Tara Watkins
Biochem 218
Final Project
June 4, 2004

Full-Atom Refinement Methods: The Key to High Resolution Protein Structure Prediction?

1. Introduction

The number of novel proteins discovered each year far outweighs the number of structures that can be resolved in the same amount of time, and while the discovery of new proteins is certainly exciting, it is difficult to put this new knowledge to use without further information on the structure of the protein. Theoretical structure prediction methods could resolve this difficulty, if reliable methods can be developed to predict high-resolution protein structures from amino acid sequences.

Currently, three basic methods of structure prediction are employed, depending on the similarity of the target protein to proteins with known structure. Comparative modelling and fold recognition methods depend on structural information from template proteins: the 3D structure of the target protein is based on the structure of these previously resolved template proteins. *Ab initio* methods rely entirely on information from the amino acid sequence to generate the structure. And while topological structure prediction is now quite consistent, particularly in the first two methods, many applications require much more specific structural information to predict enzyme interactions or design novel drugs.

In the case of *ab initio* methods, considerably more effort is required before they reach even the accuracy of the comparative modelling and fold recognition methods, much less experimental accuracy. And such prediction methods may soon become the most important, since they are the only way to resolve the structures of very distant proteins. One could imagine that were life discovered under the icy crust of Jupiter's moon Europa, the structure of those proteins could not be predicted based on evolutionarily similar proteins here on earth.

It is thus necessary to develop refinement methods employing full atom representations, including side chain interactions. Such refinement methods would be employed following topological prediction to further improve the accuracy of the predicted structure. In the case of *ab initio* methods, refinement algorithms would likely bring about vast improvements in their predictive power, since they lack the "clues" provided by template proteins, relying mainly on atomic interactions to predict even the topological structure. This paper focuses on the necessity of such refinement methods. Following a brief background on the three methods of structure prediction, several recent refinement methods will be presented: their successes, failures, potential improvements, and suggestions for future work.

2. Background

As mentioned in the introduction, there are three basic methods for predicting protein structure, depending on the similarity of the target protein to previously resolved proteins. Finding such template proteins is extremely important, since there is currently no method to reliably predict protein structure from sequence alone. Modelling from template proteins is based on the assumption that sequence homology indicates structural similarity, which is, to a certain extent, true. However, these models are truly only approximations (Baker and Sali, 2001) and even small errors can render the models virtually useless if those errors occur in critical regions. Nonetheless, they have proved relatively successful in the past, and will certainly form the backbone for successful structure prediction algorithms in the future. Since *ab initio* models do not have the benefit of template proteins for making their predictions, such methods use only information from the sequence itself, creating a much more complex problem.

2.1 Comparative Modelling

Comparative modelling relies the most on sequence similarity of the three methods. A protein, or proteins, with known structure and high homology to the target protein is found; these proteins will act as the templates. The target protein is then aligned with the template protein(s), and a basic 3D structure of the target protein is generated from the aligned regions (that is, the regions of the target protein which align with the template are assumed to have the same structure as those regions in the template protein). Estimated structures of the unaligned regions are then filled in appropriately. Refinement methods should then be employed to further increase the accuracy of the alignments and find the lowest possible energy.

2.2 Fold Recognition

Fold recognition methods are based on a similar approach, but in this case, no reliable template protein exists; since there appear to be only a finite number of folds possible, it is not surprising that dissimilar sequences may still have the same folds. Fold recognition takes advantage of this fact, searching for remote homologues. Multiple possible 3D structures are then generated based on these remote homologues, which can then be evaluated using an energy function to determine which structure is the most accurate. Some methods use additional structural information to predict the structures of the target protein, but the extent to which this improves the predictions is somewhat questionable, since some very successful algorithms use only sequence based information (Rychlewski et al., 2000). Again, accurate refinement methods would be beneficial, since it is not reasonable to expect perfect conservation of structure over large evolutionary distances.

2.3 *Ab Initio* methods

With no available template structure, modelling must be attempted based solely on sequence information. Since attempting to model protein folding based on all the

atomic interactions would be an extremely complex (not to mention impossibly long) process, *ab initio* methods employ reduction algorithms to simplify the problem. The amino acid sequence is normally resolved into a predicted secondary structure sequence, and various algorithms test the potential conformations of these secondary structures to find the lowest energy positions. While several methods such as the Rosetta algorithm (Simons et al., 1999) and the Touchstone program (Kihara et al., 2001) have produced correct topologies, these methods are far from achieving the required accuracy for complex applications. The reduction of the problem to secondary structures limits the possible accuracy unless a reliable method of reverting back to an atomic structure can be applied following the initial topological prediction, after which accurate refinement methods are required to account for the atomic interactions.

2.4 General Discussion

For Comparative Modelling and Fold Recognition methods, accurate alignment appears to be a key feature; since important conserved areas implies that the structure may also have been conserved, recognition of these key areas is necessary. It follows then, that using multiple sequence alignments (Bates et al., 2001) or even using segmental protein templates (that is, multiple protein templates corresponding to different sections of the target protein) (Venclovas, 2001) should improve accuracy. Nonetheless, in order to gain higher resolution structures with accuracy on the atomic level, further refinement will be necessary following initial structure prediction.

Unfortunately, full-atom refinement methods have not performed well in the past. Their shortcomings had become so well known that in the 2002 CASP competition, all the top modelling groups chose not to employ any refinement methods at all, since previous use of such methods had actually made the predicted structures worse (Schonbrun et al., 2002). Clearly this does not bode well for *ab initio* prediction methods, which lack the accurate initial prediction power of the other two methods. And while Comparative Modelling and Fold Recognition methods may continue to improve in accuracy with better alignment algorithms and increases in the experimentally solved protein structures to act as templates, *ab initio* methods will certainly fall behind without improvement in refinement methods. Luckily for the future of *ab initio* protein structure prediction, although in practice many groups have effectively given up on the currently available refinement algorithms, considerable effort is still being conducted in finding new, more reliable methods.

3. Refinement Methods

In terms of atomic interaction protein folding in general, the ideal solution would of course be to numerically simulate the folding process under reversible conditions. Clearly this is impossible, so more feasible approaches like threading and evaluation of free energy functions as a basis for selection of conformation have been employed (Fain et al, 2002; Feig and Brooks, 2002). Unfortunately, research has indicated that these methods usually result in only a portion of the protein being correctly modelled (Fan and Mark, 2004). Despite past failure, considerable effort (and hope) is still focused toward molecular dynamics simulations, while other efforts have concentrated on variable

constraints. Other approaches involve more specific attempts to alter the protein microenvironment to create more favourable folding conditions or computationally mimic natural folding. Here, we discuss two constraint refinement methods, two molecular dynamics simulations, and one novel microenvironmental approach.

Note: In general, the accuracy of a predicted structure and the success of the refinement methods are measured in terms of the root mean square deviation (RMSD) of the predicted (and/or refined) structure from the native structure. This is not always the best measure of success, particularly where refinement is concerned, since it represents an average deviation of the entire structure. Slight changes in critical regions can impact the accuracy of a prediction greatly and not affect the RMSD significantly. In instances when the RMSD does not accurately represent the achieved refinements, other measures will be discussed, otherwise only changes in the RMSD are included.

3.1 Constraint Methods

3.1.1 Local Constraint Refinement

This constraint method, tested by Lu and Skolnick (2003), takes advantage of the fact that many of the initial structures predicted are already topologically correct and are relatively close to the true structure of the protein. Thus, in the refinement process, it would deleterious to allow local portions of the protein structure to deviate far from the initial configuration; not only would it be a computational waste of time, but the likelihood of finding a lower energy conformation outside an initial range is much less likely. However, while it seems logical to constrain the protein segments to an area around their initial configuration, for any refinement to occur they must be given some ability to rotate and reposition. To achieve this, Lu and Skolnick added an additional potential term to the Monte Carlo sampling program which applied a penalty when a local structure deviated 2 Å from the initial configuration. The additional term effectively biased the local structures to the original conformation, while giving them a limited space in which to search for more favourable positions. Following the Monte Carlo simulation, the structures were rebuilt with atomic detail and evaluated for the structure with the lowest atomic potential.

The results indicated that the applied penalty was appropriate, giving the structures enough flexibility to overcome barriers separating the initial structure from potentially better conformations, but constraining the segments already in correct or nearly correct positions. The method was tested on 24 proteins (see Appendix, Table 1), and in 14 cases, the RMSD improved by more than 0.3 Å and in one case was as high as 1.8 Å. In two cases, the RMSD was measurably worse, by 0.4 and 0.6 Å. The remaining 8 cases had no significant change (the refined structures were within 0.3 Å of the original predicted structure).

3.1.2 Reduced Contact Refinement

In the *ab initio* prediction program used by Lu and Skolnick (2003), the potential contains a term based on threading-based contact predictions which are extremely useful

in finding the correct topology for the target protein. The Reduced Contact Refinement method takes advantage of the fact that several of these predicted contacts are clearly incorrect. The process involves reviewing and calculating the distances between each pair of predicted contacts in the proposed structures. For about 25% of the contact pairs, the distances between them were longer than 10 Å, and they were thus considered inconsistent with the structure. These contacts were then removed as constraints and the Monte Carlo sampling program is rerun with the adjusted potential, using only the contacts that could potentially be satisfied during refinement. The simulation tested 20 proteins (see Appendix, Table 2); in 8 cases the structure improved by more than 3 Å, 11 proteins did not see significant change (the refined structures were within 3 Å of the original structure predicted), and the refined structure of one protein was more than 3 Å from the original structure.

3.1.3 Discussion

Both these methods met with relative success; if not able to actually improve the accuracy of the predicted structure, the methods at least, in most cases, didn't make them worse. Although this may seem a bit ridiculous, the results indicate that it is at least worth applying these refinement methods, since the likelihood of worsening the predicted structure is small. On the other hand, approximately 50% of the structures actually improved a noticeable amount.

More interesting is that, since the same research group ran the two tests, many of the same proteins were used in the studies. Comparing the results for each protein showed that the results of the two refinement methods did not entirely overlap. Five proteins were improved by both methods, Local Constraint Refinement improved 6 proteins not improved by Reduced Contact Refinement, and Reduced Contact Refinement improved 3 proteins not improved by Local Constraint Refinement (only 20 proteins were run in the Reduced Contact simulation versus 24 in the Local Constraint simulation, so whether the reduced contact method would have improved the remaining 4 proteins is obviously unknown). Since the two methods were successful with different proteins, it appears that it would be beneficial to combine the two methods and use them both in the same refinement simulations. Additionally, the proteins affected negatively by the refinement methods individually might be effectively "cancelled out", or at least the negative effects lessened, by combining the two methods, since the two proteins made worse by the Local Constraint Refinement were not significantly effected by Reduced Contact Refinement, and the protein noticeably worsened by Reduced Contact Refinement was actually improved using the Local Constraint method.

3.2 Molecular Dynamics (MD) Methods

3.2.1 Classical MD Simulation

In this study, Fan and Mark (2004a) used 15 proteins to assess the efficiency of classical MD simulation techniques for use in refinement of protein structural models. Atomic-based empirical force fields in explicit solvents were used for refinement of the structures and four models were generated for each of the 15 proteins (giving 60 total

models), used in conjunction with 2 controls for each protein from the experimentally determined structures. The simulations were performed using the GROMACS (Groningen Machine for Chemical Simulation) package with the GROMOS96 43al force field for condensed phases; further details of the simulation method can be found in the Fan and Mark paper and will not be discussed here, as this paper is more concerned with the effectiveness of refinement methods in general.

Initial simulations were run on the structures for 5 nsec. A rough approximation of significance was taken as a 10% change in the RMSD. Results from the first simulations showed that 11 of the 60 models had improved structures, 18 of the models had gotten worse, and 31 of the models had experienced no significant change. Further simulations were performed to test the effect of increasing the length of the simulation time on the degree of refinement. Three models were selected, two of the same protein A and one from a different protein B; model 1 of protein A was simulated for 100 nsec, while model 2 of protein A and protein B were simulated for 400 nsec.

Model 1 of protein A had an initial predicted structure that was very close to the native structure (2.6 Å difference), and after 100 nsec the final refined structure was within experimental uncertainly (1.2 Å). The other two models tested had initial predicted structures significantly further from the experimental structure (both 8.7 Å), and were selected to determine if extended length refinement (this time running the simulations for 400 nsec) would be successful on more distant models. Model 2 of protein A reached a relatively stable configuration after about 200 nsec, with an RMSD of 6.8 Å. Protein B experienced large fluctuations over the entire period (see Appendix, Figure 3) of the simulation, ending the simulation at 7.0 Å (by shear coincidence; had the simulation ended 5 nsec before or after 400 nsec the result would have been completely different).

The authors also performed an additional study to test the effects of elevated temperature as a means to avoid becoming trapped in local minima. The previous simulations had been performed (over the majority of the simulation time) at 300 K (a brief relaxation period at 250 K was also included). This study tested selected models at 300 K and 325 K for comparison. The most accurate model from 14 of the 15 proteins was selected (one of the proteins was unstable at 300 K so was not used in this study), with four additional models which were successfully refined after 5 nsec of simulation in the initial study. The five models which were initially closest to the native structure showed lower RMSDs at 325 K than at 300 K; in contrast, the four models with the largest initial deviations from the native structure had lowers RMSDs at 300 K than at 325 K. The authors also commented that further tests performed at even higher temperatures did not lead to better results; rather, in the case of protein structures initially far from the true native structure, increased temperature simply led to unfolding.

3.2.2 Statistical Potential Guided MD

This method (Lu and Skolnick, 2003) uses a combined MD and atomic statistical potential method. 20 MD simulations, each 50 picoseconds long are conducted on the initial proposed structure, producing 20 individual structures. These structures are then evaluated with an atomic statistical pair potential and the lowest energy structure is selected. This new structure is then used as the starting point for a new series of

simulations; this iterative procedure is repeated until the atomic potential converges. Due to problems with structure collapse, additional constraints were imposed on the simulation to retain the secondary structures.

Ten proteins were tested using this method, with mixed success. For all-α proteins, improvements were generally observed; though the majority of these improvements were only modest, two cases saw improvements of more than 0.5 Å. For all-β and mixed α/β proteins the method did not show any significant improvement over the initial structure, and in most cases actually made the structure worse.

3.2.3 Discussion

Clearly the two methods achieved vastly different results. The results of the Statistical Potential Guided MD simulation were in keeping with the results obtained in earlier studies (Lee et al., 2001) showing that simulations of length shorter than 1 nsec are not long enough to show significant refinement; additionally, the Classical MD Simulation discussed here indicates that simulations may take as long as 10 nsec before results are achieved. Perhaps increasing the simulation time of the Statistical Potential Guided MD simulations would increase the number of positive results. Additionally, the method described does not appear to have accounted for hydrogen bonding interactions. Although this further complicates the simulation, it may produce better results by more accurately accounting for the determining factors.

Several interesting results were achieved in the Classical MD Simulation study. The first is that all three long-term simulations show an initial rise in RMSD at the start of the refinement process (see Appendix, Figures 3 and 4), due to large interatomic forces as the force field is applied caused by small errors in packing (Schonbrun et al., 2002). The resulting distortion in the structure appears to have previously corresponded with the end of the simulation period, leading to the conclusion that MD methods cannot accurately refine the protein structure.

Another interesting result is that during the extended-period simulations, extensive spontaneous rearrangements were observed, with substantial regions of the protein unfolding and refolding. This phenomenon was evidenced in the large fluctuations in RMSD experienced by protein B over the course of the simulation. This situation might be a perfect application for the Local Constraint Refinement method discussed above, which prevents gross homology changes. On the other hand, these large fluctuations may have been occurring because the initial predicted structure deviated significantly from the native structure, in which case it would not necessarily be beneficial to constrain the structure into the initial predicted conformation.

Finally, the briefly mentioned results found at temperatures higher than 325 K are intriguing. Fan and Mark mention that no further improvement was made in the structures refined at higher temperatures, and that in cases when the initial deviations in the predicted structure were large, the proteins tended to unfold. This indicates that structures close to the native structure suffered no deleterious effects when refined at higher temperatures, while the distant structures were affected drastically. If this could be proved as a general result, it might provide an easy way to roughly predict how close a predicted structure is to the native structure.

3.3 Microenvironment Alteration: Mimicking the Action of Folding Chaperones for Structural Refinement

While this method might not justifiably be called a "microenvironment alteration" since the method is actually mimicking a more active form of alteration (that is, the invention of other proteins), technically speaking the intervention of the chaperones does significantly alter the active protein environment. As can be deduced from the title of this section, this refinement method is based on mimicking the effect of molecular chaperone proteins. The primary role of these proteins is to assist in the folding of other proteins, since proteins can readily adopt metastable partially folded states and, unhindered, might take a significant amount of time to fold into their native structure (Feldman and Frydman, 2000). Chaperones are effective through several different methods, one of which is by binding to unfolded or partially folded surfaces and disrupting non-native interactions. The misfolded protein can then be released (often into a protected environment where other binding interactions will not occur) providing an opportunity for the protein to refold in the correct manner. If this folding is not successful, the chaperones will simply rebind to the protein, repeating the process until the correct conformation is achieved (Takagi et al., 2003, Martin, 1998).

Fan and Mark (2004b), hypothesized that since initial protein structure predictions are similar to misfolded proteins *in vivo* (mainly possessing packing errors and incorrect arrangements of the secondary structure elements), perhaps mimicking the effect of molecular chaperones will induce more accurate folding. Rather than providing a surface on which the misfolded protein could bind (such as the chaperone creates), the authors opted to modify the solvent environment, cycling the polarity of the solvent. Initially increasing the polarity will promote limited unfolding as increased energy allows the protein to overcome local free energy barriers. Following a short period of increased polarity, the polarity will then be decreased to promote refolding. By restricting the length of time during which unfolding can occur, the method should primarily affect the misfolded regions.

Three proteins were tested using this refinement method. Comparison of the RMSD of the initial models with models following 5 cycles of refinement, Protein A was unchanged at 5.9 Å, Protein B decreased from 5.5 Å to 4.6 Å, and Protein C experienced a vast improvement from 8.7 Å to 5.5 Å. Particularly for Protein A, the RMSD does not accurately represent the success of the refinement. In the original predicted structure, the N terminus projects in precisely the opposite direction from the native structure, and a small, incorrectly located triple-stranded β-sheet is present, compared with the five-stranded β-sheet in the native structure. After refinement, the N terminus points in the correct direction, and a five-stranded β-sheet in its correct position has begun to form (see Appendix, Figure 1). Plotting the intramolecular potential energy as a function of the simulation time shows a clear systematic decrease with time (see Appendix, Figure 2); the energies of the refined structures were significantly lower than the energies of the initial models.

This approach certainly seems to hold some promise, particularly since it is the only method presented here which is designed for application to predicted protein structure with major folding errors; the previous refinement methods either required nearly correct predicted structures or worked best on structures with small deviations

from the native structure. And although this method was only tested on 3 sample proteins, all three refined structures were significantly improved, giving hope that larger sample sizes will show at least somewhat effective results.

The method might be further improved by optimising the solvent environment depending on which protein is being refined; the current method used the same solvent environments for all three proteins, but individual proteins will respond differently in distinct environments. Additionally, the refinement was only achieved over 5 cycles, during which the structures continued to improve. Additional cycles might achieve additional refinement. Furthermore, although the cycle times were selected to be short enough not to promote unfolding of correctly folded regions, the authors did not attempt to optimise the times. It would appear to be beneficial to perform additional tests to determine a cut-off time of sorts, over which the greatest number of misfolded regions will be effected without causing correctly folded areas to unfold.

4. Conclusions

While none of the methods discussed above met with complete success, all the approaches (save perhaps the Statistical Potential Guided MD) achieved a measure of success and have at least paved the way for further research in these areas. Constraints in particular lend themselves to easy use: they can be combined with each other and also potentially with the other refinement methods. Molecular dynamics simulations have received so many negative reports in the recent past that researchers had virtually given up hope that they would ever be useful. And although the Statistical Potential Guided MD simulation was not very successful, the authors' use of a more complex system might still provide a useful technique, especially if used in conjunction with a longer simulation time, as indicated in the Classical Simulation. Meanwhile, mimicking chaperones is an entirely novel approach, and may prove to be the first of many refinement methods based on attempts to more accurately model in vivo protein folding.

It seems that there may be many beneficial combinations to these varied refinement methods. For example, if further work could be done to prove that inaccurately predicted structures unfold at high temperatures, predicted structures could be separated into two classes, inaccurate and accurate predictions. The inaccurate predictions could then be subjected to longer-term chaperone mimicking refinements to attempt to correct the gross errors, then subjected to the temperature test again, iterating until a better structure were achieved. The structures already judged as more accurate could be simulated in a shorter series of chaperone mimicking refinements, and/or any of the other refinement methods. Certainly this is not to say that these methods are necessarily ready for use, since their results were largely unpredictable even when the sample size was large enough to draw any conclusions. Nonetheless, although there is much to be done, both in improving the methods discussed here, and in realising new refinement methods, the future of protein structure refinement seems promising, giving hope for the possibility of high-resolution theoretical models in the near future.

5. References

Baker, D. and Sali, A. 2001. Protein structure prediction and structural genomics. *Science* **294**: 93-96.

Bates, P.A., Kelley, L.A., MacCallum, R.M., and Sternbern, M.J. 2001. Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins* **45**(suppl 5):39-46.

Fain, B., Xia, Y., and Levitt, M. 2002. Design of an optimal Chebyshev-expanded discrimination function for globular proteins. *Protein Sci.* **11**:2010-2021.

Fan, H. and Mark, A.E. 2004a. Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Sci.* **13**:211-220.

Fan, H. and Mark, A.E. 2004b. Mimicking the action of folding chaperones in molecular dynamics simulations: Application to the refinement of homology-based protein structures. *Protein Sci.* **13**:992-999.

Feig, M. and Brooks III, C.L. 2002. Evaluating CASP4 predictions with physical energy functions. *Proteins* **49**:232-245.

Feldman, D.E. and Frydman, J. 2000. Protein folding in vivo: The importance of molecular chaperones. *Curr. Opin. Struct. Bio.* **10**:26-33.

Kihara, D., Lu, H., Kolinski, A., and Skolnick, J. 2001. TOUCHSTONE: an ab initio protein structure prediction method that uses threading-based tertiary restraints. *Proc Natl Acad Sci* **98**(18):10125-30.

Kihara, D., Zhang, Y., Lu, H., Kolinski, A., and Skolnick, J. 2002. Ab initio protein structure prediction on a genomic scale: application to the Mycoplasma genitalium genome. *Proc Natl Acad Sci* **99**(9):5993-8.

Lee, M.R., Tsai, J., Baker, D., and Kollman, P.A. 2001. Molecular dynamics in the endgame of protein structure prediction. *J. Mol. Bio.* **313**:417-430.

Lu, H. and Skolnick, J. 2003. Application of statistical potentials to protein structure refinement from low resolution *ab initio* models. *Biopolymers* **70**:575-584.

Martin, J. 1998. Protein folding assisted by GroEL/GroES chaperonin system. *Biochemistry* (*Mosc*) **63**:374-381.

Rychlewski, L., Jaroszewski L., Li, W., and Godzik, A. 2000. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci.* **9**: 232-241.

Schonbrun, S., Wedemeyer, W.J., and Baker, D. 2002. Protein structure prediction in 2002. *Current Opinion in Structural Biology* **12**:348-354.

Simons, K.T., Bonneau, R., Ruczinski, I., and Baker, D. 1999. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins* **37**(Suppl 3): 171-176.

Takagi, F., Koga, N., and Takada, S. 2003. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. *Proc. Natl. Acad. Sci*. **100**:11367-11372.

Venclovas, C. 2001. Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment. *Proteins* **45**(suppl 5):47-54.

6. Appendix

Table 1: Results from Local Constraint Refinement (Lu and Skolnick, 2003)

Name	Length	Class	Initial RMSD (Å)[a]	Best RMSD (Å)[b]	Final RMSD (Å)[c]
1ixa	39	Small	4.5	3.5	**4.1**
1rpo	61	Small	3.8	2.7	**3.2**
1c5a	66	α	5.7	4.6	5.5
1pou	71	α	3.2	2.8	3.0
1pou	71	α	4.0	3.2	**3.5**
1kjs	74	α	4.8	4.6	5.2
1aoy	78	α	4.5	3.4	**3.8**
2af8	86	α	4.4	3.8	4.5
2lfb	100	α	5.8	5.4	6.0
256bA	106	α	3.4	3.1	3.3
1hlb	138	α	3.2	2.6	**2.7**
1pgx	56	α/β	2.4	2.2	2.5
2ptl	60	α/β	2.6	2.2	**2.2**
2fmr	65	α/β	4.0	3.4	3.8
1life	91	α/β	10.3	9.6	10.3
1shaA	103	α/β	4.1	3.4	**3.7**
1cewl	108	α/β	5.7	5.5	5.9
1gpt	47	β	3.5	2.7	*4.1*
1shg	57	β	4.9	4.3	**4.6**
1vif	60	β	3.7	2.8	**3.3**
1csp	64	β	5.0	3.4	**4.4**
1sro	66	β	4.6	4.0	**4.3**
1sro	66	β	6.6	4.4	**4.8**
1ah9	71	β	5.4	4.8	**4.9**
1tit	89	β	2.6	1.9	2.4
1ksr	100	β	5.6	4.4	**5.1**
ave			4.6	3.8	4.3

[a] Initial RMSD is the RMSD of the starting structure that was generated with TOUCHSTONE program.
[b] Best RMSD is the best structure generated during the sampling in the refinement. Of course, without the knowledge of native structures, we cannot always pick out the best structures.
[c] Final RMSD is the RMSD of the output of the refinement procedure. These are the lowest energy structures selected by the pairwise atomic statistical potential. The cases where the RMSD improves by more than 0.3 Å are shown in bold; those where the RMSD deteriorates by more than 0.3 Å are shown in italic.

Table 2: Results from Reduced Contact Refinement (Lu and Skolnick, 2003)

Name	Length	Class	Initial RMSD (Å)[a]	Best RMSD (Å)[b]	Final RMSD (Å)[c]
1ixa	39	Small	4.5	3.8	4.3
1rpo	61	Small	3.8	2.9	**3.3**
1c5a	66	α	5.7	4.7	**5.2**
1kjs	74	α	4.8	4.6	4.9
1aoy	78	α	4.5	3.3	**4.1**
2af8	86	α	4.4	4.0	4.4
2lfb	100	α	5.8	5.2	**5.4**
256bA	106	α	3.4	3.0	3.5
1hlb	138	α	3.1	2.9	3.2
1pgx	56	α/β	2.4	2.3	2.6
2ptl	60	α/β	2.6	2.3	2.5
1shaA	103	α/β	4.1	3.1	**3.5**
1cewI	108	α/β	5.7	5.2	5.6
1gpt	47	β	3.5	2.9	3.3
1shg	57	β	4.9	4.5	5.1
1csp	64	β	5.0	3.8	**4.3**
1sro	66	β	6.7	4.6	**5.5**
1ah9	71	β	5.4	5.1	*5.9*
1tit	89	β	2.8	2.3	**2.4**
1ksr	100	β	5.6	5.2	5.5
ave			4.4	3.8	4.2

[a] Initial RMSD is the RMSD of the starting structure that was generated with TOUCHSTONE program.
[b] Best RMSD is the best structures generated during the sampling in the refinement. Of course without the knowledge of native structures, we cannot always pick out these best structures.
[c] Final RMSD is the RMSD of the output of the refinement procedure. These are the lowest energy structures selected by the pairwise atomic statistical potential. The cases where the RMSD improves by more than 0.3 Å are shown in bold; those where RMSD deteriorates by more than 0.3 Å are shown in italic.

Figure 1: Protein A: Initial predicted model (left), refined structure after 5 cycles of chaperone simulation (center), and experimentally determined structure (right).

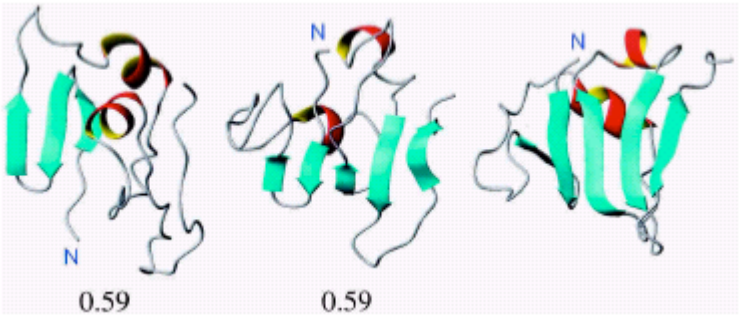

Figure 2: Time evolution of the protein energy and RMSD for Protein A (left), Protein B (center) and Protein C (right).

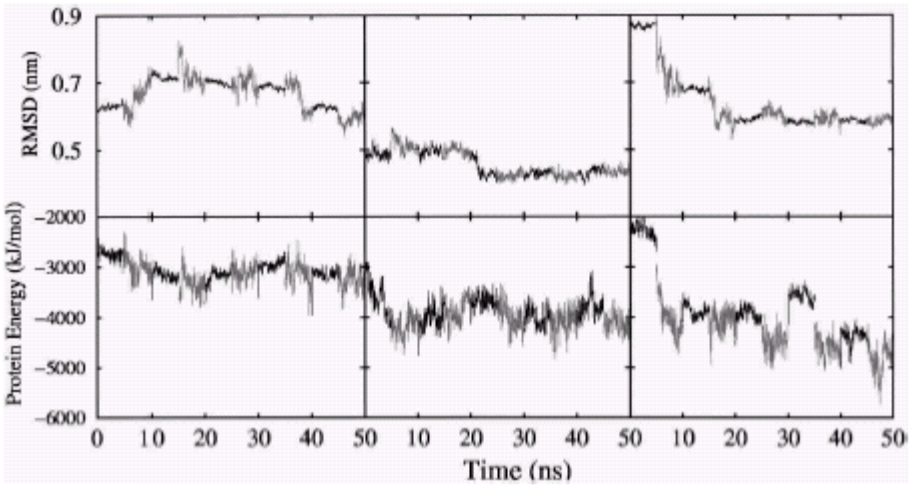

Figure 3: RMSD over the simulation period for Model 2 of Protein A (black line) and Protein B (grey line)

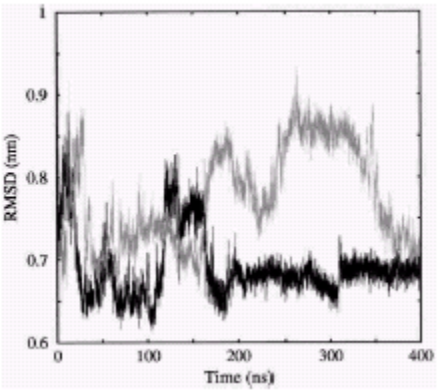

Figure 4: RMSD over the simulation period for Model 1 of Protein A