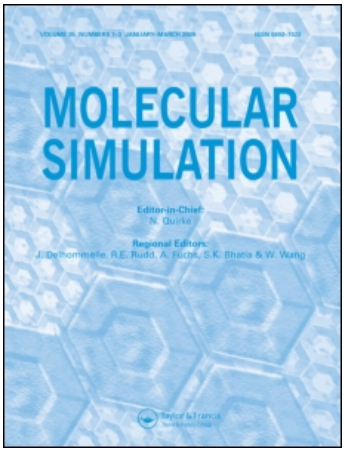

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

A novel protocol of energy optimisation for predicted protein structures built by homology modelling

Tao Xu[a]; Lujia Zhang[a]; Xuedong Wang[a]; Dongzhi Wei[a]
[a] State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai, P.R. China

Online publication date: 24 November 2010

To cite this Article Xu, Tao , Zhang, Lujia , Wang, Xuedong and Wei, Dongzhi(2010) 'A novel protocol of energy optimisation for predicted protein structures built by homology modelling', Molecular Simulation, 36: 13, 1104 — 1109
To link to this Article: DOI: 10.1080/08927022.2010.513771
URL: http://dx.doi.org/10.1080/08927022.2010.513771

Molecular Simulation
Vol. 36, No. 13, November 2010, 1104–1109

Taylor & Francis
Taylor & Francis Group

A novel protocol of energy optimisation for predicted protein structures built by homology modelling

Tao Xu, Lujia Zhang, Xuedong Wang and Dongzhi Wei*

State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, P.R. China

(*Received 15 March 2010; final version received 1 August 2010*)

Homology modelling was applied to predict the three-dimensional (3D) structures of six sets of lipase proteins. Sequence identities between the target and template were 34.6, 44.9, 57.4, 69.9, 79.0 and 86.2%, respectively. Then, eight different protocols including three optimising factors [periodically bounded cell (PBC) water, molecular dynamics (MD) simulation, 'grade-unpacking' strategy or 'combinatorial' strategy] were used to refine the initial model of each system. By comparing the energy-optimised models with the true 3D structure of the target protein in terms of all backbone atoms' root mean square deviation, we determined a novel but all-purpose protocol for model refinement. The protocol refined a homology model by adopting the 'grade-unpacking' strategy for energy minimisation while the model was solvated in PBC water. Furthermore, by comparing the influence of each single optimising factor on the accuracy of the refined structure, we found that introducing the MD simulation into the model refinement method would decrease the accuracy of the final protein structure while methods with either PBC water or the 'grade-unpacking' strategy would increase the accuracy of the final model.

Keywords: homologue modelling; molecular dynamics simulation; model refinement; grade-unpacking

1. Introduction

Knowledge of three-dimensional (3D) protein structures is crucial to answering many biological questions; however, the rapidly growing number of sequenced genes and genomes is heavily outpacing the number of experimentally determined structures [1]. Despite the significant progress in X-ray crystallography and high-field NMR spectroscopy for solving protein structures [2], protein modelling is now becoming the technique of choice for routine structure 'determinations' to reduce the rapidly widening gap between the number of known sequences and the number of experimentally determined 3D structures. Nowadays, more and more researches are being carried out on the basis of a predicted structure [3–5], suggesting the necessity as well as the popularity of protein structure modelling.

With the promotion of the critical assessment of structure prediction (CASP) experiment, significant progress has been made in the field of protein structure prediction [6–8]. Moreover, homology or template-based modelling has been the most successful method for protein structure prediction in CASP experiments [9,10]. Homology protein structure modelling builds a 3D model for a protein of unknown structure (the target) based on one or more related proteins of known structure (the templates). In practice, homology modelling is a multi-step process that can be summarised in seven steps [11]:

(1) Template recognition and initial alignment
(2) Alignment correction
(3) Backbone generation
(4) Loop modelling
(5) Side-chain modelling
(6) Model optimisation
(7) Model validation

It is generally believed that the key element of homology modelling lies in the ability to detect the structural similarities based on amino acid sequence, i.e. sequence alignment. Thus, most theoretical studies are primarily focused on how to improve the accuracy of the alignment [12,13], especially for models based on less than 30% sequence identity to the templates. Compared with the successfulness in sequence alignment, little work has been done on model refinement.

Model refinement is used to eliminate the structural errors introduced in the model building step by a few hundred steps of energy minimisation. Research on this field has been started in recent years [14,15], and it always revolves around the development of a more precise energy function for energy minimisation [16,17]. However, research on whether an all-purpose method for model refinement exists or not is still absent. Based on the study of Lip8, Zhang et al. [18] proposed a general refinement method for homology models. However, there are two main drawbacks in that research:

*Corresponding author. Email: dzhwei@ecust.edu.cn

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927022.2010.513771
http://www.informaworld.com

(1) Only traditional methods, i.e. Procheck [19], Errat [20] and Verify3D [21], were used to judge the accuracy of the protein structure during the model validation steps. This could introduce extra error in the validation process, because all these three methods themselves introduce inaccuracy to some extent;
(2) Sequence identity between Lip8 and its template was low (30%), the method for model refinement based on such single target may not be suitable for other targets, such as cases in high- or medium-resolution models.

In this paper, our goal is to establish an all-purpose protocol for theoretical model optimisation based on more rigid implementations. For this purpose, a protein of known crystal structure (CS) was treated as one without structure and then used as the target for modelling. After building the homology model, we used eight different protocols to refine the initial model (IM). Finally, by comparing the refined models with the true 3D structure of the target protein in terms of all backbone atoms' root mean square deviation (RMSD), we could determine a best protocol which deviated the refined model least from its CS. Such blind test of model refinement methods was similar to the principles of CASP [22]. The study in this paper differs from the previous one in three aspects:

(1) Instead of using the model validation programs for judging the accuracy of the final model (FM), RMSD of all backbone atoms between the protein model and its CS was used. This ideally ruled out all errors in the model validation steps;
(2) To reduce the possibility that the best refinement protocol found in this paper may perform badly in other cases, six sets of protein spanning from low- to high-resolution models were used (sequence identities between the target and template are 34.6, 44.9, 57.4, 69.9, 79.0 and 86.2%, respectively);
(3) Influence of the optimising factor used in this study on the model accuracy was described extensively.

2. Materials and method

A flow chart of the experiment performed in this study is shown in Figure 1. For each target sequence, MODELLER [23], one of the most popular packages in the past few CASPs [24], was used to build the IM. Then, eight different protocols including three refinement factors ['grade-unpacking' strategy or 'combinatorial' strategy periodically bounded cell (PBC) water, molecular dynamics (MD) simulation] were used to optimise the same IM. Finally, all the resulting theoretical models were aligned to their native CS in terms of the RMSD of all backbone atoms [25]. The FMs that deviated least from its CS suggested that the protocol used in the refinement of this model

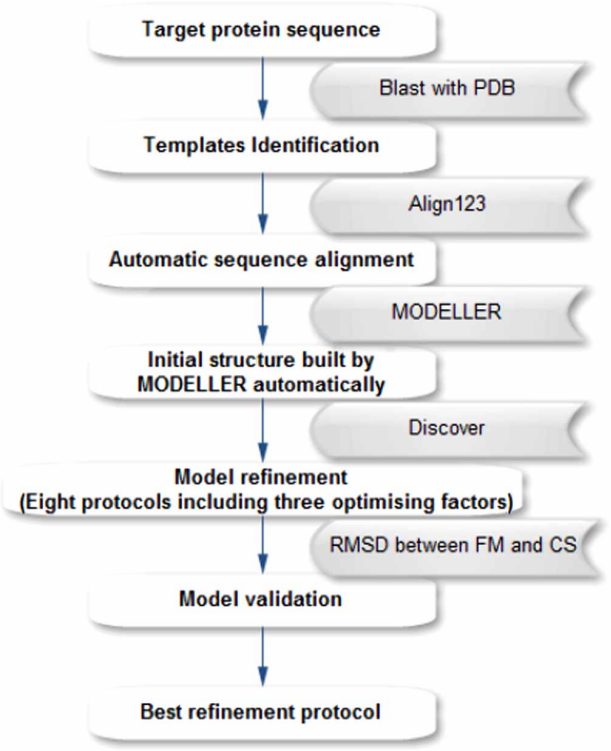

Figure 1. Flow chart of the model building and its subsequent refinement process in this study.

outperformed others. After the analysis of all six sets of target protein, a best but all-purpose model refinement protocol was determined.

Target protein was named according to its corresponding PDB ID in RCSB [26]. Symbol '-A' or '-B' denotes the A or B chain of the proteins. Besides, target 1LGY-A can actually find the template with sequence identity higher than 30%. But to meet the need of the study, 1LGY-A_30 and 1LGY-A_50 were treated as different targets despite their same amino acid sequence. '_30' denoted that the sequence identity between the target and its template was around 30%, while '_50' suggested that the sequence identity was about 50%.

Figure 1 is described in detail in the following sections.

2.1 Theoretical model building by MODELLER

For each sequence, multiple templates used for modelling were identified by blasting the query sequence to PDB (http://www.rcsb.org/pdb/home) [26]. Then, multiple sequence alignments were carried out automatically by Align123 to identify the possible structurally conserved regions. It was ensured that no manual corrections were made to the alignment results to avoid human bias. PAM120 was used as the protein weight matrix. Other parameters for alignment are kept default (for the alignment result, see Figures 1–6 of the Supplementary Material, available online). Then, MODELLER was used

to build five different models, each of which generated three different loop regions. The optimisation level was medium and no hydrogen atoms were built into the structures. Since MODELLER automated the modelling process, no other human intervention was made in order to reduce the human bias. Among the resulting 15 theoretical models, the one with the lowest PDF values was used as the IM for the following study.

2.2 Eight different protocols for model refinement

For each IM, hydrogen atoms were added at pH=7. Then, eight different protocols including three optimising factors [namely PBC water, MD simulation, 'grade-unpacking' strategy or 'combinatorial' strategy] were used to refine the IM, resulting in eight FMs. Detailed description of the eight protocols is listed below, and the protocols are named similarly to the names of their corresponding FM with a character 'P' (abbreviation of 'protocol') in the front:

PFM1: PBC water was added to IM. Then, the 'combinatorial' strategy was applied to minimise its energy (2400 iterations). The system was further refined by a 300 ps MD simulation at 300 K. Finally, the 'combinatorial strategy' (2400 iterations) was used to reduce the structural errors introduced during the MD simulation.
PFM2: All the parameters and steps are set the same as for FS1 except that no PBC water was added to IM.
PFM3: PBC water was added to IM. Then, the 'grade-unpacking' strategy was applied to minimise its energy (300×8 iterations).
PFM4: 'Grade-unpacking' strategy was applied to minimise the energy of IM (300×8 iterations).
PFM5: PBC water was added to IM. Then, the 'grade-unpacking' strategy was applied to minimise its energy (300×8 iterations). The system was further refined by a 300 ps-long MD simulation at 300 K. Finally, the 'grade-unpacking' strategy was applied to minimise its energy (300×8 iterations) to reduce the structural errors introduced during the MD simulation.
PFM6: All the parameters and steps are set the same as for FS5 except that no PBC water was added to IM.
PFM7: PBC water was added to IM. Then, the 'combinatorial strategy' (300×8 iterations) was applied to minimise its energy.
PFM8: 'Combinatorial' strategy was applied to minimise the energy of IM (300×8 iterations).

Water cells of different targets are presented in Table 1 of the Supplementary Material, available online. The 'grade-unpacking' strategy involves the following steps: first, heavy atoms were fixed, and then the steepest descent (SD) and conjugate gradient (CG) methods were used sequentially to minimise the energy; second, the backbone atoms were fixed, and the SD and CG methods were sequentially used for energy minimisation; third, C-α atoms were fixed, and the SD and CG methods were used sequentially for energy minimisation; finally, the SD and CG methods were used sequentially for energy minimisation without atom fixation. Such a delicate minimisation strategy was believed to reduce the structural errors that might be brought forward by energy minimisation. 'Combinatorial strategy' means that SD and CG were used sequentially for energy minimisation.

All energy minimisations and MD simulations were carried out by Discover module of InsightII. The CVFF force field was used. A non-bonded cut-off distance of 9.5 Å and a distance-dependent dielectric function with a scaling factor of 1 was used in all the calculations. For protocols with MD simulation, the system was first run through an MD warm-up phase to a temperature of 300 K in a series of six steps at 50 K intervals, where the simulated duration of each interval was 1.5 ps. Then, an NTV ensemble (constant number of atoms, temperature and volume), a temperature of 300 K and a time step of 1 fs were used in MD simulations.

2.3 Model validation

For the validation of the IM, RMSDs of all backbone atoms were calculated between the IM and its corresponding CS. A large RMSD value suggested the low accuracy of the IM. For the evaluation of the FM, we first measured its RMSD with CS. Then, the effect of the energy minimisation on the deformation to the IM was calculated by subtracting the RMSD value of the IM from that of the FM. A positive value denoted that the FM was more different from its CS than IM was, suggesting that the accuracy of the IM decreased after the model refinement. A negative value, however, indicated that IM would become more precise after the energy minimisation.

3. Results and discussion

3.1 Initial structure

As shown in Table 1, the RMSD value decreased while the sequence identity increased. This suggested that the IM would be more accurate as the sequence identity increases. And this was in accordance with Baker's findings of the relationship between the sequence identity and the homology model accuracy [27]. However, there was an exception. The IM of 1THG had a larger RMSD value than 1LGY-A_30 whose sequence identity was about 10% less than that of 1THG. It was probable that the sequence length of the target would be the reason. Since the sequence length of 1THG was much longer than that of 1LGY-A_30, automatic modelling of such long sequence without manual corrections may lower the accuracy of theoretical

Table 1. RMSD between the IM and its CS.

Target protein	Highest sequence identity with template	Sequence length	PDB code of template proteins	RMSD of the backbone atoms between the IM and its CS (Å)
1LGY-A_30	34.6% (1TIB)	269 aa	1TIB, 2HL6-A	3.846
1THG	44.9% (1CLE-A)	544 aa	1CLE-A, 1UKC-A, 1F6W-A	5.169
1LGY-A_50	57.4% (4TGL-A)	269 aa	4TGL-A, 1TIB	2.031
1GPL	69.9% (1ETH-A)	449 aa	1ETH-A, 1HPL-B, 1BU8-A, 2OXE-A	1.593
1BU8-A	79.0% (2OXE-A)	449 aa	2OXE-A, 1W52, 1ETH-A, 1LPA-B	1.365
1LPB-B	86.2% (1ETH-A)	449 aa	1ETH-A, 1HPL-B, 1GPL-A	0.924

models. In this sense, the sequence length might play another role in the accuracy of predicted structures.

3.2 Determination of a best protocol for model refinement

Table 1 of the Supplementary Material shows the differences between the FM and its CS. However, if we used these data to judge directly the performance of the eight refinement protocols, we would probably get the wrong answer. This is because the IM for refinement deviated differently from the CS between the six targets. Only by subtracting the RMSD value of the IM from that of the FM would we know whether the model refinement process would deteriorate or improve the accuracy of the IM.

Table 2 describes the subtracted results. The results were unexpected but reasonable. For the low-resolution models, namely 1LGY-A_30 and 1THG, each of the eight protocols for model refinement could improve the accuracy of the FM (both systems showed an average RMSD value of -0.41 and -0.47 Å, respectively). For the medium-resolution models, namely 1LGY-A_50, many protocols (namely PFM1, PFM3, PFM4, PFM5, PFM7, PFM8) increased the veracity of the IM by a small margin. But protocols PFM2 and PFM6 even deteriorated the accuracy of the IM. The most surprising result came from the high-resolution models (1GPL, 1BU8-A, 1LPB-B). Nearly all the refinement protocols brought down the accuracy of the IM. As described in Section 3.1, the IMs of the medium- or high-resolution models were very close to the true structure, and had very few structural errors. Model refinement would probably deform some correct parts of the IM rather than fix its errors. So a more accurate energy function that was used to minimise the energy of the IM was urgently needed [17]. However, the situation for the low-resolution models was different. More errors existed in these models, so energy minimisation could easily correct the errors in the IM and made the IM more precise.

The above analysis demonstrated that most of the protocols in this study were not suitable for refining the medium/high-resolution models, but PFM3 was found to improve the accuracy of all six sets of targets. In this sense, PFM3 could be the only protocol which was suitable for the refinement of models for all target sequences.

3.3 Effect of a single optimising factor on the accuracy of the model

Although Table 2 identifies an all-purpose protocol for model refinement, it was not suitable for describing how the three optimising factors influenced the accuracy of the IM. Therefore, a detailed comparative analysis of different optimising factors was necessary. To solve the problem, we made a subtraction recalculation of the raw data in Table 1 of the Supplementary Material. The RMSD value of the FM refined by the protocol with the right factor that was intended to study was used to subtract the RMSD value of the corresponding FM refined by the protocol whose other two factors were the same, e.g. RMSD of PFM1 minus that of PFM2 was used to study the effect of PBC water on the IM. The average value of all six targets was calculated as well.

Table 3 demonstrates the influence of the MD factor. As shown in the table, most of the values were positive. Since the subtrahend involved protocols with the MD factor (PFM1, PFM2, PFM5 and PFM6) and the minuend

Table 2. RMSD between the FM and the IM.

	FS1 (Å)	FS2 (Å)	FS3 (Å)	FS4 (Å)	FS5 (Å)	FS6 (Å)	FS7 (Å)	FS8 (Å)
1LGY-A_30	−0.45	−0.07	−0.54	−0.55	−0.55	−0.10	−0.49	−0.58
1THG	−0.36	−0.65	−0.52	−0.52	−0.33	−0.46	−0.54	−0.40
1LGY-A_50	−0.06	0.30	−0.22	−0.16	−0.09	0.21	−0.18	−0.20
1GPL	0.28	0.67	−0.06	0.07	0.28	1.10	0.10	0.10
1BU8-A	0.26	0.55	−0.04	0.04	0.23	−0.38	0.02	0.03
1LPB-B	0.96	2.75	−0.09	0.11	0.93	2.00	0.14	0.25

Table 3. Influence of the MD factor on the accuracy of the FM.

	FS1 − FS7 (Å)	FS2 − FS8 (Å)	FS5 − FS3 (Å)	FS6 − FS4 (Å)	Average (Å)
1LGY-A_30	0.04	0.51	−0.01	0.45	0.25
1THG	0.18	−0.25	0.19	0.06	0.05
1LGY-A_50	0.12	0.5	0.13	0.37	0.28
1GPL	0.18	0.57	0.34	1.03	0.53
1BU8-A	0.24	0.52	0.27	−0.42	0.15
1LPB-B	0.82	2.5	1.02	1.89	1.56

Table 4. Influence of the PBC water factor on the accuracy of the FM.

	FS1 − FS2 (Å)	FS3 − FS4 (Å)	FS5 − FS6 (Å)	FS6 − FS8 (Å)	Average (Å)
1LGY-A_30	−0.38	0.01	−0.45	0.09	−0.18
1THG	0.29	0	0.13	−0.14	0.07
1LGY-A_50	−0.36	−0.06	−0.3	0.02	−0.18
1GPL	−0.39	−0.13	−0.82	0	−0.34
1BU8-A	−0.29	−0.08	0.61	−0.01	0.06
1LPB-B	−1.79	−0.2	−1.07	−0.11	−0.79

Table 5. Influence of the 'grade-unpacking' or 'combinatorial' strategy factor on the accuracy of the FM.

	FS3 − FS7 (Å)	FS4 − FS8 (Å)	FS5 − FS1 (Å)	FS6 − FS2 (Å)	Average (Å)
1LGY-A_30	−0.05	0.03	−0.1	−0.03	−0.04
1THG	0.02	−0.12	0.03	0.19	0.03
1LGY-A_50	−0.04	0.04	−0.03	−0.09	−0.03
1GPL	−0.16	−0.03	0	0.43	0.06
1BU8-A	−0.06	0.01	−0.03	−0.93	−0.25
1LPB-B	−0.23	−0.14	−0.03	−0.75	−0.29

were the corresponding ones without the factor (PFM7, PFM8, PFM3 and PFM4), the result in this table suggested that the FMs refined by the protocols with MD would get deteriorated most of the time. In this case, the MD simulation was not suitable for model refinement. This was consistent with the work of Chen and Brooks [28].

Table 4 describes the influence of the PBC water factor. Refinement protocols with the factor (PFM1, PFM3, PFM5 and PFM6) did improve the FM than the corresponding ones without PBC water (PFM2, PFM4, PFM6 and PFM8) for more than half of the targets. For targets whose FM got worse, the magnitude of deterioration was rather limited. So solvating the model in the PBC water was preferable for model refinement [29]. In fact, we believe that adding the solvents for model refinement could simulate the real surroundings of the protein as much as possible, because most proteins were active in a solvated environment.

Table 5 shows the influence of the last factor – 'grade-unpacking' strategy or 'combinatorial' strategy. Still, only four out of the six targets were improved using the protocol with 'grade-unpacking' (PFM3, PFM4, PFM5 and PFM6), but the deterioration of other two targets was low. This suggested that the 'grade-unpacking' strategy was better for model refinement.

4. Conclusions

- An all-purpose protocol for the refinement of a homology model was established. The protocol refined the homology model by adopting the 'grade-unpacking' strategy for energy minimisation while the model was solvated in PBC water.
- The three optimising factors (PBC water, MD simulation, 'grade-unpacking' strategy or 'combinatorial' strategy) used in this study imposed a different influences on the accuracy of the FM. The PBC water factor and the 'grade-unpacking' strategy factor increased the model accuracy while the MD factor decreased the model accuracy.

Acknowledgements

This work was funded by the National Basic Research Program of China (973 Program) 2009CB724703. We are grateful to the anonymous reviewers for their helpful advice.

References

[1] K. Ginalski, *Comparative modeling for protein structure prediction*, Curr. Opin. Struct. Biol. 16 (2006), pp. 172–177.
[2] A. Nayeem, D. Sitkoff, and S. Krystek, *A comparative study of available software for high-accuracy homology modeling: From*

sequence alignments to structural models, Protein Sci. 15 (2006), pp. 808–824.

[3] T. Kalupov, M.B. Bourdet, S. Dadé, H. Serrano, J. Wartelle, N. Guyot, L. Juliano, T. Moreau, A. Belaaouaj, and F. Gauthier, *Structural characterization of mouse neutrophil serine proteases and identification of their substrate specificities: Relevance to mouse models of human inflammatory diseases*, J. Biol. Chem. 284 (2009), pp. 34084–34091.

[4] J.H. Ren, X.Q. Xiong, Y. Sha, and M.C. Yan, *Structure prediction and R115866 binding study of human CYP26A1: Homology modelling, fold recognition, molecular docking and MD simulations*, Mol. Simulat. 34 (2008), pp. 337–346.

[5] J. Thun, M. Schoeffel, and J. Breu, *Crystal structure prediction could have helped the experimentalists with polymorphism in benzamide*, Mol. Simulat. 34 (2008), pp. 1359–1370.

[6] Y. Zhang, A.K. Arakaki, and J. Skolnick, *TASSER: An automated method for the prediction of protein tertiary structures in CASP6*, Proteins 61 (Suppl. 7) (2005), pp. 91–98.

[7] A. Kolinski and J.M. Bujnicki, *Generalized protein structure prediction based on combination of fold-recognition with* de novo *folding and evaluation of models*, Proteins 61 (Suppl. 7) (2005), pp. 84–90.

[8] Y. Zhang, *Protein structure prediction: When is it useful?* Curr. Opion. Struct. Biol. 19 (2009), pp. 145–155.

[9] A. Kryshtafovych, C. Venclovas, K. Fidelis, and J. Moult, *Progress over the first decade of CASP experiments*, Proteins 61 (Suppl. 7) (2005), pp. 225–236.

[10] J. Moult, *A decade of CASP: Progress, bottlenecks and prognosis in protein structure prediction*, Curr. Opin. Struct. Biol. 15 (2005), pp. 285–289.

[11] E. Krieger, S.B. Nabuurs, and G. Vriend, *Homology modelling*, Methods Biochem. Anal. 44 (2003), pp. 509–523.

[12] Z. Xiang, *Advances in homology protein structure modeling*, Curr. Protein. Pept. Sci. 7 (2006), pp. 217–227.

[13] L.R. Forrest, C.L. Tang, and B. Honig, *On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins*, Biophys. J. 91 (2006), pp. 508–517.

[14] M.N. Offman, P.W. Fitzjohn, and P.A. Bates, *Developing a move-set for protein model refinement*, Bioinformatics 22 (2006), pp. 1838–1845.

[15] C.M. Summa and M. Levitt, *Near-native structure refinement using* in vacuo *energy minimization*, Proc. Natl Acad. Sci. USA 104 (2007), pp. 3177–3182.

[16] L. Wroblewska, A. Jagielska, and J. Skolnick, *Development of a physics-based force field for the scoring and refinement of protein models*, Biophys. J. 94 (2008), pp. 3227–3240.

[17] A. Jagielska, L. Wroblewska, and J. Skolnick, *Protein model refinement using an optimized physics-based all-atom force field*, Proc. Natl Acad. Sci. USA 105 (2008), pp. 8268–8273.

[18] L.J. Zhang, T. Xu, P.Q. Yuan, and D.Z. Wei, *Optimizing strategies research on the later stage of protein structure model*, Chem. J. Chin. Univ. 29 (2008), pp. 977–980.

[19] R.A. Laskowski, M.W. MacArthur, and D.S. Moss, *PROCHECK: A program to check the stereochemical quality of protein structures*, J. App1. Cryst. 26 (1993), pp. 283–291.

[20] C. Colovos and T.O. Yeates, *Verification of protein structures: Patterns of nonbonded atomic interactions*, Protein Sci. 2 (1993), pp. 1511–1519.

[21] D. Eisenberg, R. Lüthy, and J.U. Bowie, *VERIFY3D: Assessment of protein models with three-dimensional profiles*, Meth. Enzymol. 277 (1997), pp. 396–404.

[22] J. Moult, K. Fidelis, A. Kryshtafovych, B. Rost, T. Hubbard, and A. Tramontano, *Critical assessment of methods of protein structure prediction (CASP)-round 6*, Proteins 69 (Suppl. 8) (2007), pp. 3–7.

[23] A. Sali and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints*, J. Mol. Biol. 234 (1993), pp. 779–815.

[24] B. Wallner and A. Elofsson, *All are not equal: A benchmark of different homology modeling programs*, Protein Sci. 14 (2005), pp. 1315–1327.

[25] O. Carugo and S.A. Pongor, *A normalized root-mean-square distance for comparing protein three-dimensional structures*, Protein Sci. 10 (2001), pp. 1470–1473.

[26] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The protein data bank*, Nucleic Acids Res. 28 (2000), pp. 235–242.

[27] D. Baker and A. Sali, *Protein structure prediction and structural genomics: Unlocking biology's storehouse*, Science 294 (2001), pp. 93–96.

[28] J. Chen and C.L. Brooks, *Can molecular dynamics simulations provide high-resolution refinement of protein structure?* Proteins 67 (2007), pp. 922–930.

[29] G. Chopra, C. Summa, and M. Levitt, *Solvent dramatically affects protein structure refinement*, Proc. Natl Acad. Sci. USA 105 (2008), pp. 20239–20244.