Application of biasing-potential replica-exchange simulations for loop modeling and refinement of proteins in explicit solvent

Srinivasaraghavan Kannan and Martin Zacharias*

School of Engineering and Science, Jacobs University Bremen, Bremen D-28759, Germany

INTRODUCTION

Knowledge of the three-dimensional (3D) structure of protein molecules is essential to understand its function. Experimental structure determination is time consuming and costly and the structure of only a limited number of proteins relative to the total number of protein sequences has so far been solved. The enormous recent progress in high-throughput genome sequencing further increases the gap between the number of known protein sequences and known structures. Comparative modeling techniques are increasingly being used to generate model structures for many proteins with sequence similarity to a known template structure.1 Template based comparative (or homology) modeling methods are so far the most reliable modeling approach for modeling protein structures.1–12 However the prediction accuracy is limited by the availability of suitable template structures, the accuracy of the target-template sequence alignment, and the modeling of segments with low (or no) similarity to a template.3,12,13 Often homology models of proteins are accurate for parts of the protein with high sequence similarities to a template but are much less realistic for flexible regions such as loops that connect conserved structural elements. Loops often determine the functional specificity of a given protein structure and can contribute to enzymatic activity or ligand binding properties.14

Loop modeling is defined as construction of 3D atomic models for short protein segments that connect regular secondary structure elements. Several methods have been developed to predict loop con conformations.12–16 Generally, loop modeling methods can be classified into knowledge-based methods, de novo or ab initio methods and combined approaches. Knowledge-based approaches employ databases of experimental protein structures as a source of loop conformations.17–24 Possible loop conformations for a given protein segment are evaluated by using rule-based filters, with evaluating criteria such as geometric fit (e.g., distance of the loop termini), sequence similarity to the target segment and inclusion of knowledge-based potentials. The de novo structure prediction approaches perform conformational

ABSTRACT

Comparative or homology modeling of a target protein based on sequence similarity to a protein with known structure is widely used to provide structural models of proteins. Depending on the target-template similarity these model structures may contain regions of limited structural accuracy. In principle, molecular dynamics (MD) simulations can be used to refine protein model structures and also to model loop regions that connect structurally conserved regions but it is limited by the currently accessible simulation time scales. A recently developed biasing potential replica exchange (BP-REMD) method was used to refine loops and complete decoy protein structures at atomic resolution including explicit solvent. In standard REMD simulations several replicas of a system are run in parallel at different temperatures allowing exchanges at preset time intervals. In a BP-REMD simulation replicas are controlled by various levels of a biasing potential to reduce the energy barriers associated with peptide backbone dihedral transitions. The method requires much fewer replicas for efficient sampling compared with T-REMD. Application of the approach to several protein loops indicated improved conformational sampling of backbone dihedral angle of loop residues compared to conventional MD simulations. BP-REMD refinement simulations on several test cases starting from decoy structures deviating significantly from the native structure resulted in final structures in much closer agreement with experiment compared to conventional MD simulations.

Proteins 2010; 78:2809–2819. © 2010 Wiley-Liss, Inc.

Key words: conformational sampling; molecular dynamics simulation; loop modeling; refinement; protein structure prediction.
homology models. Misura et al. attempted to refine a set of
approaches have also been applied to the refinement of
computationally expensive with the refinement of one model
within the 10 lowest-energy models. However, the method is
found in 22 out of 39 cases a model that is closer to the
resembles de novo search methods encounter similar diffi-
cal conformations or size of the conformational space
increases with increasing loop length. 7
Similar to modeling of protein loop segments the refine-
ment of protein structures built by comparative modeling
requires both large scale conformational sampling and accu-
rate atomic energy functions. Progress has been made using
various knowledge based potentials that are augmented by
energy terms motivated by consideration of important physi-
cal interactions and sampling is often facilitated by low reso-
lation initial conformational searches. Baker and co-
workers for example, reported encouraging results for 5 of
16 small proteins using a refinement protocol in which multiple
rounds of random torsion-angle perturbation and Monte Carlo (MC)
relaxation were performed on low-resolution models built from a set of sequence homologues of the target protein using the Rosetta approach. Another method based on fragment assembly and MC simulation is Tasser, which has been applied to the refinement structures determined by NMR spectroscopy. Recently, these ab initio approaches have also been applied to the refinement of homology models. Misura et al. attempted to refine a series of homology models using Rosetta in combination with evolutionarily derived distance constraints. The authors found in 22 out of 39 cases a model that is closer to the native structure than the template over the aligned regions within the 10 lowest-energy models. However, the method is computationally expensive with the refinement of one model requiring several CPU days. In addition to the approaches outlined earlier, a number of methods that employ statistical potentials or empirical scoring functions to select near-native models from an ensemble of homology models have also been described.

Alternatively, molecular dynamics (MD) should be potentially useful for structure refinement and loop modeling. In principle, MD simulations include conformational entropy effects and could serve as tool for conformational search of the loop region and also for atomic resolution refinement of protein model. However, at room temperature affordable MD simulations can be kinetically trapped and explore only the basin of attraction near the starting structures. Several methods have been proposed to overcome the barrier crossing problems. For example, Viktor and Simmerling have used low barrier MD simulations in implicit solvent to generate loop conformations. Kirs et al. and Riemann and Zacharias used adjustable - barrier dihedral potentials for conformational search of protein loops during MD simulations. Hao and Mark carried out conventional (c)MD simulations in explicit water for the refinement of homology modeled structures. Though still a significant deviation of simulated structures from the experimental structures was observed, simulations of tens to hundreds of nanoseconds resulted on average in final conformations in closer agreement with experiment than the start structures.

The replica exchange MD (REMD) simulation method has been frequently used for enhancing the conformational sampling of biomolecules. In REMD simulations, several copies (replicas) of the system are simulated independently and simultaneously using classical MD at different simulation temperatures (or force fields: Hamiltonians). At preset intervals, pairs of replicas (neighbouring pairs) are exchanged with a specified transition probability. In most REMD simulations the temperature is used as a parameter that varies among the replicas (T-REMD). The random walk in temperature allows conformations trapped in locally stable states (at a low simulation temperature) to escape by exchanging with replicas at higher simulation temperature. T-REMD simulations in combination with a Generalized Born implicit solvent model have been used for the prediction of loop conformations and refinement. Recently, Jiang et al. have used T-REMD simulations with a statistical potential for refinement of homology modeled proteins in explicit solvent. However, a major drawback of T-REMD simulations is the rapid increase of the number of required replicas with the system size to cover a desired temperature range in turn requiring rapidly increasing computational resources.

As outlined above, the majority of loop prediction and structural refinement methods employs an implicit solvent description. In particular, in case of solvent exposed loop structure this may strongly limit the realistic prediction of loop structures. Ultimately, it is desirable to efficiently predict loop structures and perform structural refinement accounting explicitly for surrounding solvent and ions. Promising results have been achieved using very extensive conventional MD simulations but requiring large computational resources not useful as standard rapid refinement methodology.

Recently, we have proposed a Biasing Potential Replica Exchange (BP-REMD) method that specifically lowers barriers for peptide backbone transitions along the replicas. This method requires much fewer replicas than T-REMD especially in the presence of explicit solvent. In the current work we have applied BP-REMD simulations.
for modeling of loops in protein structures and to refine protein decoy structures in explicit solvent. In 13 test cases the 5 replica BP-REMD protocol is able to identify a most populated conformational cluster closer to experiment than the start structure and performed always better than five independent cMD simulations.

### MATERIALS AND METHODS

#### Test structures

The initial structures for the loop modeling and for refinement were selected from the Rosetta decoy set. For loop modeling several alpha, beta and alpha/beta protein decoys were chosen (see Table I), in such a way that the model contained already most (>90%) of the correct secondary structure elements and differed only in the connecting loop parts from the native structures. To perform realistic loop modeling, distances restraints were applied to the rest of protein molecule other then the specific loop residues. It is important to note that distances between all Cα-atoms excluding the specific loop residues, were derived from the initial decoy structure (not the native structure) and were included during both MD and BP-REMD simulations. During the simulations structures were allowed to move freely within Cα − Cα distance changes of ±0.5 Å and a quadratic penalty with a force constant of 1 kcal mol⁻¹ Å⁻² for deviations larger then 0.5 Å from the decoy start structure. The target loop regions connecting secondary structure elements were completely free during all stages of the simulations.

For the refinement simulations of complete proteins seven alpha and alpha/beta protein decoy structures from the Rosetta decoy set were used (see Table II). No restraints were used during these refinement simulations.

#### Simulation setup

Hydrogen atoms were added to the decoy structures using the xleap module of the Amber9 package. Explicit TIP3P water molecules were added to all decoys to form a truncated octahedral boxes using xleap (minimum distance between protein and box boundary: 10 Å). All MD simulations were carried out with the Sander module of the AMBER9 package in combination with the parm03 force field. The simulation systems were first subjected to energy minimization (1000 steps). During MD simulations, the protein was initially harmonically restrained (25 kcal mol⁻¹ Å⁻²) to the energy minimized start coordinates, and the system was heated up to 300 K in steps of 100 K followed by gradual removal of the positional restraints and a 1ns unrestrained equilibration at 300 K. The resulting system was used as starting structure for both BP-REMD and conventional MD simulations. During all simulations the long range electrostatic interactions were treated with the particle mesh Ewald (PME) method using a real space cutoff distance of cutoff = 9 Å (constant pressure of 1 bar). The Settle algorithm was used to constrain bond vibrations involving hydrogen atoms, which allowed a time step of 2 fs.

#### Biasing potential replica-exchange simulations

The BP-REMD method is a Hamiltonian REMD method that employs a biasing potential for the Φ and
psi peptide backbone dihedral angles. The biasing potential is based on a potential of mean force (PMF) for each of the two dihedral angles calculated for a model peptide (alanine dipeptide) in explicit solvent. Addition of the biasing potential during a simulation lowers the energy barriers for backbone dihedral transitions in a peptide or protein essentially making the protein chain more flexible. The biasing potential had been derived previously by potential-of-mean force (PMF) free energy simulations on the backbone dihedral \( \Phi \) and \( \Psi \) of Alanine dipeptide in explicit water. During a BP-REMD the level of biasing is gradually changed along the replicas such that frequent transitions are possible at high levels of biasing. Note, that the same biasing potential derived from the model system was used for all backbone dihedral in the protein (except Gly and the Pro \( \Phi \) angle).

In a BP-REMD simulation different biasing potential levels are applied in each replica (one reference replica runs without any biasing potential) and replica exchanges between neighboring biasing levels were attempted every 1000 MD steps (2ps) and accepted or rejected according to a Metropolis criterion (5000 attempted exchanges in 10 ns). An advantage compared to T-REMD is the fact that the energy differences are only affected by the force field term that changes upon going from one replica to another replica run. For the present simulations we used five replicas, and the acceptance probability for replica exchanges was in the range of 30–40%.

For each case two independent sets of BP-REMD simulations (10 simulations in total) and two sets of conventional MD simulations with five independent standard MD simulations per set were carried out at 300K using two different starting decoy conformations. In the case of loop modeling distance restraints were used during both MD and BP-REMD simulations (no restraints during full refinement).

### Test Systems and Results of Refinement Simulations

| Protein | Type   | \( N_{\text{tot}} \) | Initial Rmsd (\( \AA \)) | Lowest Rmsd (\( \AA \)) | Rmsd (\( \AA \)) cluster | BP-REMD |
|---------|--------|----------------------|--------------------------|--------------------------|--------------------------|---------|
| 1PGX    | Alpha/beta | 57                    | 3.0 (4.0)                | 2.3 (3.4)                | 2.5 (3.8)                | 1.6 (2.6) | 1.7 (2.9) |
| 1RES    | Alpha   | 35                    | 2.8 (3.5)                | 2.0 (3.2)                | 3.3 (4.1)                | 1.1 (2.1) | 1.4 (2.3) |
| 1R69    | Alpha   | 61                    | 2.4 (3.4)                | 2.0 (3.2)                | 2.1 (3.6)                | 1.2 (2.0) | 1.3 (2.2) |
| 1UBA    | Alpha   | 33                    | 3.7 (4.4)                | 2.3 (3.5)                | 3.0 (4.1)                | 1.8 (3.2) | 1.9 (3.3) |
| 1GAB    | Alpha   | 47                    | 8.1 (9.6)                | 9.1 (10.6)               | 10.2 (11.9)              | 4.3 (5.3) | 4.8 (5.9) |
| 1PGX    | Alpha/beta | 57                    | 3.0 (4.0)                | 2.3 (3.4)                | 2.5 (3.8)                | 1.6 (2.6) | 1.7 (2.9) |

See legend of Table I. The last 5ns of simulation trajectories were used for calculation of lowest Rmsd and after performing cluster analysis (C(\( \alpha \)) atom). Values in the brackets correspond to the Rmsd of heavy atoms with respect to the experimental structure. The second row for each case corresponds to the results of a second independent simulation starting from a different start structure.

### Results

#### Prediction of protein loop structure

MD loop modeling simulations were performed on three alpha helical proteins, two beta proteins and one \( \alpha \)-beta protein structures. The protein structures contained loop structures ranging in size between 3 and 7 residues. For each case two different starting structures were used with heavy atom root-mean square deviation (Rmsd) of 2.5–5.5 \( \AA \) from the native loop structure (Table I). Decoy start structures were obtained from the Rosetta protein decoy set and contained already almost correctly folded secondary structures (>90% according to the DSSP program) with the main structural variation due to loop regions that connect secondary structure elements. To stabilize the near-native secondary structure and also the spatial arrangement of secondary structure elements distance restraints (between backbone C(\( \alpha \)) atoms) were included during both MD, and BP-REMD simulations to prevent the system from unfolding or sampling states far away from the folded structure (see Methods for details). Note, that the distance restraints were derived entirely from the decoy start structures without inclusion of any knowledge of the native structure.

The distance restraints were sufficiently soft to allow distance fluctuations of \( \pm 0.5 \) \( \AA \) without any change in restraining energy. The protocol corresponds basically to a loop refinement approach that in contrast to most other approaches with fixed loop anchor site allows considerable freedom for adjustment of secondary anchor elements connected to the flexible loop segments (the loop regions were completely mobile during all simulation stages). For all the cases five independent standard MD simulations (with different initial velocities) were...
carried out from two different start structures at 300 K. The results of MD loop modeling simulations are summarized in Table I and are illustrated for pdb5znf and pdb1pft in Figure 1. The Rmsd of sampled conformations was calculated by superimposing the whole protein on to its native structure and calculating the Rmsd only of the loop part. Hence, the calculated Rmsd includes also the arrangement of the loop relative to the rest of the protein. In case of pdb1pft (loop 2, residues 13–17) during five independent 10 ns simulations (with different initial velocities the deviation from the native loop structure remained similar to the start structure [Fig. 1(a)]. For pdb5znf (loop1, residues 5–8) during three out of five cMD simulations the Rmsd from the native loop structure decreased from $\sim$4 Å to $\sim$2.5 Å [Fig. 1(d)]. Along with standard cMD simulations BP-REMD simulations with five biasing levels at 300K were carried from the same initial decoy structures (two independent BP-REMD simulations with different start structures such that the total computational demand is identical to the cMD control simulations). For both pdb1pft and pdb5znf the reference replica (without biasing) in the BP-REMD simulations sampled loop conformations significantly closer to experiment than observed during cMD simulations [Fig. 1(b,e)]. The rapidly oscillating Rmsd in the reference replica (during the whole simula-

Figure 1
Root-mean-square deviation (Rmsd of heavy-atoms) of sampled loop conformations of protein pdb1pft in explicit solvent from the native structure versus simulation time for cMD (A) and BP-REMD (B) simulations. For cMD the Rmsd during five independent simulations (different line colors) and for the BP-REMD the Rmsd of sampled structures in the reference replica are plotted. (C, left panel) Superposition of the decoy pdb1pft start structure with incorrect loop (blue) onto the experimental structure (red). (C, right panel) Superposition of centroid (structure closest to the average structure of the cluster) of the most populated cluster during the last 5 ns of the BP-REMD simulations (red) onto the native structure (blue). The loop residues are in stick representation and the rest of the protein as cartoon representation. (D–F) same as (A–C) for the pdb5znf structure.
tion) also indicates extensive exchanges between replicas during the entire BP-REMD runs.

Similar to pdb1pft and pdb5anf the BP-REMD simulations sampled loop conformations close to the native structure often with heavy atom Rmds of only ~1 to 1.5 Å (Table I). Moreover, the most populated cluster centroid was also in all cases close to the native loop structure. This indicates that the native loop topology can simply be identified by a structure representing the most populated conformational cluster.

In contrast, starting from the same incorrect decoy loop conformations cMD simulations sampled mainly conformations that remained close to the starting conformations (Table I) with an Rmds of the most populated cluster centroid close to the Rmds of the start structure (cf. column 5 and 7 in Table I).

The sampling quality of BP-REMD and cMD for sampling relevant loop structures was also compared in terms of backbone dihedrals angles for the central residues of loop1 in pdb5znf (see Fig. 2). With standard cMD simulations, the sampled dihedral states remain close to the starting dihedral state presumably because of the presence of energy barriers. The lowering of dihedral angle barriers along the replicas in the BP-REMD allows more transitions that if relevant for the sampling in the unbiased force field can exchange and improve sampling in the unbiased reference simulation (see Fig. 2). For all four loop residues a broader sampling of dihedral states

Figure 2
Comparison of backbone dihedral angle sampling of the four central loop residues of pdb5znf during cMD simulations and in the unbiased reference replica of the BP-REMD simulation (during final 5ns). The upper/lower two plots correspond to the results starting from the first/second loop decoy structure, respectively. Each dot in the Ramachandran plots corresponds to a Φ-Ψ-pair recorded every 2 ps (same number of dots for each case). Note, that in case of the cMD simulations the combined results of all five cMD simulations is shown (to achieve the same point density only every fifth data point is plotted). Start (green circle) and native (red triangle) dihedral angle combination are indicated.
was observed in case of the BP-REMD compared to cMD simulations overlapping with the native dihedral angle state.

**Refinement simulations of decoy protein structures**

MD refinement simulations were carried out on several decoy model protein structures including five alpha helical proteins and two alpha-beta protein. Similar to the loop modeling simulations, the initial structures were taken from the Rosetta protein decoy set. In contrast to the loop prediction simulations no restraints were used during the decoy refinement simulations. The decoy structures already contained partially correct secondary structures and in some cases an overall topology similar to the native structure [Fig. 3(a)]. However, for three cases the initial decoy structures were far from the native structure (Rmsd >5 Å) with partially incorrect topology [pdb1gab, 2ptl, 1bw6; illustrated in Fig. 4(a)]. The backbone Rmsd with respect to the native structure was in all cases between 2.4 and 8.1 Å (corresponding to ~3.4 to 9.6 Å Rmsd of heavy atoms; Table II). Starting from two different decoy structures five independent cMD simulations and BP-REMD simulation with five replicas were carried out at 300K in explicit solvent.

In the case of standard cMD simulations during 10 ns simulation time the cluster centroid of the most populated cluster and the lowest Rmsd (from native) structures showed overall in most cases only a slight or no improvement compared to the decoy start conformations (Table II). In contrast, application of the BP-REMD methodology starting from the same conformations

---

**Figure 3**

(A) Superposition of one protein decoy start structure (red cartoon) on the corresponding experimental structure (blue cartoon). (B) Superposition of cluster centroid (structure closest to the average structure of the cluster, red) from the most populated cluster during the last part (last 5 ns) of the cMD simulations and the experimental structure (blue). (C) same as in (B) for the most populated cluster of conformations sampled in the last 5 ns of BP-REMD simulations (unbiased reference replica).
resulted in basically all cases in finally sampled conformations in much closer agreement with experiment than the cMD simulations (Table II, Fig. 3).

Remarkably, for the three proteins with largest deviation of initial decoys from the native structure (backbone Rmsd 5.1–8.1 Å) [pdb1GAB, 2PTL and 1BW6; Fig. 4(a)] application of the BP-REMD method resulted in finally most populated conformational clusters with a deviation of 3.5 to 4.8 Å from the native conformation. Even in the final stage of the simulations frequent exchanges between the near native compact structures and more expanded structures were observed. For example, in case of the 1GAB protein the reference replica of the REMD simulation frequently sampled a near-native conformation but also less folded structures close to the start conformation in contrast to cMD simulations (see Fig. 5). This indicates also a small free energy difference between the more open state and the folded compact state of the 1GAB protein.

DISCUSSION

Depending on the degree of sequence similarity to a template structure homology modeled structures often require further refinement in particular in loop regions or segments of low target-template similarity. Even within a family of homologous proteins functional variations can arise as a consequence of structural differences which are often found on exposed loop regions. And more over the generation of more realistic models that are closer to the native state than the template structure is still a great challenge. In recent years a variety of loop modeling and structural refinement methods have been developed (see Introduction for an overview). The great majority of these methods employ implicit solvation models and often use rigid anchor conformations which may severely restrict the available conformational space of the loop segment. Preferably, loop structure prediction and refinement should be performed in explicit solvent to represent the aqueous environment of the protein as realistically as possible. MD simulations in explicit solvent have already been used to refine homology modeled proteins but require in the order of ~100ns simulation time to achieve reasonable refinement depending on the deviation of the starting conformation. The application of the BP-REMD method significantly enhanced the conformational sampling of loop structures compared to cMD simulations and transitions...
to near-native loop conformations were observed in basically all loop modeling test cases and accumulated as most populated cluster in the final phase of the simulations. In contrast to many other procedures not only the loop segment was flexible during the simulations but also the anchor regions and the rest of the protein were allowed to adjust within predefined limits. This might be important for realistic loop prediction because inaccuracies in the placement of the anchor regions (e.g., during comparative protein modeling) may interfere with realistic loop structure prediction. Similar to loop structure prediction the BP-REMD method showed also significantly better performance than standard cMD simulation during refinement of complete protein structures in explicit solvent conditions. The much smaller computational demand compared to T-REMD allows application of the methodology to a large range of proteins and modeling under the currently most accurate explicit solvent description to realistically account for protein solvation. The relatively short simulation length of ~10 ns may also allow the methodology to effectively compete in terms of speed with systematic search approaches or other MD-based methods that employ implicit solvent descriptions. An implicit solvent model may not always be sufficiently accurate to realistically describe the relative energies of different conformational states of proteins and loop structures and to correctly account for solvent entropy contributions. It should be kept in mind that not only sampling but also the accuracy of the underlying force field limits the outcome of refinement simulations. However, the improved sampling in BP-REMD simulations compared to cMD simulations could also help to systematically refine force field parameters to further improve the output of comparative modeling and refinement calculations.

ACKNOWLEDGMENTS

This work was performed using the computational resources of the CLAMV (Computer Laboratories for Animation, Modeling and Visualization) at Jacobs University Bremen and supercomputer resources of the EMSL (Environmental Molecular Science Laboratories) at the PNNL (Pacific Northwest National Laboratories).

REFERENCES

1. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 2000;29:291–325.
2. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294:93–96.
3. Vitkup D, Melamud E, Moult J, Sander C. Completeness in structural genomics. Nat Struct Biol 2001;8:539–566.
4. Zhang Y. Progress and challenges in protein structure prediction. Curr Opin Struct Biol 2008;18:342–348.
5. Zhang Y, Skolnick J. The protein structure prediction problem could be solved using the current PDB library. Proc Natl Acad Sci USA 2005;102:1029–1034.
6. Ginalska K. Comparative modeling for protein structure prediction. Curr Opin Struct Biol 2006;16:172–177.
7. Roehl CA, Strauss CE, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. Proteins 2004;55:656–677.
8. Zhang Y, Arakaki AK, Skolnick J. TASSER: an automated method for the prediction of protein tertiary structures in CASP6. Proteins 2005;61(Suppl 7):91–98.
9. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 2005;33:2302–2309.
10. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 2003;374:461–491.
11. Kryshtafovyx A, Venclovas C, Fidelis K, Moult J. Progress over the first decade of CASP experiments. Proteins 2005;61(Suppl 7):225–236.
12. Venclovas C, Zemla A, Fidelis K, Moult J. Assessment of progress over the CASP experiments. Proteins 2003;53(Suppl 1):S85–S95.
13. Tress M, Ezkurdiu I, Grana O, Lopez G, Valencia A. Assessment of predictions submitted for the CASP6 comparative modeling category. Proteins 2005;61(Suppl 7):27–45.
14. Rossi KA, Weigelt CA, Nayeem A, Krystek SR. Loopholes and missing links in protein modeling Protein Sci 2007;16:1999–2012.
15. Valencia A. Protein refinement: a new challenge for CASP in its 10th anniversary. Bioinformatics 2005;21:277.
16. Moult JA. Decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol 2005;15:285–289.
17. Jones TA, Thirup S. Using known substructures in protein model building and crystallography. EMBO J 1986;5:819–822.
18. Martin ACR, Thornton JM. Structural families in loops of homologous proteins: automatic classification, modeling and application to antibodies. J Mol Biol 1996;263:800–815.
19. Oliva B, Bates PA., Querol, E, Avilès, FX, Sternberg, MJE. An automated classification of the structure of protein loops. J Mol Biol 1997;266:814–830.
20. Ruffino SD, Donate LE, Canard LJJH, Blundell TL. Predicting the conformation class of short and medium size loops connecting regular secondary structures: application to comparative modeling. J Mol Biol 1997;267:352–367.
21. Van Vlijmen WWT, Karplus M. PDB-based protein loop prediction: parameters for selection and methods for optimization. J Mol Biol 1997;257:975–1001.
22. Wojcik J, Monrov J-P, Chomilier J. New efficient statistical sequence-dependent structure prediction of short to medium-sized protein loops based on an exhaustive loop classification. J Mol Biol 1999;289:1469–1490.
23. Deane CM, Blundell TL. CODA: a combined algorithm for predicting the structurally variable regions of protein models. Proteins Sci 2001;10:599–612.
24. Burke DF, Deane CM. Improved protein loop prediction from sequence alone. Protein Eng 2001;14:473–478.
25. Brucoleri RE, Karplus M. Conformational sampling using high-temperature molecular dynamics. Biopolymers 1990;29:1847–1862.
26. Hornak V, Simmerling C. Generation of accurate protein loop conformations through low-barrier molecular dynamics Proteins 2003;51:577–590.
27. Tappura K. Influence of rotational energy barriers to the conformational search of protein loops in molecular dynamics and ranking the conformations. Proteins 2001;44:167–179.
28. Hornak V, Simmerling C. Development of softcore potential functions for overcoming steric barriers in molecular dynamics simulations J Mol Graph Model 2004;22:405–413.
29. Carlassi L, Englander SW. The loop problem in proteins: a Monte Carlo simulated annealing approach Biopolymers 1993;33:1271–1286.
30. Brucoleri RE, Ab initio loop modeling and its application to homology modeling. Methods Mol Biol 2000;143:247–264.
31. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. Proteins 2004;55:351–367.
32. Levefelt C, Lundh D. A fold-recognition approach to loop modeling. J Mol Model 2006;12:125–139.
33. Zhu K, Pincus DL, Zhao S, Friesner RA. Long loop prediction using the protein local optimization program. Proteins 2006;65:438–452.
34. Rapp CS, Friesner RA. Prediction of loop geometries using a generalized born model of solvation effects. Proteins 1999;35:173–183.
35. Xiang Z, Soto CS, Honig B. Evaluating conformational free energies for the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA 2002;99:7432–7437.
36. de Bakker PIW, DePristo MA, Burke DF, Blundell TL. Ab initio construction of polypeptide fragments: accuracy of loop decoy discrimination by an all-atom statistical potential ant the AMBER force field with the Generalized Born solvation model. Proteins 2003;51:21–40.
37. John B, Sali A. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res 2003;31:3982–3992.
38. Rai BK, Fiser A. Multiple mapping method: a novel approach to the sequence-to-structure alignment problem in comparative protein structure modeling. Proteins 2006;63:644–661.
39. Martin AC, Cheetham JC, Rees AR. Modeling antibody hypervariable loops: combined algorithm. Proc Natl Acad Sci USA 1989;86:9268–9272.
40. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci USA 2004;101:7594–7599.
41. Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. Science 2005;309:1868–1871.
42. Zhang Y, Arakaki AK, Skolnick JR. TASSER. An automated method for the prediction of protein tertiary structures in CASP6. Proteins 2005;61:91–98.
43. Lee SY, Zhang Y, Skolnick J. TASSE. R-based refinement of NMR structures. Proteins 2006;63:451–456.
44. Misura KMS, Chivian D, Rohl CA, Kim DE, Baker D. Physically realistic homology models built with ROSETTA can be more accurate than their templates. Proc Natl Acad Sci USA 2006;103:5361–5366.
45. Wu D, Jernigan R, Wu Z. Refinement of NMR-determined protein structures with database derived mean-force potentials. Proteins 2007;68:232–242.
46. Han R, Leo-Macias A, Zerbinio D, Bastolla U, Conteras-Moreira B, Ortiz AR. An efficient conformational sampling method for homology modeling. Proteins 2008;71:175–188.
47. Jagielaska L, Wroblewska L, Skolnick J. Protein model refinement using an optimized physics-based all-atom force field. Proc Natl Acad Sci USA 2008;105:8268–8273.
48. Lu H, Skolnick J. Application of statistical potentials to protein structure refinement from low resolution ab initio models. Biopolymers 2003;70:575–584.
49. Cui F, Jernigan R, Wu Z. Refinement of NMR-determined protein structures with database derived distance constraints. J Bioinform Comput Biol 2005;3:1315–1329.
50. Fiser A, Sali A. Modeller. Generation and refinement of homology-based protein structure models. Methods Enzymol 2003;374:461–491.
51. Kidera A, Go N. Refinement of protein dynamic structure: normal mode refinement. Proc Natl Acad Sci USA 1990;87:3718–3722.
52. Hollup SM, Taylor WR, Jonassen I. Structural fragments in protein model refinement. Protein Pept Lett 2008;15:964–971.
53. Jonassen I, Klose D, Taylor WR. Protein model refinement using structural fragment tessellation. Comput Biol Chem 2006;30:360–366.
54. Kairys V, Gilson MK, Fernandes MX. Using protein homology models for structure-based studies: approaches to model refinement. Sci World J 2006;6:1542–1554.
55. Kmiecik S, Gront D, Kolinski A. Towards the high-resolution protein structure prediction. Fast refinement of reduced models with all-atom force field. BMC Struct Biol 2007;7:43.
56. Misura KM, Baker D. Progress and challenges in high-resolution refinement of protein structure models. Proteins 2005;59:15–29.
57. Riemann RN, Zacharias M. Refinement of protein cores and protein-peptide interfaces using a potential scaling approach. Protein Eng Des Sel 2003;18:465–476.
65. Stumpff-Kane AW, Maksimiak K, Lee MS, Feig M. Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. Proteins 2008;70:1345–1356.
66. Subramaniam S, Tcheng DK, Fenton JM. A knowledge-based method for protein structure refinement and prediction. Proc Int Conf Intell Syst Mol Biol 1996:4:218–229.
67. Topf M, Baker ML, Marti-Renom MA, Chiu W, Sali A. Refinement of protein structures by iterative comparative modeling and Cryo EM density fitting. J Mol Biol 2006;357:1655–1668.
68. Wroblewska L, Jagielska A, Skolnick J. Development of a physics-based force field for the scoring and refinement of protein models. Biophys J 2008;94:3227–3240.
69. Zhu J, Xie L, Honig B. Structural refinement of protein segments containing secondary structure elements: local sampling, knowledge-based potentials, and clustering. Proteins 2006;65:463–479.
70. Fan H, Mark AE. Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Sci 2004;13:211–220.
71. Fan H, Mark AE. Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy; a molecular dynamics simulation study. Proteins 2003;53:111–120.
72. Fan H, Mark AE. Mimicking the action of folding chaperones in molecular dynamics simulations: application to the refinement of homology-based protein structures. Protein Sci 2004;13:992–999.
73. Fan H, Mark AE. Mimicking the action of Gro EL in molecular dynamics simulations: application to the refinement of protein structures. Protein Sci 2006;15:441–448.
74. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins 2003;50:496–506.
75. Chopra G, Summa CM, Levitt M. Solvent dramatically affects protein structure refinement. Proc Natl Acad Sci USA 2008;105:20239–20244.
76. Swendsen RH, Wang JS. Replica Monte Carlo simulations of spin glasses. Phys Rev Lett 1986;57:2607–2609.
77. Okamoto Y. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. Protein Sci 2004:13:992–999.
78. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–151.
79. Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers 2001;60:96–123.
80. Chen J, Brooks CL, IIIrd. Can molecular dynamics simulations provide high-resolution refinement of protein structure? Proteins 2007;67:922–930.
81. Cheng X, Cui G, Hornak V, Simmerling C. Modified replica exchange simulation methods for local structure refinement. J Phys Chem B 2005;109:8220–8230.
82. Zhu J, Fan H, Periole X, Honig B, Mark AE. Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. Proteins 2008;72:1171–1188.
83. Rathore N, Chopra M, de Pablo JJ. Optimal allocation of replicas in parallel tempering simulations. J Chem Phys 2005;122:24111–24118.
84. Kannan S, Zacharias M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. Proteins 2007;66:697–706.
85. Frickenhaus S, Kannan S, Zacharias M. Efficient evaluation of sampling quality of molecular dynamics simulations by clustering of conformational dynamics models. J Comput Chem 2009;30:479–492.
86. Kannan S, Zacharias M. Folding simulations of Trp-cage mini protein in explicit solvent using biasing potential-exchange replica molecular dynamics simulations. Proteins 2009;76:448–460.
87. Tsai J, Bonneau R, Morozov AV, Kuhlman B, Rohl CA, Baker D. An improved protein decoy set for testing energy functions for protein structure prediction. Proteins 2003;53:76–87.
88. Case D, Pearlman DA, Caldwell JW, Cheatham TE, III, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Tsai V, Radmer RJ, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner PK, Kollman PA. Amber 9. University of California, 2003, San Francisco.
89. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
90. Duan Y, Wu A, Chowdhury CS, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003;24:1999–2012.
91. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 1992;13:952–962.
92. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 1992;13:952–962.
93. Okamoto Y. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. J Mol Graph Model 2004;22:425–439.