JOURNAL OF COMPUTATIONAL PHYSICS 12, 49–64 (1973)

Refinement of X Ray Data on Proteins. II. Adjustment of Structure of Specified Geometry to Relieve Atomic Overlaps*

PAUL K. WARME† AND HAROLD A. SCHERAGA‡

Department of Chemistry, Cornell University, Ithaca, New York 14850

Received May 24, 1972

In a first stage of refinement of the X ray data on proteins, the dihedral angles (ϕ_i, ψ_i, χ_i) of the backbone and side chains were adjusted by a least-squares fit to the X ray coordinates so that the computed structure conformed to the bond angles and bond lengths derived from crystal structures of its constituent amino acids. The results are used here as the starting point for the second stage of the refinement, in which atomic overlaps that persist through the first stage are relieved. The energy of the first-stage structure is reduced considerably in the second stage, and the root-mean-square deviation of the computed coordinates from the X ray ones is reduced in going from the first to the second stage of refinement. The second-stage refinement procedure is applied here to lysozyme, and the resulting structure may then be used as a starting point for conformational energy calculations in which the *total* energy is minimized.

INTRODUCTION

As a first step in the refinement of X ray data on proteins, the backbone dihedral angles (ϕ, ψ) and the side-chain dihedral angles (χ_1, χ_2, χ_3, χ_4) were adjusted to obtain a least-squares best fit [1] to the X ray coordinates. The bond lengths and bond angles used for this first stage of refinement were held fixed at the values observed in crystals of the individual amino acids, and the peptide group was maintained in the planar trans conformation. The backbone dihedral angles were adjusted first, and then the side-chain dihedral angles were adjusted (with the backbone atoms maintained in their previously adjusted positions).

In our second stage of refinement, reported here, the same geometrical constraints

* This work was supported by research grants from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312), and from the National Science Foundation (GB-28469X1).

† Postdoctoral Fellow of the National Institute of General Medical Sciences, National Institutes of Health, 1969–1971.

‡ To whom requests for reprints should be addressed.

are maintained, and the nonbonded interaction energy between all atom pairs is reduced in order to ensure that the adjusted structure is a stereochemically feasible one. The total energy function to be minimized (with respect to the dihedral angles) includes not only the nonbonded energy but also a "fitting potential" to ensure that the adjusted structure will remain close to the X ray conformation. In contrast to the first-stage refinement, the backbone and side-chain dihedral angles are adjusted simultaneously in this second stage of refinement.

The objective of this second stage of refinement is to rapidly (and, therefore, economically) relieve the atomic overlaps which persist through the first stage of refinement. This is a necessary preliminary step before undertaking the final stage of refinement, which will take into account all of the known energy contributions in proteins (including nonbonded, electrostatic, hydrogen-bond, torsional, and solvent energies). Also, the first two stages, in themselves, represent partial refinements which a crystallographer may wish to achieve for a particular application at hand, without proceeding to the full-scale refinement of the third stage.

The second-stage refinement is illustrated here for lysozyme.[1] It has also been carried out for α-chymotrypsin [2].

METHOD

The final dihedral angles resulting from the first-stage refinement of lysozyme (using the geometry described by Momany et al. [3] and the X ray coordinates supplied by Phillips [4]) were used as the initial conformation for this study [1]. In several cases, some of the side-chain atom positions were unavailable; therefore, the corresponding side-chain dihedral angles (χ_1 of residues 86, 101, 125, and 128; χ_2 of residues 18, 73, 101, 121, 125, and 128; χ_3 of residues 12, 73, 121, 125, and 128; and χ_4 of residues 1, 73, 97, 112, 125, and 128) were initially set equal to 180°.

For practical reasons, the second-stage refinement (like the first-stage one) is generally performed on twenty-residue segments of the protein, with the remainder of the molecule kept in its previously-adjusted conformation. The major reason for operating on twenty-residue segments is that the number of derivative contributions required in the minimization procedure increases approximately as the square of the number of residues to be adjusted (see Eq. 3 of Ref. [1]); hence, the computation time would be increased very significantly if all dihedral angles of

[1] A listing of the coordinates of the partially refined structure of lysozyme, described in this paper, has been deposited as Document No. NAPS-01941 with the ASIS National Auxiliary Publication Service, c/o CCM Information Corp., 909 Third Avenue, New York, NY 10022. A copy may be secured by citing the document number and by remitting $2.00 for microfiche or $6.00 for photocopies. Advance payment is required. Make checks or money orders payable to: ASIS-NAPS.

the entire protein were adjusted simultaneously. Another reason for performing the computations on short segments is that, in some cases, it is desirable to confine large changes of conformation to a local region of the protein, as illustrated later by the example of residues 67–76 of lysozyme. The segments chosen for the refinement of lysozyme consisted of residues 1–20, 19–38, 37–56, 55–68, 67–76, 75–94, 93–114, and 113–129. The first two amino acid residues of each segment were taken to be the same as the last two residues of the previous segment in order to avoid the possibility of end effects. Also, while refining the ith segment, all residues in segments 1 through $i - 1$ were maintained in their previously adjusted conformations, while all residues following the ith segment were held in the conformations resulting from the first-stage refinement.

One problem associated with segmentation of the protein, as described here, is that a small gap is introduced between the last atom of segment i, which is undergoing refinement, and the first atom of segment $i + 1$, which is held fixed in a particular conformation. In order to eliminate this gap, the previously refined segment i is first lengthened to include, in addition, all residues in segment $i + 1$ in a single continuous chain. This alters the orientation of segment $i + 1$. However, segment $i + 1$ is returned to its proper orientation by readjusting its backbone dihedral angles by the first-stage refinement procedure [1] to make the backbone atoms conform as closely as possible to the coordinates computed earlier [1] in the first-stage refinement (rather than to the X ray coordinates). After this adjustment, the second-stage refinement procedure is applied to segment $i + 1$. In all cases, the effect of this preliminary additional application of the first-stage procedure proved to be a minor readjustment of the chain, affecting the dihedral angles of only the first few residues of segment $i + 1$. This procedure is carried out for all of the consecutive pairs of segments so as to produce a continuous chain without gaps after refinement of all segments of the protein.

The nonbonded energy was expressed as

$$E_{NB} = \sum_i \sum_{j \neq i} \epsilon_{ij}((R_{ij}/r_{ij})^{12} - 2(R_{ij}/r_{ij})^6), \tag{1}$$

where ϵ_{ij} is the minimum energy associated with an interaction between atoms i and j, R_{ij} is the distance between atoms i and j at the position of the energy minimum, and r_{ij} is the actual distance between atoms i and j. The parameters ϵ_{ij} and R_{ij} were those derived by Momany et al. [5] from studies of crystals of small molecules. The 10–12 hydrogen bond potential described by McGuire et al. [6] was used in place of Eq. (1) in the case of interactions between atoms which could participate in an H ··· O hydrogen bond. This potential is of the form

$$E_{HB} = \sum_{H,O} (A/r_{H\cdots O}^{12}) - (B/r_{H\cdots O}^{10}), \tag{2}$$

where $r_{H\cdots O}$ is the distance between the hydrogen and oxygen atoms. The parameters A and B in Eq. (2) were modified to compensate for the absence of electrostatic energy contributions at this stage of refinement (electrostatic energies are not introduced until the third-stage refinement). In order to yield a minimum energy of -5.0 kcal at a distance[2] of 1.83 Å, the values $A = 3.59 \times 10^4$ kcal Å^{12} and $B = 1.28 \times 10^4$ kcal Å^{10} were used. In addition, hydrogen atoms attached to carbon atoms were treated as "united atoms;" i.e., the radii of aliphatic and aromatic carbons were enlarged to reflect the presence of hydrogen atoms, as in the procedure of Gibson and Scheraga [7]. The nonbonded parameters, ϵ_{ii}, which were used for pairs of identical "united atoms," were computed from the polarizabilities (α_i) and effective numbers of electrons (N_i) summarized elsewhere [7, 8], and the R_{ii} parameters were adjusted by comparing the resulting value of E_{NB} with that obtained using the parameters of Momany et al. [5] for two methyl groups (with hydrogens) interacting in various orientations. The values of ϵ_{ii} and R_{ii} for these "united atoms" are summarized in Table I. The ϵ_{ij} and R_{ij} parameters which were used for pairs of nonidentical atoms were derived from the following equations:

$$R_{ij} = (R_{ii} + R_{jj})/2, \tag{3}$$

$$\epsilon_{ij} = \frac{3e(\hbar/m^{1/2})\,\alpha_i\alpha_j}{4R_{ij}^6[(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}]}, \tag{4}$$

where e is the electronic charge and m is the electronic mass (compare Eqs. (1)–(3) from Ref. [8]).

TABLE I

Parameters for Nonbonded Interaction Energy Between "United Atoms"[a]

Atom type	ϵ_{ii} (kcal/mole)	R_{ii} (Å)
–CH, aliphatic	0.0634	4.20
–CH_2, aliphatic	0.1028	4.20
–CH_3, aliphatic	0.1492	4.20
–CH, aromatic	0.1613	4.00

[a] The nonbonded parameters for interaction of these "united atoms" with other types of atoms were derived from the parameters in Ref. [5] together with Eq. (3) and (4).

[2] Although the 10–12 hydrogen bond potential used here is similar in form to the one described by McGuire et al. [6], the parameters have been changed to allow for a strong hydrogen bond without the inclusion of the electrostatic energy contribution.

Equations (1) and (2) were applied to compute the nonbonded interaction energy between pairs of atoms whose relative positions are governed by rotations about two or more intervening single bonds. For atoms whose relative positions are affected by rotation about only one intervening single bond (so-called 1–4 interactions), the following special treatment was used. The *total* interaction energy (nonbonded plus electrostatic plus intrinsic torsional energy [5]) was computed in 10° increments for all atom pairs whose relative positions are affected by variation of only that particular dihedral angle. The resulting dependence of energy on dihedral angle was fit by a Fourier series of twelve terms for backbone dihedral angles, and six terms for side-chain dihedral angles. The Fourier coefficients derived in this way were then used to compute the total interaction energy at any given dihedral angle for all atoms whose positions depend only on that particular dihedral angle. The sum of all such rotational energy contributions is termed E_{ROT}. Since hydrogen atoms had been included in the computation of the energy (which was subsequently fit by a Fourier series), this approach accurately simulates the interactions between hydrogen atoms on adjacent carbons, even though hydrogen atoms do not appear explicitly on the "united atoms."

In order to ensure that the atomic positions in the refined structure would be close to the X ray coordinates, a fitting potential E_{FP}, of the form,

$$E_{\mathrm{FP}} = W \sum_i D_i^2, \tag{5}$$

was included in the function to be minimized, where W is a weighting factor adjusted to give the desired balance between E_{FP} and the other energy terms, and D_i is the distance between the calculated atom position and its corresponding X ray position. The value of W used in the calculations on lysozyme was 50 kcal/mole Å^2, which allowed some atoms to move far enough to relieve overlaps with other atoms, but not far enough to make the overall fit to the X ray structure deteriorate. The summation in Eq. (5) is carried out over all atoms i in the segment under consideration. The method for evaluating the derivatives of E_{FP} with respect to the dihedral angles was described earlier [1]. The disulfide loop-closing potential, E_{SS}, described by Gibson and Scheraga [9] was used to preserve disulfide bonds between the corresponding cysteine residues.

The energy ($E_{\mathrm{NB}} + E_{\mathrm{HB}} + E_{\mathrm{ROT}} + E_{\mathrm{FP}} + E_{\mathrm{SS}}$) was minimized with respect to the backbone *and* the side-chain dihedral angles with the Fletcher–Powell modification [10] of the Davidon [11] minimizer. Ten cycles of minimization were carried out on each segment except the fifth (residues 67–76), for which 15 cycles were allowed because of the very high initial energy of this segment.

The positions of the hydroxyl hydrogens of serine, threonine, and tyrosine are undetermined by the X ray data, and, in most cases, the amide oxygen and nitrogen atoms of asparagine and glutamine are indistinguishable in the electron density

map. Since these residues are capable of forming strong hydrogen bonds, which may have an important influence on the conformation of the protein, the energies associated with various alternative side-chain conformations were first calculated for all residues of these types. For serine and threonine residues, energies were compared at $\chi_2 = -60°$, 60°, 180°, and for tyrosine residues at $\chi_3 = 0°$ and 180°. The energies of asparagine (or glutamine) residues were computed using the χ_2 (or χ_3) value calculated from the X ray coordinates as well as the conformation obtained by a change of 180° in these dihedral angles; the latter conformation corresponds to a reversal of the positions of the amide oxygen and nitrogen atoms. Starting at the *N*-terminus of the protein, and holding the remainder of the protein fixed, the energies of the alternative conformations listed above were computed individually for each of the above five amino acid types in the order in which they were encountered in the amino acid sequence. Then each side chain was placed in its most favorable conformation before going on to the next. After placing all of the 37 serine, threonine, tyrosine, asparagine, and glutamine side-chains in their most favorable conformations, the preceding procedure was repeated in order to determine whether any further adjustments were needed. The fact that no further changes were necessary during the second iteration suggests that these side-chain rotations are not strongly interdependent. This adjustment to optimize hydrogen bonds was performed both before adjusting the twenty-residue segments and again after all of the segments had been adjusted.

Results

After the first-stage refinement of lysozyme [1], the most severe overlap in the structure was between the backbone amide hydrogen of serine residue 72 and the backbone carbonyl oxygen of proline residue 70 (see Fig. 1). The overlap could be relieved only by a substantial change in the conformation of residues 70–74, and it seemed desirable to confine the required changes in dihedral angles to as small a local region as possible, especially since the X ray coordinates of residues 70–74 were still under investigation according to Phillips [4]. Therefore, we chose to refine residues 67–76 as a ten-residue segment, instead of a twenty-residue one, in order to avoid transmitting the local uncertainty to more remote areas of the protein. After the second-stage refinement, the values of ϕ for residues 72 and 74 and the values of ψ for residues 69, 70, 71, and 73 each differed by more than 40° from the angles computed from the X ray coordinates, as shown in Table II. In spite of these large changes in the conformations of residues 69–74, the positions of residues 75 and 76, at the end of this segment, were hardly affected (see Fig. 1). Thus, this readjustment of the chain did not seriously affect the conformation of the rest of the protein. Furthermore, this segment lies on the surface of the

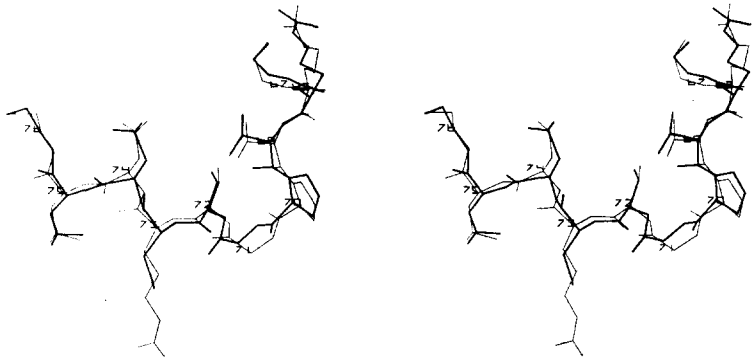

FIG. 1. Stereo comparison of the X ray conformation (heavy line) of lysozyme residues 67–76 with the conformation resulting from the second-stage refinement (light line). All atoms larger than hydrogen are shown except the side-chain atoms of arginine residue 73, for which some of the X ray coordinates were not available. The residues are numbered near their C^{α} atoms; 67 = gly, 68 = arg, 69 = thr, 70 = pro, 71 = gly, 72 = ser, 73 = arg, 74 = asn, 75 = leu, 76 = cys. The deviations from the X ray coordinates are much greater than average in this segment.

protein, so that this local change is less likely to influence other sections of the structure.

The final dihedral angles resulting from this second stage of refinement are listed in Table II along with the deviation of each angle from the value calculated from the X ray coordinates. The total nonbonded energy ($E_{NB} + E_{HB} + E_{ROT}$) of each residue (taken as the sum over all pair atomic interactions within a residue and between all atoms of the given residue and the rest of the protein molecule), obtained after energy minimization, is plotted by residue number in Fig. 2, and the combined RMS deviation of all atoms in each residue from their corresponding X ray coordinates is plotted by residue number in Fig. 3. The results of the first and second stages of refinement are compared in Table III. The large decrease in energy (10^7 kcal to 10^3), obtained in the second stage of refinement, arises primarily from the reduction of the nonbonded energy and the hydrogen bond energy [it should be noted that the hydrogen bond potential (Eq. 2) will yield high positive energies when steric overlaps between hydrogen and oxygen atoms are present]. For example, about half of the initial nonbonded energy resulted from several close contacts between the side chains of arginine residues 45 and 68, and essentially all of the initially high hydrogen bond energy was caused by the overlap between the backbone amide hydrogen of residue 72 and the carbonyl oxygen of residue 70. After refinement, most of the close contacts have been eliminated and, at the same time, the RMS deviation from the X ray coordinates has been reduced significantly (primarily in the side-chain coordinates). The most unfavorable interactions remaining after the second stage of refinement (see Fig. 2) are between *adjacent*

TABLE II

Second-Stage Refined Dihedral Angles of Lysozyme and Deviations from X Ray Values[a,b]

	Residue	Dihedral angles (degrees)					
		ϕ	ψ	χ_1	χ_2	χ_3	χ_4
1	lys	−180(****)	120(2)	−160(15)	−96(67)	−175(−12)	−180(****)
2	val	−109(−7)	135(8)	−11(180)			
3	phe	−85(−13)	178(11)	−46(13)	52(−29)		
4	gly	−105(8)	167(2)				
5	arg	−34(25)	−82(−7)	−121(46)	−152(53)	66(−2)	−153(146)
6	cys	−50(−18)	−77(−30)	−75(28)			
7	glu	−21(54)	−71(−40)	−179(22)	171(−23)	−172(−39)	
8	leu	−41(5)	−47(6)	176(−3)	48(−10)		
9	ala	−47(21)	−70(−16)				
10	ala	−43(19)	−72(−50)				
11	ala	−25(21)	−77(−10)				
12	met	−38(30)	−54(−29)	−64(1)	−157(23)	180(****)	
13	lys	−44(−3)	−64(11)	−130(21)	−114(56)	174(−8)	109(−83)
14	arg	−52(−36)	−14(46)	−161(146)	−178(5)	72(169)	74(−94)
15	his	−84(−4)	−29(−33)	−92(−19)	89(−3)		
16	gly	92(5)	34(47)				
17	leu	−99(−44)	−15(41)	−94(−29)	−180(−11)		
18	asp	−69(4)	133(−6)	−156(16)	179(****)		
19	asn	56(−9)	25(28)	−86(−23)	*115*(177)		
20	tyr	−83(3)	117(−21)	−176(−10)	70(7)	−179(****)	
21	arg	77(27)	8(−28)	−103(−10)	−55(6)	180(−18)	127(−16)
22	gly	96(29)	20(−7)				
23	tyr	−112(−13)	124(−19)	−82(−5)	−63(3)	*0*(****)	
24	ser	−59(12)	159(24)	161(−10)	−180(****)		
25	leu	−70(−33)	−35(−3)	−176(−45)	−174(−21)		

[a] Dihedral angles which were undefined by the X ray coordinates are indicated by (****). The Side-chain angles which were altered to optimize hydrogen bonds are set in italics.

[b] The values in parentheses are the deviations from the dihedral angles computed from the X ray coordinates.

Table continued

TABLE II (*continued*)

Residue		Dihedral angles (degrees)					
		ϕ	ψ	χ_1	χ_2	χ_3	χ_4
26	gly	−66(−21)	−4(33)				
27	asn	−92(−11)	−20(30)	−90(17)	−73(30)		
28	trp	−106(−41)	6(49)	−45(18)	62(−32)		
29	val	−104(−19)	−38(3)	158(−6)			
30	cys	−91(−9)	−13(−11)	157(−15)			
31	ala	−81(1)	−51(1)				
32	ala	−80(40)	−2(21)				
33	lys	−81(−6)	−60(4)	156(4)	70(−35)	−163(12)	−163(27)
34	phe	−84(−21)	2(31)	−77(4)	−58(−25)		
35	glu	−91(−14)	−63(−22)	−157(−42)	55(56)	94(−53)	
36	ser	−109(10)	−31(−9)	62(−30)	*−60*(****)		
37	asn	99(27)	13(−5)	−144(−19)	*−165*(165)		
38	phe	95(20)	−6(−5)	−76(−20)	177(5)		
39	asn	−91(13)	120(14)	−152(−4)	−18(−5)		
40	thr	−87(−27)	−4(12)	56(1)	−180(****)		
41	gln	−94(−10)	−27(5)	−104(−25)	−173(−11)	*122*(−150)	
42	ala	−35(20)	153(15)				
43	thr	−165(−36)	123(−31)	72(9)	*60*(****)		
44	asn	−128(21)	154(45)	−81(−17)	−77(31)		
45	arg	−124(−17)	133(−6)	174(−22)	−169(53)	−142(7)	158(−29)
46	asn	−82(17)	−146(−2)	−55(−8)	*−63*(−172)		
47	thr	−110(−16)	−16(2)	−25(−3)	180(****)		
48	asp	−97(1)	36(−4)	74(−5)	134(14)		
49	gly	68(−32)	−49(27)				
50	ser	−34(−34)	153(39)	122(−18)	*60*(****)		
51	thr	−144(−23)	155(−8)	−82(−19)	*60*(****)		
52	asp	−109(36)	146(46)	−65(21)	174(−19)		
53	tyr	−147(−61)	156(14)	−39(13)	−111(−19)	*1*(****)	
54	gly	95(5)	180(5)				
55	ile	−80(−29)	−17(16)	−175(−11)	144(65)		
56	leu	−98(39)	−17(−43)	−115(27)	13(−21)		

Table continued

TABLE II (*continued*)

	Residue	Dihedral angles (degrees)					
		ϕ	ψ	χ_1	χ_2	χ_3	χ_4
57	gln	64(30)	78(23)	−77(−34)	−68(18)	−69(7)	
58	ile	−93(−24)	144(16)	−43(−16)	170(9)		
59	asn	−81(−5)	155(−11)	−150(22)	*62*(157)		
60	ser	−113(−3)	28(14)	71(20)	−180(****)		
61	arg	−95(−16)	−16(3)	164(−14)	114(−56)	−167(−2)	−166(40)
62	trp	−169(−34)	−6(41)	8(35)	60(−20)		
63	trp	−88(0)	−75(−30)	−78(−31)	161(47)		
64	cys	−120(22)	158(11)	50(8)			
65	asn	−105(−10)	138(−14)	175(−6)	46(20)		
66	asp	−117(45)	−21(−48)	37(−40)	−168(38)		
67	gly	88(0)	11(20)				
68	arg	−119(6)	−15(−31)	137(69)	147(−16)	90(−74)	164(−7)
69	thr	−100(15)	125(41)	−55(12)	177(****)		
70	pro	−58(−13)	−105(−79)				
71	gly	−49(15)	70(87)				
72	ser	−103(−46)	137(7)	−99(−22)	*−70*(****)		
73	arg	−132(−6)	78(−41)	−91(−14)	180(****)	−180(****)	180(****)
74	asn	−50(66)	102(9)	−134(53)	−13(11)		
75	leu	−94(−24)	−24(15)	−100(−23)	−47(18)		
76	cys	−97(−2)	−5(−36)	−103(−49)			
77	asn	42(−11)	75(35)	−138(7)	−60(−3)		
78	ile	−175(−38)	133(−15)	−174(−2)	84(17)		
79	pro	−58(27)	127(−6)				
80	cys	−49(−33)	−42(23)	−54(20)			
81	ser	−49(−10)	−40(8)	85(−25)	180(****)		
82	ala	−64(−26)	5(54)				
83	leu	−104(−55)	1(29)	−73(−12)	−180(2)		
84	leu	−101(−6)	−18(−10)	−57(−29)	−175(6)		
85	ser	−51(−24)	153(0)	84(−20)	*−60*(****)		
86	ser	−77(14)	18(7)	180(****)	*−60*(****)		
87	asp	−125(−5)	102(−7)	−149(4)	96(−24)		

Table continued

TABLE II (*continued*)

	Residue	Dihedral angles (degrees)					
		ϕ	ψ	χ_1	χ_2	χ_3	χ_4
88	ile	−73(−8)	40(34)	64(−8)	132(−29)		
89	thr	−84(−39)	−47(31)	−89(−22)	*60*(****)		
90	ala	−73(−73)	−42(43)				
91	ser	−57(4)	−66(−6)	−60(−2)	180(****)		
92	val	−50(7)	−55(−22)	158(14)			
93	asn	−47(37)	−40(−20)	−99(10)	−3(−1)		
94	cys	−75(0)	−55(0)	−173(−12)			
95	ala	−56(−18)	20(73)				
96	lys	−106(−77)	−25(17)	−81(−23)	−109(55)	−180(−19)	97(−34)
97	lys	−106(−7)	6(30)	−61(5)	−168(−10)	124(−4)	−180(****)
98	ile	−103(−15)	−28(1)	−61(−1)	162(2)		
99	val	−97(−12)	3(11)	65(6)			
100	ser	−111(14)	−2(−30)	−81(−19)	*60*(****)		
101	asp	−65(50)	−59(−22)	−24(****)	178(****)		
102	gly	148(−34)	20(29)				
103	asp	−128(2)	1(0)	−84(−29)	132(37)		
104	gly	52(5)	−148(−12)				
105	met	−75(10)	4(−8)	−52(12)	−140(27)	51(−17)	
106	asn	−69(9)	−22(11)	−91(4)	−58(6)		
107	ala	−68(−19)	−5(11)				
108	trp	−109(−7)	94(10)	−81(−24)	−75(26)		
109	val	−33(1)	−63(−11)	−172(132)			
110	ala	−39(28)	−107(−64)				
111	trp	0(42)	−79(−20)	178(15)	78(−26)		
112	arg	−39(17)	−66(4)	171(−11)	−169(3)	−42(20)	−179(****)
113	asn	−37(30)	−77(−30)	−93(−9)	164(11)		
114	arg	−121(−16)	88(87)	1(30)	122(−78)	70(26)	142(−50)
115	cys	176(−43)	−53(−20)	−102(−47)			
116	lys	−71(2)	116(−3)	−179(−3)	154(6)	138(−9)	−170(−5)
117	gly	101(19)	28(12)				
118	thr	−145(−35)	155(−22)	83(14)	−*60*(****)		

Table continued

TABLE II (*continued*)

	Residue	Dihedral angles (degrees)					
		ϕ	ψ	χ_1	χ_2	χ_3	χ_4
119	asp	−74(27)	90(12)	80(26)	−65(−6)		
120	val	−79(−24)	9(45)	−63(4)			
121	gln	−76(−32)	−23(−1)	−67(−25)	180(****)	180(****)	
122	ala	−83(−6)	10(25)				
123	trp	−99(−35)	−23(36)	−84(−30)	114(5)		
124	ile	−115(−32)	15(32)	53(−25)	162(−10)		
125	arg	−79(−19)	124(0)	−180(****)	−180(****)	180(****)	180(****)
126	gly	93(−13)	−1(11)				
127	cys	−91(−15)	142(20)	−62(−24)			
128	arg	−92(−27)	95(−5)	−180(****)	−180(****)	180(****)	180(****)
129	leu	−116(12)	8(0)	−40(14)	158(−25)		

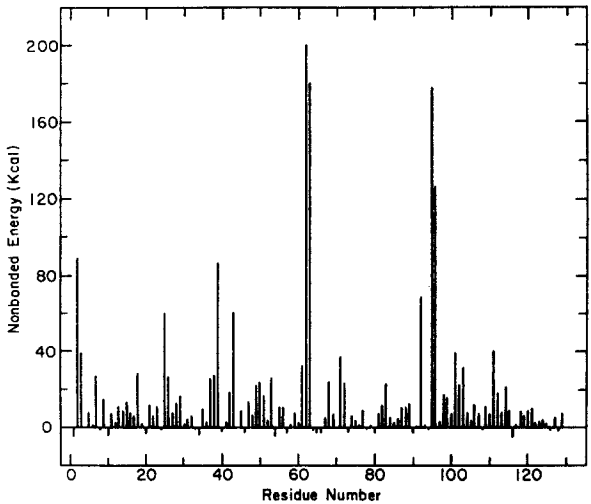

FIG. 2. Total nonbonded energies for all atoms in each residue of lysozyme after the second stage of refinement.

residues (62 and 63, 95 and 96); therefore, these should not cause long-range disruption of the conformation during the final full-energy minimization. As shown in Fig. 4, the overall structure resulting from this second stage of refinement is very similar to the X ray structure of lysozyme.

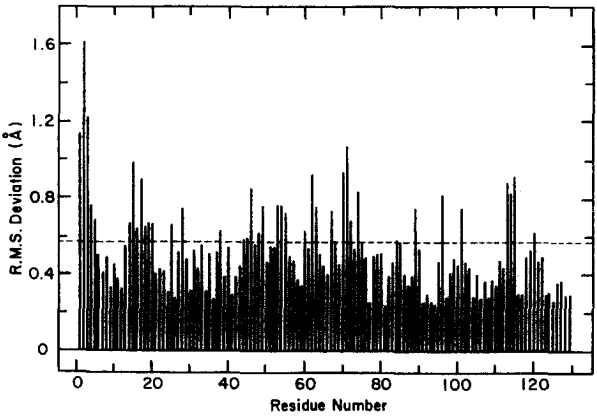

FIG. 3. Root-mean-square deviations for all atoms in each residue of lysozyme after the second stage of refinement. The overall root-mean-square deviation is shown as a dashed line.

TABLE III

Summary of Results from Stage I and Stage II Refinements

	Stage I	Stage II
RMS deviations from X ray coordinates		
Backbone atoms	0.48 Å	0.53 Å
Side-chain atoms	0.92 Å	0.64 Å
All atoms	0.66 Å	0.57 Å
Average deviations from X ray dihedral angles		
Backbone angles	20.2°	20.8°
Side-chain angles	22.5°	24.3°
Energy contributions[a]		
Nonbonded (E_{NB})	1.73×10^4 kcal	1.31×10^3 kcal
Hydrogen bond (E_{HB})	1.56×10^7 [b]	-5.88×10^1
Rotational (E_{ROT})	8.54×10^2	8.40×10^2
Disulfide loop (E_{SS})	1.67×10^2	4.10×10^1
Total	1.56×10^7	2.13×10^3

[a] These energy contributions are summed over all 129 residues of lysozyme.

[b] This high initial value of E_{HB} arises from the steric overlap between the amide carbonyl oxygen of residue 70 and the amide hydrogen of residue 72. The distance between these atoms increased from 0.60 Å to 1.58 Å in the course of refinement.

FIG. 4. Stereo comparison of the final structure after the second stage of refinement (heavy line) with the X ray structure (light line). Only the alpha carbon positions are shown.

About half (17 out of 36) of the serine, threonine, tyrosine, asparagine, and glutamine side chains had adopted one of the alternative side-chain conformations, discussed earlier, either in the initial or the final optimization of hydrogen bonds, The angles affected by these changes are italicized in Table II.

All disulfide bridges in the structure resulting from this second stage of refinement are intact, having *S–S* bond distances ranging between 2.2 and 2.3 Å (equilibrium value 2.2 Å) and bond angles and dihedral angles within acceptable ranges except for the *C–S–S* bond angle between half-cystine residues 30 and 115, which was 141° (equilibrium value 104°).

The computation time for each residue on an IBM 360/65 computer was about 30 sec, and the total time for the second-stage refinement of all 129 amino acid residues of lysozyme was less than one hour.

DISCUSSION

The summary of the results in Table III shows that our objective of relieving atomic overlaps, while maintaining a close fit to the X ray coordinates of lysozyme, was achieved. In general, the deviations of the coordinates of the side-chain atoms decreased while those of the backbone atoms increased, resulting in a net decrease of 14% in the overall RMS deviation from the X ray coordinates. The trend toward improvement of the side-chain fit at the expense of the backbone fit probably results from the procedure used for the first-stage refinement, in which the backbone dihedral angles were first adjusted to fit the backbone atoms (N, C^{α}, C^{β}, C′, and O) while neglecting the side-chain atoms beyond the C^{β} atom; the

backbone atoms were then held fixed while adjusting the side-chain dihedral angles to fit the side-chain atoms. This approach places greater weight (in the first stage) on the X ray coordinates of the backbone atoms to reflect the fact that the side-chain atoms are often less accurately defined in the electron density map. However, in this second stage of refinement, the fitting potential weights backbone and side-chain atom deviations equally so that, where sterically possible, the side-chain atoms can move closer to their X ray positions at the expense of a slight increase in the deviations of the coordinates of the backbone atoms.

One deficiency of the second-stage refinement procedure described here is that the amino acid residues in the first segment interact with many residues further along in the chain which have not yet been adjusted. If the residues in the later segments subsequently move during refinement of those segments, the amino acids in the first segment have no opportunity to readjust. The approximation of holding the remainder of the protein fixed becomes better for later segments because the amino acids in previous segments have already been adjusted. However, it turns out that most atoms move very little in this second stage of refinement, so that this approximation is quite valid even for the first few segments. This deficiency could be corrected, in part, by applying the procedure iteratively, cycling through the protein more than once, although our results after only one iteration on lysozyme were judged to be a satisfactory starting conformation for the final stage of refinement, as discussed later. In any case, the structure resulting from this second-stage refinement is not a global energy minimum because, in addition to the foregoing considerations, electrostatic and solvent energy contributions were neglected, and only ten cycles of minimization were performed on each segment. The fact that the minimization was not carried to completion was probably beneficial in the present context, because some of the bad contacts were not completely relieved by movement of residues in the early segments but were effectively passed on to later segments, where movement of the other residue involved in the bad contact could take place.

The intermediate structure of lysozyme resulting from this second stage of refinement is an appropriate starting conformation for full energy minimization, taking into account nonbonded, hydrogen bond, electrostatic, torsional, and solvent energy contributions. In this final stage of refinement, all dihedral angles will be minimized simultaneously and the fitting potential should no longer be required, since the repulsive energy contributions are now almost balanced by attractive forces.

It is of interest to compare the procedure described in this series of papers with the refinement procedure of Levitt and Lifson [12]. In their procedure, all bond lengths, bond angles, and dihedral angles are treated as variables, whereas, we maintain the bond lengths and bond angles fixed at the values observed in crystal structures of the individual amino acids. The added flexibility gained by allowing

variation of bond lengths and bond angles permits a closer fit to the X ray coordinates but has the disadvantage of requiring about eight times as many variables to completely describe the structure. The number of variables is of crucial importance during energy minimization, because the number of cycles (iterations) of minimization required to reach the minimum is generally greater than or equal to the number of variables. We have, therefore, chosen to adjust only dihedral angles, with the expectation that the bond lengths and bond angles encountered in proteins will not deviate appreciably from the corresponding values in crystal structures of the amino acids; in any event such variations can be allowed for in a later stage of energy minimization, if it proves to be required.

Unfortunately, a direct comparison of our results for lysozyme (RMS = 0.57 Å) with the results of Levitt and Lifson (RMS = 0.22 Å) is impossible, because they did not include nonbonded energies in the calculation which yielded the preceding result. Levitt and Lifson did include nonbonded energies in one of their calculations on myoglobin, and reported an increase in the overall RMS deviation from the X ray coordinates when nonbonded energies were included (RMS = 0.15 Å) over the value obtained by omission of the nonbonded energy contributions (RMS = 0.09 Å). However, these RMS deviations were based on the initial coordinates, which had been refined previously by Diamond's method [13], and not on the X-ray coordinates. These RMS values would, no doubt, be much higher if based on the X ray coordinates. Furthermore, they did not report their final nonbonded energy for this calculation on myoglobin, so that again their results cannot be compared with our results on lysozyme.

REFERENCES

1. P. K. WARME, N. GO, AND H. A. SCHERAGA, *J. Computational Phys.* **9** (1972), 303.
2. K. E. B. PLATZER, P. K. WARME, F. A. MOMANY, AND H. A. SCHERAGA, *Int. J. Peptide Protein Research* **5** (1973), in press.
3. F. A. MOMANY, R. F. MCGUIRE, AND H. A. SCHERAGA, in preparation.
4. D. C. PHILLIPS, private communication, January, 1970.
5. F. A. MOMANY, L. M. CARRUTHERS, R. F. MCGUIRE, AND H. A. SCHERAGA, in preparation.
6. R. F. MCGUIRE, F. A. MOMANY, AND H. A. SCHERAGA, *J. Phys. Chem.* **76** (1972), 375.
7. K. D. GIBSON AND H. A. SCHERAGA, *Proc. Natl. Acad. Sci. U. S. A.* **58** (1967), 420.
8. R. A. SCOTT AND H. A. SCHERAGA, *J. Chem. Phys.* **45** (1966), 2091.
9. K. D. GIBSON AND H. A. SCHERAGA, *Proc. Natl. Acad. Sci. U. S. A.* **58** (1967), 1317.
10. R. FLETCHER AND M. J. D. POWELL, *Computer J.* **6** (1963), 163.
11. W. C. DAVIDON, "AEC Research and Development Report," ANL-5990, 1959.
12. M. LEVITT AND S. LIFSON, *J. Mol. Biol.* **46** (1969), 269.
13. R. DIAMOND, *Acta Cryst.* **21** (1966), 253.