PROTEINS: Structure, Function, and Genetics 15:183–190 (1993)

A Novel Parameterization Scheme for Energy Equations and Its Use to Calculate the Structure of Protein Molecules

Mark E. Snow
University of Michigan Information Technology Division, Scientific Computation Group, Ann Arbor, Michigan 48103-4943

ABSTRACT **A novel scheme for the parameterization of a type of "potential energy" function for protein molecules is introduced. The function is parameterized based on the known conformations of previously determined protein structures and their sequence similarity to a molecule whose conformation is to be calculated. Once parameterized, minima of the potential energy function can be located using a version of simulated annealing which has been previously shown to locate global and near-global minima with the given functional form. As a test problem, the potential was parameterized based on the known structures of the rubredoxins from *Desulfovibrio vulgaris*, *Desulfovibrio desulfuricans*, and *Clostridium pasteurianum*, which vary from 45 to 54 amino acids in length, and the sequence alignments of these molecules with the rubredoxin sequence from *Desulfovibrio gigas*. Since the *Desulfovibrio gigas* rubredeoxin conformation has also been determined, it is possible to check the accuracy of the results. Ten simulated-annealing runs from random starting conformations were performed. Seven of the 10 resultant conformations have an all-C_α rms deviation from the crystallographically determined conformation of less than 1.7 Å. For five of the structures, the rms deviation is less than 0.8 Å. Four of the structures have conformations which are virtually identical to each other except for the position of the carboxy-terminal residue. This is also the conformation which is achieved if the determined crystal structure is minimized with the same potential. The all-C_α rms difference between the crystal and minimized crystal structures is 0.6 Å. It is further observed that the "energies" of the structures according to the potential function exhibit a strong correlation with rms deviation from the native structure. The conformations of the individual model structures and the computational aspects of the modeling procedure are discussed.**

Key words: protein structure, homology modeling, sequence homology, simulated annealing, energy minimization, molecular modeling

INTRODUCTION

With rapid advances in molecular biology and biochemistry, the number of interesting proteins of known sequence but unknown structure is increasing rapidly. While the number of protein molecules whose structure has been determined, either crystallographically or by nuclear magnetic resonance spectroscopy, continues to grow, experimental structure determination continues to be a slow and difficult process. The idea of homology modeling, calculating the structure of a protein based on the known structures of homologous molecules, has been explored extensively,[1–15] but no single approach to the problem has gained wide usage or success. With the growing databases of both determined structures and sequences, the possible applications of useful homology calculations are rapidly expanding.

Recently, it has been shown that data from sequence alignments could be used to produce distance and chirality constraints which could serve as input to distance geometry programs to produce model structures.[15] This approach made it possible to utilize the information available in determined structures and sequence alignments while avoiding the pitfalls and inherent bias present in models built by manual methods. In the present work, data from determined structures and sequence alignments are used in a new way to parameterize a type of energy equation. With the problem posed this way, distances in the model structure are not constrained but are subject to energy *restraints*, with the magnitude of each individual energy term varied automatically in accordance with its importance or reliability as indicated by the structural and alignment

Received January 20, 1992; revision accepted June 20, 1992.
Address reprint requests to Dr. Mark E. Snow, Scientific Computation Group, University of Michigan Information Technology Division, 535 West William Street, Ann Arbor, MI 48103-4943.

data. The implementation of the problem as an energy equation opens the possibility of combining or superimposing the 'modeling' potential on conventional molecular mechanics or other types of energy equations.

The functional form of the energy equation has previously been used in specially designed "protein folding potentials,"[16,17,19,20] and issues related to the minimization of the functional form have been studied extensively.[16–20] It has been shown that, if the determined crystal structure is used in the parameterization, it is possible to parameterize such an equation so as to have a unique global minimum corresponding to the experimentally determined conformation.[20] In test problems, such a global minimum has been located by simulated annealing from a variety of starting conformations.[20]

For the homology modeling problem, of course, the determined structure of the target molecule is not available for use in the parameterization. This paper puts forward a method to parameterize the potential based on sequence alignments and determined structures of homologous molecules. The crystallographically determined conformations of rubredoxins from four different bacterial species[21–24] offer an excellent test system for the method.

METHODS

The Potential

Although the method of parameterization is new, the functional form of the potential has been used in models of protein structures.[16,17,19] The functional form is also familiar as a "10–12" form of the Lennard–Jones potential[30] which has been used to represent van der Waals interactions in some molecular mechanics force fields.[31,32] The form is chosen for efficiency of computation and because a great deal of work has been done concerning its global optimization.[16–20] Amino acid residues are modeled as single points centered on the α-carbon. The form of the potential is

$$E = \sum_i \sum_{j>i} (A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-10})$$

where r_{ij} is the distance between atoms i and j, and A and B are parameters. It is conceptually useful to implement the change of variables

$$r^o = (6/5)^{1/2}(A/B)^{1/2}, \qquad \epsilon = \frac{B^6}{A^5}[(5/6)^6 - (5/6)^5]$$

giving

$$E = \sum_i \sum_{j>i} \epsilon_{ij} \left[5\left(\frac{r_{ij}^o}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^o}{r_{ij}}\right)^{10} \right].$$

Here it can be seen that r^o represents an *optimum* separation for residues i and j, and that ϵ represents the energy corresponding to achieving the optimum separation. If the crystal structure of the molecule whose structure is to be calculated were known, then the potential could be parameterized by simply setting each of the r^os equal to the value observed in the crystal structure. It has been previously shown that such a potential has a unique global minimum corresponding to the crystal structure, and that this minimum can be located from numerous starting points using the annealing algorithm.[20] In the case where the structure to be calculated is not known, we would like to set the r^os to a best guess value, and to choose the ϵs to reflect our confidence in the r^os. The determined structures of homologous proteins provide a basis to do this.

Rubredoxin Sequences and Structures

The amino acid sequence of the target molecule, the rubredoxin from *Desulfovibrio gigas,* and the sequences of the homologous rubredoxins from *Desulfovirio vulgaris, Desulfovibrio desulfuricans,* and *Clostridium pasteurianum* were obtained from the Protein Identification Resource.[25] The coordinates of the crystallographically determined conformations[21–24] were obtained from the Brookhaven Protein Data Bank.[26]

Sequence Alignments

The first step in building a parameterization from the structures of the homologues is to perform a sequence alignment of the target molecule and all the homologues. In this study, alignments were performed using the algorithm of Altschul and Erickson[27] as implemented in the EuGene program package.[28] A Dayhoff cost matrix as used with a gap penalty of 2.5 and an incremental penalty of 0.5.

Parameterization

In addition to performing a sequence alignment, it is necessary to calculate distance matrices for each of the homologues of known structure. With sequence alignments and distance matrices in hand, the guidelines for building the potential are as follows: For each pair of atoms in the target molecule, there may exist (as determined by the sequence alignments) a corresponding atom pair in one or more of the homologues. The interresidue distance, ρ, for each such pair can be found in the distance matrices. The ϵ and r^o values in the potential are set based on the frequency with which homologues have a corresponding residue pair, the observed ρ values and the range of observed ρ values. r^o is set according to

$$r_{ij}^o = \left[\frac{\sum \rho_{i'j'}^{-10}}{\sum \rho_{i'j'}^{-12}} \right]^{1/2}$$

where the sum is over all homologues which exhibit a residue pair which (according to the sequence alignments) corresponds to the pair ij in the target molecule. The indices ij may not be equivalent to

TABLE I. Values for ϵ_{ij} as a Function of r^0_{ij}, the Range in $\rho_{i'j'}$, and n_{ij}*

r^0_{ij} (Å)	ρ-range (Å) max $\rho_{i'j'}$ − min $\rho_{i'j'}$	ϵ_{ij}
< 5.0	< 0.2	$200n_{ij}$
	0.2–0.8	$100n_{ij}$
	0.8–2.5	$50n_{ij}$
	> 2.5	0
5.0–10.0	< 0.2	$100n_{ij}$
	0.2–0.8	$50n_{ij}$
	0.8–2.5	$20n_{ij}$
	> 2.5	0
> 10.0	< 0.2	$50n_{ij}$
	0.2–0.8	$20n_{ij}$
	0.8–2.5	$5n_{ij}$
	> 2.5	0

*n_{ij} represents the number of homologues which have a residue pair $i'j'$ which, according to the sequence alignments, corresponds to the ij-pair in the target molecule. Other variable names are defined as in the text.

$i'j'$, depending on the alignment of the sequences. The corresponding ϵ value is set based on the value of r^0 and the number and range of observed ρ as laid out in Table I. The idea is that residue pairs which exhibit a very tight distribution in observed distance separation will have a strong energy term while those which exhibit a very wide distribution in distance separation will have little or no contribution to the total energy. Further, interactions which appear consistently in all homologues will have a larger energy term than those that do not.

Two additions to the potential are necessary in order to deal with residue pairs in the target molecule for which there are no corresponding pairs in the homologues. It is important that there be some energy term representing the interaction between sequential (nearest neighbor) residues along the chain, even if there is no corresponding residue pair in the homologues. Any consecutive residue pair which cannot be parameterized according to the scheme in Table I is given an ϵ of 200 units and an r_0 of 3.8 Å. This term prevents the chain from breaking. Chain breakage during simulated annealing was observed in some runs when a potential which did not contain this term was used (unpublished results). It is also important for residue pairs that may be widely separated along the chain to have at least a small repulsive potential. Were this not the case, it would be conceivable, for example, for two domains to fold very well individually, but to end up intertwined with one another due to a shortage of interaction terms between them. Nonneighbor residue pairs which cannot be parameterized according to Table I are given a parameterization with A of 50 and B of 0.

This strategy represents a complete parameterization of the energy equation for molecules containing no posttranslational modifications, metal ions, or other prosthetic groups. Rubredoxins, however, are iron–sulfur proteins containing an Fe–S(4) cluster. The issue of the iron atom was handled in the following way: It was observed that the four cysteine residues that are ligands to the Fe atom align correctly for the three determined structures, and further that these cysteine residues all align with cysteines in the *Desulfovibrio gigas* sequence as well (Fig. 1). Thus each of the six relevant Cys–Cys residue pairs not only occur in all homologues, but the interresidue distances are highly conserved as well. It thus seemed likely that the Fe binding site would retain its position whether or not the Fe was explicitly included in the model. The decision was made to omit the Fe atom from the potential, but to double the ϵ values on each of the six Cys–Cys interaction terms.

Minimization

Minima of the potential are located using a previously described simulated annealing algorithm.[20] Starting conformations were generated by choosing dihedral angles at random between −180 and +180°, bond angles at random between +90 and +180°, and fixing C_α–C_α bond lengths at 3.8 Å. Annealing was performed in the space of dihedral and bond angles. The initial step-size range was 30° for both types of angles. Three hundred sweeps through the angles were performed per iteration of the program, and an initial "temperature" of 1200 was used. Both the temperature and the step-size range were subject to update schedules, with the temperature being doubled on acceptance rates less than 0.4, and halved on acceptance rates greater than 0.6. On iterations completed without an energy drop, the step-size range was halved and the temperature was multiplied by 0.4. After the temperature reached a value of 0.001 or after 180 iterations, annealing was terminated and minimization was completed with conjugate gradients minimization.[29] This final minimization was performed in Cartesian-coordinate space rather than angle space, thus allowing bond lengths to achieve their equilibrium values. The derivative-based conjugate-gradients algorithm is much more efficient than simulated annealing once the temperature is low enough that the conformation is within the radius of convergence of a given minimum.

Computational Issues

The programs for calculating distance matrices and for developing a parameterization of the energy equation from sequence alignments and distance matrices were written in fortran 77, and run on a Sun SPARCstation IPC. The output of these programs is a single file containing atom pairs, r^0 and ϵ values for the target problem. This file is read by a

```
      Desulfovibrio gigas:  MDIYVCTVCGYEYDPAKGDPDSGIKPGTKFEDLPDDWACPVCGASKDAFE..KQ
   Desulfovibrio vulgaris:  -KK-------------E----N-V----S-D---A--V------P-SE--..AA
 Clostridium pasteurianum:  -KK-T------I---ED----D-VN---D-K-I----V--L--VG--E--evEE
Desulfovibrio desulfuricans: -QK---N---------EH-.......NVP-DQ-----C-----V---Q-S..PA
```

Fig. 1. Aligned sequences of the four rubredoxins. One-letter amino acid symbols are used. Uppercase letters indicate aligned nonidentical residues, lowercase letters indicate unaligned residues, dashes (- - -) indicate aligned identical residues, and dots (...) indicate gaps.

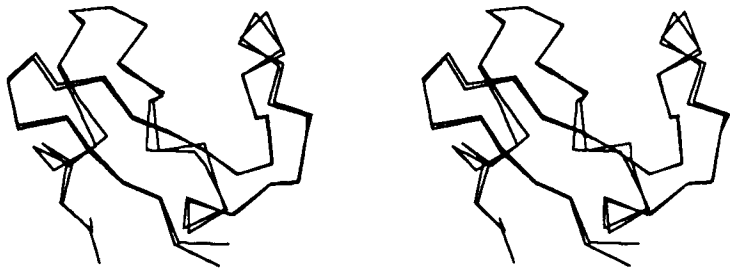

Fig. 2. Stereo pair showing the superimposed conformations of the crystal structure and the minimized crystal structure. The structure moves 0.6 Å (all-C_α rms) during minimization.

TABLE II. Energies and All-C_α rms Deviations of the 10 Model Structures vs. the Experimentally Determined (X-Ray) Structure and the Minimized X-Ray Structure*

Structure	Energy	rms vs. native	rms vs. minimized
0	−144836.76	0.70	0.38
1	−138526.34	1.38	1.34
2	−144836.76	0.56	0.60
3	−144836.76	0.56	0.60
4	−119249.01	3.85	3.73
5	−142185.05	0.72	0.74
6	−101209.90	8.15	8.04
7	−144836.76	0.75	0.50
8	−134241.02	1.68	1.57
9	−111974.77	3.47	3.41

*The minimized X-ray structure differs from the experimental structure by 0.60 Å rms, and has an energy of −144836.76.

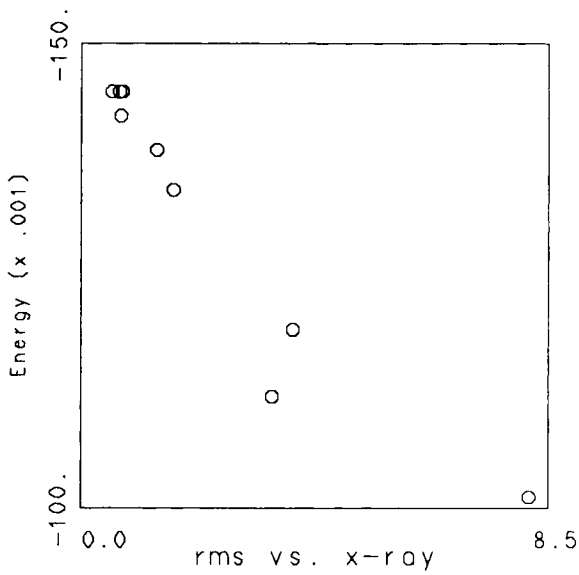

Fig. 3. The energy of the 10 structures is plotted against the all-C_α rms deviation vs. the X-ray structure (in Å). There is an apparent correlation, with the lower energy structures showing greater similarity to the crystallographically determined conformation.

program which generates random conformations and performs the minimizations to produce the model structures. These calculations were performed on a SPARCstation IPC, on an IBM RS/6000 Model 540, and on a single processor of a Cray Y-MP.

RESULTS

The alignment of the rubredoxin sequences from *Desulfovibrio vulgaris, Desulfovibrio desulfuricans, Clostridium pasteurianum,* and *Desulfovibrio gigas* is shown in Figure 1.

The potential was parameterized as described in Methods and 10 simulated annealing runs were performed from random starting conformations. The final energies after minimization ranged from −101209.90 to −144836.76. Four of the 10 annealing runs resulted in conformations with the lowest observed energy of −144836.76. No other energy was observed more than once.

Since the determined crystal structure of *Desulfovibrio gigas* rubredoxin is available, it was possible to use this structure as a starting point for

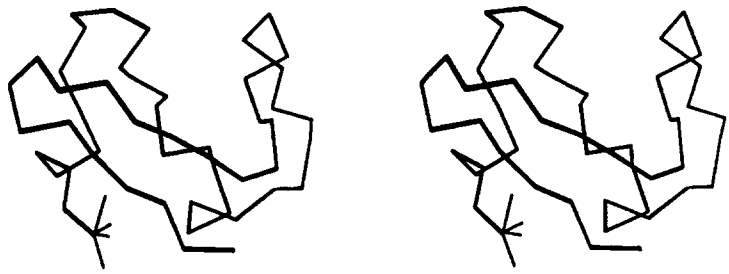

Fig. 4. Stereo pair showing the superimposed conformations of the minimized crystal structure of *Desulfovibrio gigas* rubredoxin and the four structures located by simulated annealing which are isoenergetic with it to eight significant figures. The only notable difference between the structures is the position of the C-terminal residue glutamine 52.

energy minimization to generate a conformation for comparison with the model structures generated by simulated annealing. This minimization was performed using conjugate gradients only. The minimized crystal structure exhibits an energy of −144836.76, equivalent to the lowest energy that was observed in 4 of the 10 model structures. The all-C_α rms deviation between the crystal structure and the minimized crystal structure is only 0.60 Å (Fig. 2).

The energies of the 10 model structures, and their all-C_α rms deviations versus both the crystal structure and the minimized crystal structure are shown in Table II. Seven of the 10 structures have energies within 10% of the lowest observed energy while the other three lie in local minima of considerably higher energy. Of the seven low energy structures, all exhibit an all-C_α rms deviation versus the native conformation of less than 1.7 Å, and five of the seven are within 0.8 Å. There is an apparent correlation between the energy and the rms deviation as shown in Figure 3.

It was surprising to find that while the minimized native structure and four of the model structures have energies which are identical to eight significant figures, these structures are not conformationally identical (Table II). Superimposing these five structures (Fig. 4) reveals that they are identical everywhere except the position of residue 52. An observation of the potential energy equation reveals that the only terms involving residue 52 are a strong interaction with residue 51 which has an r^o of 3.81 Å, a very weak interaction with residue 1 which has an r^o of 10.70 Å, and the general weak repulsive term with the rest of the molecule. The effect of this is that, as long as the distance between residues 51 and 52 remains fixed, residue 52 can rotate quite freely with virtually no effect on the total energy. This can occur because residue 52 is a terminal residue (a rotation at a location other than a terminal residue would involve the movement of multiple residues). Residue 1 does not exhibit this behavior because there are many more strong interactions involving residue 1. Residues 51 and 52 have very few interaction terms because of the break in the sequence alignments between residues 50 and 51 (Fig. 1).

It is informative to look at the conformations of the other model structures (Fig. 5). Structure 5 has an rms deviation vs. the native structure of the same order as the four isoenergetic structures, but an energy which is slightly higher. When aligned with these structures, it exhibits a slight difference in the second strand in addition to the difference at residue 52. Structures 1 and 8 have slightly higher energies and exhibit rms deviations in the 1.3 to 1.7 Å range. In each of these cases, the structural differences vs. the minimized native are localized to one region of the molecule (aside from the difference at residue 52). Structure 1 exhibits a difference in the turn between strands three and four and at the beginning of strand four. Structure 8 exhibits a difference at the end of strand one and the turn into strand two. The remaining three (high energy) structures differ considerably from the minimized native structure. Structure 6 is so different that it cannot even be meaningfully aligned. Structure 9 (Fig. 5d) can be seen to follow the same general fold for strands one through four, but to diverge at the C-terminal end of the molecule. Given the range in the quality of the model structures, the strong correlation between the rms and energy is particularly useful.

Computational Issues

CPU times for an identical run were 147 min on the Cray, 502 on the IBM, and 46 hr on the SUN. Because of the nature of simulated annealing, times vary from run to run. The test run which was performed with an identical seed on all three machines was slightly long, but not atypical (runs on the RS/6000 ranged from 485 to 503 min). These times could be improved considerably by using a less rigorous annealing schedule, but this would undoubtedly result in a higher percentage of structures lying in high-energy local minima.

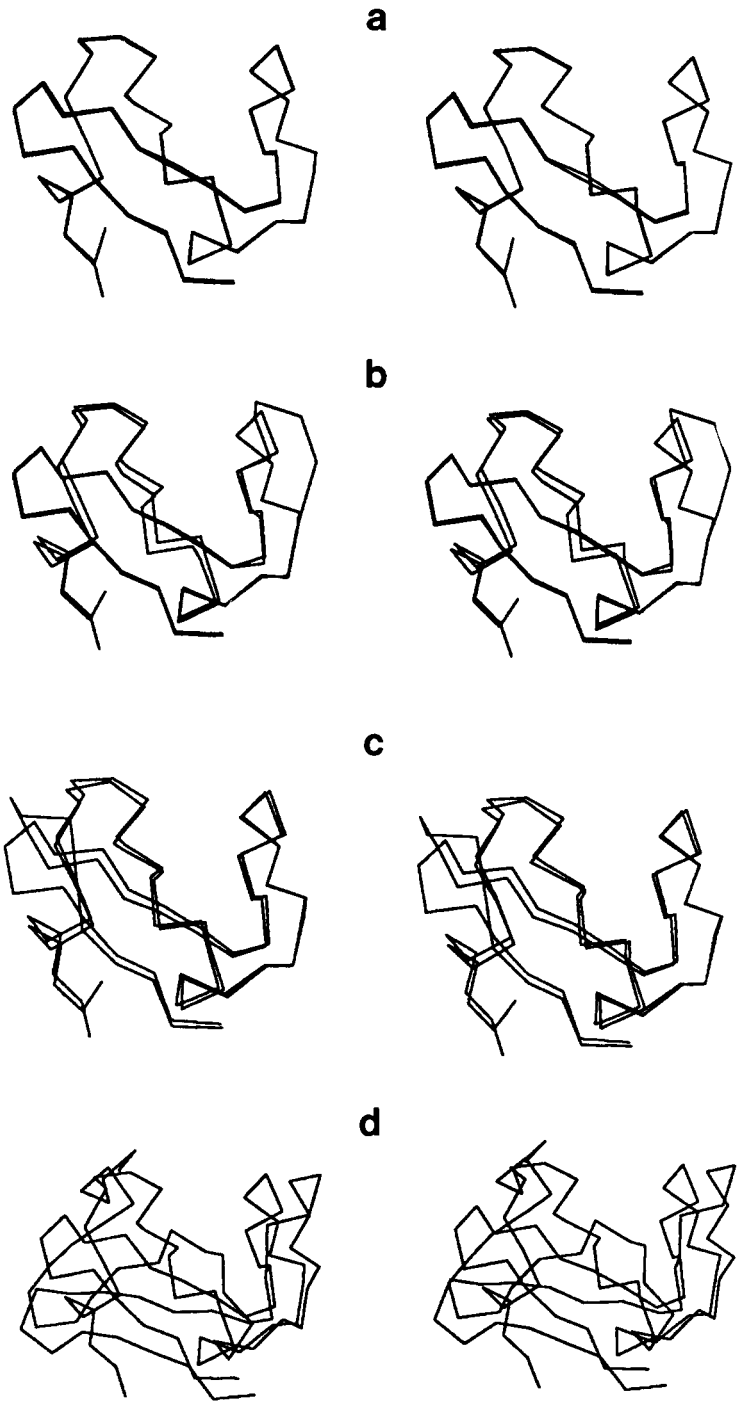

Fig. 5. Stereo pairs showing some of the conformations residing in higher energy minima superimposed on the minimized crystal structure. (**a**) Structure 5 has only slightly higher energy than the minimized X-ray structure. It exhibits a difference at residue 52 and a slight difference in the second strand. (**b**) Structure 1 exhibits a difference in the turn between strands three and four and at the beginning of strand four. (**c**) Structure 8 exhibits a difference at the end of strand one and the turn into strand two. (**d**) Structure 9 is of higher energy. While the N-terminal portion of the molecule follows the same general fold as the minimized crystal structure, the C-terminal portion diverges considerably.

Comparison With Traditional Homology Modeling

It may be informative to compare this procedure with more conventional homology modeling approaches. Toward this end, a model of the *Desulfovibrio gigas* rubredoxin was also built using the protein design software from Polygen Corporation.[33] The Polygen software produced a structure-based sequence alignment which differed from the sequence-based alignment in two respects. The two residue

gaps in the *Desulfovibrion gigas, Desulfovibrio vulgaris,* and *Desulfovibrio desulfuricans* sequences occurred after residue 52 rather than between residues 50 and 51 (see Fig. 1), and the large gap in the *Desulfovibrio desulfuricans* sequence occurred between H19 and D20 rather than between D20 and N21. As might be expected, the structure-based alignment appears to be better than the sequence-based alignment. In the future, it may be desirable to use structure-based alignments with the new homology-based parameterization scheme as well.

Pairwise alignments with the Polygen software indicated that residues 1 through 34 of the model should be based on the *Desulfovibrio vulgaris* structure and residues 35 through 52 on the *Clostridium pasteurianum* structure. This was carried out, and sidechains were placed following the Polygen protocol.[33] The geometry was then regularized to produce the final model. This model differed by 0.95 Å (all-C_α rms) from the determined X-ray structure of the *Desulfovibrio gigas* protein, and by 0.93 Å from the structure which had been minimized using the potential energy equation parameterized from the homologues.

DISCUSSION

This would appear to be an effective method for dealing with the homology modeling problem. Those regions of the target structure which can reasonably be calculated based on the experimentally determined structures of homologous molecules appear to be correctly modeled in an automatic way. Regions which cannot be well determined exhibit multiple minima which correspond to geometrically reasonable conformations consistent with the available data. In the rubredoxin case, the data from the three homologues were sufficient to produce a parameterized potential with an apparent global energy minimum which agrees with the minimized native structure everywhere except the carboxy-terminal residue. Indeed 7 of 10 model structures lie within 1.7 Å rms of both the native and minimized native conformations. Further, these seven model structures could easily be separated from the other three on the basis of energy criteria alone. This is important if the method is to be useful in producing models for molecules of undetermined structure.

While the size of the rubredoxins (45 to 54 amino acids) is smaller than many proteins of interest, both the size and the sequence alignments in this problem are nontrivial. It thus seems quite possible that the method will be useful for larger more complex problems. Work is under way to investigate the applicability of the methods to larger systems and to systems where there is a lower degree of structural and/or sequence homology.

The nature of the energy calculations and the simulated annealing should lend itself to implementation on parallel machine architectures. This possibility is under investigation. A parallel implementation of the program may become important if larger model systems are to be addressed.

Expressing the homology-modeling problem in this form, as the minimization of an energy equation, offers a number of advantages over other model-building strategies. Clearly, the automated generation of structures from random starting conformations should eliminate the sorts of human bias that can occur with manual model-building efforts. The variable ϵs of the energy terms allows information to be weighted in accordance with its reliability or importance. The fact that the model building procedure is expressed in the form of minimizing an energy equation means that the method could easily be combined with or superimposed upon conventional molecular mechanics methods. The representation of the energy terms as sums of even powers of distances between atoms, a form that already exists in most molecular mechanics potentials, should make this particularly straightforward.

ACKNOWLEDGMENTS

The author is grateful to Dr. Gordon Crippen and to Dr. Timothy Havel for valuable discussions regarding this work. Parts of this work were made possible by a grant of computing time from the San Diego Supercomputing Center.

REFERENCES

1. Swenson, M.K., Burgess, A.W., Scheraga, H.A. Conformational analysis of polypeptides: Application to homologous proteins. In: "Frontiers in Physiochemical Biology." Pulman, B. ed. New York: Academic Press, 1978: 115–142.
2. Warme, P.K., Momany, F.A., Rumball, S.V., Tuttle, R.W., Scheraga, H.A. Computation of structures of homologous proteins. alpha-lactalbumin from lysozyme. Biochemistry 13:768–782, 1974.
3. Greer, J. Comparative model-building of the mammalian serine proteases. J. Mol. Biol. 153:1027–1042, 1981.
4. Blundell, T.L., Bedarkar, S., Rinderknecht, E., Humbel, R.E. Insulin-like growth factor: A model for tertiary structure accounting for immunoreactivity and receptor binding. Proc. Natl. Acad. Sci. U.S.A. 75:180–184, 1978.
5. Blundell, T., Sibanda, B.L., Pearl, L. Three-dimensional structure, specificity and catalytic mechanism of renin. Nature (London) 304:273–275, 1983.
6. Greer, J. Model structure for the inflammatory protein C5a. Science 228:1055–1066, 1985.
7. Palmer, K.A., Scheraga, H.A., Riordan, J.F., Vallee, B.L. A preliminary three-dimensional structure of angiogenin. Proc. Natl. Acad. Sci. U.S.A. 83:1965–1969, 1986.
8. Jurasek, L., Olafson, R.W., Johnson, P., Smillie, L.B. Miami Winter Symp. 11:93–123, 1976.
9. de la Paz, P., Sutton, B.J., Darsely, M.J., Rees, A.R. Modeling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. EMBO J. 5:415–425, 1986.
10. Shih, H.H-L., Brady, J., Karplus, M. Structure of proteins with single-site mutations: A minimum perturbation approach. Proc. Natl. Acad. Sci. U.S.A. 82:1697–1700, 1985.
11. Snow, M.E., Amzel, L.M. Calculating three-demensional changes in protein structure due to amino-acid substitutions: The variable regions of immunoglobulins. Proteins Struct. Funct. Genet. 1:267–279, 1986.
12. Snow, M.E., Amzel, L.M. A molecular-mechanics study of the conformation of the interchain disulfide of human immunoglobulin G4. Mol. Immunol. 25:1019–1024, 1988.

13. Blundell, T.L., Carney, D., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J., Singh, D.A., Sibanda, B.L., Sutcliffe, M. Knowledge-based protein modeling and design. Eur. J. Biochem. 172:513–520, 1988.
14. Feldmann, R.J., Bing, D.H., Potter, M., Mainhart, C., Furie, B., Furie, B.C., Caporale, L.H. On the construction of computer models of proteins by the extension of crystallographic structures. In: "Macromolecular Structure and Specificity: Computer-Assisted Modeling and Applications," Vol. 439. Venkataraghavan, B., Feldmann, R.J., eds. New York: Annals of the New York Academy of Sciences, 1985: 12–43.
15. Havel, T.F., Snow, M.E. A new method for building protein conformations from sequence alignments with homologues of known structure. J. Mol. Biol. 217:1–7, 1991.
16. Crippen, G.M., Snow, M.E. A 1.8 Angstrom resolution potential function for protein folding. Biopolymers 29:1479–1489, 1990.
17. Crippen, G.M., Ponnuswamy, P.K. Determination of an empirical energy function for protein conformational analysis by energy embedding. J. Comput. Chem. 8:972–981, 1987.
18. Crippen, G.M., Havel, T.F. Global energy minimization by rotational energy embedding. J. Chem. Inf. Comput. Sci. 30:222–227, 1990.
19. Seetharamulu, P., Crippen, G.M. A potential energy function for protein folding. J. Math. Chem. 6:91–110, 1991.
20. Snow, M.E. Powerful simulated-annealing algorithm locates global minimum of protein-folding potentials from multiple starting conformations. J. Comput. Chem. 13: 579–584 1992.
21. Watenpaugh, K.D., Sieker, L.C., Jensen, L.H. Crystallographic refinement of rubredoxin at 1.2 Angstroms resolution. J. Mol. Biol. 138:615–633, 1980.
22. Sieker, L.C., Stenkamp, R.E., Jensen, L.H., Pickril, B., LeGall, J. Structure of rubredoxin from the bacterium *Desulfovibrio desulfuricans*. FEBS Lett. 208:73–76, 1986.
23. Adman, E.T., Sieker, L.C., Jensen, L.H. Structure of rubredoxin from *Desulfovibrio vulgaris* at 1.5 Angstroms resolution. J. Mol. Biol. 217:337–352, 1991.
24. Frey, M., Sieker, L., Payan, F., Haser, R., Bruschi, M., Pepe, G., LeGall, J. Rubredoxin from *Desulfovibrio gigas*. A molecular model of the oxidized form at 1.4 Angstroms resultion. J. Mol. Biol. 197:525–541, 1987.
25. Protein Identification Resource release 28. National Biomedical Research Fondation, 1991.
26. Protein Data Bank Database release 58. Brookhaven National Laboratory, 1991.
27. Altschul, S.F., Erickson, B.W. Optimal sequence alignment using affine gap costs. Bull. Math. Bio. 48:603–616, 1986.
28. EuGene Sequence Analysis Package. Copyright 1989 Baylor College of Medicine.
29. Fletcher, R., Reeves, C.M. Function minimization by conjugate gradients. Comput. J. 7:149–154, 1964.
30. Lennard-Jones, J.E. "On the Determination of Molecular Fields." Proc. R. Soc. London, Ser. A 106:441–462, 1924.
31. Weiner, S.J., Kollman, P.A., Case, D.A. Singh, U.C. Ghio, C., Alagona, G., Profeta, S., Weiner, P. "A New Force Field for Molecular Mechanical Simulation of Nuclear Acids and Proteins." J. Am. Chem. Soc. 106:765–784, 1984.
32. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." J. Comput. Chem. 4:187–217, 1983.
33. Polygen Corporation. Protein Design Users Guide, June 1991.