##
#
MSKCC Document Delivery Services
#
Monday, November 14, 2005
#
###

Request ID: DDS35866
User: Gangi-Dino, Rita
Location: MSK
Requested on: 11/12/2005
Needed by: 11/17/2005
Journal Title: Proteins
ISSN: 0887-3585
Article Author(s): Chinea G
Article Title: The use of position-specific rotamers in model building by homology.
Year: 1995 Nov
Volume: 23
Issue: 3
Pages: 415-21
PMID: 8710834

PROTEINS: Structure, Function, and Genetics 23:415–421 (1995)

The Use of Position-Specific Rotamers in Model Building by Homology

G. Chinea,[1] G. Padron,[1] R.W.W. Hooft,[2] C. Sander,[2] and G. Vriend[2]
[1]*Center for Genetic Engineering and Biotechnology, Havana, Cuba and* [2]*EMBL, Heidelberg, Germany*

ABSTRACT **In this study we concentrate on replacing side chains as a subtask of model building by homology. Two problems arise. How to determine potential low energy rotamers? And how to avoid the combinatorial explosion that results from the combination of many residues for which multiple good rotamers are predicted? We attempt to solve these problems by choosing position-specific rather than generalized rotamers and by sorting the residues that have to be modelled as a function of their freedom in rotamer space. The practical advantages of our method are the quality of the models for cases of high backbone similarity, the small amount of human intervention needed, and the fact that the method automatically estimates the reliability with which each residue has been modeled. Other methods described in this issue are probably more suitable if large backbone rearrangements or loop insertions and deletions need to be modeled.**

© 1995 Wiley-Liss, Inc.

Key words: **model building by homology, position specific rotamers, model evaluation**

INTRODUCTION

Model building by homology normally follows distinct steps:

- A sequence alignment between the protein to be modeled and one or more proteins of known structure is established. In case of high (>75%) homology the alignment is trivial, but for low (<40%) homology cases it is mainly the quality of the alignment that determines the quality of the model.
- Based on this alignment a backbone is generated. This can be the backbone of the most homologous structure or a hybrid backbone.
- Side chains are placed in the model. The choice of rotamer is simple for strictly conserved residues, as their side chain conformations are usually conserved. Nonconserved residues often can occur in several favorable rotamers, leading to an exponentially rising number of combinations. This combinatorial difficulty arises from the fact that spatially close residues interfere with one another, so that the choice of "best" rotamers for one residue depends on that for neighboring residues. Techniques such as Monte Carlo procedures,[5] tree searching algorithms,[11] etc., address this problem.
- If there are insertions or deletions, loops have to be remodeled, or modeled ab initio. Database searches for loops with similar anchoring points in the structure are often used to build these loops, but energy-based ab initio modeling techniques have also been employed (see elsewhere in this issue).
- The final model needs to be optimized. Energy minimization, sometimes combined with molecular dynamics, is normally used for final structure optimization.
- The quality needs to be estimated. Several techniques have recently been described that can be used to estimate the quality of protein models, but visual inspection may still be required to verify that certain structural aspects of the model do not contradict what is known about the functional aspects of the molecule.

Here we concentrate on the problem of grafting new side chains onto the backbone scaffold of the known structure, without building new loops, and without adjusting the backbone. We introduce two new techniques:

- Usage of position-specific rotamers rather than a standard rotamer library improves the quality of positioning side chains.
- Careful selection of the order in which the residues are added to the model can overcome the combinatorial problem.

The combination of these two techniques leads to a fast and robust side chain modeling method. If the similarity between the backbone of the structure and the backbone of the model is high, models can be produced which are equally good, or better—as judged by an all atom root mean square deviation (RMSd)—as models produced using other techniques.

POSITION-SPECIFIC ROTAMERS

Studies on side chain torsion angles[1–4] have greatly enhanced our understanding of protein

---

Received April 3, 1995; revision accepted July 3, 1995.
Address reprint requests to G. Vriend, EMBL, Meyerhofstrasse 1, Heidelberg, Germany.

© 1995 WILEY-LISS, INC.

MEMORIAL SLOAN-KETTERING CANCER CENTER

structure. Residues tend to cluster their chi-1 torsion angle around 60°, 180°, and 300°. Depending on the local secondary structure the population of these three states can differ enormously. For example, 60° is hardly ever observed for any residue in an α-helix, except for serine. Residues in a β-strand show the strongest preference for 300°, except for valine which prefers 180°. Most of these preferences are easily understood from an analysis of local steric effects.[1,3]

The main problem with rotamer prediction is that the body of solved protein structures is still too small to statistically extract the rules needed to uniquely determine the preferred conformation for each side chain. We are therefore forced to neglect certain aspects of protein structure. In a previous study[6] we analyzed wild-type and mutant protein structures and asked how much the local backbone conformation determines the side chain conformation. Position-specific rotamer distributions were determined while residues distant in the sequence and all nearby side chains were neglected. Surprisingly, more than 80% of all mutant structures could be predicted correctly, and for the majority of the other 20% no prediction was made rather than a wrong prediction.

A rotamer distribution for a certain residue type at a certain position, called a position-specific rotamer distribution, is determined by extracting from a database of nonredundant protein structures[7] all suitable fragments of 5 or 7 residues (7 in helix and strand, 5 in case of irregular local backbone). The database fragment extraction technique has been described before,[6,8] and will here be summarized only briefly. Suitable fragments are those that have a local backbone conformation similar to the one around the evaluated position, and have the same residue type at the central position. Backbone similarities can be determined extremely fast by C_α–C_α distance comparisons.[9] In these analyses, the RMS deviation of the α-carbon positions between the structure and the database fragment was kept below 0.5 Å. Rotamers are rejected if they would lead to severe van der Waals' clashes when placed in the model. A severe van der Waals' clash is defined as the distance between two atom being shorter than the sum of the van der Waals' radii minus 0.5 Å.

In the modeling process, position-specific rotamers are used twice. They are first used to determine the freedom in rotamer space of a residue, and thereafter to decide how to place the side chains.

ROTAMER FREEDOM

A big problem for most modeling methods is that they are based on a cost or energy function that includes contacts between residues distant in the sequence, but close in space. This implies that the whole molecule needs to be built before any selected rotamer can be evaluated. This leads to a classical "chicken and the egg" problem. In order to model a side chain correctly, all other side chains have to be modeled correctly first. So, where to start? Several techniques have been described to overcome this problem. Monte Carlo procedures[5,10] seem the most appropriate for this purpose, but Desmet et al.[11] already indicated that other solutions might exist.

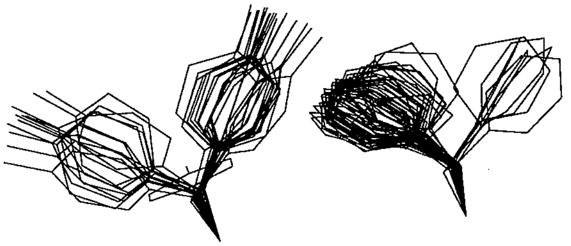

Fig. 1. Example of two overlapping rotamer distributions.

Figure 1 shows two overlapping rotamer distributions. The tyrosine has two almost equally probable rotamers, but the phenylalanine has a clear preference for one rotamer. In the absence of other interactions the optimal solution is, of course, to model the phenylalanine in its most preferred rotamer, and the tyrosine in its second best rotamer. This solution can be found either by complete enumeration of all possibilities, or by simply modeling the phenylalanine side chain first.

Figure 2 shows several examples of position-specific rotamer distributions. Figure 2A shows an example where the position-specific rotamer distribution is extremely narrow. If such a case would occur in a modeling study one could model this residue immediately, and never worry about it again. Figure 2D shows an example of a very wide rotamer distribution. Such a residue should obviously be modeled late in the modeling procedure, because it can much more easily adapt to the space left to it after all other side chains have been placed. In 2A the rotamer is probably entirely determined by the local backbone, whereas in 2D the side chain seems to adapt to its three dimensional environment.

We base our modelling strategy on simple probability principles. The narrower the rotamer distribution, the higher the probability that it corresponds to the rotamer in the X-ray structure. To quantify rotamer distributions we define rotamer freedom as $E = (P/P_{\text{tot}})\ (F/F_{\text{tot}})$ in which P is the population with χ_1 within 75° of the most populated of the three standard χ_1 values (60°, 180°, 300°), P_{tot} is the sum of all rotamers that fall within 75° of any of these three standard χ_1 values, F is the total number of rotamers in this distribution, and F_{tot} is the maximal number of rotamers searched for at any position (we chose $F_{\text{tot}} = 40$ in this study). The term F/F_{tot} reduces the importance of residues for which only very few position specific rotamers are found. E is maximal if 40 (almost) identical rotamers are

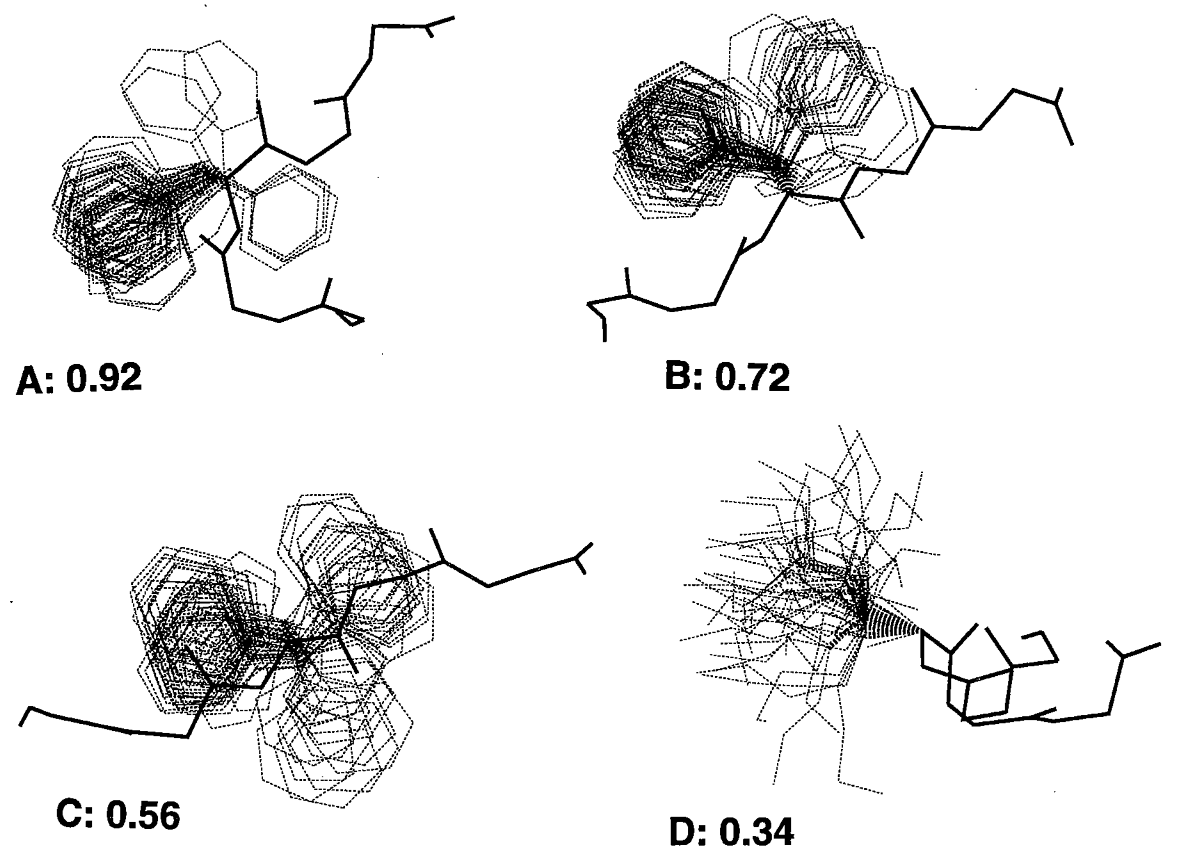

Fig. 2. Examples of rotamer distributions with different rotamer freedom. Rotamers are shown, superposed on the corresponding local backbone structure. The rotamer freedom is indicated (1.0 means no freedom; 0.33 means maximal freedom).

found. The larger the spread and the smaller the number of hits found in the database, the smaller E becomes. In Figure 2 the derived rotamer freedom for four examples is indicated.

PLACING SIDE CHAINS

For most residue positions 40 rotamers could be extracted from the database. Per position all observed rotamers were scored using the function:

$$S=\omega_1\cdot P_{\mathrm{loc}}/P_{\mathrm{tot}}-\omega_2\cdot R_{\mathrm{f}}-\omega_3\cdot f_{\mathrm{p}}+\omega_4\cdot Q+\omega_5\cdot H-\omega_6\cdot B$$

where S = score for one rotamer

P_{loc} = population in the χ_1 bin of this rotamer (see above)

R_{f} = RMS deviation of the fit of the database C_αs on the corresponding C_αs in the backbone of the model

f_{p} = function of the difference between $\phi\varphi$ in the structure and in the database hit. f_{p} is zero if both ϕ and φ are $< 30°$ wrong. f_{p} is a quadratic function of the errors in ϕ and φ which becomes the dominant term when ϕ or φ is $> 90°$ wrong.

Q = quality of packing[12]

H = number of hydrogen bonds formed

B = function of the van der Waals' clashes of the side chain

ω_{1-6} = weights that are set such that all six terms on average contribute equally much to the spread in S.

The rotamer with the highest S is selected, and put in the model. The first three terms are dominated by information from the local backbone, the terms 4–6 are dominated by information about contacts with residues further away in the sequence.

The combination of this scoring scheme with the sorting of residues as function of the rotamer freedom has an additional advantage. Early in this modeling process the residues are built that have a very narrow rotamer distribution, which indicates that the conformation is mainly determined by the local backbone, and the absence of many not yet modeled residues is not a disadvantage. Residues with wider rotamer distributions, which therefore are more influenced by the rest of the molecule, are built later, when more residues are already completed. We used the following modeling protocol:

- First a sparse model is generated. In this sparse model all conserved residues are left untouched, but other residues are mutated into alanine, unless they will become glycine or proline. Glycines, alanines, and prolines are placed immediately, without the

MEMORIAL SLOAN-KETTERING CANCER CENTER

use of position-specific rotamers. No backbone adaptations are made at this stage.

- Determine the rotamer freedom for the residues still to be modeled.
- Mutate as described above in order of decreasing freedom.

After completion several possibilities exist to finetune the model. Energy minimizations are routinely used as the last step in modeling experiments. We tried several optimisation protocols and got the best result using the following steps:

- Patch up the altered residues. One by one all residues that got altered are mutated into alanine, and directly afterward mutated back into the correct residue, using the mutation protocol described above. This way the environment of the residues is more complete than in the previous step. Most times this results in slightly different solutions for the rotamer selection problem, but sometimes rather different rotamer choices are observed, especially for surface residues. This second round uses the same order of residue positions as the first Ala → X mutation round.
- Do a brute force search in torsion space to get rid of severe van der Waals' clashes. In this search all side chain torsion angles of all residues in the model are free to change up to maximally 5°.
- Do a second run in which all torsion angles of all residues that still show clashes are left completely free.
- Do a short (100 steps) energy minimization in vacuum (using GROMOS[13]).

TESTING THE METHOD

The methods were tested in several different ways. The most realistic test was the Asilomar meeting, but in the months before we have been optimizing the method by building models of known structures. For this purpose a set of pairs of proteins was collected that is representative for the present contents of the PDB.[14] This test set contains hardly any insertions or deletions. Roughly equally many all-α, all-β and mixed type protein pairs were selected. We made sure that several molecules had cofactors, and that our test cases were spread evenly over the 35–95% sequence identity range. Our training set consisted of 1cseI 37 2ci2I; 1bpt 45 1aapA; 4pfk 54 1pfk; 4xiaA 67 9ximA; 8dfr 75 2dhf; 1lpf 83 3lad; 2gbp 94 3gbp; 1ifb 33 2hmb; 5hvp 48 1ivp; 1fkf 57 1yat; 4azuA 63 1azcA; 1hsaA 73 1vaaA; 8i1b 79 1i1b; 2lal 90 2ltn; 5pal 45 1omd; 2cro 52 2or1; 2ycc 62 5cyt;1babB 73 1fdhG; 4p2p 83 2bpp; 2cts 94 4csc (pairs of PDB identifiers with chain identifier; the percentage sequence identity is given between the proteins).

As a second test we converted these proteins into poly-alanine, and modeled them back to the original molecule using the protocol described above. In contrast to the previous test case, there are no conserved residues (except a few alanines). Models reconstructed this way had a lower all atom RMSd than models built by homology, partly because all backbone atoms are perfect. However, the side chain atom RMSd often was better too (despite the larger number of side chains modeled). The number of core residues with χ_1 more than 45° wrong was about the same or lower. This clearly indicates the importance of the quality of the backbone coordinates for our modeling strategy. The subset of the database used for the reconstruction never held homologs to the molecule being modeled.

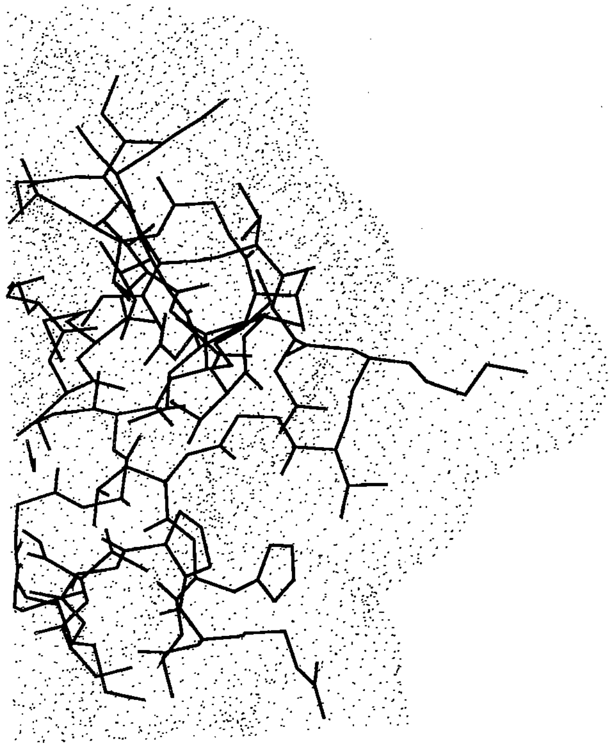

Fig. 3. The hyperexposed lysine-11 in HPR. The water-accessible surface is indicated by dots.

We tried to optimize our modeling protocol by systematically changing the parameters and some of the algorithms. For this purpose a single number is needed that indicates the quality of a model. The RMSd is not a good indicator because of four reasons: surface residues, water molecules, errors in structures, and crystal contacts. These four points will be described below.

Surface residues. It is never certain whether the experimentally determined 3D structure is correct in all details. In many cases multiple side chain conformations exist, but only one is experimentally observed. Figure 3 shows, as an example, residue lysine-11 in HPR. This residue sticks out into the solvent and its side chain does not make intramolecular contacts. Differences between the structure and the model therefore are irrelevant, but this lysine nevertheless has a large impact on the RMSd.

Water molecules. Hydrogen bonds are an important factor in our rotamer scoring function. However, the hydrogen bonding pattern of a molecule can be determined correctly only if all crystal waters are available. We cannot yet model water molecules, and it is thus to be expected that errors are made in residues for which the side chain conformation is partly determined by water molecules.

Crystallographic errors. Protein structures contain errors. For example, about 20% of all asparagines, glutamines, and histidines in the PDB need a 180° rotation in their sidechain to make proper hydrogen bonds. [The HPR structure contained only one such case (glutamine-88) and is in this respect one of the best files in the PDB.] In our experience it is important to completely verify all aspects of the structure from which the model will be built. Also, the real structure should be fully verified before RMS values between structure and model are determined.

Crystal contacts. In lysine-11 in HPR (see above) all side chain atoms make contact with a neighbor molecule in the crystal, which explains its conformation, but makes its prediction rather difficult. A typical modeling test in which we model one known structure from the other would give the following results: RMSd for all core atoms: 0.8 Å; RMSd for residues involved in crystal contacts either in of the two structures: 1.8 Å; RMSd for the other surface residues: 1.5 Å.

Based on these facts we decided not to use the RMSd on all atoms, but to exclude atoms in the side chains of residues that were hyperexposed (i.e., did not make intramolecular contacts) or that were involved in symmetry contacts.

DISCUSSION

What did we learn from Asilomar, what went right, and what went wrong? As you can see in Mike James' contribution in this same volume, we participated with three models. All three were made fully automatic, using standard procedures in the WHAT IF[15] program. Using the appropriate options every user would arrive at these same models. Manual intervention took place only for those three residues in the CRABP model where the program "requested" it.

In the CRABP case the alignment was difficult, we used an incorrect alignment, and the model produced therefore was of low quality. If a side chain is placed in a wrong conformation, it can be rotated to a better position later during the modeling process, but the possibility of modifying the alignment during the modeling has not yet been implemented. Better sequence alignment methods, or the possibility to add knowledge to the alignment procedure, are required to model cases like CRABP more reliably.

NDK was supposed to be a simple task. The percentage sequence identity with the known structure (1NDL[16]) is 77%, so a high backbone similarity could be expected.[17] However, WHAT IF modeled this structure poorly (all atom RMS deviation is 2.40 Å). Figure 4 shows the differences between the model and the real structure. All major errors are located in the C-terminal arm, surface residues, and tryptophan-75.

The C-terminal arm has rather different conformations in 1NDL (the template) and NDK (the model). We still adhere to complete conservation of backbone coordinates, and therefore are not able to deal with such large rearrangements. Not only is the C-terminal arm modeled very poorly, its misplacement also causes several other residues to be modeled incorrectly.

Several residues at the surface are modeled incorrectly. As discussed above, often this is not a serious problem. However, in a few cases WHAT IF incorrectly folds residues back onto the surface of the protein to form a hydrogen bond, or, vice versa, sticks a residue into the solvent that actually should fold back onto the surface to form a hydrogen bond. This clearly indicates the importance of a high quality hydrogen bond force field. At the time of the Asilomar meeting we treated hydrogen bonds binary; a hydrogen bond exists or it does not. We have now derived a hydrogen bond specific force field using the inverse Bolzmann relation on hydrogen bonds observed in small molecules. This improves the modeling, but not as much as hoped for (2 out of 5 relevant cases are now modeled correctly, but one residue previously modeled correctly is now modeled wrong). Clearly, improvements are needed in this direction.

The HPR project was straightforward in our hands: no insertions or deletions and a high backbone similarity. Consequently, we did reasonably well (see Mosimann et al., this issue, Table IV). However, there is still room for improvement, as shown in Figure 5.

The loop around histidine 15 shows differences up to 2.0 Å. We cannot predict such loop displacements. In this case the changes are caused by intensive crystal contacts in the HPR structure that are not present in 2HPR.[16] However, often such loop displacements are an intrinsic property of the structure, and ought to be predictable.

Aspartic acid-10 shows two preferred rotamers. The correct one is the most populated one, and thus this residue gets modeled correctly. There is only one preferred rotamer for valine-12, and this is also the conformation found in the real structure. This rotamer is also the preferred one in all existing standard rotamer libraries, and would also have been modeled correctly without the use of position specific rotamers. Figure 6 shows the rotamers for aspartic acid-66. All position-specific rotamers point in the same direction, and the side chain thus is placed

MEMORIAL SLOAN-KETTERING CANCER CENTER

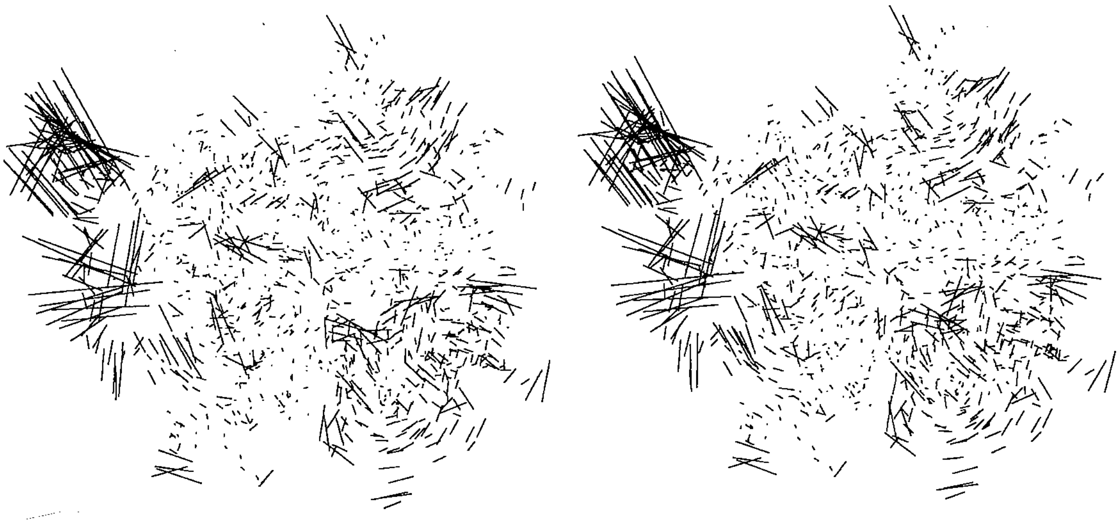

Fig. 4. Errors in the NDK model. The lines connect the positions of corresponding atoms in the X-ray structure and in our model. Model and structure are superposed on all but the 10 C-terminal α-carbons.

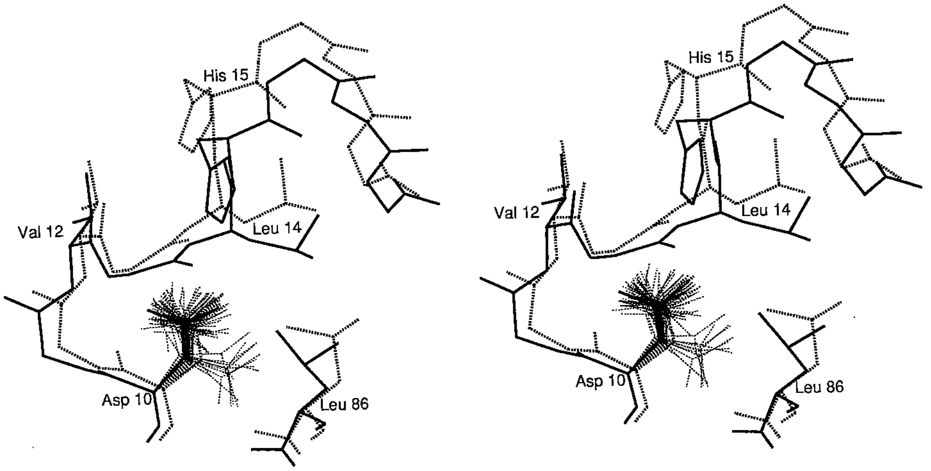

Fig. 5. Representative part of the HPR model and the real structure. Structure and model are superposed using all but the C-terminal α-carbon.

accordingly. In the structure, however, this aspartic acid forms a poor hydrogen bond with the backbone N of alanine-2 and a hydrogen bond with a tightly bound water molecule. Until we can determine the relative importance of hydrogen bonds at the surface of proteins, we will have to live with such errors.

Leucine-86 is "the wrong way around" (χ_1 40° off and χ_2 90° off) in the model (Fig. 5). This leads to a large RMSd for the two C_δ atoms, however, the C_δ atoms in the model and in the structure occupy the same space. Both C_δ atoms make the same contacts in the model and the structure. We therefore think that errors like these are not very important.

The examples mentioned in the previous paragraphs are only a small subset of all residues that can be used to illustrate the advantages and disadvantages of our methods. The modeling protocol that we used was the best we could derive so far. Several alternative protocols were attempted, but they either did not improve the models, or the improvement was too small to warrant the extra CPU time needed.

Conserved residues do not always have the same rotamer. The largest RMSd values for conserved residues are seen for methionines and lysines. We attempted to treat these residues as if they were not conserved, i.e., convert them to alanine, and mutate them back. However, this made the RMSd between the models and the real structures on average a little bit worse.

The rotamer freedom is based on rotamers that could be modeled given the sparse model. After putting several of the side chains with low freedom into this sparse model, the rotamer freedom of the residues still left to be modeled can have changed. Periodic updating of the rotamer freedom (e.g., after

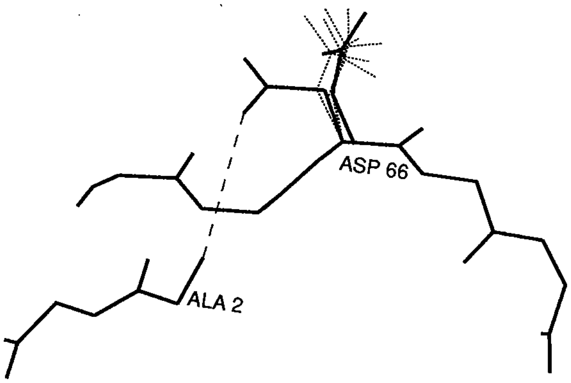

Fig. 6. Rotamer distribution and real position for aspartic acid-66 in HPR. The hydrogen bond between residue 66 and the backbone of residue 2 is indicated. The position-specific rotamers for residue 66 are shown in dashed lines.

putting in 25, 50, and 75% of all side chains to be modeled) did not significantly improve the RMS deviation between the models and the real structures, but costs extra CPU time.

The best results are obtained if the final model is subjected to a very short energy minimization. Several energy minimization protocols were tried. Longer energy minimizations, or energy minimizations combined with short low temperature molecular dynamics runs (all in vacuum), normally increased the RMSd. We have not yet tried to perform molecular dynamics runs in water.

CONCLUSIONS

Where do we stand to date, and where should we go from here? Where do the errors come from, and what needs to be improved most urgently?

There are several main areas for improvement. The techniques described in this article probably still can be improved. Alterations in the backbone should be predicted and insertions should be modeled. We need more and better techniques to detect errors in models. We need to get a better definition of what is right and what is wrong, e.g., by a contact-based model quality indicator rather than by an RMSd. We need to scale the individual contributions to the "energy" of a rotamer better, e.g., when is a good hydrogen bond more important than a good torsion angle? But most importantly, we need a better understanding of the why and how of protein structure so that better force fields for energy minimization can be developed, and even better models can be presented at the next Asilomar protein structure prediction meeting.

ACKNOWLEDGEMENTS

First we want to thank John Moult and his team for organizing the Asilomar meeting. This modeling exercise should, of course, have been seen as a milestone in science and not as a competition. However, the competitive aspect of this comparative modeling made us work hard on this topic.

We thank the homology modeling assessors, Mike James and his team, for doing a heroic job. We also want to thank Anna Tramontano, Reinhard Schneider, Ruben Abagyan, David Thomas, and Liisa Holm for stimulating discussions. Hans Döberling and his team kept the computers running that we used for the uncountable test runs.

We are especially grateful to the many crystallographers who make the coordinates of protein structures available to the scientific community either for prediction purposes, or by deposition in the PDB. WHAT IF is available from "Vriend@EMBL-Heidelberg.DE".

REFERENCES

1. McGregor, M.J., Islam, S.A., Sternberg, M.J.E. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. J. Mol. Biol. 198: 295–310, 1987.
2. Dunbrack, R.L., Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein side-chains. Nature Struct. Biol. 1:334–339, 1994.
3. Morris, A.L., MacArthur, M.W., Hutchinson, E.G., Thorton, J.M. Stereochemical quality of protein structure coordinates. Proteins 12:345–364, 1992.
4. Ponder, J.W., Richards, F.M. Tertiary templates for proteins use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol. 193: 775–791, 1987.
5. Holm, L., Sander, C. Database algorithm for generating protein backbone and side-chain co-ordinates from a C_α trace, application to model building and detection of coordinate errors. J. Mol. Biol. 218:183–194, 1991.
6. Filippis, V. De, Sander, C., Vriend, G. Predicting local structural changes that result from point mutations. Prot. Eng. 7:1203–1208, 1994.
7. Hobohm, U., Scharf, M., Schneider, R., Sander, C. Selection of representative protein data sets. Prot. Sci. 1:409–417, 1992.
8. Levitt, M. Accurate modeling of protein conformation by automatic segment matching. J. Mol. Biol. 226:507–533, 1992.
9. Jones, T.A., Thirup, S. Using known substructures in protein model building and crystallography. EMBO J. 5:819–823, 1986.
10. Abagyan, R., Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J. Mol. Biol. 235:983–1002, 1994.
11. Desmet, J., Mayer, M. de, Hazes, B., Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature (London) 356:539–542, 1992.
12. Vriend, G., Sander, C. Quality control of protein models: Directional atomic contact analysis. J. Appl. Cryst. 26:47–60, 1993.
13. Gunsteren, W. van, Berendsen, H. BIOMOS, Biomolecular Software, Laboratory of Physical Chemistry, Univ. Groningen, The Netherlands, 1987.
14. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1972.
15. Vriend, G. WHAT IF: A molecular modeling and drug design program. J. Mol. Graph 8:52–56, 1990.
16. Worthylake, D., Meadow, N., Roseman, S., Liao, D-I, Herzberg, O., Remington, S. Three-dimensional structure of the *Escherichia coli* phosphocarrier protein III. Proc. Natl. Acad. Sci. U.S.A. 88:10382–10386, 1991.
17. Chothia, C., Lesk, A.M. The relation between the divergence of sequence and structure in proteins. EMBO J. 5:823–826, 1986.