Refinement of the X-Ray Structure of Lysozyme by Complete Energy Minimization†

Paul K. Warme‡ and Harold A. Scheraga*

ABSTRACT: A procedure is described for minimizing the total potential energy of a protein, starting from atomic coordinates derived from X-ray diffraction studies. By first adjusting the dihedral angles (ϕ_i, ψ_i, χ_j) of the backbone and side chains, the polypeptide chain is constrained to adopt planar trans amide groups and to conform to the bond lengths and bond angles derived from crystal structures of its constituent amino acids. Then the conformational energy of the protein is minimized, treating only the dihedral angles as variables. Energy contributions for nonbonded interactions, hydrogen-bonding interactions, electrostatic interactions, torsional interactions, and for closing disulfide loops are included in the computations. A "fitting potential" is employed to ensure that the computed minimum-energy structure coincides closely with the observed X-ray coordinates. The calculations are carried out in three successive stages, the first two of which have been described previously. We report here the third stage refinement of lysozyme, and discuss the optimized conformation in terms of the magnitudes of the various energy contributions. The final structure of lysozyme exhibits a very low conformational energy and yet maintains a close resemblance to the original X-ray structure.

The atomic coordinates from X-ray diffraction studies of protein crystals at a resolution of ~2 Å are subject to uncertainties which, though quite small, may be of the order of several tenths of an angstrom unit. If the X-ray structure is to be used to investigate, say, the details of enzyme–substrate interactions (Platzer *et al.*, 1972), it is necessary that the enzyme be in a low-energy conformation so that deviations induced by substrates can be examined. Since conformational energy calculations are very sensitive to slight errors in atomic positions, particularly when these errors result in atomic overlaps, the minimization of the conformational energy of a protein should lead to a structure in which such overlaps are eliminated.

We have developed a series of refinement procedures to improve the atomic coordinates of a protein. For reasons of economy of computer time and also ease of computation, we have carried out the refinement in three stages. In stage I (Warme *et al.*, 1972), the dihedral angles of the backbone and side chains of the polypeptide chain are adjusted by a least-squares procedure to produce a structure conforming as closely as possible to the X-ray coordinates but, at the same time, one which is constrained to have the fixed bond lengths, bond angles, and planar trans amide groups that are characteristic of crystal structures of amino acids and small peptides (Momany *et al.*, 1973).[1] At the end of stage I, which is solely a geometric fitting procedure (without the computation of energies), the polypeptide chain has acquired "standard" bond lengths and bond angles and planar trans amide groups. However, the constraints of fixed geometry have introduced slight movements of the atomic positions from their X-ray coordinates, which may produce atomic overlaps (in addition to those inherent in the original X-ray structure). The primary purpose of stage I is to introduce a standard geometry characteristic of *each* type of amino acid residue; it then becomes relatively easy to formulate the matrix operations to rotate about single bonds during the course of subsequent energy minimization. We have found it most economical to carry out the subsequent refinement in two stages (II and III). In stage II (Warme and Scheraga, 1973), the effort is made to relieve most of the major atomic overlaps by minimizing only a part of the total energy. The function which is minimized in stage II includes only nonbonded, hydrogen bond, torsional, disulfide loop-closing energies, and a "fitting" potential. In order to conserve computer time during the initial energy minimization, the electrostatic energy was omitted in stage II, and the minimization was terminated after the energy of every amino acid residue had been reduced below 200 kcal. In this paper, we report the procedure and results for carrying out stage III in which we minimize a function which includes not only the energy contributions listed above but also the electrostatic energy; in stage III, we also use a more refined hydrogen-bond potential (compatible with the electrostatic energy introduced at this stage) and carry out the minimization of the total conformational energy through many more cycles, until a plateau in the energy surface is reached.

The stage III refinement procedure is applied to the structure of lysozyme which had already been carried through stages I and II (Warme *et al.*, 1972; Warme and Scheraga, 1973), starting with the X-ray coordinates of Phillips (1970). The amino acid sequence, to which these coordinates apply, is that recently reported by Rees and Offord (1972). The final structure of lysozyme has a low energy (compared to that obtained at the end of stage II), and retains a very close resemblance to the original X-ray structure.

Procedure

A. Reorientation of Coordinate System. Before the energy can be calculated, all atomic positions must be specified in a

† From the Department of Chemistry, Cornell University, Ithaca, New York 14850. *Received June 29, 1973.* This work was supported by research grants from the National Science Foundation (GB-28469X2), from the National Institute of General Medical Sciences of the National Institutes of Health, U. S. Public Health Service (GM-14312), from the Eli Lilly Grants Committee, and from Walter and George Todd.

‡ Postdoctoral Fellow of the National Institute of General Medical Sciences, National Institutes of Health, 1969–1971. Present address: Department of Biochemistry, The Pennsylvania State University, University Park, Pa. 16802.

[1] Data also given in Lewis *et al.* (1973b).

fixed coordinate system. In stages I and II, this coordinate system was introduced by placing the N, C^{α}, and C^{β} atoms of residue 1 (the N-terminal residue) as closely as possible to the corresponding X-ray coordinates. However, since the accuracy of the X-ray coordinates of these three (or any other set of three) atoms is apparently rather low, greater-than-average deviations of the computed coordinates from the X-ray ones persisted throughout stages I and II for the first several residues of the chain. In order to improve the fit of these first three or four residues to the X-ray coordinates, a new coordinate system was chosen by increasing the number of atoms used to define the coordinate system, *i.e.*, by placing atoms N, C^{α}, C^{β}, C′, and O of residue 1 and atoms N and C^{α} of residue 2 as closely as possible to the corresponding X-ray coordinates. The best superposition was computed by an iterative sequence of rigid-body rotations and translations of this group of atoms, without changing dihedral angle ψ from the X-ray value. A new fixed coordinate system for lysozyme was then obtained by placing atoms N, C^{α}, and C^{β} of residue 1 at these optimized positions. The overall RMS deviation of all atoms in the first three residues of lysozyme from the corresponding X-ray coordinates was reduced from 1.33 to 0.49 Å (without affecting the fit of the rest of the chain) by the use of this new fixed coordinate system. A preliminary readjustment of the conformation of the first 15-residue segment was then performed, as described earlier (Warme and Scheraga, 1973), in order to restore residues 2–15 as closely as possible to the positions they occupied before realignment of the first residue. The resulting conformation was then used as the starting point for the further refinement of residues 1–15, as described in section B.

B. Reiteration of Stage II Refinement Procedure. In developing the stage III procedure, it was found that premature elimination of the fitting potential (which constrains the structure to remain close to the X-ray coordinates) resulted in many deviations of the computed structure from the X-ray one which were larger than 1 Å. It become evident that only a few remaining close atomic overlaps, producing moderately high nonbonded energies, dominated the course of energy minimization in such a way as to cause these large deviations. To circumvent this problem, the magnitude of the fitting potential (*i.e.*, W in eq 5 of Warme and Scheraga (1973)) was reduced in two steps, instead of removing it all at once, while conducting further energy minimization by the stage II and III refinement procedures. Since the stage III refinement is more costly in terms of computer time than stage II, because of the inclusion of electrostatic energies and the longer minimization time allowed in stage III, we chose to reiterate stage II using a reduced value of $W = 10$ kcal/(mol Å^2), in contrast to the higher value of 50 kcal/(mol Å^2) used in the stage II procedure reported earlier (Warme and Scheraga, 1973). As before, the stage II calculations were performed on 20-residue segments of the protein in order to reduce the number of partial-derivative calculations and thereby conserve computer time. In this reiteration the 20-residue segments were chosen differently from the first application of stage II, *viz.*, by producing a segment shift of approximately five residues; thus, the ends of the 20-residue segments in the reiteration fell in the interior of the segments chosen for the first application of stage II, thereby tending to smooth out end effects. As before, successive 20-residue segments overlapped the previous segment by two residues to reduce end effects. The overlapping segments chosen for this reiteration of the stage II refinement were 1–15, 14–33, 32–51, 49–68, 67–86, 85–104, 103–122, and 121–129.

One of the most troublesome regions of lysozyme during the first application of stage II was the segment containing amino acid residues 67–76. It was found that rather large local conformational changes were required in order to achieve a low-energy structure for this region, as reported earlier (Warme and Scheraga, 1973). While performing the reiteration of the stage II refinement, it was noted that a 180° rotation of the CO–NH group joining residues Pro-70 and Gly-71 permitted a closer fit to the X-ray coordinates, while simultaneously reducing the conformational energy. Therefore, this alteration was incorporated in all subsequent energy minimizations. This change is equivalent to a transition from a type I to a type II β bend (Venkatachalam, 1968; Lewis *et al.*, 1971).

C. Energy Parameters for Stage III Refinement. 1. *Nonbonded Energy.* The nonbonded energy function (Lennard-Jones potential) E_{NB} and its parameters for –CH, $-CH_2$, and $-CH_3$ groups (treated as "united atoms") are those of Warme and Scheraga (1973). The nonbonded energy parameters for C′, N, and O (and also for hydrogen atoms attached to nitrogen or oxygen atoms) are those documented by Lewis *et al.* (1973b). In order to avoid calculating the weak nonbonded interaction energies between atoms separated by more than the sum of their van der Waals radii *plus* 2.0 Å, a cutoff function [eq 3 of Gibson and Scheraga (1967a)] was employed. Those interactions between atoms whose relative positions are affected by rotation about only one intervening single bond (designated as 1-4 interactions) were treated as torsions (E_{ROT}), using the Fourier series method described by Warme and Scheraga (1973). The Fourier coefficients were optimized to fit the conformational energy curves calculated for rotations about bonds of various types, using the energy parameters of Momany *et al.* (1973).[1] The curves which are fitted by this procedure include energy contributions for all atom pairs (including hydrogen atoms) whose relative positions are effected only by rotations about that particular bond. The value of E_{ROT} was computed as the sum of contributions for each dihedral angle, α, by

$$E_{ROT}(\alpha_k) = a_k + \sum_{i=1}^{n-1} b_{ik} \sin i\alpha + \sum_{j=1}^{n} c_{jk} \cos j\alpha \quad (1)$$

where k denotes a particular type of dihedral angle, and a_k, b_{ik}, c_{jk} are the Fourier coefficients for each type of angle (listed in Table I). For backbone dihedral angles a 12-term expansion ($n = 6$) was used, and for side-chain dihedral angles a 6-term expansion ($n = 3$) was used.

2. *Disulfide Loop-Closing Potential.* The procedure and parameters of Gibson and Scheraga (1967b), involving a loop-closing potential E_{SS}, were used to preserve disulfide bonds within the correct pairs of half-cystine residues. This potential ensures that the S–S bond length, the C^{β}–S–S bond angles, and the $C^{\alpha}C^{\beta}$–SS and C^{β}S–SC^{β} dihedral angles remain close to their standard values.

3. *Fitting Potential.* As in stage II, a fitting potential was used to prevent the atoms from departing too far from their X-ray positions during energy minimization. It is of the form

$$E_{FP} = W \sum_i D_i^2 \quad (2)$$

where W is a weighting factor which maintains a 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 summation is taken over all backbone

TABLE I: Fourier Coefficients for Rotational Energies.[a]

A. Backbone Dihedral Angles[b]

Dihedral Angle	Type[c]	a	b_1	b_2	b_3	b_4	b_5	c_1	c_2	c_3	c_4	c_5	c_6
ϕ	I	2.94	0.00	0.00	0.00	0.00	0.00	−2.96	3.40	−2.82	1.47	−0.81	0.48
ϕ	II	9.27	−10.19	−9.23	−1.20	2.76	2.06	3.78	−0.36	−9.23	−1.14	−0.43	1.56
ψ	I	2.41	0.24	0.00	0.00	0.00	0.00	−0.18	1.86	−1.40	0.64	−0.32	0.20
ψ	II	78.97	81.71	61.65	0.00	−17.45	−4.53	0.21	1.06	−68.46	0.89	−0.67	−2.57
ψ	III	3.56	3.56	2.79	−0.24	−1.05	−0.41	0.74	−0.04	−3.20	0.24	0.01	0.46

B. Side-Chain Dihedral Angles[b]

Amino Acid Type[d]	Coefficients used for χ_1	χ_2	χ_3	χ_4	Angle Type	a	b_1	b_2	c_1	c_2	c_3
Ser	1	2			1	2.52	1.05	0.11	−2.09	0.35	−0.70
Thr	3	4			2	0.67	0.00	0.00	−0.41	0.26	−0.56
Val	5				3	4.17	−0.37	−0.34	−1.78	−0.21	−2.18
Leu	6	7			4	2.46	2.31	1.37	−0.42	0.12	−1.70
Ile	5	8			5	2.80	−1.20	−0.69	0.53	−0.27	−2.12
Met	6	9	10		6	2.82	−0.77	−0.66	−2.07	1.37	−2.28
Asn	11	12			7	2.59	1.63	1.15	−0.84	0.81	−1.78
Gln	6	13	12		8	4.56	4.06	2.99	1.55	0.87	−1.94
Phe	6	14			9	0.48	0.00	0.00	−0.09	0.08	−0.49
Trp	6	14			10	1.47	0.00	0.00	−0.72	0.50	−1.33
Tyr	6	14	15		11	3.43	−0.14	0.59	−2.13	0.23	−1.60
Cys	6				12	0.25	0.00	0.00	0.00	0.28	0.00
Asp	11	12			13	2.62	0.00	0.00	−2.85	1.85	−1.94
Glu	6	13	12		14	1.52	0.00	0.00	0.00	1.96	0.00
His	6	16			15	1.75	0.00	0.00	0.00	−1.75	0.00
Lys	6	13	13	13	16	0.43	0.00	0.00	−0.36	0.51	−0.22
Arg	6	13	17	18	17	2.11	0.00	0.00	−2.14	1.37	−1.60
					18	14.90	0.00	0.00	−18.77	14.88	−10.95

[a] Energy is in kcal/mole, and includes nonbonded and electrostatic energies which are affected by rotation about the particular bond. [b] In conjunction with eq 1, these coefficients will produce a torsional energy curve which is shifted by 180° relative to the position corresponding to the IUPAC–IUB angle convention (1970). Therefore, 180° should be added to each angle before applying eq 1. [c] ϕ type I is used only for glycine; ϕ type II is used for all other residue types. ψ type I is used for glycine except when followed by proline. ψ type II is used for any residue (except glycine) when followed by proline. ψ type III is used for all other cases. These distinctions allow for the absence of a β carbon in glycine and for the presence of a carbon atom attached to the backbone nitrogen in proline. [d] The numbers listed under χ_1, χ_2, χ_3, and χ_4 refer to the angle types for which Fourier coefficients are listed to the right.

and side-chain atoms. Two factors influence the selection of the value of W, *viz.* (1) E_{FP} must be small enough to allow atoms to move to positions of lower energy, and (2) E_{FP} must also be large enough to prevent unduly large departures from the X-ray coordinates. Throughout stage III, the value of W was reduced (by a factor of 10 relative to stage II) to a level of 1.0 kcal/(mol Å^2) in order to allow atoms to move comparatively freely to positions of lower energy.

4. *Electrostatic Energy*. One of the major differences between the stage II and III refinement procedures is that electrostatic energy contributions are taken into account in stage III. The partial electrostatic charges used in this work were derived from CNDO/2 quantum mechanical calculations on amino acids (carried out by Momany *et al.* (1973) and reported by Lewis *et al.* (1973b)). Since the computational algorithms used in this work treat aliphatic and aromatic hydrogens as part of the central carbon atom ("united atom" treatment), the partial charges of these "united atoms" were computed by summing the partial charges of the hydrogen atoms with the charge of the central carbon atom. Any atoms bearing charges less than 0.05 charge unit were rounded down to zero charge, and the charges on adjacent atoms were adjusted accordingly to maintain over all electroneutrality of each amino acid residue.

Another approximation used in the present calculations was to set to zero the overall net charge of the carboxylate groups of aspartate and glutamate residues ($C^{\gamma,\delta}$, +1.0; $O_{1,2}$, −0.5), the ϵ-amino groups of lysine residues (C^{ϵ}, +0.1; N, −1.0; $H_{1,2,3}$, +0.3), and the guanidino groups of arginine residues (C^{δ}, +0.12; N^{ϵ}, −0.35; H^{ϵ}, +0.17; C^{ζ}, +0.48; $N^{\eta}_{1,2}$, −0.42; $H^{\eta}_{1,2,3,4}$, +0.105). There are two reasons for setting the overall charges of these groups to zero. First, in every case, these groups are on the surface of lysozyme, and thus are in a medium of high dielectric constant. Second, the net charges on these groups of atoms are probably balanced by counterions from the external medium, which would also tend to attenuate their long-range electrostatic energies (Hesselink *et al.*, 1973).

In order to reduce the time for computing electrostatic energies as much as possible, consistent with accuracy, a

cutoff distance was imposed, beyond which electrostatic energies were considered to be negligible. To a good approximation, the backbone atoms (C′, O, N, H, and C^{α}) have a composite net charge of zero. Similarly, the side-chain atoms of polar residues which bear significant electrostatic charges may be considered as clusters or groups of atoms having a composite net charge of zero. For accurate calculation of electrostatic energies, it is important to ensure that, when one atom of a given cluster interacts electrostatically with an atom from another cluster, *all* possible pairwise electrostatic interactions (*i.e.*, each atom in the first cluster with each atom in the second cluster) must be computed, for otherwise an unbalanced electrostatic energy will result. Thus, the cutoff distance referred to above applies to the distance between backbone or side-chain *clusters* of atoms, as measured from the central atom of a cluster to the central atom of the interacting cluster. Since the cutoff distance used for electrostatic interactions is much larger than that applied for nonbonded interactions, a much larger number of atomic interactions must be considered, thereby substantially increasing the computation time when electrostatic energy contributions are included in the calculation. The number of atoms within the cutoff distance from a central atom increases approximately as the cube of the cutoff distance, and therefore a considerable saving of computation time is possible by using a short cutoff distance. On the other hand, the cutoff distance must be large enough to ensure that the computed electrostatic energy is accurate. By computing the electrostatic energy of lysozyme as a function of the cutoff distance, it has been found that the electrostatic energy is essentially constant (within 1%) for cutoff distances greater than 9 Å; thus, we have used this value for the present work. With these stipulations, the electrostatic energy is calculated for all pairwise cluster interactions by the Coulomb equation

$$E_{EL} = \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}} \tag{3}$$

where q_i and q_j represent partial charges, ϵ is the effective dielectric constant [2.0 in the present work (McGuire *et al.*, 1972)], and r_{ij} is the distance between atoms i and j.

5. *Hydrogen Bond Energy.* A second major difference between the stage II and III refinement procedures is in the treatment of hydrogen bonds. In the stage II refinement, an approximate hydrogen bond potential was used because of the omission of electrostatic energies at that stage. In stage III, a more refined general hydrogen bond potential (GHB), described by McGuire *et al.* (1972), was used. This potential is of the form

$$E_{HB} = \sum_{H,O} (A_k/r^{12}_{H\cdots X} - B_k/r^{10}_{H\cdots X}) \tag{4}$$

where A_k and B_k are constants for hydrogen bonds of a given type k, $r_{H\cdots X}$ is the distance between the hydrogen bonding atoms, and X denotes either an oxygen atom or a histidine nitrogen atom. The summation is taken over all hydrogen bonding atom pairs, and these same atom pairs are necessarily excluded from the normal nonbonded energy but not from the electrostatic energy.

6. *Free Energy of Solvation.* When the free energy of solvation [computed by the method of Gibson and Scheraga (1967a)] is included, minimization of the energy of lysozyme tends to produce large-scale deviations (many exceeding 1 Å) of the coordinates of polar side-chain atoms from their corresponding X-ray coordinates. This effect is most pronounced for the side-chain atoms of lysine, aspartic acid, glutamic acid, and arginine, which contribute very large positive solvation free energies when their water of hydration is displaced by surrounding atoms. For example, complete removal of all solvent from the ϵ-amino group of lysine contributes 63 kcal/mol to the free energy of solvation, and thus a lysine side chain will tend to move a considerable distance from its original position in order to avoid dehydration, even if other favorable, but weaker, interactions are thereby broken. Furthermore, the theroetical and experimental foundations of the calculation of the free energy of solvation are not as well established as those of the other energy contributions (E_{NB}, E_{HB}, E_{ROT}, and E_{EL}); therefore, we have chosen to omit the solvation free energy contributions during energy minimization in the present work.

D. Segmentation of the Polypeptide Chain. In the stage III refinement, the conformational energy of the protein chain is minimized in segments of 10 amino acid residues, in contrast to the longer segment length of 20 residues used in stages I and II. However, because of the extra fitting constraint applied to the atoms N, C^{α}, and C^{β} of the amino acid at the C-terminus of each segment (see below), a segment was never terminated with a glycine residue since it has no β-carbon atom. For this reason, any 10-residue segment which would have ended with a glycine residue was altered to a 9- or 11-residue segment. Since the number of partial derivative contributions to the total gradient increases approximately as the square of the number of residues per segment, the use of shorter segments in stage III partially offsets the greatly increased computation time arising from the inclusion of electrostatic energies. The more complete energy minimization carried out in stage III also substantially increases the computation time, because it involves a greatly increased number of energy calculations as compared to the limited minimization of stage II. As in stages I and II, an overlap of two residues between successive segments was maintained. For the first iteration of the stage III refinement, the segments chosen were 1–10, 9–18, 17–25, 24–34, 33–42, 41–50, 49–58, 57–66, 65–74, 73–82, 81–90, 89–98, 97–106, 105–114, 113–122, and 121–129. For the second iteration, the segments were chosen to overlap the segments of the first iteration so that the ends of the segments used in the second iteration would fall near the centers of the segments used in the first iteration, thus reducing possible end effects. The segments whose energies were minimized in the second iteration were 1–6, 5–14, 13–21, 20–29, 28–37, 36–45, 44–53, 52–62, 61–70, 69–78, 77–86, 85–94, 93–103, 102–111, 110–119, and 118–129. As in the stage II procedure, the backbone and side-chain dihedral angles of only the segment currently undergoing energy minimization are altered, while all other atoms of the protein are retained in the positions which resulted from the previous segment-by-segment minimization. However, interaction energies are computed from all atoms within the current segment to all other atoms of the whole protein (but only within the cutoff distances mentioned above), including atoms outside the current segment.

One of the problems associated with the stage II refinement procedure described earlier was that a small gap was introduced at the carboxyl end of each segment during energy minimization, thus necessitating a preliminary readjustment of the conformation of the subsequent segment in order to restore it to its original position. Since the conformational energy was already very low after two iterations of the stage II refinement (see Table II), very small adjustments of atomic

TABLE II: Summary of Energy Contributions at Each Stage of Refinement of Lysozyme.

Stage	Iteration	Energy (kcal/mol)[a]					
		NB	HB	EL	ROT	SS	Total
I[b]		1.7×10^4	5.3×10^6	−1061	854	167	5.3×10^6
II	1[c]	1310	37	−1057	840	41	1171
II	2[c]	176	−65	−1057	805	9	−130
III	1[d]	−335	−76	−1072	605	4	−875
III	2[d]	−416	−86	−1086	542	3	−1042

[a] Per mole of protein, calculated with the energy functions used in stage III. [b] This is the energy computed for the structure obtained after the *geometrical* refinement of stage I. [c] These correspond to the first (Warme and Scheraga, 1973) and second (this paper) applications of the stage II refinement. The hydrogen bond energies (HB) given here for stages I and II-1 conflict with the values reported earlier because of the different hydrogen bond function used in the present work. [d] These correspond to the first and second applications of the stage III refinement.

positions were anticipated during stage III. Therefore, three atoms (N, C^α, and C^β) of the C-terminal residue of each segment are constrained to remain nearly fixed in position by applying a very powerful fitting potential only to these atoms. The potential used for this purpose was identical to eq 2, with the value of W set to 2.5×10^5 kcal/(mol Å^2). The magnitude of W was designed to ensure that the indicated atoms would remain within 0.001 Å of their positions from the previous stage. Despite this large value of W, the fitting energy is actually very small because the end atoms do not deviate from their desired positions. The advantage of this approach is that, when the polypeptide chain is extended to include the amino acid residues of the subsequent segment, the newly generated coordinates are always very close (within 0.01 Å) to their positions from the previous stage, thus eliminating the need for any preliminary readjustment of the coordinates. A further advantage of this approach is that an automated procedure for stepping from one segment to the next segment is possible, without requiring any intervening readjustment step. Of course, this procedure restricts the movement of atoms near the carboxyl end of each segment; however, these atoms are always free to move in the succeeding segment, which overlaps the previous segment by two residues. Furthermore, during the second iteration of the stage III refinement, the terminal residues of the segments from the previous iteration are always in the interior of the new segments, where they have greater freedom to move. Since the N-terminus of each segment is fixed in the stage III procedure, and the C-terminus is constrained by a large value of W, most of the movement is confined to the interior of any ten-residue segment.

E. Minimization Procedure. The minimization technique used for the stage III refinement differs substantially from that used in stage II. The variable metric method of Fletcher (1970, 1971) was used, with a slight modification to limit the change of any dihedral angle to 0.1 radian within any single iteration of minimization, and with another modification to set the anticipated step size for any iteration equal to the average of the actual step size and the anticipated one in the previous iteration of minimization. The objective of the first modification is to avoid unreasonably large changes of dihedral angles within any single iteration, which might cause an irreversible jump out of the current potential well. The objective of the second modification is to attenuate the rate of change of step size from one iteration to the next and thereby prevent the step size from becoming too small subsequent to an iteration which terminates with a very small step size. The Fletcher minimizer utilizes only first derivatives, but is even more efficient [by a factor of about 50%, as judged by the same tests made in stage I (Warme *et al.*, 1972) and also by energy calculations on some sample octapeptides] than the Fletcher–Powell (1963) modification of the Davidon (1959) minimizer, which was used in stage II. For each segment, $2N$ iterations of minimization were performed, where N is the number of variable dihedral angles in the segment. In cases where a minimum was approached, as judged by a slow decrease in the energy, before $2N$ iterations had been performed, the minimization was terminated after 150 energy evaluations because additional calculations did not produce any further significant decrease in energy. Although this minimization procedure is quadratically convergent, the energy minimum (again, as judged by the slow decrease in energy) was not attained in some cases, even after $2N$ iterations. However, in all cases, the final average decrease in energy for each additional iteration was less than 0.01 kcal per iteration (per segment), and in most cases the successive decreases were much smaller, thus indicating that a solution very close to the local energy minimum had been attained.

Because of the very strong fitting constraint on the N, C^α, and C^β atoms of the C-terminal residue of each segment, even a small movement of any one of these atoms will cause a large energy increase (*e.g.*, a deviation of 0.01 Å produces an energy of 25 kcal). Since very small deviations of the backbone dihedral angles will affect the positions of these atoms at the carboxyl end of the segment, the partial first derivatives of the energy with respect to the backbone dihedral angles of the *whole* segment were rather large (the maximum value for any angle within a segment ranged from 10 to 220 kcal/deg for various segments), even at the point of termination of minimization. In other words, the potential well in which this fitting is carried out has very steep sides, and thus the derivatives are quite large even for very small deviations from the minimum. In spite of this tight constraint on the backbone dihedral angles, they are sufficiently free to move about 7°, on the average, per iteration of stage III refinement, as shown by Table V. This flexibility arises from the possibility of coordinated movements of multiple backbone dihedral angles, which can lead to reduction of the conformational energy while yet holding the C-terminal end nearly fixed in position. The partial derivatives of the energy with respect to the *side-chain* dihedral angles were quite low (the maximum value for any angle within a segment ranged from 0.002 to 0.12 kcal/deg

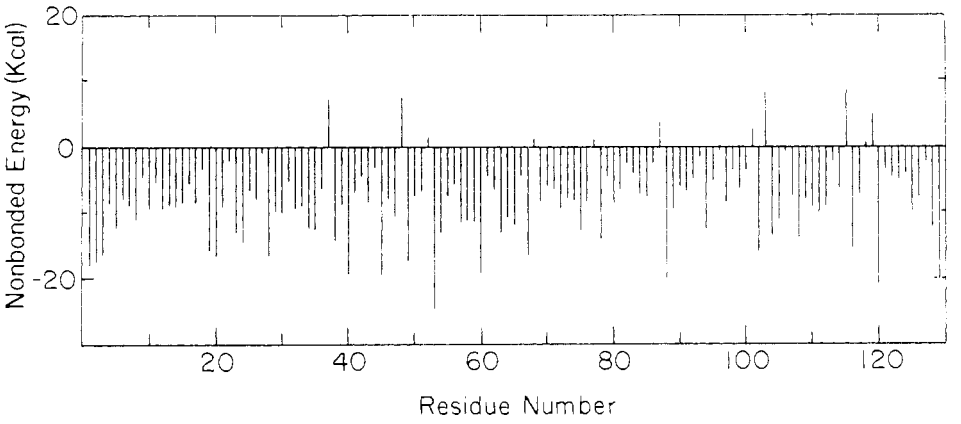

FIGURE 1: Plot of total energy ($E_{NB} + E_{HB} + E_{ROT} + E_{EL}$) of each residue of lysozyme after stage III refinement, calculated as one-half of the sum of the energies of interaction with all other atoms of the entire protein which are within the cutoff distance.

for various segments), because these derivatives are not influenced by the strong fitting constraint on the C-terminal residue.

The computation time for one complete iteration of stage III refinement on all segments of lysozyme (*i.e.*, the whole protein), using a CDC 6600 computer, was about 120 min. For the sake of comparison, the computation time for stages I and II would be approximately 5 and 15 min per iteration, respectively, on the same computer.

Results

A. Energy Contributions at Various Stages of Refinement. The magnitudes of the total energy E_{TOTAL} and its various contributions (E_{NB}, E_{HB}, E_{EL}, E_{ROT}, and E_{SS}) at each stage of refinement are summarized in Table II. It is important to realize that the various contributions to E_{TOTAL} are calculated with respect to (different) arbitrary reference states; hence, the net changes in the energy contributions in going from one stage of refinement to the next are more significant than the magnitudes of the energy contributions.

The nonbonded energy decreases significantly at each successive stage until a very negative nonbonded energy is achieved in the final structure. However, 108 pairs of atoms are still closer than the usual minimum contact distances, after the second iteration of stage III. Of these, 81 close contacts contribute unfavorable energies of less than 1 kcal, which are probably counterbalanced by favorable energy contributions from other associated interactions (54 of these are interactions between backbone atoms of adjacent residues). Of the remaining 27 close contacts (with energies ≥ 1 kcal), 16 are caused by interactions between nitrogen or oxygen atoms bearing hydrogen atoms and the oxygen atom to which the hydrogen atom is hydrogen bonded. In such cases, these unfavorable energies are generally offset by very favorable hydrogen bond and electrostatic energies in the same region. Six other close contacts contributing nonbonded energies greater than 1 kcal are caused by interactions of the branched chain γ carbons of valine and isoleucine with the backbone amide hydrogen atom of the same residue. The remaining five close contacts follow no general pattern, but the highest energy of all is only 5.4 kcal; thus, it seems reasonable to conclude that essentially all atomic overlaps have been eliminated by these energy minimization procedures.

From the third column of Table II, it is evident that the hydrogen bond energy also steadily becomes more favorable during the course of minimization. The values of E_{HB} given in this table differ somewhat from those reported earlier (Warme and Scheraga, 1973) in connection with the stage II refinement, because of the different hydrogen bond potential used in this work. It should also be noted that the total energy of a hydrogen bond is distributed between E_{HB} and E_{EL}. Thus, the values of E_{HB} given in this table are not to be compared directly with those computed using other hydrogen bond potential functions.

The electrostatic energy decreases only slightly during the stage III energy minimization, as shown in column 4 of Table II. This result justifies, in part, our omission of electrostatic energy during the stage II refinement. The other reason for this omission was of a more practical nature, *viz.*, that inclusion of the electrostatic energy significantly increases the computation time. In any case, it is important to include the electrostatic energy in the final energy calculations in order to treat the hydrogen-bonding interactions correctly.

The rotational energies given in column 5 of Table II decrease significantly at all stages of refinement, but particularly in the third stage. The explanation for the concentration of these changes in the later stages of refinement is that the energy to be gained is distributed quite evenly over all of the dihedral angles in the structure. From the fact that there are 493 variable dihedral angles in lysozyme, it is clear that the average energy contribution for any given dihedral angle is of the order of 1 kcal, since they are fairly uniformly distributed. Since the minimization procedure operates primarily on the highest energy interactions in the molecule, the close atomic contacts are the first to be operated on, and only when these are essentially completely eliminated will the minimizer reduce the already comparatively low rotational energies.

The total energy ($E_{NB} + E_{HB} + E_{EL} + E_{ROT}$) for each residue of lysozyme is plotted in Figure 1. The net energy for almost every residue is negative, the worst case being residue Cys-115, which has an energy of 8.4 kcal. This high energy results from an unfavorable nonbonded interaction between the β methylene group and the backbone amide hydrogen of this residue.

All four disulfide bonds of the final structure for lysozyme are very close to their ideal values, as indicated by the values in the sixth column of Table II.

B. Deviations from X-Ray Coordinates. Another important criterion for judging the refined structure of lysozyme is its deviation from the original X-ray structure. The RMS deviations of the backbone atoms (including C^{β}), side-chain atoms, and the composite figure for all atoms are presented in Table III. The reasons for the substantial decreases of the RMS

TABLE III: Summary of RMS Deviations from X-Ray Coordinates at Each Stage of Refinement of Lysozyme.

Stage	Iteration	RMS Deviation[a] (Å) Backbone Atoms	Side-Chain Atoms	All Atoms	Fitting Constant W (kcal/(mol Å^2))	ΣD_i^2 (Å^2)	E_{FP}[b] (kcal/mol)
I		0.48	0.92	0.66		418	
II	1	0.53	0.64	0.57	50	312	15,600
II	2	0.53	0.51	0.52	10	259	2,590
III	1	0.59	0.70	0.63	1	380	380
III	2	0.58	0.67	0.62	1	368	368

[a] The RMS deviation is computed as the square root of the sum of the squares of the deviations of the computed coordinates from the corresponding X-ray coordinates, divided by the number of atoms compared. [b] Since the C-terminal atoms of each segment are very close to their desired positions, the magnitude of the fitting energy (for these atoms, using $W = 2.5 \times 10^5$ kcal/(mol Å^2)) is less than 1 kcal.

deviations observed in going from stage I to stage II have already been discussed (Warme and Scheraga, 1973). In the second iteration of stage II refinement, the overall fit to the X-ray coordinates improves dramatically, for much the same reasons, in spite of a decrease of the fitting parameter W. However, when W is further decreased to a value of 1 kcal/(mol Å^2) in stage III, the atoms are allowed much greater freedom to deviate from the X-ray coordinates if the overall energy of the structure can thereby be lowered. Thus, the RMS deviations increase significantly during the first iteration of the stage III energy minimization. However, by the end of the second iteration of stage III refinement, the energy to be gained by further deviation is effectively balanced by the small fitting constraint, and therefore, the RMS deviation decreases just slightly during the second iteration. This example obviates the importance of choosing an appropriate value for W, since a low value of W will allow atoms to deviate excessively from the X-ray coordinates, while a high value will not allow sufficient freedom of movement to achieve a low conformational energy. Ideally, the fitting parameter W should be set to zero for the final energy minimization. However, because we have not minimized the free energy of solvation during the present energy calculations, it is likely that the resulting distortion of the effective potential energy would give rise to some rather large deviations from the X-ray coordinates in the absence of the weak fitting potential. In other words, if one or more terms of the total energy expression is inaccurate (or omitted entirely), the potential energy is unbalanced and would probably lead to an incorrect conformational energy minimum. The imposition of a fitting constraint effectively biases the solution toward a conformation which is similar to the X-ray structure.

The fact that the fitting energy (E_{FP}) is still 368 kcal at the end of stage III (see Table III) should not be interpreted as indicating that this term is dominating during the energy minimization. At the end of stage II, when W is reduced from 10 to 1 kcal/(mol Å^2), E_{FP} assumes a value of 259 kcal/mol, which is its lowest value throughout the course of all three stages of refinement. Because of the constraints of fixed geometry, it is doubtful whether E_{FP} could be reduced much below the value of 259 kcal/mol. During stage III, E_{FP} increases by 109 from 259 to 368 kcal/mol; however, at the same time, the other energy terms decrease by a total of 912 from -130 to -1042 kcal/mol (see Table II). Thus, one must not consider the magnitudes of the various energy contributions in the final structure, but rather the relative changes in these energy terms from one stage of refinement to the next, in order to determine which energy terms are having the greatest effect on the energy minimization. By this criterion, the fitting energy (E_{FP}) contribution in stage III is quite minor in comparison to the other energy contributions ($E_{NB} + E_{HB} + E_{EL} + E_{ROT} + E_{SS}$). This follows from the data of Tables II and III, which show that the changes in E_{FP} during stages III-1 (+0.94 kcal/residue) and III-2 (-0.093 kcal/residue) are substantially less than the changes in the total energy during stage III-1 (-5.91 kcal/residue) and during III-2 (-1.18 kcal/residue). Thus, although the fitting constraint is not removed entirely in stage III, it is sufficiently weak to permit reduction of the energy by conformational readjustments while yet preventing unduly large deviations from the X-ray coordinates.

The overall RMS deviations for all atoms (both backbone and side chain) of each residue of lysozyme are plotted in Figure 2. It is apparent that the deviations are quite evenly distributed throughout the molecule, rather than being concentrated in any particular area. However, we have noted some trends in surveying the deviations for various types of atoms. For example, there is a definite tendency for the backbone carbonyl oxygen atoms and the C^β atoms to deviate from the X-ray coordinates more than the other backbone atoms. The RMS values for the final minimized structure are 0.49, 0.53, 0.64, 0.47, and 0.74 Å for the N, C^α, C^β, C′, and O atoms, respectively. Another trend which is apparent from the data in Table III is that the side-chain atoms tend to deviate from the X-ray coordinates more than the backbone atoms. This was why the procedure of stage I, which places greater emphasis on accurate fitting of the backbone atoms, was adopted (Warme *et al.*, 1972). It is of interest that the overall RMS deviation of the side-chain atoms decreases to about the same level as that of the backbone atoms during stage II, although, in stage III, the side-chain RMS value increases about three times as much as that of the backbone atoms. This behavior is related to the reduction in the value of the fitting parameter W in stage III, which allows atoms to deviate to a greater extent from their corresponding X-ray coordinates if the conformational energy can thereby be reduced. In both stages II and III, a uniform fitting constraint is applied to all backbone and side-chain atoms; however, the strong fitting potential energy tends to dominate the total energy in stage II whereas, in stage III, the effective magnitude of this energy contribution is quite small relative to the nonbonded, hydrogen bond, rotational, and electrostatic energy contributions (*cf.* Tables II and III). It is thus significant that,

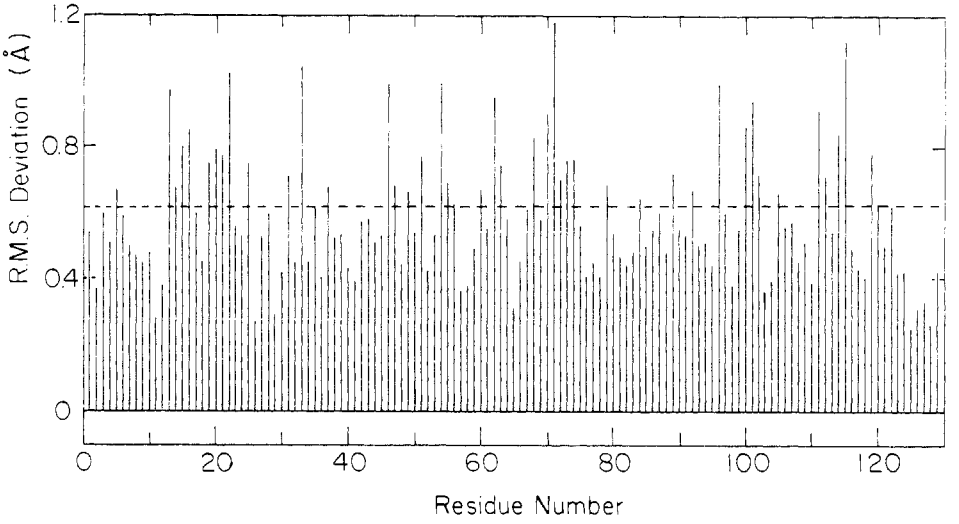

FIGURE 2: Plot of RMS deviation for each residue of lysozyme after stage III refinement. The overall RMS deviation is shown as a dashed line.

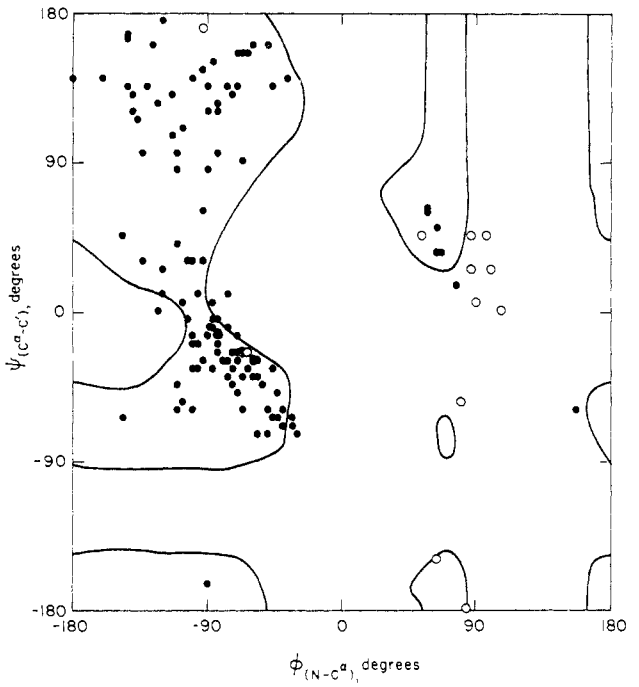

FIGURE 3: Plot of refined backbone dihedral angles of lysozyme after stage III refinement. The open circles denote glycine residues, whereas all other residues are represented by solid circles. The 10-kcal energy contours for an alanine dipeptide are included for comparison (Ponnuswamy *et al.*, 1973). The open circles fall in allowed regions for a glycine dipeptide.

when the fitting constraint is relaxed, the side-chain atoms deviate more than the backbone atoms, in line with our expectations.

It is also apparent from our results that the interior atoms of lysozyme tend to deviate from their corresponding X-ray coordinates less than the exterior atoms. The static accessibility data of Lee and Richards (1971) were used to classify each atom of lysozyme as internal (if not accessible to solvent) or external (if accessible to solvent). The overall RMS deviation for 461 internal atoms was 0.57 Å (average deviation of 0.50 Å), whereas the RMS deviation of the 496 external atoms was 0.66 Å (average deviation of 0.58 Å); this suggests that there is less thermal motion in the interior, thereby enabling a better fit to be achieved in this region of the protein. The fact that the overall RMS deviation for all atoms (0.62 Å) is much larger than the average deviation for all atoms (0.54 Å) reflects the fact that the relatively few atoms which deviate widely from their corresponding X-ray coordinates contribute disproportionately to the root mean square value.

C. Variations of Dihedral Angles. The backbone and side-chain dihedral angles for the fully minimized structure of lysozyme are listed in Table IV, along with their deviations from the initial values computed directly from the X-ray coordinates. Table V provides a summary of the average deviations of the dihedral angles from the X-ray angles, as well as the average deviations from the angles of the previous stage of refinement. Although the angles change significantly from one stage to the next, the overall deviation from the X-ray values remains approximately constant. It is also of interest that, during stage III, the backbone dihedral angles are sufficiently free to move 6 or 7° per iteration on the average, in spite of the very strong fitting constraint on the C-terminal residue of each segment. The ϕ–ψ plot for the final backbone dihedral angles (Figure 3) bears a remarkable resemblance to the plot based on the X-ray data (Blake *et al.*, 1967). However, the plotted points are segregated into the low-energy regions to a greater extent in the final minimum-energy structure presented here. Nearly all residues other than glycine fall within the 10 kcal contours for an alanine dipeptide (Ponnuswamy *et al.*, 1973). The fact that many points lie in the area between the α-helix region and the β region confirms that 3_{10}-helical and other sharp-bend conformations are allowed conformations in proteins and are not just artifacts related to inaccuracies of the X-ray data (Lewis *et al.*, 1973a).

The Cartesian coordinates of the computed structure (stage III) of lysozyme have been deposited.[2]

Discussion

The atomic coordinates of proteins deduced from X-ray diffraction studies are subject to several types of uncertainties.

[2] Document No. NAPS 02265 with the ASIS National Auxiliary Publication Service, c/o Microfiche Publications, 305 East 46th Street, New York, N. Y. 10017. A copy may be secured by citing the document number and by remitting $1.50 for microfiche or $5.00 for photocopies. Advance payment is required. Make checks or money orders payable to Microfiche Publications.

TABLE IV: Third-Stage Refined Dihedral Angles of Lysozyme and Deviations from X-Ray Values.[a]

Residue	φ	ψ	X_1	X_2	X_3	X_4
1 LYS	-180(****)	142(24)	-171(6)	-174(-11)	174(-22)	176(****)
2 VAL	-110(-7)	96(-32)	-178(13)			
3 PHE	-62(10)	155(-11)	-88(-28)	88(8)		
4 GLY	-94(19)	168(3)				
5 ARG	-47(12)	-64(11)	-171(-4)	165(10)	73(6)	78(17)
6 CYS	-67(-36)	-60(-13)	-74(30)			
7 GLU	-30(46)	-73(-42)	-170(31)	175(-19)	-160(-27)	
8 LEU	-43(3)	-63(-10)	-178(2)	70(12)		
9 ALA	-33(35)	-63(-9)				
10 ALA	-50(13)	-62(-39)				
11 ALA	-39(7)	-70(-3)				
12 MET	-44(24)	-48(-23)	-80(-15)	-171(9)	174(****)	
13 LYS	-54(-13)	-47(27)	-147(5)	-90(80)	-178(0)	76(-115)
14 ARG	-59(-43)	-28(33)	-161(146)	-172(11)	74(172)	79(-89)
15 HIS	-78(2)	-28(-33)	-81(-7)	76(-16)		
16 GLY	95(9)	46(58)				
17 LEU	-102(-48)	-3(53)	-108(-43)	-169(0)		
18 ASP	-83(-10)	125(-14)	167(-22)	-132(****)		
19 ASN	64(-1)	49(52)	-79(-16)	128(-170)		
20 TYR	-108(-22)	110(-28)	-176(-10)	75(12)	-178(****)	
21 ARG	75(25)	17(-19)	-95(-3)	-54(6)	-170(-8)	84(-59)
22 GLY	87(20)	44(18)				
23 TYR	-142(-43)	118(-26)	-80(-3)	-73(-7)	1(****)	
24 SER	-59(12)	162(26)	147(-24)	177(****)		
25 LEU	-71(-34)	-29(3)	179(-51)	103(-104)		
26 GLY	-64(-19)	-24(13)				
27 ASN	-83(-1)	-13(37)	-93(14)	-62(41)		
28 TRP	-93(-29)	-31(12)	-69(-5)	92(-3)		
29 VAL	-77(8)	-40(1)	172(8)			
30 CYS	-71(11)	-25(-23)	-179(8)			
31 ALA	-73(8)	-44(8)				
32 ALA	-92(28)	32(55)				
33 LYS	-105(-31)	-56(7)	173(21)	54(-51)	-159(16)	175(5)
34 PHE	-105(-42)	28(57)	-60(20)	-60(-27)		
35 GLU	-110(-34)	-59(-18)	-161(-46)	71(72)	81(-66)	
36 SER	-123(-4)	-2(20)	67(-25)	-77(****)		
37 ASN	67(-6)	37(19)	-158(-34)	-151(179)		
38 PHE	62(-13)	35(36)	-76(-19)	145(-27)		
39 ASN	-122(-18)	124(18)	-165(-17)	-8(5)		
40 THR	-85(-24)	-24(-9)	49(-7)	161(****)		
41 GLN	-83(1)	-16(17)	-79(0)	-168(-7)	101(-171)	
42 ALA	-45(10)	135(-3)				
43 THR	-143(-13)	133(-21)	38(-24)	75(****)		
44 ASN	-139(10)	130(21)	-74(-10)	-74(34)		
45 ARG	-100(7)	140(1)	-165(-1)	-180(43)	-170(-21)	-180(-6)
46 ASN	-89(10)	-166(-22)	-77(-29)	-84(167)		
47 THR	-102(-7)	-14(3)	-49(-27)	162(****)		
48 ASP	-101(-3)	29(-11)	77(-2)	123(3)		
49 GLY	80(-21)	-54(22)				
50 SER	-38(-37)	140(26)	103(-37)	88(****)		
51 THR	-121(0)	177(13)	-40(23)	157(****)		
52 ASP	-129(15)	137(37)	-73(13)	163(-30)		
53 TYR	-142(-56)	167(26)	-54(-2)	-91(2)	-11(****)	
54 GLY	84(-7)	-179(6)				
55 ILE	-81(-30)	-29(4)	-167(-2)	146(57)		
56 LEU	-87(50)	-7(-33)	-97(46)	-32(-56)		
57 GLY	57(23)	59(4)	-75(-32)	-72(14)	-77(0)	
58 ILE	-75(-6)	134(6)	-61(-34)	129(-33)		
59 ASN	-68(7)	153(-13)	-163(9)	11(106)		
60 SER	-111(-1)	49(26)	60(9)	170(****)		
61 ARG	-109(-30)	-45(-27)	167(-11)	123(-47)	-174(-9)	-178(28)
62 TRP	-146(-11)	46(93)	33(60)	50(-31)		
63 TRP	-147(-59)	-63(-18)	-71(-24)	129(15)		
64 CYS	-126(16)	162(14)	59(17)			
65 ASN	-115(-19)	128(-24)	170(-11)	54(27)		
66 ASP	-99(63)	-35(-62)	38(-39)	-153(54)		
67 GLY	107(20)	1(10)				
68 ARG	-132(-7)	32(16)	76(9)	146(-17)	155(-8)	-175(14)
69 THR	-135(-20)	114(29)	-69(-2)	158(****)		
70 PRO	-69(-25)	154(180)				
71 GLY	53(117)	46(63)				
72 SER	-87(-30)	151(22)	-154(-77)	172(****)		
73 ARG	-135(-10)	95(-24)	-77(0)	174(****)	-179(****)	180(****)
74 ASN	-68(49)	92(-1)	-140(47)	-13(11)		
75 LEU	-86(-17)	-37(2)	-99(-23)	-51(14)		
76 CYS	-85(11)	-17(-48)	-75(-22)			
77 ASN	57(5)	61(20)	-145(0)	-62(-6)		
78 ILE	-160(-23)	142(-6)	-164(8)	88(20)		
79 PRO	-69(16)	137(5)				
80 CYS	-58(-42)	-30(35)	-61(14)			
81 SER	-62(-23)	-35(14)	64(-46)	179(****)		
82 ALA	-68(-31)	-13(36)				
83 LEU	-89(-39)	-16(12)	-59(2)	175(-4)		
84 LEU	-82(13)	-22(-13)	-67(-39)	170(-8)		
85 SER	-50(-23)	161(8)	106(2)	-74(****)		
86 SER	-78(12)	-12(-23)	174(****)	-72(****)		
87 ASP	-92(28)	119(10)	-158(-5)	153(32)		
88 ILE	-92(-27)	59(53)	59(-13)	139(-22)		
89 THR	-100(-55)	-58(21)	-84(-17)	70(****)		
90 ALA	-72(-71)	-27(58)				
91 SER	-68(-8)	-52(8)	-62(-4)	-179(****)		
92 VAL	-66(-10)	-41(-8)	155(11)			
93 ASN	-57(26)	-40(-21)	-91(18)	-8(-6)		
94 CYS	-74(1)	-33(21)	-167(-6)			
95 ALA	-78(-41)	8(61)				
96 LYS	-100(-70)	-21(20)	-78(-20)	-109(55)	177(-23)	81(-49)
97 LYS	-107(-9)	3(26)	-78(-12)	-180(-22)	174(46)	179(****)
98 ILE	-95(-7)	-34(-4)	-65(-4)	162(2)		
99 VAL	-88(-2)	-11(-3)	66(8)			
100 SER	-88(37)	-12(-40)	-78(-16)	75(****)		
101 ASP	-58(57)	-28(9)	159(****)	-90(****)		
102 GLY	100(-81)	23(32)				
103 ASP	-119(11)	8(7)	-112(-58)	146(51)		
104 GLY	64(16)	-149(-12)				
105 MET	-97(-11)	10(-2)	-69(-5)	-160(7)	67(-1)	
106 ASN	-66(12)	-23(10)	-72(23)	-70(-6)		
107 ALA	-45(4)	-37(-21)				
108 TRP	-82(20)	118(34)	-67(-11)	-88(12)		
109 VAL	-61(-27)	-42(10)	174(118)			
110 ALA	-55(12)	-73(-29)				
111 TRP	-33(9)	-71(-13)	-178(19)	84(-19)		
112 ARG	-39(17)	-61(8)	-160(18)	-168(5)	-55(5)	-178(****)
113 ASN	-49(18)	-75(-29)	-82(2)	158(5)		
114 ARG	-114(-9)	106(105)	-39(-9)	135(-65)	80(35)	-179(-11)
115 CYS	157(-63)	-58(-25)	-89(-34)			
116 LYS	-72(1)	129(10)	-174(1)	165(17)	85(-62)	172(-23)
117 GLY	88(5)	27(11)				
118 THR	-143(-33)	167(-10)	87(18)	-56(****)		
119 ASP	-84(17)	94(16)	62(8)	-50(9)		
120 VAL	-82(-28)	-7(29)	-59(8)			
121 GLN	-66(-22)	-23(-1)	-77(-34)	177(****)	-89(****)	
122 ALA	-88(-11)	7(22)				
123 TRP	-95(-31)	-22(37)	-90(-36)	117(8)		
124 ILE	-119(-36)	27(44)	53(-25)	123(-50)		
125 ARG	-89(-29)	133(9)	-168(****)	179(****)	180(****)	180(****)
126 GLY	89(-17)	4(15)				
127 CYS	-95(-18)	145(22)	-62(-24)			
128 ARG	-92(-26)	86(-14)	-170(****)	-179(****)	179(****)	180(****)
129 LEU	-109(18)	86(78)	-54(0)	158(-25)		

[a] Dihedral angles which were undefined by the X-ray coordinates are indicated by (****). The values in parentheses are the deviations from the dihedral angles computed from the X-ray coordinates.

TABLE V: Summary of Average Deviations[a] of Dihedral Angles from X-Ray Values and from Previous Stages of Refinement of Lysozyme.

Stage	Iteration	Average Deviation (deg)			
		Backbone Angles		Side-Chain Angles	
		Dev from X-ray	Dev from Stage N-1	Dev from X-ray	Dev from Stage N-1
I		20.2		22.5	
II	1	20.8	3.2	24.3	1.3
II	2	22.3	7.6	21.2	7.0
III	1	22.4	7.1	23.2	12.7
III	2	22.7	5.6	23.5	4.2

[a] The average deviation is computed as the sum of the differences between the computed dihedral angles and the comparison set of dihedral angles, divided by the number of angles compared. As a point of reference, a rotation of 10° produces a movement of 0.25 Å for an atom attached (at a distance of 1.5 Å) to a tetrahedral carbon atom situated at one end of the bond about which rotation takes place.

Firstly, the structure of the protein in the crystal may be perturbed by neighboring protein molecules, by high ionic strength of the crystallizing medium, or by the presence of the heavy atom in isomorphous derivatives. Secondly, there are frequently some ambiguities in the interpretation of the electron density map arising from disorder in the crystal or from general lack of definition of atomic positions. Since small groups of atoms are visible in the density map rather than individual atoms, some latitude is exercised in building the skeletal wire model. Finally, the act of measuring the coordinates of each atom from the model may introduce uncertainties. At a scale of 2 cm/Å, an uncertainty of only 2 mm in measurement of a coordinate corresponds to an uncertainty of 0.10 Å for that coordinate, thus introducing an uncertainty of about 0.17 Å in the position of that atom if its other two coordinates are measured to the same accuracy. Computer methods are now available for determining the positions of atoms more accurately by direct reference to the electron density map without the intervening step of building a physical model (Diamond, 1971). However, most of the atomic coordinates of proteins now available, including the coordinates of lysozyme used in the present work, were not determined by the latter method, but were measured from a physical model. Unfortunately, a precise indication of the accuracy of the reported coordinates of each atom is not generally available. Refinement techniques based on conformational energy calculations provide a means for alleviating some of the uncertainties inherent in the X-ray coordinates, and thus may produce a more accurate description of the molecule.

One of the basic assumptions of the present refinement technique is that the polypeptide chain may be represented in terms of fixed bond lengths and bond angles characteristic of crystal structures of small peptides and amino acids. The computational advantages of maintaining fixed geometry have already been described (Warme and Scheraga, 1973). Our approach to the refinement of X-ray coordinates of proteins is similar in some respects to that of Nishikawa *et al.* (1972). Their first stage of refinement, like ours, is designed to adjust a polypeptide chain to correspond as closely as possible to the X-ray structure, while maintaining fixed bond lengths, fixed bond angles, and planar trans peptide groups. They also introduce nonbonded energy calculations in a second stage of refinement (Nishikawa and Ooi, 1972), although, in contrast to our procedure, they omit consideration of all side-chain atoms (beyond C^β) and all hydrogen-bonding atoms. The best structure which they obtained for lysozyme exhibited an RMS deviation of 0.47 Å (for N, C^α, C^β, C′, and O only) and a nonbonded energy of −270 kcal/mol. Since the calculations presented in the present paper include hydrogen bond, electrostatic, rotational, and disulfide energies, in addition to nonbonded energy, and both backbone and side-chain atoms are included, a more exact comparison of results is not possible.

A discussion of the similarities and differences between our method and that of Levitt and Lifson (1969) has already been presented elsewhere (Warme and Scheraga, 1973). They take a somewhat different approach, in that they allow both bond lengths and bond angles to vary somewhat from their equilibrium values while minimizing the energy of the protein.

There is a danger in prematurely allowing freedom for deviation of bond lengths and bond angles during refinement of protein structures. If an atomic overlap must be relieved, this relief may be obtained in a variety of ways. One way to separate the overlapping atoms would be to allow some bond angles in the immediate vicinity to deviate from their equilibrium values. However, the overlap may just as well reflect a problem of larger scope, perhaps relating to some error in the orientation of an amino acid one or two residues away from the site of the overlap. By the present procedure, which maintains fixed bond lengths and bond angles, the likely course of action to eliminate the overlap would be to move amino acid residues on both sides of the overlap, as well as the amino acid directly involved in the overlap. Of course, if bond angles which deviate appreciably from their normal equilibrium values actually exist in proteins, the same argument leads to the conclusion that the local deformation will be transmitted to neighboring amino acid residues when fixed bond lengths and bond angles are maintained. This effect will be manifested by larger-than-average deviations of the refined coordinates from the X-ray coordinates or, alternatively, as a high local conformational energy. In the present approach, fixed geometry is maintained during the refinement in order to determine whether a low-energy conformation can be attained within these limitations. If it becomes necessary, flexible geometry can then be introduced in a subsequent stage of refinement (Niu *et al.*, 1973).

In summary, we have described a practical procedure for the complete energy refinement of X-ray structures of proteins, using recently improved empirical energy functions. The procedure has been applied to lysozyme, with the result that the final conformation at the plateau of the conformational energy surface is very similar to the X-ray structure, indicating that the energy functions used here are reliable. Since fixed bond lengths and bond angles were used in the computations, the close resemblance of the final low-energy structure to the X-ray structure of lysozyme would suggest that it is not necessary to introduce extensive stretching of bonds or bending of bond angles in order to describe the conformation of a protein within the limits of experimental uncertainties. Now that a minimum-energy conformation of lysozyme has been achieved (within the context of the energy functions and other constraints used in the calculations), a reference state is available to which the energies of other, slightly differing conformations of lysozyme can be compared.

The refined structure of lysozyme has been used as a model for comparison with a computed structure of α-lactalbumin, as described in the accompanying paper (Warme *et al.*, 1974).

Acknowledgments

We are indebted to Drs. F. A. Momany, P. K. Ponnuswamy, and D. Rasse for helpful discussions.

References

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1967), *Proc. Roy. Soc., Ser. B 167*, 365.

Davidon, W. C. (1959), AEC Research and Development Report, ANL-5990.

Diamond, R. (1971), *Acta Crystallogr., Sect. A 27*, 436.

Fletcher, R. (1970), *Comput. J. 13*, 317.

Fletcher, R. (1971), A Survey of Algorithms for Unconstrained Optimisation, IMA/NPL Symposium on Numerical Methods for Unconstrained Optimisation, Jan 1971.

Fletcher, R., and Powell, M. J. D. (1963), *Comput. J. 6*, 163.

Gibson, K. D., and Scheraga, H. A. (1967a), *Proc. Nat. Acad. Sci. U. S. 58*, 420.

Gibson, K. D., and Scheraga, H. A. (1967b), *Proc. Nat. Acad. Sci. U. S. 58*, 1317.

Hesselink, F. Th., Ooi, T., and Scheraga, H. A. (1973), *Macromolecules 6*, 541.

IUPAC-IUB Commission on Biochemical Nomenclature (1970), *Biochemistry 9*, 3471.

Lee, B., and Richards, F. M. (1971), *J. Mol. Biol. 55*, 379.

Levitt, M., and Lifson, S. (1969), *J. Mol. Biol. 46*, 269.

Lewis, P. N., Momany, F. A., and Scheraga, H. A. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 2293.

Lewis, P. N., Momany, F. A., and Scheraga, H. A. (1973a), *Biochim. Biophys. Acta 303*, 211.

Lewis, P. N., Momany, F. A., and Scheraga, H. A. (1973b), *Isr. J. Chem. 11*, 121.

McGuire, R. F., Momany, F. A., and Scheraga, H. A. (1972), *J. Phys. Chem. 76*, 375.

Momany, F. A., McGuire, R. F., and Scheraga, H. A. (1973), *J. Phys. Chem.*, to be submitted.

Nishikawa, K., and Ooi, T. (1972), *J. Phys. Soc. Jap. 32*, 1338.

Nishikawa, K., Ooi, T., Isogai, Y., and Saito, N. (1972), *J. Phys. Soc. Jap. 32*, 1331.

Niu, G. C. C., Go, N., and Scheraga, H. A. (1973), *Macromolecules 6*, 91.

Phillips, D. C. (1970), private communication.

Platzer, K. E. B., Momany, F. A., and Scheraga, H. A. (1972), *Int. J. Peptide Protein Res. 4*, 201.

Ponnuswamy, P. K., Warme, P. K., and Scheraga, H. A. (1973), *Proc. Nat. Acad. Sci. U. S. 70*, 830.

Rees, A. R., and Offord, R. E. (1972), *Biochem. J. 130*, 965.

Venkatachalam, C. M. (1968), *Biopolymers 6*, 1425.

Warme, P. K., Go, N., and Scheraga, H. A. (1972), *J. Comput. Phys. 9*, 303.

Warme, P. K., Momany, F. A., Rumball, S. S., Tuttle, R. W., and Scheraga, H. A. (1974), *Biochemistry 13*, 768.

Warme, P. K., and Scheraga, H. A. (1973), *J. Comput. Phys. 12*, 49.