experimental data provides a direct test of the simulation results. Helices I and IV have a high probability of being present in the TSE of Im7 (~80% and ~70%, respectively). On the contrary, helix II and helix III are not very well structured in the TSE. In Im9, helices I and IV are both formed in the TSE, while helix III is mostly absent. Helix II is present in about 50% of the Im9 structures, a fraction higher than that for helix II in Im7. This is a surprising result since the ϕ^{exp} values for this helix are lower in Im9 than those of Im7 (average values 0.17 and 0.38, respectively). Presumably, this reflects the fact that ϕ^{exp} values probe mainly the tertiary interactions between helices, rather than local backbone conformations. This is in accord with the results from the Ala to Gly mutations performed to probe the degree of helix formation in the TSE.[22] The mutation of eight solvent exposed alanine residues to glycine in Im9 results in ϕ^{exp} values for helices I, II and IV of 0.5–1.0, 0.7–1.5 and 0.4–0.7, respectively,[22] indicating a high probability of helix formation. Corresponding mutations have not been made extensively in Im7. However, mutations of A13, A77, and A78 to Gly suggest that helices I and IV are also substantially formed in both the intermediate and the TSE of Im7. The ϕ values in the TSE for these Ala-to-Gly mutations are 1.25, 0.86, and 0.78, respectively.[19] In accord with these results, recent hydrogen exchange experiments have demonstrated that the three long helices are fully formed in the Im7 folding intermediate (S.A. Gorski, A.P. Capaldi, C. LeDuff , G.R. Moore and S.E.R., unpublished results).

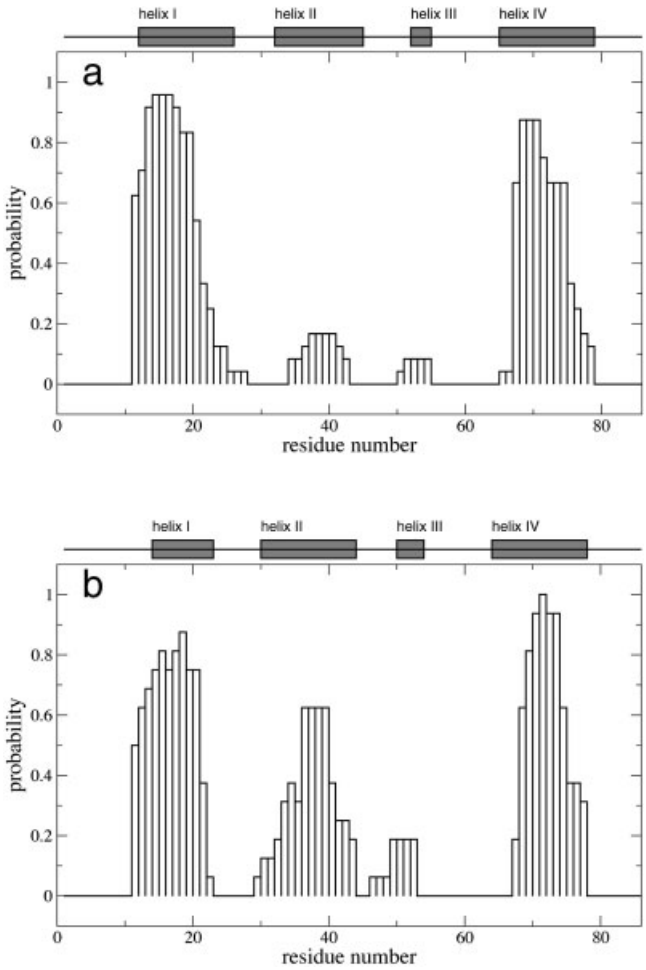

Fig. 12. Average secondary structure according to DSSP[43] for the TSEs of Im7 (**a**) and of Im9 (**b**). Only α-helices are shown.

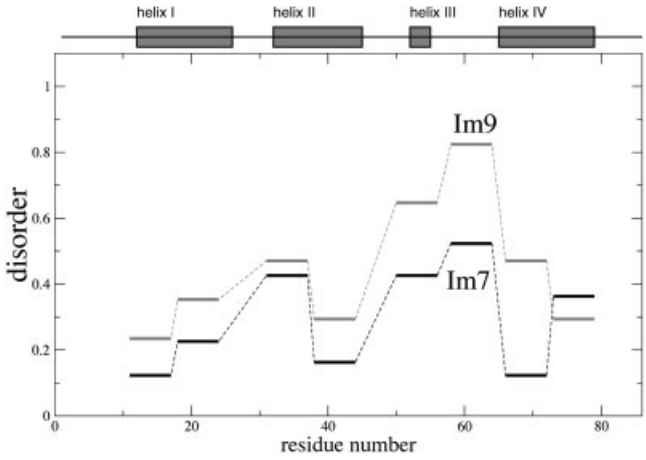

Fig. 13. Fraction of different conformations for eight distinct seven-residue segments in the TSE structures of Im7 and Im9.

To analyze the degree of local structure formation in each TSE we considered eight seven-residue fragments that are positioned along the sequence of each protein (see Fig. 13). For each fragment, we used a clustering analysis to select the most representative structure among those that we determined for the TSE. We then measured the fraction γ of structures at more than 1 Å RMSD from the representative one. When γ is close to zero all the structures are similar while when $\gamma = 1$ the structures are rather heterogeneous. The results, shown in Figure 13, are in agreement with those already discussed and indicate that the most disordered fragments in both Im7 and Im9 correspond to those in the region of helix III and the loop between helices III and IV. The two most ordered fragments are in the N-terminal region of helix I and in the C-terminal region of helix II. Helix II is only partially formed but it is in a region which is not particularly heterogeneous; i.e., in the structures contributing to TSE this segment assumes similar conformations, although the typical α helical hydrogen-bonding pattern is not always present. In addition, the N-terminal region of helix IV is highly structured in the TSE of both Im7 and Im9. Overall, Im7 and Im9 show the same trend as to which regions are more or less ordered, although the TSE of Im9 is significantly more heterogeneous, consistent with the results described above.

CONCLUSIONS

To complement the protein engineering experiments of Friel et al.[22] and of Capaldi et al.[19] for folding of the bacterial immunity proteins Im7 and Im9, we determined sets of structures representing the transition state ensembles of each protein at atomic resolution by using the experimental ϕ values as restraints and sampling the accessible conformational space with all-atom molecular dynamics simulations. Consistent with the experimental results, the calculated TSEs of both proteins show (1) that helices I, II, and IV are well formed in the TSE, (2) that helix III docks after the transition state is traversed, (3) that the hydrophobic collapse to form the TSE is specific, and (4) that the packing of the three helices is native-like in the region where helices I and IV are in contact.

As in previous all-atom simulations of the TSE with experimental ϕ value restraints (e.g., acylphosphatase[14]), the main purpose of this type of analysis is to obtain structural information concerning the TSE that goes beyond that available directly from the experimental studies. For example, the simulations have demonstrated the heterogeneity of the TSEs of both Im7 and Im9 and that a wide range of structures is compatible with the ϕ^{exp} values. Most importantly, we have shown that the TSE of Im7 is much narrower than that of Im9, although the two ensembles have similar overall architecture or folds, i.e., the typical distance between any two members of the TSE of Im9 is much larger than that of Im7. This is a surprising result, since the TSEs of the two proteins have similar experimental β_T values that exceed 0.85 and similar radii of gyration and solvent-accessible areas (see Table I). The results show that for both Im7 and Im9 a broad range of structures can satisfy a set of relatively low ϕ^{exp} values and yet be consistent with high β_T values; the latter was not introduced as a restraint on the ensembles. Moreover,

the TSEs of both Im7 and Im9 contain substantial numbers of non-native interactions as well as native ones, with both contributing to the stability of the ensembles. The compatibility of compactness and a high degree of heterogeneity of the TSE in these two α-helical proteins can be explained by the variability in the angles between the α helices in the TSE, even though (both from experiments and simulations presented here) the three long helices are highly formed in the TSE of both proteins. The results provide a structural characterization of the TSE's that is consistent with a diffusion-collision mechanism for the folding of these proteins.[19,32]

Our analysis provides a detailed characterization of the TSE which is not possible from the ϕ^{exp} values alone. For example, we showed that the accessible surface area of the fluorescent probes, W75 in Im7 and W74 in Im9, has a very broad distribution in the TSE, ranging from 10% to 100% exposure and that the distribution of the distances between the single Trp and the His that quenches its fluorescence in the native protein is very broad for the TSE of both proteins. These data thus highlight the breadth of the folding landscape, even at the level of a TSE which is 90% as compact as the native structure. The data also demonstrate the importance of dissecting the nearly properties of individual structures from the averaged parameters available from ϕ^{exp} values.

The synergy between theory and experiment[33–36] is an important element in the rapid progress that has been made recently in understanding the protein folding reaction. We have shown in this paper how the determination of the TSE consistent with experimental ϕ values can provide information not available from the ϕ values, per se. Im7 and Im9, which have a similar contact order, high sequence identity (60%), and essentially the same native structures, fold in a corresponding manner (i.e., the TSE's have the same fold), although they differ significantly in the breadth of the transition state ensemble that they traverse.

METHODS

Proteins

Im7[10] (PDB code 1AYI) and Im9[11] (PDB code 1IMQ) are homologous 87(86)-residue four-helix bundle proteins. In all the calculations done in this work, the respective PDB structures, shown in Figure 1, have been used as starting and reference structures. The ϕ values, which we have used for restraining the TSE ensemble, have been determined for 20 residues of Im9[22] and 22 residues of Im7.[19] In both cases the ϕ value refer to the folding from the denatured state. (i.e., $\phi_{TS} = \Delta\Delta G_{U-TS}/\Delta\Delta G_{N-U}$).

Microscopic Definition of ϕ Values

For a configuration at time t we define the calculated ϕ value of residue I as

$$\phi_I^{calc}(t) = \frac{N_I(t)}{N_I^{nat}} \tag{1}$$

where N_I, the number of contacts made by residue I, is

$$N_I = \sum_{i\in I}^{M}\sum_{j\notin I}^{M}\theta(r_{ij} - r_c)\Delta_{ij}(Q) \tag{2}$$

M is the number of side-chain atoms in the protein, r_{ij} the distance between the atoms i and j and $\Delta_{ij}(Q) = 1$ if atoms i and j are closer than a cut-off distance r_c in the native structure and belong to residues at least Q residues apart in the sequence, and zero otherwise. A definition based on side-chains is appropriate since experimental ϕ values are primarily a measure of the loss of side-chain contacts at the transition state, relative to the native state. It has been shown that there is a good correlation between loss of stability and loss of side-chain contacts within about 6 Å on mutation.[37]

For the set of structures that represent the TSE the experimental ϕ values, $\phi^{exp}{}_i$, can be compared with those calculated by averaging over the TSE; that is,

$$\langle\phi_i^{calc}\rangle = \frac{(N_i)^{TSE}}{N_i^{nat}} \tag{3}$$

The comparison between the $\phi_i{}^{calc}$and the $\phi_i{}^{exp}$ values provides a measure of the compatibility of the experimental transition state and an ensemble of structures determined by simulations. A good correlation between the two sets of values indicates that the TSE determined by molecular dynamics is consistent with the experiments. If enough $\phi_i{}^{exp}$ values are available the calculated TSE resembles the true transition state. Thus, a detailed study of the properties of the transition state and the interactions between different regions of the polypeptide chain can be carried out.

Bias for the Determination of the Transition State

The method for determining the TSE is implemented by introducing a small perturbation that forces the system to follow trajectories which, starting at the native state, lead to decreasing deviations between the experimental and calculated ϕ values. For this purpose we introduce the quantity

$$\rho(t) = \frac{1}{N_\phi}\sum_{i\in E}(\phi_i^{calc}(t) - \phi_i^{exp})^2 \tag{4}$$

where E is the list of the N_ϕ available experimental ϕ values, $\phi_i{}^{exp}$. The quantity ρ is the mean square deviation between the ϕ^{exp} and the ϕ^{calc} values. The TSE is reached by introducing a bias in the dynamics that slowly decreases the quantity ρ to about zero within 2 ns. Starting from this TS-like conformation, the TSE is sampled by performing 1 ns simulations at room temperature and at increasingly high temperatures with a harmonic term in the energy restraining the ϕ^{calc} values around the ϕ^{exp} values.

The conformational space compatible with the experimental restraints is rather large. To increase the efficiency of the sampling, we introduced a temperature-like parameter υ. This parameter is the temperature corresponding to the artificial energy used in the simulations. By varying

the value of υ we modify the computation time required to generate conformations compatible with the experimental restraints. Increasing the value of υ decreases the role played by the actual physico-chemical interactions, while keeping the restraints satisfied. Five values of υ were used, ranging from 300 K to 780 K. At high values of υ only restraints and steric interactions determine the structures sampled. These structures are rather non-native although their architecture is well defined and native-like. For this reason we use a supplementary criterion to select structures which have an average ϕ^{calc} (computed over all residues) close to the experimental average ϕ^{exp} value. This criterion excludes most of the structures sampled at υ = 300 K and at υ = 780 K. The pseudo-temperature υ would correspond to the real temperature only in the absence of restraints. As the ϕ^{exp} values have been measured every few positions along the polypeptide chains,[22] there are no structural regions that are locally unrestrained. Therefore in the simulations that we carried out there is no part of the system that is sampled at very high effective temperatures. The quantity ρ (i.e., the RMSD between the ϕ^{exp} and the ϕ^{calc} values) never exceeds 0.01; i.e., the experimental restraints are satisfied at least within their experimental error. The correlation between the ϕ^{exp} and the ϕ^{calc} values is larger than 0.95 (Im7) and 0.90 (Im9) in all simulations.

The 5,000 conformations (one each ps along the five trajectories) were clustered using a pairwise RMSD (using a 3 Å cut-off) and the central structure of clusters with more that five members were selected. The TSE discussed in this work consists of all those which have $\langle\phi^{calc}\rangle$ (computed for all residues, not only those experimentally measured) between $0.7\langle\phi^{exp}\rangle$ and $\langle\phi^{exp}\rangle$, where the average is computed for the residues for which ϕ^{exp} have been measured. This criterion is introduced to account for the fact that the residues mutated experimentally are mainly hydrophobic, so that their average ϕ value may represent an upper limit for the average ϕ value of all residues.

Other criteria to identify the putative transition state among the structures determined by the restrained MD might be devised; for example one can use the accessible surface area by invoking an approximate relation between solvent-accessible surface and the experimental m value.[21,38,39]

Molecular dynamics simulations were performed with the CHARMM program[40] using an all-atom model of the protein[41] and an implicit model for the solvent.[42] The potential function used, EEF1, provides a potential-of-mean-force description of the solvent. The EEF1 solvation free energy can be decomposed into a sum of pairwise interactions,[17,42] i.e., the effective energy, which includes the solvated contribution, can be written as a sum over pairs of residues as

$$E(EEF1) = \sum_{I} \sum_{J \geq I} E_{IJ} \qquad (5)$$

For both proteins, the sets of transition state structures is available upon request.

Conformational Averaging

The ϕ^{exp} values are measured as an average over a large ensemble of molecules in solution. Eq. (1) considers only one molecule and therefore it may create a distribution of ϕ^{calc} values that is artificially unimodal because the bias of Eq. (4) forces the instantaneous ϕ^{calc} value to be close to the ϕ^{exp} value. To address this problem, Davis et al.[23] introduced a Monte Carlo method in which two replicas of a protein molecule were simulated in parallel and the ϕ^{calc} values were calculated as an average over the two molecules. For Im7 and Im9 we have generalized this Monte Carlo method to an arbitrary number (M) of replicas and shown that for M ranging from 1 to 100 the distribution of ϕ^{calc} values were unimodal for all the residues, thus showing that the method of Eq. (1) can be applied with M = 1 in the case of Im7 and Im9. This is not a trivial result, since Davis et al.[23] showed that in the case of a small three-stranded β peptide, the distributions of ϕ^{calc} values is multimodal.

The *B* score

We recently proposed a measure, the "betweenness" (or B score), that detects the centrality of a residue in the TSE.[26] Residues with the largest B score in the TSE are those most important in the folding process. This method to identify important residues is complementary to the one that selects the minimal set of ϕ values to be used as a bias for the determination of the structure of the TSE.[21]

ACKNOWLEDGMENTS

We thank Christopher M. Dobson for advice and support while this work was written. We also thank Amedeo Caflisch, Jane Clarke, Andrew Capaldi, and Graham Spence for stimulating discussions. We are grateful to Urs Haberthür for the system management of the Beowulf cluster at the University of Zurich where all the simulations were performed. EP acknowledges financial support from Forschungskredit der Universität Zürich. C.T.F. and S.E.R. were supported by the Biotechnology and Biological Sciences Research Council (BBSRC), the University of Leeds and the Wellcome Trust. S.E.R. is a BBSRC Professorial Fellow. S.E.R. and C.T.F. are members of the Astbury Centre for Structural Molecular Biology, which is part of the North of England Structural Biology Centre and is supported by BBSRC. M.K. was supported in part by a grant from the NIH. M.V. was supported by a Royal Society University Research Fellowship.

REFERENCES

1. Goldenberg DP. Finding the right fold. Nat Struct Biol 1999;6:987–990.
2. Plaxco KW, Larson S, Ruczinski I, Riddle DS, Thayer EC, Buchwitz B, Davidson AR, Baker D. Evolutionary conservation in protein folding kinetics. J Mol Biol 2000;298:303–312.
3. Fowler SB, Clarke J. Mapping the folding pathway of an immunoglobulin domain: Structural detail from ϕ-value analysis and movement of the transition state. Structure 2001;9:355–366.
4. Martñez JC, Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat Struct Biol 1999;6:1010–1016.
5. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I,

Baker D. Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol 1999;6:1016–1024.
6. Northey JG, DiNardo AA, Davidson AR. Hydrophobic core packing in the SH3 domain folding transition state. Nat Struct Biol 2002;9:126–130.
7. Guerois R, Serrano L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. J Mol Biol 2000;304:967–982.
8. Ferguson N, Capaldi AP, James R, Kleanthous C, Radford SE. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. J Mol Biol 1999;286: 1597–1608.
9. Gorski SA, Capaldi AP, Kleanthous C, Radford SE. Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. J Mol Biol 2001;312:849–863.
10. Dennis CA, Videler H, Pauptit RA, Wallis R, James R, Moore GR, Kleanthous C. A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. Biochem J 1998;333: 183–191.
11. Osborne MJ, Breeze AL, Lian LY, Reilly A, James R, Kleanthous C, Moore GR. Three-dimensional solution structure and 13C nuclear magnetic resonance assignments of the colicin E9 immunity protein Im9. Biochemistry 1996;35:9505–9512.
12. Wales DJ. A microscopic basis for the global appearance of energy landscapes. Science 2001;293:2067–2070.
13. Fersht AR. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. New York: W.H. Freeman & Co.; 1999.
14. Paci E, Vendruscolo M, Dobson CM, Karplus M. Determination of a transition state at atomic resolution from protein engineering data. J Mol Biol 2002;324:151–163.
15. Paci E, Clarke J, Steward A, Vendruscolo M, Karplus M. Self-consistent determination of the transition state for protein folding. Application to a fibronectin type III domain. Proc Natl Acad Sci USA 2003;100:394–399.
16. Fersht AR. Protein folding and stability: The pathway of folding of barnase. FEBS Lett 1993;325:5–16.
17. Paci E, Vendruscolo M, Karplus M. Native and non-native interactions along protein folding and unfolding pathways. Proteins 2002;47:379–392.
18. Paci E, Vendruscolo M, Karplus M. Validity of Gō models: Comparison with a solvent-shielded empirical energy decomposition. Biophys J 2002;83:3032–3038.
19. Capaldi AP, Kleanthous C, Radford SE. Im7 folding mechanism: misfolding on a path to the native state. Nat Struct Biol 2002;9: 209–216.
20. Lindorff-Larsen K, Paci E, Serrano L, Dobson CM, Vendruscolo M. Calculation of mutational free energy changes in transition states for protein folding. Biophys J 2003;85:1207–1214.
21. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a protein folding transition state. Nature 2001;409:641–645.
22. Friel CT, Capaldi AP, Radford SE. Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and differences in the folding of homologous proteins. J Mol Biol 2003;326:293–305.
23. Davis R, Dobson CM, Vendruscolo M. Determination of the structures of distinct transition state ensembles for a β-sheet peptide with parallel folding pathways. J Chem Phys 2002;17: 9510–9517.
24. Ternström T, Mayor U, Akke M, Oliveberg M. From snapshot to movie: Analysis of protein folding transition states taken one step further. Proc Natl Acad Sci 1999;96:14854–14859.
25. Bastolla U, Vendruscolo M, Knapp EW. A statistical mechanical method to optimize energy functions for protein folding. Proc Natl Acad Sci USA 2000;97:3977–3981.
26. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. A small-world view of the amino acids that play a role in protein folding. Phys Rev E 2002;65:061910.
27. Girvan M and Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci USA 2002;99:7821–7826.
28. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 2002;320:369–387.
29. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 1992;224:771–782.
30. Horovitz A, Fersht AR. Cooperative interactions during protein folding. J Mol Biol 1992;224:733–740.
31. Wallis R, Leung KY, Osborne MJ, James R, Moore GR, Kleanthous C. Specificity in protein-protein recognition: conserved Im9 residues are the major determinants of stability in the colicin E9 DNase-Im9 complex. Biochemistry 1998;37:476–485.
32. Karplus M, Weaver DL. Protein-folding dynamics. Nature 1976; 260:404–406.
33. Fersht AR, Daggett V. Protein folding and unfolding at atomic resolution. Cell 2002;108:573–582.
34. Radford SE, Dobson CM. From computer simulations to human disease: Emerging themes in protein folding. Cell 1999;97:291–298.
35. Dobson CM, Karplus M. The fundamentals of protein folding: Bringing together theory and experiment. Curr Opin Struct Biol 1999;9:92–101.
36. Vendruscolo M, Paci E. Protein folding: Bringing theory and experiment closer together. Curr Opin Struct Biol 2003;13:82–87.
37. Cota E, Hamill SJ, Fowler SB, Clarke J. Two proteins with the same structure respond very differently to mutation: The role of plasticity in protein stability. J Mol Biol 2000;302:713–725.
38. Tanford C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. Adv Protein Chem 1970;24:1–95.
39. Lazaridis T, Karplus M. "New View" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
40. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. J Comp Chem 1983;4: 187–217.
41. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular dynamics system. J Chem Phys 1996;105: 1902–1921.
42. Lazaridis T, Karplus M. Effective energy function for protein dynamics and thermodynamics. Proteins 1999;35:133–152.
43. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometric features. Biopolymers 1983;22:2577–2637.
44. Koradi R, Billeter M, Wüthrich K. MOLMOL: A program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.