Protein modeling with reduced representation: statistical potentials and protein folding mechanism

Dariusz Ekonomiuk, Marcin Kielbasinski and Andrzej Kolinski✉

Faculty of Chemistry, Warsaw University, Warszawa, Poland;
✉e-mail: Kolinski@chem.uw.edu.pl

A high resolution reduced model of proteins is used in Monte Carlo dynamics studies of the folding mechanism of a small globular protein, the B1 immunoglobulin-binding domain of streptococcal protein G. It is shown that in order to reproduce the physics of the folding transition, the united atom based model requires a set of knowledge-based potentials mimicking the short-range conformational propensities and protein-like chain stiffness, a model of directional and co-operative hydrogen bonds, and properly designed knowledge-based potentials of the long-range interactions between the side groups. The folding of the model protein is cooperative and very fast. In a single trajectory, a number of folding/unfolding cycles were observed. Typically, the folding process is initiated by assembly of a native-like structure of the C-terminal hairpin. In the next stage the rest of the four-ribbon β-sheet folds. The slowest step of this pathway is the assembly of the central helix on the scaffold of the β-sheet.

Computer simulations of protein folding with all-atom representation of protein conformational space are still not practical. The enormous number of internal degrees of rotational freedom and the very complex intramolecular and intermolecular interactions, especially when the surrounding solvent is treated in an explicit fashion, prohibit molecular dynamics (MD) study of long time-scale processes in proteins (Lee *et al.*, 2001). Indeed, globular proteins fold in milliseconds to seconds (Anfinsen & Scheraga, 1975) (although there are exceptions of extremely fast and extremely slow folding processes), which is far beyond the capabilities of contemporary MD computational technology. Thus, it is necessary to simplify the problem by reducing protein representation and by designing a less rugged energy landscape of model proteins (Kolinski & Skolnick, 1996; 2004; Boniecki *et al.*, 2003). Ideally, these simplifications should not decrease the resolution of a model to a level where a reliable reconstruction of at least some atomic details becomes impossible. Such a protein model has been described recently in great details and its force field is available from our homepage http://biocomp.chem.uw.edu.pl for viewing and downloading. The CABS model (Kolinski, 2004) assumes a united atom representation of polypeptide chains with up to four centers of interactions per residue: alpha carbon (CA), beta carbon (B), the center of mass of the remaining portion of the side (S) group, and the center of virtual $C\alpha–C\alpha$ bond. Its force field consists of knowledge-based statistical potentials derived from a careful analysis of structural regularities seen in experimentally determined protein structures. The effect of the solvent is treated in an implicit fashion. It has been demonstrated that the CABS model and its earlier versions are very efficient tools for protein modeling allowing *de novo* structure prediction, extension of the range of applicability of comparative modeling, and other related applications in computational proteomics and protein biophysics. Probably the best proof of the versatility and capability of this new modeling tool was provided by the results of the 6th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction – CASP6, which was conducted in the summer of 2004. The blind predictions of protein structures by the predictor group Kolinski and Bujnicki

were scored second best among about two hundred groups participating in the experiment. This group used the CABS model for molecular simulations of almost all the target proteins. A compilation of the CASP6 results may be found either at the CASP6 homepage http://predictioncenter.llnl.gov/casp6/ or in the form of a short summary at our homepage http://biocomp.chem.uw.edu.pl. A detailed analysis and assessment of CASP6 results is provided in a special 2005 issue of the journal *Proteins: Structure, Function, and Bioinformatics*.

In this work we analyze the effects of various types of molecular interactions on the conformational properties of a CABS model protein and attempt to relate these effects to specific physical interactions in real systems. Then, we demonstrate that, in spite of the statistical origin of the interaction scheme, the CABS model reproduces semi-quantitatively the mechanism of protein folding. The model system selected for this study is the B1 immunoglobulin-binding domain of Streptococcal protein G (PDB code 2gb1) (Gronenborn *et al.*, 1991; Wikstrom *et al.*, 1994). Figure 1 shows a schematic drawing of the protein's native conformation. Although this protein is very small (56 amino acids) its fold is very stable and well-defined. At the same time the fold topology is quite complicated for a protein of this size and consists of a four-ribbon β-sheet with a helix on top of it. Thus, it is a perfect system for rapid testing of folding algorithms.

CABS MODEL: REPRESENTATION, FORCE FIELD AND SAMPLING SCHEME

Detailed description of the design of the CABS model may be found in the recent publication (Kolinski, 2004). Here an outline is provided for the reader's convenience. The framework of a polypeptide chain representation is its Cα-trace, restricted to the underlying simple cubic lattice with the mesh size equal to 0.61 Å. The virtual bonds between alpha carbons belong to a set of 800 vectors of the type $\mathbf{v} = [i, j, k]$ with integer coordinates i, j, k and with the vector's length defined by the following inequalities: $29 \leq |v|^2 \leq 49$. Such choice allows representation of crystallographic structures with an accuracy of about 0.35 Å cRMSD as measured on the alpha carbons after the best superposition. The Cα-trace provides a convenient reference system for defining the positions of the remaining united atoms, which are not restricted to the lattice. Relative positions of the Cβs and the centers of the side groups are derived from the statistics of known protein structures. Only two rotamers (the most probable ones) are allowed for the side groups. Cαs, Cβs, and centers of the Cα–Cα virtual bonds are treated as hard spheres, while the side groups are soft spheres. Fig-

ure 2 shows a schematic illustration of a short fragment of a CABS model polypeptide chain.

The model force field contains several components simulating physical interactions in real proteins. First, there are short-range conformational biases which impose a protein-like chain stiffness. In the spirit of polymer physics terminology short-range means here interactions between united atoms that are close to each other along the chain contour, i.e., are separated by one, two or three residues. The long-range means interactions between the united atoms, which are close in space, although separated by at least two segments along the chain (Kolinski, 2004). The second type of interactions is a set of sequence specific interactions which determine short-range conformational propensities. The third component of the force field is the highly directional potential simulating the main-chain hydrogen bonds. Model hydrogen bonds are defined in respect to mutual orientation and distance of the alpha carbons and the centers of the Cα–Cα virtual bonds. The pairwise interactions between the side groups are "context-dependent" and they change with changing mutual orientation of the interacting side groups and with changing conformation of the corresponding two-bond segments of the main chain. Twelve distinct geometrical contexts are taken into account. The pairwise potential, being derived for folded structures, accounts in an implicit way for the averaged effect of the surrounding solvent. For single domain globular proteins an additional component of the force field was added — the one body centrosymmetric potential, describing preferences of finding particular amino-acid residues at a given relative (measured in units of the expected radius of gyration of the native structure) distance from the center of gravity of the chain. All the potentials
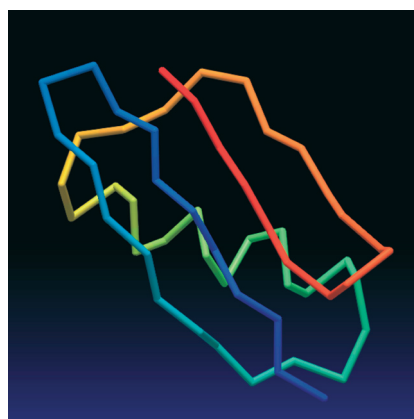

Figure 1. High resolution experimental structure of the B1 immunoglobulin-binding domain of Streptococcal protein G – 2gb1.
Only the alpha carbon trace is shown for the sake of clarity. The color code changes from blue at the N-terminus to red at the C-terminus.

were derived from the statistical regularities seen in a non-redundant database of high-resolution protein structures.

All simulations were performed at isothermal conditions using classical asymmetric Metropolis (Metropolis *et al.*, 1953) scheme, with a set of local conformational updates described elsewhere (Kolinski, 2004). The local character of the conformational transitions selected in a random mechanism leads to qualitatively correct long time dynamics of the model system. Long trajectories were generated and carefully analyzed for each case of selected subsets of model potentials. We inspected the average conformational properties of short protein fragments, average conformations of the entire protein and its dynamics along the trajectory.

ROLE OF VARIOUS INTERACTIONS IN PROTEIN FOLDING

In order to elucidate the effect of particular components of the model force field a series of simulations were performed with increasing "completeness" of the interaction scheme. With a single exception (discussed later) all the computational experiments were done at a reduced dimensionless temperature (measured in energy of interaction units) T = 1.5. This temperature is close to the collapse transition temperature, which was determined in a separate set of simulations. The relative scaling of particular interactions follows that obtained for optimized potentials described in previous publications (Boniecki *et al.*, 2003; Kolinski, 2004). All studied systems had the same excluded volume interactions and soft-core repulsions (scaling factor 4) of the side groups.

System with one-body centrosymmetric potential (S1)

Such a system simulates a simple hydrophobic collapse model of protein folding. The strength of the centrosymmetric potential is assumed to be equal to 1, as in all of the remaining experiments described below. The chain is completely flexible. While the hydrophobic residues tend to be buried inside the forming globule, the polar and charged residues tend to be exposed. This simulation addresses the long-standing question: Can the hydrophobic collapse itself induce formation of elements of secondary structure? (Dill *et al.*, 1995). The answer is that the effect is very small, if not negligible. Analysis of the simulation trajectories shows that the expanded random coil and more compact, partially collapsed, chains have very similar average local conformations, which can be described as loosely defined and fluctuating extended fragments punctuated by more flexible bent fragments. The
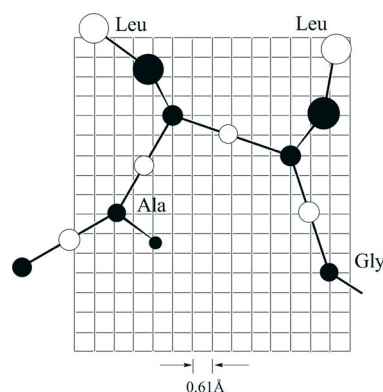

Figure 2. An example of a short fragment of CABS model polypeptide chain.
The solid circles correspond to the alpha carbon united atoms (restricted to the vertices of the underlying simple cubic lattice with the mesh size equal to 0.61 Å) and beta carbons (off lattice). The open circles indicate the centers of the side groups and centers of the virtual Cα–Cα bonds.

fragment which corresponds to the helix in the native structure has also mostly extended conformations. Interestingly, the bent fragments correlate weakly with the turn regions in the native structure. This is not surprising as these turn regions are more hydrophilic than the rest of the sequence. Therefore, the centrosymmetric potential favors their average location on the surface of the collapsed globular states. This is illustrated in Fig. 3, where the average cRMSD from the corresponding native structure for seven-residue fragments is plotted as a function of the fragment position along the chain. The value
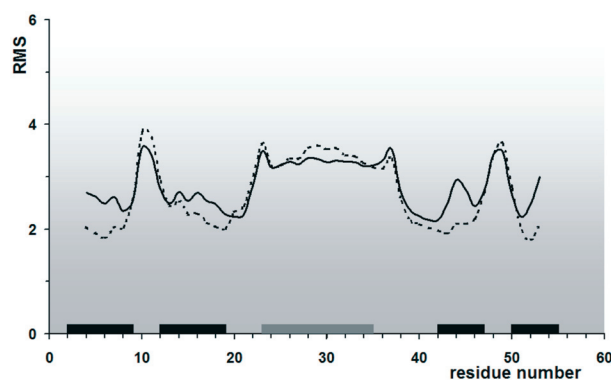

Figure 3. Average values of the cRMSD from the corresponding native fragments for the Cα-trace of seven-residue fragments (i-3, i+3) as a function of the position along the chain i.
The black horizontal bars indicate positions of the β-strands in the native structure. The gray bar indicates the position of the central helix. The solid line corresponds to simulations of the system S1 with excluded volume and one-body centrosymmetric interactions only. The dashed lines (system S2) is for simulations with the S1 interactions and the short range (generic and sequence specific) interactions. In both cases the reduced temperature of the model systems was T = 1.5.

of cRMSD for the i-th residue corresponds to the fragment spanning residues from i–3 to i+3. Since most of the molecule (on average) adopts irregular extended conformations the values of cRMSD for the β-strand regions in the native structure (marked by black bars) is systematically lower (about 2.5 Å) than for the helical (grey bar) and turn fragments, where it is in the range of 3.5–4.0 Å, which is characteristic for essentially random correlations.

System with one-body centrosymmetric potential and short-range interactions (S2)

Introduction of short-range interactions does not change the general picture described in the previous section (see also Fig. 3). Generic protein-like stiffness (scaling factor 1.0) and sequence-specific short range conformational propensities (scaling factor 0.5) influence essentially only the putative β-sheet regions, where the average distance from the native structure decreases to about 2.0 Å. The putative helix forms very rarely, most of the time remaining in a loosely defined and rapidly changing extended state. Why the short-range interactions enhance the β-strands but not the helix? The most likely explanation is related to the different entropy of the two types of local structure. The extended basin is wider. There is a huge number of possible geometrical arrangements of disordered expanded structures, while the helical basin is narrow with a
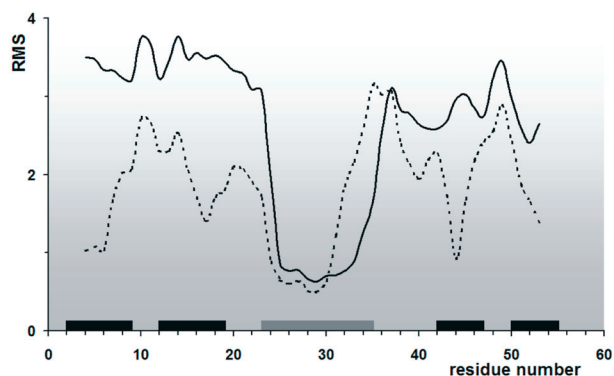

Figure 4. Average values of the cRMSD from the corresponding native fragments for seven-residue fragments (i–3, i+3) as a function of the position along the chain i. The black horizontal bars indicate positions of the β-strands in the native structure. The gray bar indicates the position of the central helix. The solid line corresponds to simulations of the system S3 with excluded volume, one-body centrosymmetric potential, the short range interactions and main chain hydrogen bonds. The dashed lines (system S4) is for simulations with complete force field, i.e., with the pairwise interactions of the side groups added to the interaction scheme of the S3 system. In both cases the reduced temperature of the model systems was T = 1.5. For the complete force field (S4) the picture is similar at the somewhat higher temperature T = 1.75, although the mobility of the system is much higher.

much smaller number of possible helix-like conformations.

System with hydrogen bonds (S3)

In these simulations, on top of the interactions of the S2 system, a model of main chain hydrogen bonds was turned on (scaling factor –2.5). The results are illustrated in Fig. 4 (solid line). In these simulations the central helix is very well defined and dissolves only occasionally for short periods of time. The average cRMSD for the helix is below 1.0 Å. For the remaining portions of the model chain the average cRMSD from the native structure is large. These regions are defined more poorly than in the system without hydrogen bonds (S2). Again, the reason is of entropic nature. The short range helical hydrogen bonds are easy to form and propagate. The loss of entropy per one hydrogen bond is small. In contrast, to form even a single long range hydrogen bond in a β-sheet requires a huge reduction of the conformational freedom of the two involved strands. Large values of cRMSD for the non-helical parts of the chain are due to frequently formed small helical fragments, even in the putative β-strand regions.

Simulations with complete force field (S4)

Adding pairwise interactions of the side groups to the set of interactions of the S3 system completes the force field of the CABS model. The results are illustrated in Fig. 4 (dashed line). The pairwise interactions change dramatically (when compared to the S3 system) the properties of the putative strands, while the effect on the helical region is less significant. Here the average accuracy of the helix (measured as the cRMSD from the native helix in the best superposition) approaches the resolution limit of the model and is about 0.5 Å. Also the β-strands frequently adopt the native-like conformations, especially the C-terminal hairpin. The average cRMSD is in the neighborhood of 1.0 Å. The largest deviations from the native local conformation are observed for the second strand, which is the N-terminal edge strand of the sheet. This is not surprising as buried strands in sheets usually have a more regular structure and are structurally more stable than the edge strands. The entire protein frequently adopts a loosely defined native fold, however, plentiful of misfolded conformations are observed along the trajectory. These misfolded structures have essentially correct elements of secondary structure, but their topology (for example the order of strands in the sheet) is frequently incorrect. Also the packing angle of the helix on top of the sheet may happen to be wrong in numerous "partly" misfolded structures. The mechanism of the fold assembly and some additional aspects of fold stability are discussed in the

next section. Here, it should be just noted that the proposed force field of the CABS model is capable of a very rapid assembly of a native-like structure of the 2gb1 protein domain. It should be pointed out that the input data for the simulations did not contain any information about the native or predicted secondary structure. Therefore, these simulations could be considered as purely *de novo in silico* folding experiments within the framework of a knowledge-based approach to the design of the force field. Other small proteins behave in a similar fashion. The more complex the topology and the larger the protein the less frequently the native fold is observed. Also the required simulation time grows rapidly with the size of the system studied.

FOLDING MECHANISM OF THE B1 DOMAIN OF PROTEIN G

Studies of the folding mechanism were performed at the reduced temperature T = 1.75, which is slightly above the folding transition temperature. The increased temperature increases significantly the system's mobility. Consequently the folding-unfolding events could be observed many times during a single long isothermal simulation. Figure 5 shows the cRMSD for the entire structure as a function of the simulation time. One can see a wide range of cRMSD values, from about 3 Å, which corresponds to low-resolution native-like conformations, to about 14 Å, the value for completely misfolded (or unfolded) states. A more careful inspection of the flowchart indicates a periodical behavior. System tends to stay for a period of time near the native conformation and then there is a rapid transition to an unfolded/misfolded structure. This suggests a cooperative character of the folding process (Finkelstein & Sha-
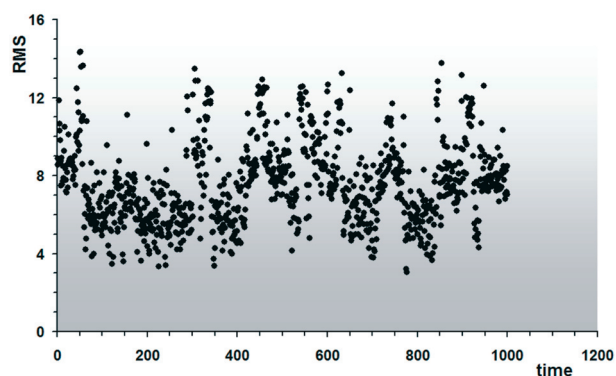

Figure 6. Flowchart for the N-terminal hairpin from the simulation run illustrated in Fig. 5.
cRMSD is calculated for residues 2–19.

khnovich, 1989; Kolinski *et al.*, 2003). A number of folding/unfolding cycles are clearly visible. Indeed, according to the vast body of experimental evidence such small one-domain monomeric proteins should fold rapidly and reversibly in a cooperative fashion (Finkelstein & Shakhnovich, 1989).

Figures 6–8 show corresponding flowcharts where the cRMSD is measured for parts of the structure: the N-terminal hairpin, the helix and the C-terminal hairpin, respectively. Comparison of these flowcharts leads to several interesting observations. First, the helix and the C-terminal hairpin fold to a very high resolution of about 0.5 Å, while the N-terminal hairpin folding is less accurate and the best folded structures are about 1 Å from the native. Second, the folding of both hairpins is much faster than the folding of the entire protein. No periodic behavior could be detected in the timescale of the plot — the folding time is less than the assumed time-unit of the simulations. In contrast, a clear signature of periodicity could be observed in the helix flowchart. This indicates that the assembly of the helix is the rate-controlling folding event. Indeed, the periodicity of the entire protein folding flowchart and the flowchart of the helix are clearly correlated. Third,

Figure 5. Flowchart from an isothermal (T = 1.75) simulation of the B1 domain of protein G.
The cRMSD from the native structure for the Cα-trace after the best superposition is plotted as a function of simulation time measured in arbitrary units. One time unit corresponds to several thousands of attempted local conformational transitions per residue.
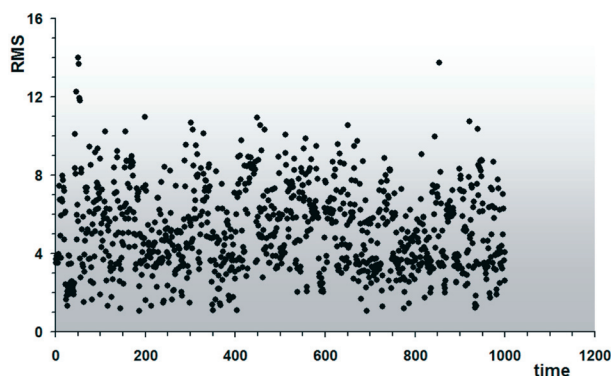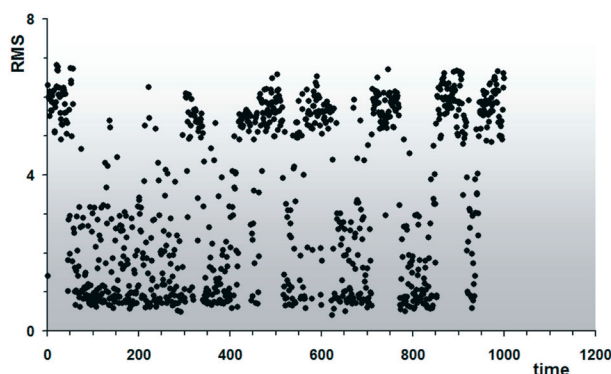

Figure 7. Flowchart for the central helix from the simulation run illustrated in Fig. 5.
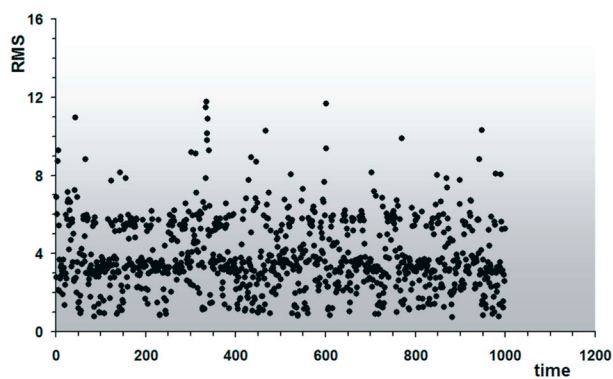cRMSD is calculated for residues 23–35.

Figure 8. Flowchart for the C-terminal hairpin from the simulation run illustrated in Fig. 5.
cRMSD is calculated for residues 42–55.

folding of the helix and the C-terminal hairpin is cooperative. There is a clear gap in the population of intermediates, the majority of structures have values of cRMSD characteristic for either near-native or completely unfolded states. A crude comparison of the relative populations of these two types of structures shows that at the experiment temperature T = 1.75 the relative frequency of folded structures is larger for the C-terminal hairpin than for the helix. Thus, the hairpin is more stable than the helix. It is known from several experiments (Kuszewski *et al.*, 1994; Frank *et al.*, 1995) and from computational studies (Munoz *et al.*, 1997; Sheinerman & Brooks, 1998; Kolinski *et al.*, 1999; Klimov & Thirumalai, 2000; Zhou *et al.*, 2001; Karanicolas & Brooks, 2002; Bolhuis, 2003; Ma & Nussinov, 2003; Brown & Head-Gordon, 2004) that the peptide identical with the C-terminal hairpin of protein G behaves as a small globular protein; it is stable in solution and folds in a cooperative fashion. There is one more interesting feature of the C-terminal flowchart. The most populated are structures with the cRMSD from native one in the range of 2.5–4.0 Å. This actually could be quite physical. This hairpin folds more frequently than the entire protein does. The folded isolated hairpin does not need to have exactly the same structure as the hairpin in the native fold of protein G. Very likely the twist of this minimal sheet would be somewhat different, however, as the plot in Fig. 8 indicates, such isolated mini-protein has a very well-defined structure, with the same long-range interaction pattern as the one present in the protein G hairpin.

An additional insight into the mechanism of protein G folding could be gained from analysis of the series of snapshots shown in Fig. 9. All the structures are compact, although the average radius of gyration is by about 30% larger than that of the native 2gb1 structure. In the majority of the snapshots the N-terminal and the C-terminal hairpins are intact. A completely assembled helix is observed

less frequently. On several snapshot, the contacting N-terminal and C-terminal strands have the correct, native-like, orientation. In general, most of the snapshots have several features of the molten globule state.

Compilation of the observations from the density of states and the periodicity seen in the flowcharts (Fig. 5–8) and from the representative snapshots from the simulation trajectory (much more snapshots have been inspected) leads to the formulation of the following most probable folding pathway. The nucleation occurs frequently *via* assembly of a slightly distorted C-terminal hairpin (Honda *et al.*, 2000). Then, the N-terminal hairpin assembles on the C-terminal hairpin scaffold, starting usually by zipping-up the N-terminal strand with the C-terminal one. Subsequently, the resulting β-sheet serves as a scaffold for the helix assembly and slow adjustment of its orientation in respect to the sheet, accompanied by simultaneous deformation of the β-sheet towards a more native-like twist. The last events appear to be the most time-consuming stages of the fold assembly. It is unclear how long it takes to fine-tune the interaction of the side chains. Anyway, this aspect of the folding process could be significantly biased in our simulation due to a rather crude representation of the side chains (two united atoms and only two rotamers available). Nevertheless, the picture of the fold assembly mechanism emerging from these simulations seems to be in at least qualitative agreement with known experimental facts (Kuszewski *et al.*, 1994; Blanco & Serrano, 1995; Kobayashi *et al.*, 1995; Honda *et al.*, 2000; McCallister *et al.*, 2000).

Finally, a bit speculative (albeit plausible) conclusion could be drawn from the present studies. It is now very well recognized that the knowledge-based approaches are very powerful tools for protein structure prediction. The present studies strongly suggest that the denatured state and the folding mechanism could also be studied with the help of the knowledge based force field built on properly designed statistical potentials. This conclusion is not so surprising in the context of recent studies indicating a much higher level of native-like local structure and compactness in the denatured state of proteins than assumed previously.

CONCLUSIONS

In this work we have shown that a minimal set of knowledge based potentials for a model assuming reduced, united atom, representation of protein conformational space requires short-range interactions mimicking protein-like stiffness of the model chain, a model of main chain directional hydrogen bonds and properly designed amino acid specific
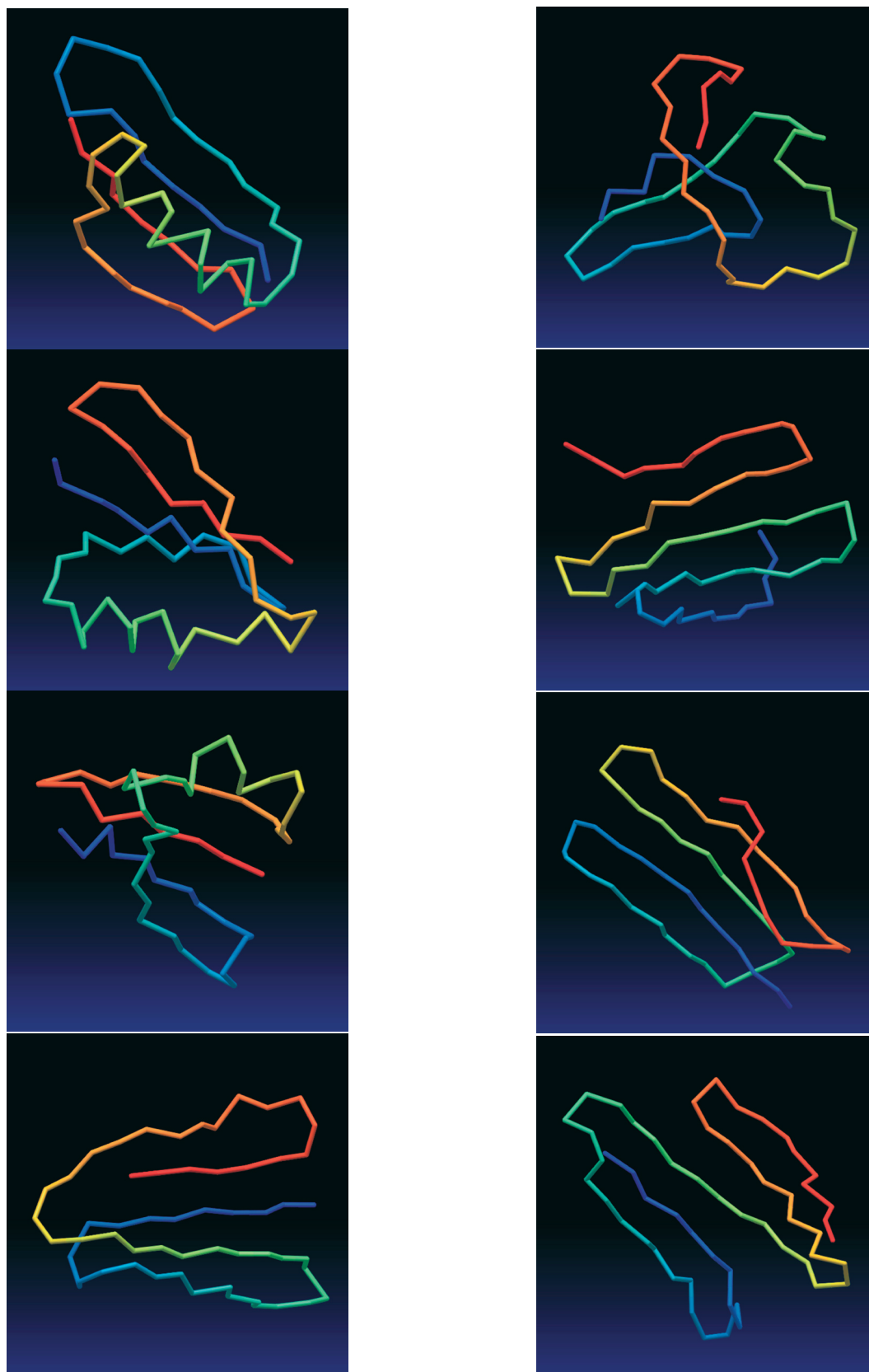
Figure 9. Snapshots of protein G conformations from the isothermal simulation at T = 1.75.
The snapshots were taken at equal intervals of the simulation time t: a) at $t = 125$, b) at $t = 250$, ..., and h) at $t = 1000$.

long-range pairwise interactions. Such a set of interactions enables *de novo* structure prediction and reproduces main features of protein folding mechanism, as it was demonstrated on the example of *in silico* folding of the B1 immunoglobulin-binding domain of streptococcal protein G.

Acknowledgement

REFERENCES

Anfinsen CB, Scheraga HA (1975) Experimental and theoretical aspects of protein folding. *Adv Prot Chem* **29:** 205–300.

Blanco FJ, Serrano L (1995) Folding of protein G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. *Eur J Biochem* **230:** 634–649.

Bolhuis PG (2003) Transition-path sampling of beta-hairpin folding. *Proc Natl Acad Sci USA* **100:** 12129–12134.

Boniecki M, Rotkiewicz P, Skolnick J, Kolinski A (2003) Protein fragment reconstruction using various modeling techniques. *J Comput Aided Mol Des* **17:** 725–738.

Brown S, Head-Gordon T (2004) Intermediates and the folding of proteins L and G. *Protein Sci* **13:** 958–970.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS (1995) Principles of protein folding — A perspective from simple exact models. *Prot Sci* **4:** 561–602.

Finkelstein AV, Shakhnovich EI (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* **28:** 1681–1694.

Frank MK, Clore GM, Gronenborn AM (1995) Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci* **4:** 2605–2615.

Gronenborn A, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253:** 657–660.

Honda S, Kobayashi N, Munekata E (2000) Thermodynamics of a beta-hairpin structure: evidence for cooperative formation of folding nucleus. *J Mol Biol* **295:** 269–278.

Karanicolas J, Brooks III CL (2002) The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci* **11:** 2351–2361.

Klimov DK, Thirumalai D (2000) Mechanisms and kinetics of beta-hairpin formation. *Proc Natl Acad Sci USA* **97:** 2544–2549.

Kobayashi N, Honda S, Yoshii H, Uedaira H, Munekata E (1995) Complement assembly of two fragments of the streptococcal protein G B1 domain in aqueous solution. *FEBS Lett* **366:** 99–103.

Kolinski A (2004) Protein modeling and structure prediction with a reduced representation. *Acta Biochim Polon* **51:** 349–371.

Kolinski A, Skolnick J (1996) *Lattice Models of Protein Folding, Dynamics and Thermodynamics*. R.G. Landes, Austin, TX.

Kolinski A, Skolnick J (2004) Reduced models of proteins and their applications. *Polymer* **45:** 511–524.

Kolinski A, Gront D, Pokarowski P, Skolnick J (2003) A simple lattice model that exhibits a protein-like cooperative all-or-none folding transition. *Biopolymers* **69:** 399–405.

Kolinski A, Ilkowski B, Skolnick J (1999) Dynamics and thermodynamics of β-hairpin assembly: Insight from various simulation techniques. *Biophys J* **77:** 2942–2952.

Kuszewski J, Clore GM, Gronenborn AM (1994) Fast folding of a prototypic polypeptide: The immunoglobin binding domain of streptococcal protein G. *Protein Sci* **3:** 1945–1952.

Lee MR, Tsai J, Baker D, Kollman PA (2001) Molecular dynamics in the endgame of protein structure prediction. *J Mol Biol* **313:** 417–430.

Ma B, Nussinov R (2003) Energy landscape and dynamics of the beta-hairpin G peptide and its isomers: Topology and sequences. *Protein Sci* **12:** 1882–1893.

McCallister EL, Alm E, Baker D (2000) Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* **7:** 669–673.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* **51:** 1087–1092.

Munoz V, Thompson PA, Hofrichter J, Eaton WA (1997) Folding dynamics and mechanism of β-hairpin formation. *Nature* **390:** 196–197.

Sheinerman FB, Brooks III CL (1998) Calculations on folding of segment B1 of streptococcal protein G. *J Mol Biol* **278:** 439–456.

Wikstrom M, Drakenberg T, Forsen S, Sjobring U, Bjorck L (1994) Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. *Biochemistry* **33:** 14011–14017.

Zhou R, Berne BJ, Germain R (2001) The free energy landscape for beta hairpin folding in explicit water. *Proc Natl Acad Sci USA* **98:** 14931–14936.