A POTENTIAL FUNCTION FOR PROTEIN FOLDING

P. SEETHARAMULU* and Gordon M. CRIPPEN✩

College of Pharmacy, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

One of the most difficult problems in computational chemistry is the prediction of the three-dimensional structure of a protein molecule given only its amino acid sequence. Although there are several programs for calculating the empirical or quantum mechanical energies, and there are more programs for either minimizing the energy as a function of conformation or for simulating the dynamics of a system of molecules, these approaches generally fail either to locate the known native conformation of small proteins and/or show that the native structure is preferred over alternative conformations. In this paper, we present the latest extension of our ongoing effort to devise an empirical potential function that correctly discriminates between the native and essentially all other conformations for more than one protein. Furthermore, the potential incorporates such a simplified description of the polypeptide chain that there is hope for locating nearly the global minimum as a means of predicting globular protein conformation. The current potential function has been parameterized to agree with the crystal structures of crambin and avian pancreatic polypeptide and the parameters thus derived are able to correctly predict the native conformations of apamin and mellitin. The key to this accomplishment is a novel nondifferentiable optimization approach to solving the nonlinear program for determining the parameters.

1. Introduction

A long-standing goal of both experimental and theoretical studies is to predict the folding pathway of a protein or at least its three-dimensional structure given only the sequence [1–5]. Certainly, there are quantum mechanical methods for calculating the energy of small molecules as a function of atomic coordinates, as well as accurate empirical classical potential energy functions for even larger molecules, such as proteins [6–8]. Although these potential functions accurately represent molecules and their conformational energetics, they necessarily have a number of local minima that increases exponentially with the molecular size. Presumably, the experimentally observed conformations correspond to minima nearly as deep as the global one, but

the complete search of the high dimensional potential energy surface is a difficult task even for a tetrapeptide [9]. Even in a local sense, the well developed computer programs that are used for such calculations seem to drive the same starting structure to different final energy minima [10]. One way to reduce the global minimization obstacle is to adopt a simplified representation of the polypeptide chain, where each amino acid residue is represented by only a few united atoms [11–15] or simplifying all 20 residue types to either glycine or alanine [16], for example.

Aside from all the questions of local agreement with known crystal structures, global minimization, and simplified representation, we view the crux of the matter to be that a potential function should favor the native conformation of a protein *globally* over all alternative conformations, something that at least some standard force fields fail to do [17,18]. We have shown earlier [19] that it is in general not possible to construct a potential where a given crystal structure corresponds to exactly a minimum, so our more precise objective is to find a potential applicable to any protein such that the global minimum for each protein is close to the experimentally known structure, and all other local minima are noticeably worse. In that same work, we showed how we adjusted the parameters of the potential function to produce apparently a global minimum within 1.8 Å root mean square interresidue distance of the crystal structure of avian pancreatic polypeptide. Subsequently, we were able to use this same potential together with a new global optimization algorithm to predict the conformation of apamin with a deviation of only 1.85 Å from the experimentally determined structure [20]. Now, we have expanded the scope to include more proteins.

2.   Methods

As in our earlier paper [19], we chose the simple representation of a polypeptide chain where each residue is reduced to a single "united atom" centered at the C^α, thus keeping the virtual bond length between the adjacent residues nearly constant at 3.8 Å. For the purposes of evaluation of the parameters, we selected crambin (1cm) and avian pancreatic polypeptide (1ppt), for the purposes of prediction, we chose apamin (1ap1) and mellitin (1mlt), and for generating alternative structures, we chose serine proteinase (1tgt) and pancreatic trypsin inhibitor (5pti). The coordinates of 1cm [21], 1ppt [22], 1mlt [23], 1tgt [24], and 5pti [25], which were all determined by X-ray crystallography, we obtained from the Brookhaven Protein Data Bank. The coordinates of 1ap1 [26] are from 2D-NMR studies [27].

2.1.   FORMULATION OF THE PROBLEM

For any particular protein, represented as n residue points in a single chain, there will be an experimentally determined native conformation *nat*, the nearest local minimum of the potential function *pert*, and many alternative conformations *alt*, possibly lying very far away in conformation space. For a metric on conformation

space, we have chosen the root mean square (rms) deviation among interresidue distances δ, as defined in eq. (1):

$$\delta(nat, pert) = \left[\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (d_{nat,i,j} - d_{pert,i,j})^2}{n(n-1)/2} \right]^{1/2}, \tag{1}$$

where $d_{nat,i,j}$ and $d_{pert,i,j}$ are the distances between the residues (i and j) in the native and perturbed structures, respectively. For a given functional form and a given set of values for the adjustable parameters, any conformation of the protein has a potential function value $E(\ )$, which we can calculate. As we have already mentioned, in general $\delta(pert, nat) > 0$, but we would at least like to minimize that distance while requiring that the greater $\delta(pert, alt)$ becomes, then the greater $E(alt) - E(pert)$ should be.

The functional form of the potential is a sum of pairwise interactions between amino acid residue points, depending on the distance through space between them, their sequence separation along the chain, and their amino acid types. Each term, therefore, is a function of the distance, a strength parameter ε, and an optimal separation ρ. Both ε and ρ depend in the same way on the sequence separation and amino acid types. We employ two functional forms for the terms: a harmonic stretching between sequentially adjacent residues or cross-linked Cys residues

$$e_b (d_{i,i+1}; \varepsilon_b, \rho_b) = \varepsilon_b \left(\rho_b^2 - d_{i,i+1}^2 \right)^2 \tag{2}$$

and for all other interactions

$$e(d_{i,j}; \varepsilon, \rho) = \varepsilon \left[\frac{3}{(d_{i,j} - \rho)^8} - \frac{4}{(d_{i,j} - \rho)^6} \right]. \tag{3}$$

For most interactions, we want $e \to 0$ asymptotically as $d_{ij} \to \infty$, but if the chain should somehow become broken, the harmonic terms in eq. (2) produce a strong restoring force. We changed from our earlier 12–10 term to eq. (3) because it made the parameter adjusting process easier.

As before, we grouped all interactions into five classes: virtual bond stretching interactions (b) between all residues i and $i + 1$, short range interactions (s) between residues i and $i + 2$, medium range interactions (m) between residues i and $i + 3$, long range interactions (l) between residues i and $j > i + 3$, and disulfide interactions (ss) between bridged Cys residues. The ε and ρ for a given interaction depend on which of the five classes it belongs to and on the residue types t_i and t_j involved. The ith

residue has numerical type t_i corresponding to its index in the ordered list {G, A, V, L, I, C, M, F, P, Y, H, W, S, T, K, R, D, N, E, Q}. In order to keep the number of parameters to a reasonable level, we grouped residue types into classes $c(t_i)$, depending on the interaction class, and then medium and long range ε's and ρ's are selected according to the pair class $p(c(t_i), c(t_j))$:

$$p(c(t_i), c(t_j)) = \frac{(c(t_i) - 1)c(t_i)}{2} + c(t_j) \quad \text{for} \quad c(t_i) > c(t_j). \tag{4}$$

Thus, the total potential function is

$$E = \sum_{i=1}^{n-1} e_b(d_{i,i+1}; \varepsilon_b, \rho_b) + \sum_{i=1}^{n-2} e\left(d_{i,i+2}; \varepsilon_s(t_{i+1}), \rho_s(t_{i+1})\right)$$

$$+ \sum_{i=1}^{n-3} e\left(d_{i,i+3}; \varepsilon_m(p(c(t_{i+1}), c(t_{i+2}))), \rho_m(p(c(t_{i+1}), c(t_{i+2})))\right)$$

$$+ \sum_{i=1}^{n-4} \sum_{j=i+4}^{n} e\left(d_{i,j}; \varepsilon_l(p(c(t_i), c(t_j))), \rho_l(p(c(t_i), c(t_j)))\right)$$

$$+ \sum_{i=1}^{n_s} e_b(d_{s_i}; \varepsilon_{ss}, \rho_{ss}). \tag{5}$$

The residue classification scheme (table 1) for medium and long range interactions was derived from our previous one [19] by splitting classes into helix forming and helix breaking residues [28,29]. As is clear from eq. (5) and table 1, this results in one class for bonded interactions, twenty for short range, seven for medium range, fourteen for long range, and one for disulfide interactions. Since we have an ε and a ρ for each pair class, there are only $2 + 40 + 2(7 \times 8)/2 + 2(14 \times 15)/2 + 2 = 310$ adjustable parameters, compared to 134 for our earlier classification scheme or 864 for all residues being distinguishable.

2.2. ADJUSTMENT

The fundamental problem in this approach is to adjust the 310 parameters such that the *pert* conformer is favored over all *alt*'s. However, as the parameters are changed, the locations of the local minima in general change simultaneously, such that the *pert* and the *alt*'s may no longer be local minima unless their residue coordinates are shifted to compensate for changes in the energy surface. Previously, we had added a constraint that the gradients at *pert* and the *alt*'s should be zero, and then treated all parameters and coordinates as variables. Unfortunately, there was a tendency to

Table 1

Classification scheme for medium and long range interaction

Medium range interactions	
Class	Residue types
1	W
2	M, R
3	A, L, C, Y, E, Q
4	I, F
5	G, P, S
6	V, D
7	H, T, K, N

Long range interactions	
Class	Residue types
1	R
2	A
3	L, F
4	I, C
5	W
6	M, E
7	Y
8	Q
9	G, P
10	N
11	T
12	V, S
13	H, D
14	K

change key local minima into saddle points as the portential surface varied. Instead, we now vary only the potential parameters directly, and then after each small parameter change, we reminimize the $E(pert)$ and the $E(alt)$'s with respect to their sets of residue coordinates:

$$\text{minimize}_{\text{parameters}}\left[\max_{conf \in \{pert,\,alts\}}(E(nat) - E(conf) + \delta(nat, conf)) \right]$$

$$\text{subject to} \begin{cases} E(conf) \text{ is minimum} & \forall\, conf \in \{pert,\, alt's\}, \\ 0 < \varepsilon_i < 10 & \forall\, i, \\ 0 < \rho_i < 10 & \forall\, i. \end{cases} \quad (6)$$

Solving eq. (6) requires a procedure similar to nondifferentiable optimization algorithms [30], namely the following:

Given a starting set of parameter values and a small set of *alt*'s,
For a limited number of outer cycles
    Evaluate the objective function (first line, eq. (6)).
    If the objective is worse,
        If *pert* is the active *conf* in the objective function,
            Accept the worse objective due to the newly discovered minimum
        Else some *alt* is the active *conf*.
            Halve the step size.
            Retract half way to the previous vector of parameters.
            Never permit the ε's and ρ's outside their bounds.
            Return the *pert* and the *alt*'s to their previous conformation.
    Else the objective is better
        Quit if the improvement in objective is too small.
        Double the step size.
        Add current *pert* to list of *alt*'s only if it was active.
    Calculate the (sub)gradient with respect to parameters at the current position.
    If the objective has been improving,
        Take a step along the normalized negative gradient.
        Never permit the ε''s and ρ's outside their bounds.

Since every evaluation of the objective function entails reminimizing with respect to their coordinates the potential of *nat* to produce the latest *pert* and all old *alt*'s to produce the new *alt*'s, we keep a list of only four alternative structures, continually eliminating duplicates, and substituting any new important minima in place of the least active *alt*.

2.3.   INITIAL SET OF PARAMETERS

Of course, the solution found by the algorithm outlined above depends on the starting values of the variables. We were fortunately able to start with the earlier parameters [19] for bonded, short range, and disulfide interactions, but the reclassification of residue types for medium and long range interactions led us to find these starting parameters differently. Experiments on the helix–coil transition of polypeptides show that polyalanine prefers the helical state [31]. Therefore, we let the right-handed α-helical conformation of $(Ala)_{20}$ be the *nat* and the fully extended conformation be the one *alt*. Then the nonlinear program produced a set of medium and long range interaction parameters for Ala–Ala that were applied to all helix former residues. By reversing the roles of the helical and extended conformations, an initial set of parameters for all helix breakers was calculated using a model homopolymer of polyglycine [32].

3. Results and discussion

We began with adjusting our starting set of parameters for 1cm, chosen because it has an accurate crystal structure showing substantial amounts of secondary structure, yet it has only 46 residues. Rotational energy embedding (ROENEMB) [20] is an algorithm that acts as nearly a global optimizer for conformational search problems. Given our starting parameters, it was able to find an *alt* for 1cm substantially better than the *pert*, but then solving the nonlinear program of eq. (6) gave a new set of feasible parameters. Alternating ROENEMB searches with parameter adjustment generated a library of more than 500 alternative structures of 1cm (of which only the four most active ones were needed for the next round of parameter adjustment). The resulting parameter set satisfied $E(pert) < E(alt)$ for all these *alt*'s of 1cm, but since there are some parameters used by 1ppt but not by 1cm, it was not surprising to find that ROENEMB could locate an *alt* of 1ppt with a better potential than its *pert*. Additional cycling between ROENEMB and parameter adjustment and between 1cm and 1ppt eventually produced a library of over 500 *alt*'s of 1ppt and a final parameter set that satisfied both proteins. This is the parameter set shown in table 2.

Our experience has been that ROENEMB is very good at locating excellent local minima, but it is not infallible. As an additional check of the global discriminatory powers of the final parameter set, we generated approximately 300 additional starting conformations for both 1cm and 1ppt and minimized the potential from all of these with respect to residue coordinates. The starting coordinates were obtained from 46 and 36 residue contiguous segments, respectively, of 1tgt and 5pti. The result is that 1cm and 1ppt minimizations began from a variety of protein-like conformers exhibiting realistic secondary and tertiary structural motifs. Altogether then, we had a database of over 800 alternative conformations for 1ppt and another 800 for 1cm.

Table 3 shows the result of using our final parameters in local minimizations from every starting conformation in the database and by searching using ROENEMB. Since our potential is very empirical and is not related to any dynamical experiments, the units must be viewed as completely arbitrary. For 1cm, the best minimum was the *pert* ($\delta = 1.56$ Å, $E = -308.02$), compared to a best *alt* having $\delta = 2.36$ Å and $E = -303.97$. Figures 1 and 2 show that both are quite close to the native and differ mostly by distortions of the loop and the ends of the helices, while keeping the overall folding and main secondary structural features intact. ROENEMB finds a very different, compact structure of substantially worse potential. The result is essentially the same for 1ppt (figs. 3 and 4), with *pert* the best ($\delta = 2.05$ Å, $E = -229.27$), compared to the best *alt* ($\delta = 2.58$ Å, $E = -225.94$) and a much worse and compact ROENEMB structure. Once again, the overall folding and secondary structure is preserved in *pert* or the best *alt*, with the main variations located in the extended strand.

Of course, the real test of the potential is its predictive power. Table 3 shows that for the small proteins 1ap1 and 1mlt – neither of which had been involved in any way in adjusting the parameters – the corresponding *pert*'s are favored over some 300 *alt*'s of each generated as before from 1tgt and 5pti. Figures 5–8 show the close

Table 2

Final set of parameters with interaction labels corresponding to eq. (5) and table 1. Interaction type b is for bond stretching; s,t is short range with intervening residue type t in the sequence {G,A,V,L,I,C,M,F,P,Y,H,W,S,T,K,R,D,N,E,Q} (e.g. 10 refers to the amino acid Y); m,p is medium range with intervening residue pair class p calculated using eq. (4); l,p is long range between residues having pair class p calculated using eq. (4); ss is the disulfide interaction.

Interaction	ε	ρ	Interaction	ε	ρ
b	2.271	3.836	m,5	5.530	3.614
s,1	3.172	4.490	m,6	3.020	4.119
s,2	0.936	4.324	m,7	2.000	4.141
s,3	2.447	4.407	m,8	0.662	4.178
s,4	6.258	4.112	m,9	5.580	4.004
s,5	3.480	4.136	m,10	1.659	3.831
s,6	3.725	4.367	m,11	2.000	4.514
s,7	3.900	4.020	m,12	5.558	4.222
s,8	3.996	4.004	m,13	3.023	4.048
s,9	5.582	4.279	m,14	1.660	3.141
s,10	5.581	4.600	m,15	2.380	4.515
s,11	0.080	4.625	m,16	2.000	4.040
s,12	2.000	4.120	m,17	2.000	3.607
s,13	3.020	4.400	m,18	4.227	4.120
s,14	4.067	3.877	m,19	6.322	4.008
s,15	1.600	4.528	m,20	4.932	4.044
s,16	5.072	4.001	m,21	0.697	4.263
s,17	2.394	4.112	m,22	2.000	4.141
s,18	5.548	4.236	m,23	0.399	4.084
s,19	3.051	4.387	m,24	0.496	3.993
s,20	1.075	4.620	m,25	6.346	3.994
m,1	2.000	4.514	m,26	3.020	3.998
m,2	2.000	4.297	m,27	4.952	3.993
m,3	2.400	3.928	m,28	1.650	4.131
m,4	2.000	3.800	l,1	0.399	4.084
			l,2	0.496	3.415

Table 2 (continued)

Interaction	ε	ρ	Interaction	ε	ρ
1,3	3.660	4.645	1,29	1.034	3.154
1,4	3.889	3.436	1,30	1.035	3.782
1,5	3.659	4.415	1,31	0.913	3.457
1,6	4.940	4.001	1,32	1.035	3.878
1,7	2.299	3.238	1,33	2.000	4.341
1,8	2.287	3.719	1,34	1.035	4.435
1,9	1.660	4.140	1,35	1.032	3.753
1,10	0.406	3.570	1,36	0.732	5.459
1,11	2.000	3.374	1,37	0.020	2.987
1,12	2.000	3.579	1,38	3.003	3.414
1,13	2.000	4.042	1,39	1.660	3.436
1,14	2.000	3.942	1,40	1.009	3.232
1,15	2.000	4.115	1,41	2.000	3.374
1,16	1.813	3.207	1,42	0.020	3.306
1,17	4.220	4.083	1,43	1.660	2.664
1,18	0.577	4.536	1,44	1.035	2.579
1,19	1.216	4.315	1,45	0.020	2.987
1,20	2.000	3.746	1,46	1.660	3.416
1,21	1.600	3.974	1,48	1.580	4.414
1,22	3.660	3.651	1,49	1.659	3.813
1,23	1.216	3.963	1,50	2.000	3.579
1,24	2.020	4.072	1,51	1.056	4.082
1,25	3.017	3.381	1,52	1.216	3.964
1,26	2.000	3.748	1,53	1.021	3.730
1,27	2.340	4.402	1,54	1.660	3.417
1,28	1.216	3.608	1,55	1.282	4.646

Table 2 (continued)

| Interaction	ε	ρ		Interaction	ε	ρ
1,56	1.856	3.232		1,82	2.015	4.312
1,57	1.856	3.816		1,83	2.000	3.746
1,58	0.496	4.134		1,84	1.216	3.974
1,59	1.535	3.714		1,85	1.216	4.439
1,60	2.000	3.942		1,86	1.115	4.508
1,61	0.020	4.314		1,87	1.580	3.308
1,62	0.220	3.375		1,88	0.020	4.082
1,63	0.990	3.897		1,89	0.020	4.315
1,64	0.220	3.231		1,90	2.220	3.746
1,65	1.216	3.814		1,91	0.704	3.364
1,66	1.056	3.966		1,92	2.600	3.451
1,67	1.214	3.374		1,93	1.800	3.963
1,68	1.189	3.579		1,94	2.200	4.272
1,69	1.660	4.042		1,95	2.400	3.375
1,70	1.657	3.939		1,96	2.000	3.748
1,71	2.000	4.115		1,97	2.400	4.539
1,72	0.020	3.746		1,98	3.320	3.808
1,73	1.016	3.748		1,99	2.760	4.140
1,74	0.991	4.359		1,100	2.800	3.751
1,75	0.220	3.374		1,101	2.400	3.963
1,76	0.934	3.580		1,102	2.200	3.375
1,77	1.215	3.942		1,103	2.000	3.748
1,78	1.662	4.116		1,104	2.600	4.439
1,79	1.856	3.306		1,105	2.400	3.608
1,80	0.576	4.082		ss	1.230	5.200
1,81	0.020	4.536				
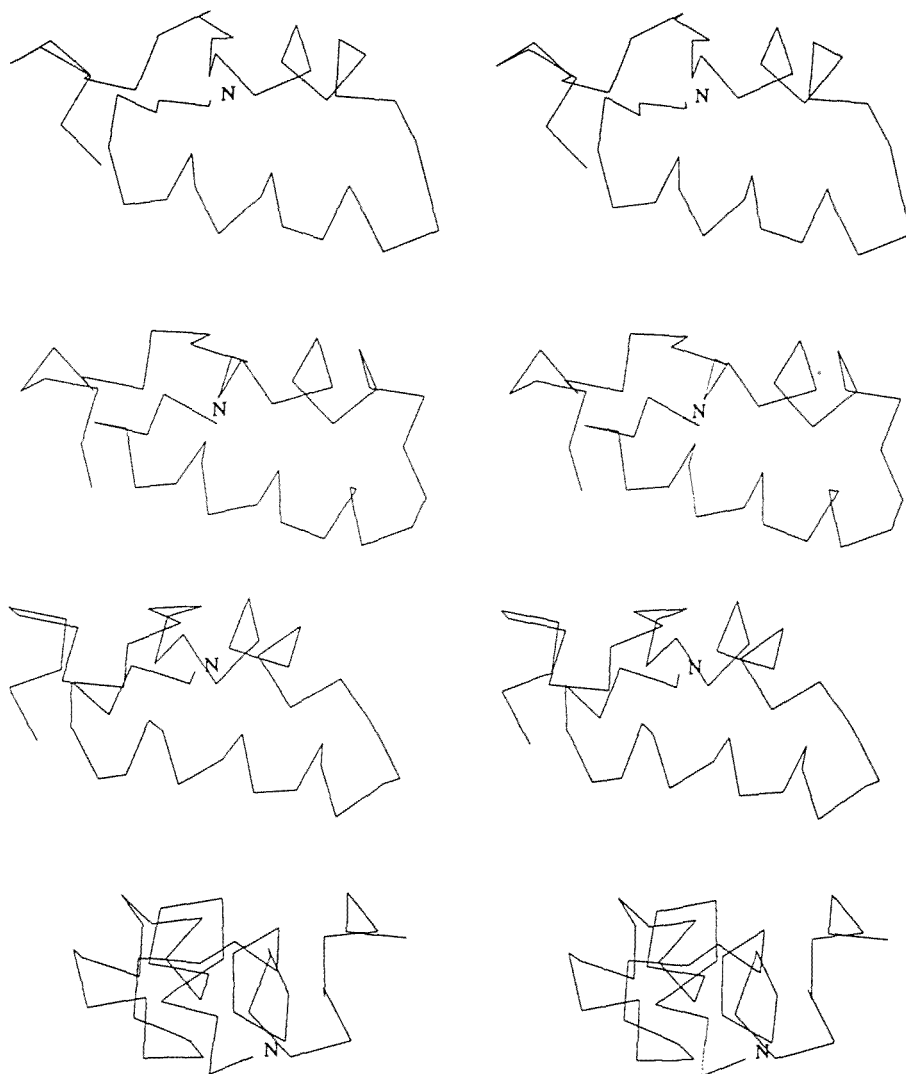
Fig. 1. Stereo pairs illustrating the comparison of C^α traces of different conformations of 1crn. Structures are shown with only the first residue marked for chain tracing purposes. From the top, the figures are the native structure, the near native structure, the best alternative structure, and the conformation favored by ROENEMB.
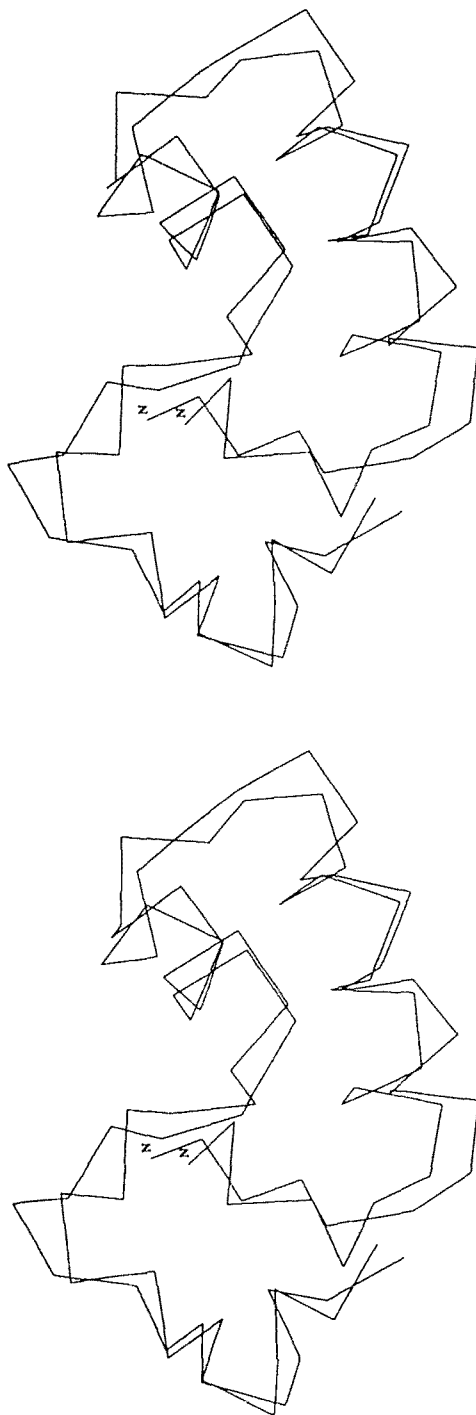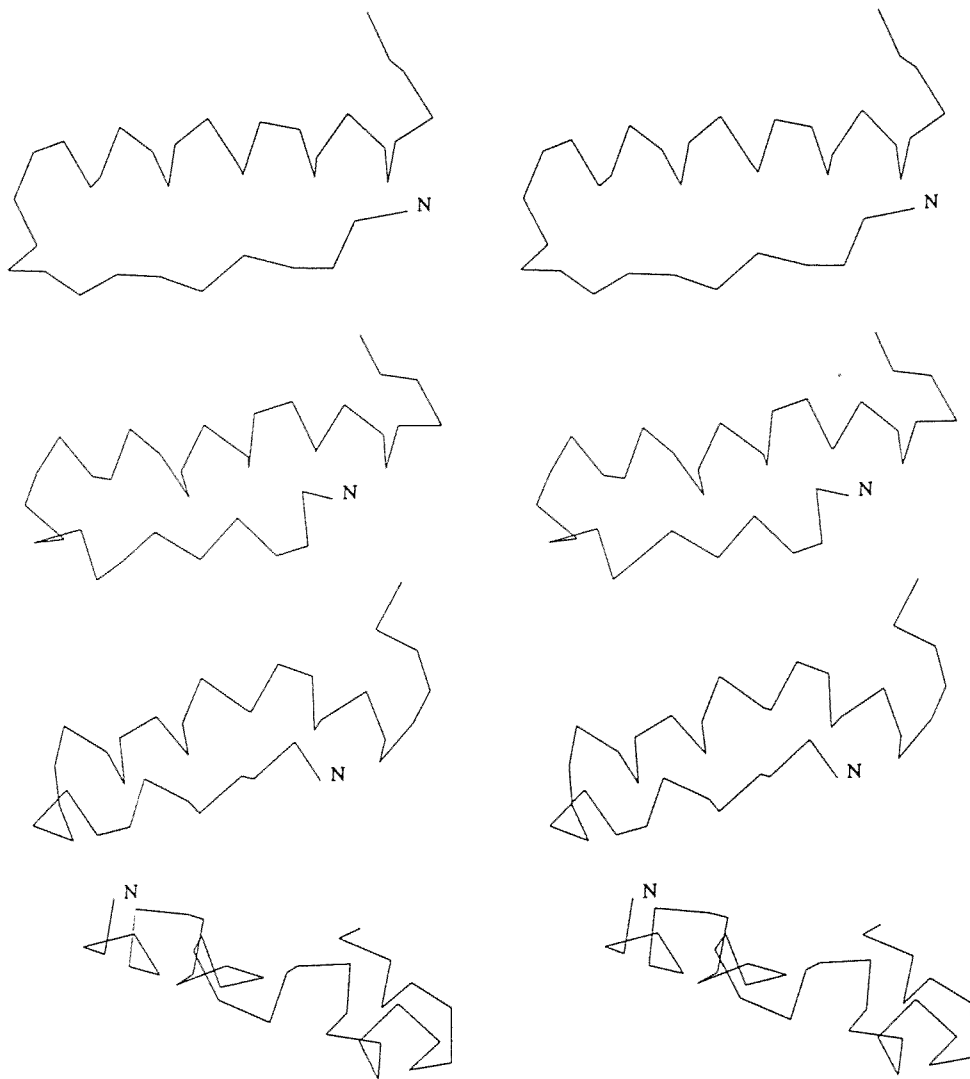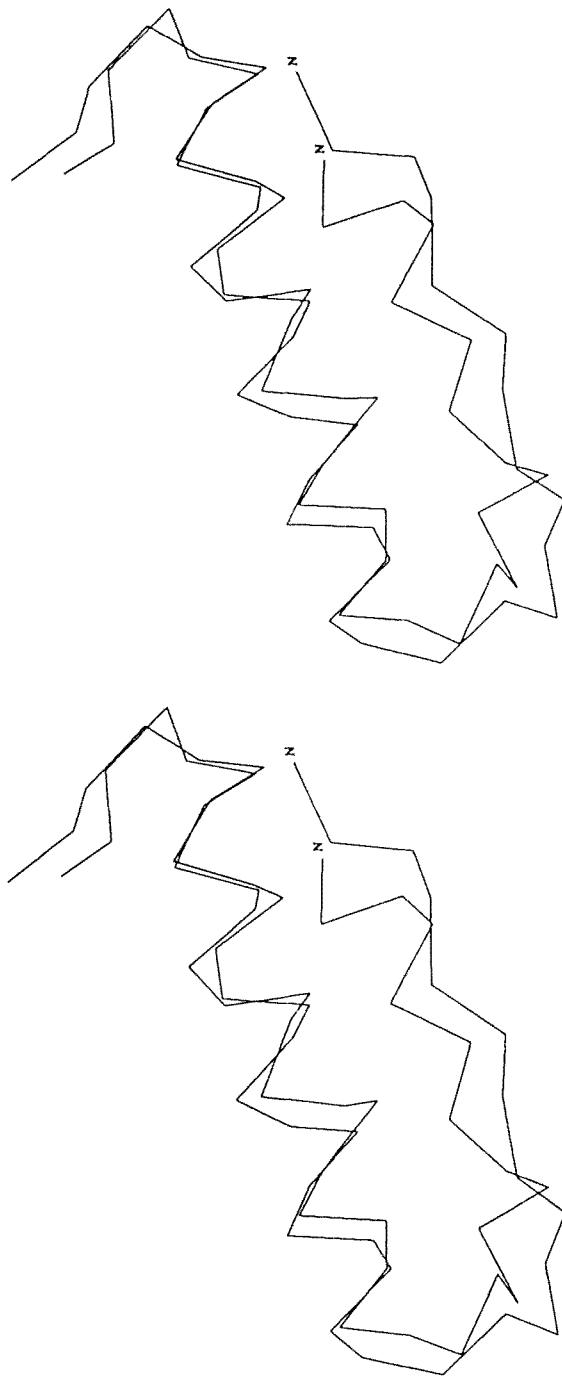
Fig. 2. A stereo pair illustrating the C^a trace of native structure of 1cm with its near native structure superimposed. Only the first residue is marked for chain tracing purposes.
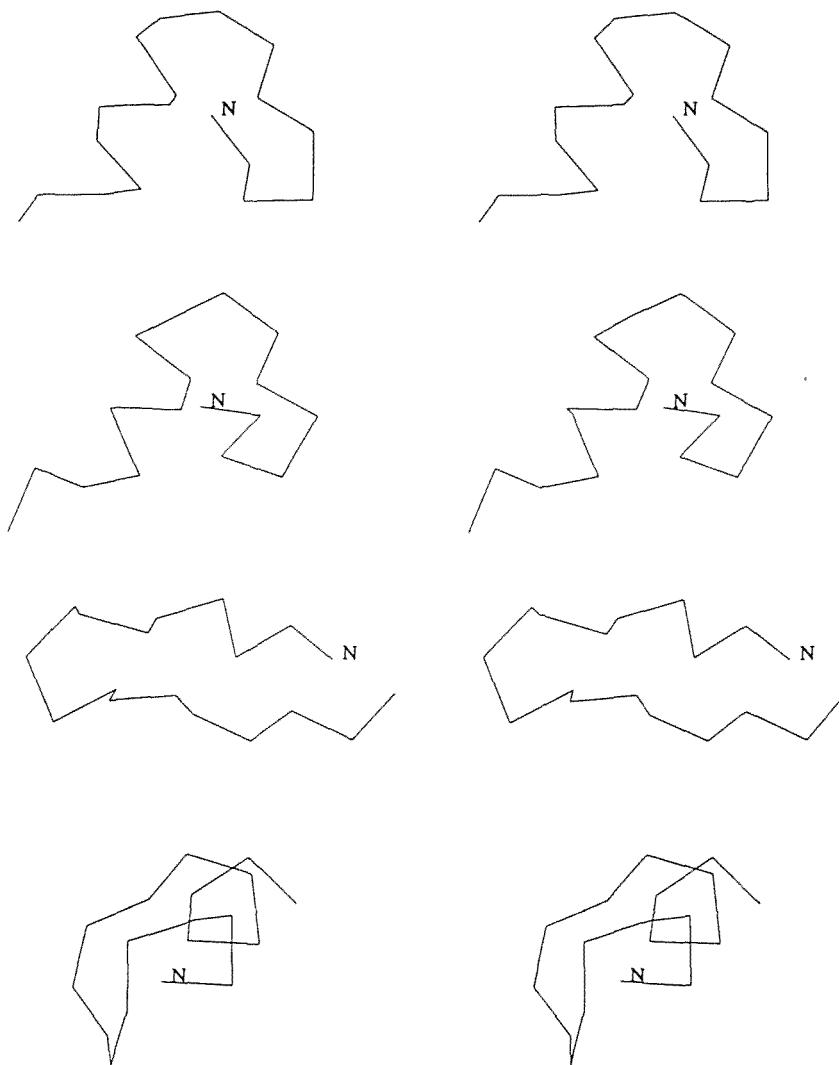
Fig. 3. Stereo pairs illustrating the comparison of C^α traces of different conformations of 1ppt. Structures are shown with only the first residue marked for chain tracing purposes. From the top, the figures are the native structure, the near native structure, the best alternative structure, and the conformation favored by ROENEMB.
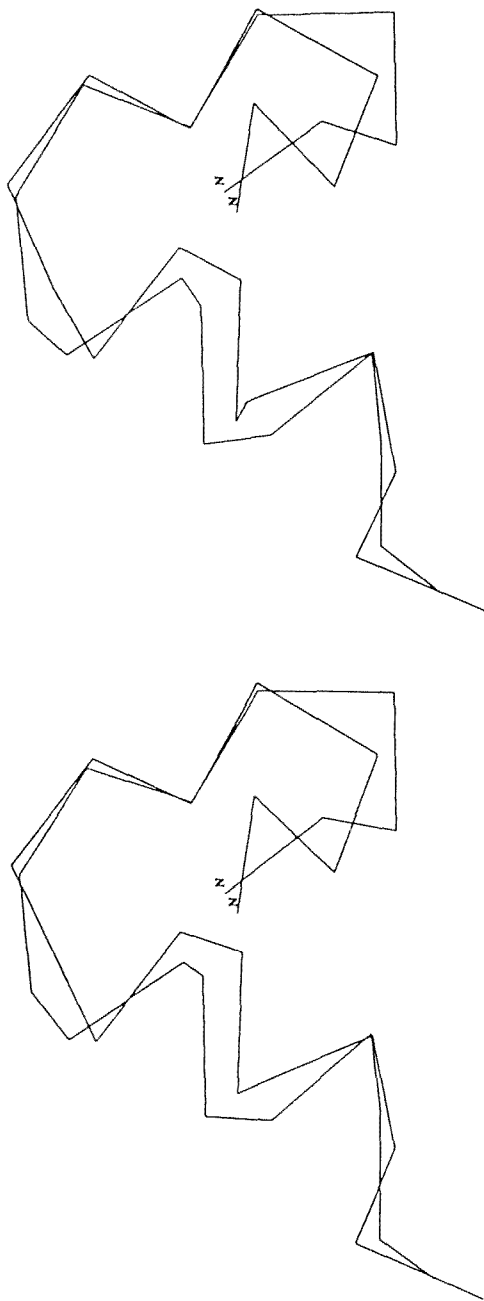
Fig. 4. A stereo pair illustrating the C^α trace of native structure of 1ppt with its near native structure superimposed. Only the first residue is marked for chain tracing purposes.

Fig. 5. Stereo pairs illustrating the comparison of C^α traces of different conformations of 1ap1. Structures are shown with only the first residue marked for chain tracing purposes. From the top, the figures are the native structure, the near native structure, the best alternative structure, and the conformation favored by ROENEMB.

Fig. 6. A stereo pair illustrating the C$^\alpha$ trace of native structure of 1ap1 with its near native structure superimposed. Only the first residue is marked for chain tracing purposes.
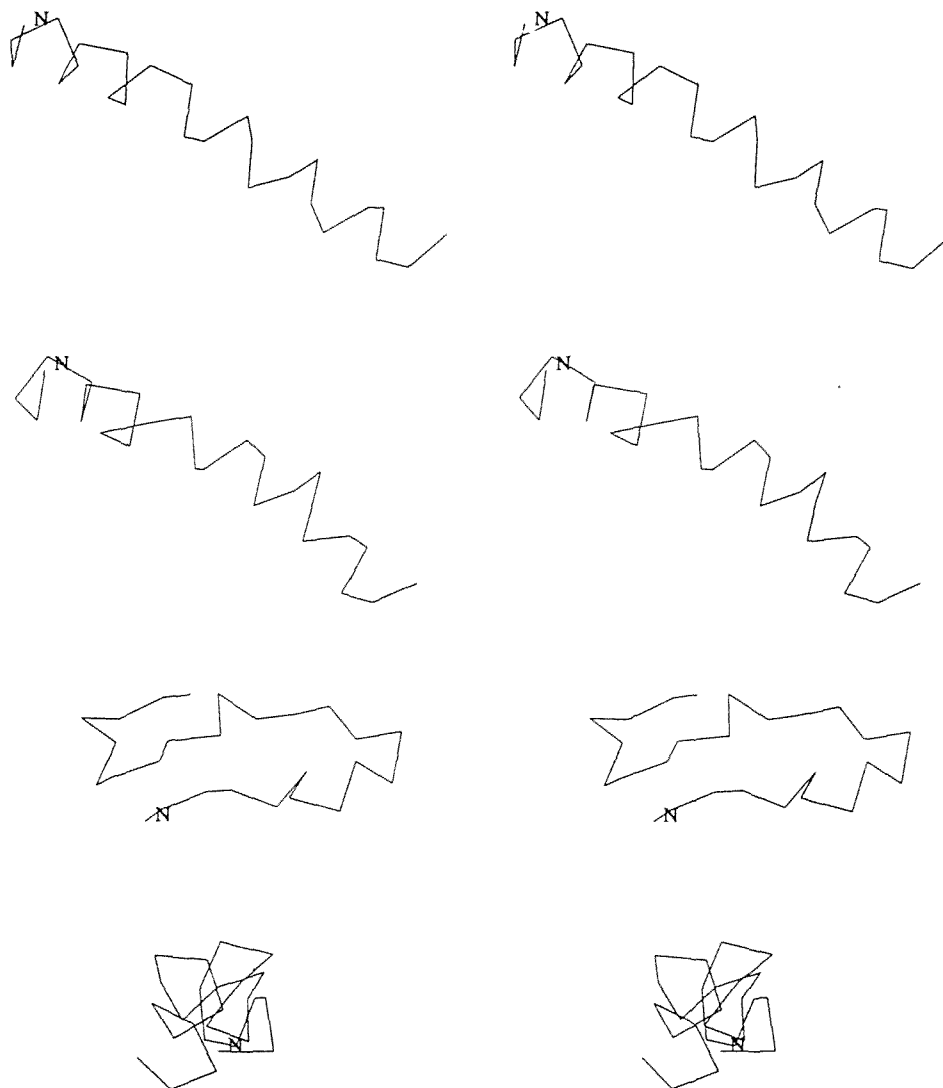
Fig. 7. Stereo pairs illustrating the comparison of C^α traces of different conformations of 1mlt. Structures are shown with only the first residue marked for chain tracing purposes. From the top, the figures are the native structure, the near native structure, the best alternative structure, and the conformation favored by ROENEMB.
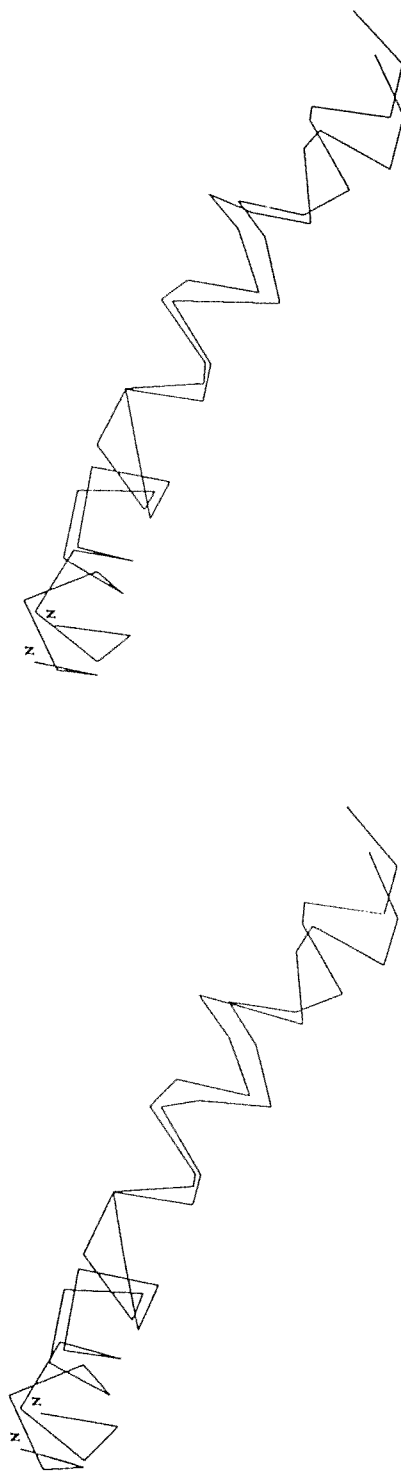
Fig. 8. A stereo pair illustrating the C^α trace of native structure of 1mlt with its near native structure superimposed. Only the first residue is marked for chain tracing purposes.

<div align="center">Table 3</div>

Comparison of energy values and rms deviation from native structure for various conformations of 1crn, 1ppt, 1ap1 and 1mlt

Protein	Native structure	Near native structure		Best alternative structure		Structure calculated from ROENEMB	
	$E^{\star}$	$\delta^{\star}$	$E^{\star}$	$\delta^{\star}$	$E^{\star}$	$\delta^{\star}$	$E^{\star}$
1crn	− 15.95	1.56	− 308.02	2.36	− 303.97	5.92	− 300.57
1ppt	− 69.19	2.05	− 229.27	2.58	− 225.94	7.73	− 199.88
1ap1	2719.11	1.14	− 89.05	3.07	− 84.61	2.55	− 76.73
1mlt	− 61.90	1.62	− 149.03	8.73	− 147.29	10.84	− 139.10

* Potential function value in arbitrary units.

* Root mean square distance deviation (eq. (1)) from the native structure in Å.

resemblance between *nat* and *pert*. Now both the ROENEMB conformers and the best *alt*'s are relatively far from the corresponding *nat*'s. The 1ap1 conformational searches can be viewed as an improvement over our earlier potential [19] in that the difference between *pert* and *nat* is only 1.14 Å compared to the earlier [20] 1.65 Å. However, for prediction of conformation directly from amino acid sequence (without any starting coordinates at all) we have to use ROENEMB, which used to get within 1.85 Å, compared to 2.55 Å using the latest parameter set.

For determining the initial values of the parameters, polyalanine and polyglycine had been used to favor helix over extended conformations and vice versa, but the subsequent fitting of 1ppt and 1crn had greatly changed the parameters. Returning now to polyalanine with helix as *nat*, we found that still the *pert* ($\delta = 1.35$ Å, $E = -120.53$) was favored over the extended conformation ($\delta = 16.95, E = -0.1$) and over the best of some 300 *alt*'s generated as above ($\delta = 7.24, E = -113.38$), but unfortunately ROENEMB found a better conformation ($\delta = 6.96, E = -147.66$). Taking the extended conformation as *nat* for polyglycine, ROENEMB found the best conformation to be the α-helix ($\delta = 15.59, E = -97.56$), followed by the best *alt* generated in the usual way ($\delta = 18.44, E = -80.93$), distantly followed by *pert* ($\delta = 0.0, E = -0.5$).

We conclude that finding some kind of potential function that favors the native structure of even a few proteins over probably all alternative conformations is apparently a more difficult task than simply ensuring there is a local minimum near the native. However, it can be done even using such a simplified model of the polypeptide chain as one point per residue. This suggests that although detailed steric representation may be essential to reproducing the fine conformational features of small molecules, the overall folding of globular proteins does not depend sensitively on even the presence of side chains! One should be careful not to ask too much of this potential. It in no way attempts to simulate the dynamics or the real forces experienced by real

proteins in solution, but rather it attempts only to guide a global minimization search toward a minimum located near the native. Whether this sort of scheme can be extended to a longer list of proteins and more successful "predictions" remains the subject of future research.

Acknowledgements

References

[1] T.E. Creighton, Proc. Natl. Acad. Sci. USA 85(1988)5082.
[2] R.L. Baldwin, Discussion of the pathway of protein folding, in: *Protein Structure Folding, Des. 2*, vol. New Ser. 69, UCLA Symp. Mol. Cell. Biol. (1987), pp. 313–320.
[3] K. Kuwajima, Proteins 6(1989)87.
[4] D. Goldberg, R. Frieden, J. Haack and T. Morrison, Nature 338(1989)127.
[5] T. Noguti and N. Go, Proteins 5(1989)97.
[6] J. Sprague, J. Tai, Y. Yuh and N. Allinger, J. Comp. Chem. 8(1987)581.
[7] S. Weiner, P. Kollman, D. Nguyen and D. Case, J. Comp. Chem. 7(1986)230.
[8] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus, J. Comp. Chem. 4(1983)187.
[9] G.M. Crippen, J. Comp. Phys. 18(1975)224.
[10] I. Roterman, K. Gibson and H. Scheraga, J. Biomol. Struct. Dynamics 7(1989)391.
[11] M. Levitt and A. Warshel, Nature 253(1975)694.
[12] S. Tanaka and H. Scheraga, Macromolecules 9(1976)945.
[13] I.D. Kuntz, G.M. Crippen, P.A. Kollman and D. Kimelman, J. Mol. Biol. 106(1976)983.
[14] C. Wilson and S. Doniach, Proteins 6(1989)193.
[15] G.M. Crippen and V. Viswanadhan, Int. J. Peptide Protein Res. 25(1985)487.
[16] A. Hagler and B. Honig, Proc. Natl. Acad. Sci. USA 75(1978)554.
[17] S. Bryant and L. Amzel, Int. J. Peptide Protein Res. 29(1987)46.
[18] J. Novotny, A.A. Rashin and R.E. Bruccoleri, Proteins: Struct., Funct., Genet. 4(1988)19.
[19] G.M. Crippen and M. Snow, Biopolymers 29(1990)1479–1489.
[20] G.M. Crippen and T. Havel, J. Chem. Inf. Comp. Sci. 30(1990)222–227.
[21] W. Hendrickson and M. Teeter, Nature (London) 290(1981)107.
[22] I. Glover, I. Haneef, J. Pitts, S. Wood, D. Moss, I. Tickle and T. Blundell, Biopolymers 22(1983)293.
[23] T. Terwillinger and D. Eisenberg, J. Biol. Chem. 257(1982)6010.
[24] J. Walter, W. Steigemann, T. Singh, H. Bartuik, W. Bode and R. Huber, Acta Cryst. Sect. B 38(1982)1462.
[25] A. Wlodawer, J. Walter, R. Huber and L. Sjolin, J. Mol. Biol. 180(1984)307.
[26] J. Pease and D. Wemmer, Biochemistry 27(1988)8491.
[27] D. Wemmer, private communication (1989).
[28] M. Vasquez, G. Nemethy and H. Scheraga, Macromolecules 16(1983)1043.
[29] M. Sueki, S. Lee, S. Powers, J. Denton, Y. Konishi and H. Scheraga, Macromolecules 17(1984)148.
[30] R. Fletcher, *Practical Methods of Optimization* (Wiley, New York, 1987).
[31] K. Platzer, V. Ananthanarayanan, R. Andreatta and H. Scheraga, Macromolecules 5(1972)177.
[32] V. Ananthanarayanan, R. Andreatta, D. Poland and H. Scheraga, Macromolucules 4(1971)417.