*J. Mol. Biol.* **(1969) 46, 269-279**

# Refinement of Protein Conformations using a Macromolecular Energy Minimization Procedure

**MICHAEL LEVITT AND SHNEIOR LIFSON**

*Weixmann Institute of Science*
*Rehovot, Israel*

*(Received 3 December 1968, and in revised form 29 July, 1969)*

**This paper presents a rapid refinement procedure capable of deriving the stable conformation of a macromolecule from experimental model co-ordinates. All the degrees of freedom of the molecule are allowed to vary and all parts of the structure are refined simultaneously in a general force-field.**

**The procedure has been applied to myoglobin and lysozyme. The deviations of peptide bonds from planar conformation and of various bond angles from their respective average values are found to contribute significantly to the retied protein conformation. Hydrogen atoms are not included in the present refinement.**

**A set of non-bonded potential functions, applicable to the equilibrium of a folded protein in an aqueous medium, is described and tested on myoglobin.**

## 1. Introduction

The Cartesian co-ordinates of a protein molecule have been obtained from measurements on a rigid wire model, built according to electron density maps derived from X-ray diffraction measurements. The errors inherent in measuring a mechanical model give rise to highly strained bond lengths and angles. Much of this strain can be relieved without affecting the relative orientation of parts of the molecule.

Diamond (1966) proposed a co-ordinate refinement procedure which varied certain dihedral and bond angles to give the best fit to the rough measured co-ordinates. The fixed bond lengths and angles of each amino acid residue were obtained from crystallographic studies of small molecules. This method has been extensively used to refine the co-ordinates of several protein structures determined by X-ray methods. (Perutz, Muirhead, Cox & Goaman, 1968; Watson, unpublished data; Blake, Mair, North, Phillips & Sarma, 1967).

Diffraction studies of small molecules indicate that all bond lengths and angles are close to their average values; however, some variation must be allowed, since bonds can stretch and angles bend if these distortions are energetically favourable. The stable conformation of a protein has its potential energy function at a minimum with respect to all atomic co-ordinates. Starting from rough model co-ordinates, the method described here minimizes the assumed total potential energy of the protein molecule, to give a refined conformation.

## 2. Methods

### *(a) Specifying the chemical structure of a macromolecule*

The functional form of the potential energy of a general molecule depends on the topological relationship of its constituent atoms. A protein has several thousand atoms and

† Present address: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, England,

it is important for computing purposes that the structure be conveniently and efficiently specified. The present procedure uses the common chemical notation, with additional symbols to describe the atomic connectivity uniquely. Each atom is represented by a single letter; different letters are used to distinguish similar types of atoms in different bonding environments (Table **1).** Parentheses enclose atoms in side-chains; ring closure, and S-S bridge formation, are indicated by symbols preceding each of the two atoms which effect the closure (several special symbols, different from those indicating atoms, are used

**TABLE 1**

*Definition* ***of symbols for atoms***

| Symbol | Atom and bonding environment | |
|---|---|---|
| O | Oxygen | —O— |
| N | Nitrogen | - N |
| C | Saturated carbon | - C - |
| A | Carbonyl carbon | A - O (I) |
| B | Resonant carbon | - B |
| S | Sulphur | —S— |

(NC(CCCC)AONC(C(C)C)AONC(C*BBBBB*B)AONCAONC(CCCNB(N)N)AONC(C4S)AONC(CCAOO)AONC(C
C(C)C)AONC(C)AONC(C)AONC(C)AONC(CC.S)AONC(CCCCN)AONC(CCCNB(N)N)AONC(C*BNBN*B)AON
CAONC(CC(C)C)AONC(CC)AONC(CAON)AONC(C*BBBBOB*B)AONC(CCCNB(N)N)AONCAONC(C*BBBBOB*
B)AONC(CO)AONC(CC(C)C)AONCAONC(CAON)AONC(C*BBN$BBBBB$*B)AONC(C(C)C)AONC(C5S)AONC
(C)AONC(C)AONC(CCCCN)AONC(C*BBBBB*B)AONC(CCAOO)AONC(CO)AONC(CAON)AONC(C*BBBBB*B)
AONC(CAON)AONC(AOC)AONC(CCAON)AONC(C)AONC(AOC)AONC(CAON)AONC(CCCNB(N)N)AONC(CAON
)AONC(AOC)AONC(CAOO)AONCAONC(CO)AONC(AOC)AONC(CAOO)AONC(C*BBBBOB*B)AONCAONC(C(C)
CC)AONC(CC(C)C)AONC(CCAON)AONC(C(C)CC)AONC(CAON)AONC(CO)AONC(CCCNB(N)N)AONC(C*BB
NB*B)AONC(C*BBN$BBBBB$*B)AONC(C6S)AONC(CAON)AONC(CAOO)AONCAONC(CCCNB(N)N)AONC(AO
C)AC*NC(CC*C)AONCAONC(CO)AONC(CC)AONC(CAON)AONC(CC(C)C)AONC(C7S)AONC(CAON)AONC(C
(C)CC)AO*NC(CC*C)AONC(C6S)AONC(CO)AONC(C)AONC(CC(C)C)AONC(CC(C)C)AONC(CO)AONC(C)
AONC(CAOO)AONC(C(C)CC)AONC(AOC)AONC(C)AONC(CO)AONC(C(C)C)AONC(CAON)AONC(C7S)AONC
(c) AONC(CCCCN)AONC(CCCC)AONC(C(C)CC)AONC(C(C)C)AONC(CO)AONC(C)AONCAONC(CAOO)AONC
AONC(CC.3C)AONC(CAON)AONC(C)AONC(C*BBN$BBBBB$*B)AONC(CC)AONC(C)AONC(C*BBN$BBBBB$
*B)AONC(CCCN)AONC(CAON)AONC(CCCNB(N)N)AONC(C5S)AONC(CCCCN)AONCAONC(AOC)AONC(CAOO
)AONC(C(C)C)AONC(CC)AONC(C)AONC(C*BBN$BBBBB$*B)AONC(C(C)CC)AONC(C)AONCAONC(C4S)A
ONC(C)AONC(CC(C)C)AOO)

**FIG.** 1. The complete chemical formula of lysozyme. The symbols * and $ close cyclic side chains; **4, 5, 6** and **7** precede **the pairs of sulphur atoms** which **form S-S bridges.**

```
-LY1-VAL-PHE-GLY-ARG4CIS-GLU-LEU-ALA-ALA-ALA-MT1-LYS-ARG-HIS-GLY-LEU-AL1-ASN-TYR
-ARG-GLY-TYR-SER-LEU-GLY-ASN-TRY-VAL5CIS-ALA-ALA-LYS-PHE-GLU-SER-ASN-PHE-ASN-THR
-GLN-ALA-THR-ASN-ARG-ASN-THR-ASP-GLY-SER-THR-ASP-TYR-GLY-ILE-LEU-GLN-ILE-ASN-SER
-ARG-TR1-TRY6CIS-ASN-ASP-GLY-ARG-THR-PRO-GLY-SER-AL1-ASN-LEU7CIS-ASN-ILE-PRO6CIS
-SER-ALA-LEU-LEU-SER-ALA-ASP-ILE-THR-ALA-SER-VAL-ASN7CIS-ALA-LYS-LY1-ILE-VAL-SER
-ALA-GLY-ASP-GLY-MET-ASN-ALA-TRY-AL1-ALA-TRY-AR1-ASN-ARG5CIS-LYS-GLY-THR-ASP-VAL
-AL1-ALA-TRY-ILE-ALA-GLY4CIS-ALA-LEU
```

(a)

| | | | |
|---|---|---|---|
| GLY | NCAO | ALA | NC(C)AO |
| VAL | NC(C(C)C)AO | ILE | NC(C(C)CC)AO |
| LEU | NC(CC(C)C)AO | SER | NC(CO)AO |
| THR | NC(AOC)AO | PRO | *NC(CC*C)AO |
| CIS | NC(C=S)AO | MET | NC(CC.SC)AO |
| LYS | NC(CCCCN)AO | ARG | NC(CCCNB(N)N)AO |
| ASP | NC(CAOO)AO | ASN | NC(CAON)AO |
| GLU | NC(CCAOO)AO | GLN | NC(CCAON)AO |
| HIS | NC(C*BNBN*B)AO | PHE | NC(C*BBBBB*B)AO |
| TYR | NC(C*BBBBOB*B)AO | TRY | NC(C*BBN⁄BBBBB⁄*B)AO |
| AL1 | NC(CC)AO | LY1 | NC(CCCC)AO |
| TR1 | NC(C*BBNB*B)AO | MT1 | NC(CC.S)AO |
| AR1 | NC(CCCN)AO | | |

(b)

**FIG. 2. (a) Abbreviated chemical formula of lysozyme. (b) The amino acid library used to translate the abbreviated chemical formula.**

for this purpose). Figure 1 shows the full chemical formula of lysozyme using this symbolism. It should be noted that this method is efficient; an n-atomic molecule is structurally specified by less than $1{\cdot}5n$ characters.

Molecules with repeating chemical units may be specified still more compactly. Thus a protein can be represented by a sequence of the common three-letter abbreviations for various amino acid residues, and a, library specifying the structure of each amino acid is used to prepare automatically the full symbolic formula, which conveys the complete structural information to the computer (see Fig. 2).

### (b) *The molecular force-field*

The total molecular potential energy is composed of many terms. Its exact form is unknown and the expression for potential energy used in the present calculations is a gross approximation given by:

$$E = \underset{\text{all bonds}}{\Sigma \tfrac{1}{2} K_b (b - b_0)^2} + \underset{\text{all bond angles}}{\Sigma \tfrac{1}{2} K_\tau (\tau - \tau_0)^2} + \underset{\text{all dihedral angles}}{\Sigma \tfrac{1}{2} K_\theta \{1 + \cos(n\theta - \delta)\}} + \underset{\text{all non-bonded pairs}}{\Sigma \epsilon_{ij} \{(r^0_{ij}/r_{ij})^{12} - 2(r^0_{ij}/r_{ij})^6\}} + \underset{\text{all atomic co-ordinates}}{\Sigma \tfrac{1}{2} w (x_i - x_i^0)^2} \quad (1)$$

where $K_b$ is the bond force constant ; b, bond length; $b_0$, equilibrium bond length ; $K_\tau$, bond-angle bending-force constant ; $\tau$, bond angle; $\tau_0$, equilibrium bond angle ; $K_\theta$, torsional barrier; 8, dihedral angle (zero for cis-conformation) ; n, periodicity of rotational function; $\delta$, phase; $\epsilon_{ij}$, depth of non-bonded minimum ; $r_{ij}$, distance between atoms $i$ and

j; $r^0_{ij}$, distance of non-bonded minimum; w, constraining force for all atoms; $x_i$, atomic Cartesian co-ordinate and $x^0_i$, experimental co-ordinate.

The first two terms of equation (1) are harmonic potentials. Torsional barriers are introduced by the third term, and non-bonded interactions are computed using the fourth term, a Lennard-Jones potential function. The fifth term, a constraint which ensures that the final minimized conformation remains close to the initial experimental conformation compensates for inadequacies of the other terms in representing the true molecular potential energy. Tables 2, 3 and 4 list all the force-field parameters used.

### *(c) The minimization method*

The molecular potential energy is minimized using the method of steepest descent (Bixon & Lifson, 1967). The ith co-ordinate $x_i$ is shifted iteratively by $\Delta x_i$ given by

$$AX_i = -k\partial E/\partial x_i \tag{2}$$

where $E$ is the total energy given by equation (1) and $k$ is a constant which depends on the step length $L$.

$$(L)^2 = \sum_i (\Delta x_i)^2 = k^2 \sum_i (\partial E/\partial x_i)^2. \tag{3}$$

The step length, L, is chosen using a regressional strategem: if the jth value of E, the total energy, is greater than the *j-1* th value of $E$, the *j* th step length is made less than the *j-1* th otherwise the jth step is made greater than the *j-1* th. By this method one moves down the energy hypersurface at an increasing pace, but slows down on reaching an uphill section. The step length $L$ is analogous to the momentum of a ball rolling on a curved surface, hence the oscillatory convergence of the total molecular potential energy (see Figs *3* and 4). The first derivative vector of the total energy, $\partial E/\partial x_i$, is calculated analytically from equation (1).

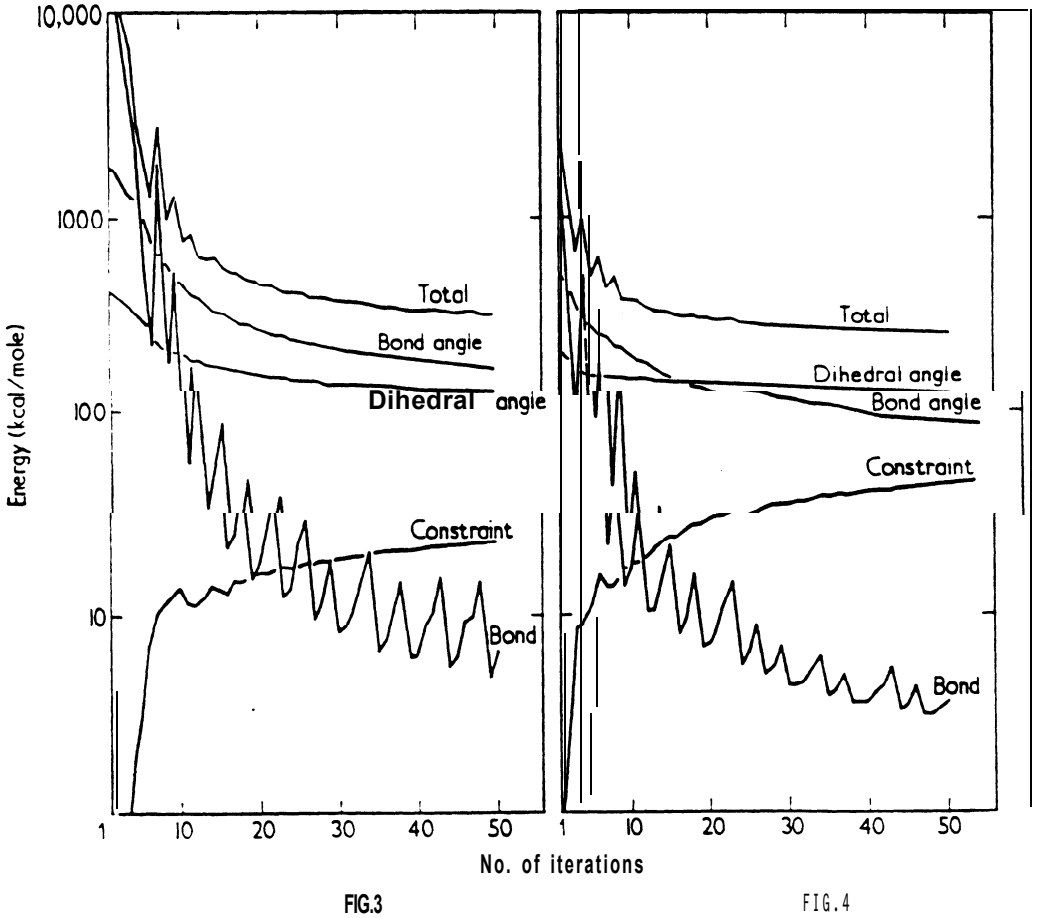


FIG.3 FIG.4

**FIG. 3. Iteration history of lysozyme.**

**FIG. 4. Iteration history of myoglobin (the constraint contribution is multiplied by ten).**

This method is not quadratically convergent but is reasonably efficient when far from the minimum. The final refinment is judged by average values and standard deviation of bond lengths and angles and by deviation from experimental co-ordinates ; and in practice final root mean square energy gradients of a few kilocalories per mole per Ångstrom unit can be obtained. The use of quadratically convergent methods (Pearson, 1969) would be impracticable with thousands of variables, since a matrix with several million elements would have to be manipulated and stored.

### 3. Results

The present method has been applied to the rough, measured co-ordinates lysozyme (Blake, Mair North, Phillips & Sarma, 1965 personal communication) and to the refined co-ordinates of myoglobin (Kendrew *et al.* 1960; Watson, unpublished results) the co-ordinates of myoglobin had been previously refined using the procedure of Diamond (1966) and are given to within 0.1Å. The atoms of those side chains of lysozyme which were not definitively located from the electron density map are not included. Non-bonded forces were not used in this refinement.

The potential function parameters in Tables 2, 3 and 4 were selected as follows. The bond-stretching force constants are approximately those suggested by Ramachandran & Sasisekharan (1968). Angle-bending force constants are all set to a nominal value of kcal/mole-radian except for bending about a sulphur atom which is assigned a value of 100 kcal/mole-radian . The equilibrium values of bond lengths and bond angles are selected from average values of the corresponding internal co-ordinates in the initial conformation of myoglobin, which had been refined from stereochemically correct residues by Diamond's method (1966).

Torsional parameters are taken from Ramachandran & Sasisekharan (1968). The torsional potentials about resonant bonds are represented by twofold periodic functions, closely representing harmonic potentials near the planar positions.

The constraining force constant $w$, which ensures an agreement between the refined and measured conformations, is varied over a wide range to find the most suitable value (see Table 6). A disadvantage of a large value of $w$ is that bond- and dihedral-angles cannot relax sufficiently; if on the other hand $w$ is too small, the defects of

TABLE 2

*Bond parameters* $E = \frac{1}{2}K_b(b - b_0)^2$

| Atoms forming bond | $\frac{1}{2}K_b$ (kcal./mole-Å²) | $b_0$ (Å) |
|---|---|---|
| CO | 300 | 1·45 |
| NC | 300 | 1·47 |
| CC | 300 | 1·53 |
| AO | 700 | 1·25 |
| AN | 550 | 1·32 |
| CA | 300 | 1·53 |
| BO | 300 | 1·33 |
| BN | 300 | 1·34 |
| BC | 300 | 1·52 |
| BB | 300 | 1·40 |
| CS | 400 | 1·76 |
| SS | 600 | 2·10 |

TABLE 3

*Bond-angle parameters* $E = \frac{1}{2}K_\tau(\tau - \tau_0)^2$

| Atoms defining bond angle | $\frac{1}{2}K_\tau$ (kcal./mole-radian²) | $\tau_0$ (degrees) |
|---|---|---|
| CNC | 30 | 112·0 |
| ANC | 30 | 123·0 |
| BNC | 30 | 122·0 |
| BNB | 30 | 108·0 |
| CCO | 30 | 111·8 |
| CCN | 30 | 110·9 |
| ccc | 30 | 112·4 |
| ACN | 30 | 112·4 |
| ACC | 30 | 113·3 |
| ACA | 30 | 111·6 |
| BCC | 30 | 11398 |
| SCC | 30 | 112·7 |
| OAO | 30 | 126·7 |
| NAO | 30 | 124·6 |
| CAO | 30 | 119·4 |
| CAN | 30 | 114·6 |
| CAC | 30 | 104·5 |
| NBN | 30 | 114·7 |
| CBN | 30 | 131·6 |
| BBO | 30 | 119·9 |
| BBN | 30 | 108·4 |
| BBC | 30 | 121·9 |
| BBB | 30 | 119·5 |
| csc | 50 | 98-7 |
| ssc | 50 | 104·0 |

TABLE 4

***Dihedral angle parameters*** $E = \frac{1}{2}K_\theta\{1 + \cos(n\theta - \delta)\}$

| Bond which is twisted | We (kcal./mole) | $n$ | $\delta$ (degrees) |
|---|---|---|---|
| N-C | 0·6 | 3 | 0 |
| C–C | 0·7 | 3 | 0 |
| A-N | 10·0 | 2 | 180 |
| G A | 0·7 | 3 | 0 |
| B-N | 20·0 | 2 | 180 |
| G B | 0·6 | 6 | 0 |
| B-B | 20·0 | 2 | 180 |
| s-c | 1·0 | 3 | 0 |
| S-S | 6·0 | 2 | 0 |

the approximate force-field would not be corrected. No constraining force would be necessary if the force-field used was that experienced by the protein molecule. After a set number of iterations, more constrained conformations are closer to the minimum. A constraint of 10 kcal./mole-Å² was finally chosen for these calculations.

The convergence method used causes the total molecular potential energy to approach the minimum in an oscillatory manner (Figs 3 and 4). Since bond stretching produces the greatest change in energy for a given displacement, the oscillations occur

TABLE 5

*The effect of the constraining weight, w*

| Weight, *w* (kcal./mole) | No. of iterations | R.M.S. co-ordinate deviation (Å) | Torsional energy (kcal./mole) | Bond-angle energy (kcal./mole) | Final R.M.S. energy gradient (kcal./mole-Å) |
|---|---|---|---|---|---|
| 100·0 | 50 | 0·453 | 194 | 399 | 0·2 |
| 10·0 | 50 | 0·205 | 137 | 166 | 0·7 |
| 1·0 | 50 | 0·213 | 130 | 147 | 2·8 |
| 0·1 | 50 | 0·215 | 129 | 145 | 2·9 |
| 10·0 | 150 | 0·217 | 118 | 133 | 0·4 |
| 0·1 | 160 | 0·255 | 103 | 108 | 1·0 |

mainly in the bond-stretching contributions to the total energy; bond-angle and dihedral-angle energies decrease monotonically. The myoglobin co-ordinates which were used had been previously refined using the method of Diamond (1966). The iteration history (Fig. 4), therefore, shows a smaller constraint contribution and lower initial total energy than that obtained for the rough co-ordinates of lysozyme (Fig. 3).

The set of bond lengths refines after about ten iterations, while bond and dihedral angles refine more gradually. In the refined conformation, bond angles of *a* given type have a spread of up to three degrees about the average value (see Tables 6 and 7).

**TABLE 6**

*Bond length refinement (lengths in Å)*

| Type of bond | Lysozyme | | Myoglobin | |
|---|---|---|---|---|
| | Average initially | Average finally | Average initially | Average finally |
| C O | 1·49±0·15 | 1·45±·003 | 1·42±·06 | 1·45±·001 |
| NC | 1·47±0·16 | 1·47±·006 | 1·47±·06 | 1·47±·003 |
| CC | 1·54±0·19 | 1·53±·005 | 1·53±·04 | 1·53±·002 |
| AO | 1·21±0·2 | 1·25±·002 | 1·25±·07 | 1·25±·002 |
| AN | 1·32±0·21 | 1·32±·003 | 1·32±·05 | 1·32±·002 |
| CA | 1·54±0·18 | 1·53±·006 | 1·53±·05 | 1·53±·004 |
| BO | 1·52±0·06 | 1·33±·002 | 1·33±·03 | 1·33±·001 |
| BN | 1·37±0·23 | 1·34±·008 | 1·34±·05 | 1·34±·004 |
| BC | 1·45±0·21 | 1·52±·006 | 1·51±·05 | 1·52±·004 |
| BB | 1941 ±0·17 | 1·40±·009 | 1·40±·05 | 1·40±·005 |
| cs | 1·81±0·16 | 1·76±·003 | 1·76±·05 | 1·76±·004 |
| ss | 2·27±0·28 | 2·10±·003 | — | — |

The bond-angle bending energy is commensurate with that of the dihedral twisting energy (Figs 3 and 4). The assumption that angles are fixed at some average value would therefore, lead to a considerably higher dihedral energy and consequently to a distorted local conformation. Bond lengths are much less scattered from the average values and could possibly be treated as fixed. The final, refined bond lengths are approximately equal to the equilibrium bond lengths of the respective energy functions, and if one assumed slightly different equilibrium lengths the average bond lengths of the refined conformation would change correspondingly. Initially, the worst bond-length and bond-angle distortions of lysozyme are 0.65 Å and 55", respec-

TABLE 7

*Bond-angle refinement*
*(angles in degrees)*

| Type of angle | Lysozyme | | Myoglobin | |
|---|---|---|---|---|
| | Average initially | Average finally | Average initially | Average finally |
| CNC | 110·6±·3 | 110·6±0·3 | 112·0±3·2 | 110·9±0·2 |
| ANC | 124·3±10·4 | 122·7±2·8 | 123·0±2·9 | 122·9±1·7 |
| BNC | 118·7±6·9 | 120·4±1·7 | 122·6±2·6 | 121·7±0·4 |
| BNB | 109·7±8·7 | 108·2±1·6 | 108·2±2·7 | 106·4±0·5 |
| CCO | 111·7±11·8 | 110·8±2·6 | 111·8±2·0 | 112·0±1·6 |
| CCN | 111·7±11·7 | 110·8±3·2 | 110·9±9·6 | 111·0±3·3 |
| ccc | 112·6±13·5 | 111·7±3·3 | 112·3±4·3 | 112·2±2·0 |
| ACN | 111·4±11·3 | 112·8±3·6 | 112·4±6·6 | 112·2±3·0 |
| ACC | 109·7±11·2 | 112·7±3·0 | 113·3±9·2 | 113·2±3·3 |
| ACA | 108·7±5·6 | 112·2±2·4 | 111·5±10·5 | 113·5±2·4 |
| BCC | 112·9±10·0 | 113·6±4·0 | 113·8±2·7 | 114·2±1·7 |
| SCC | 114·0±9·2 | 114·5±3·8 | 112·7±3·3 | 112·4±0·5 |
| OAO | 110·6±14·2 | 124·0±1·0 | 126-7±6·3 | 124·4±0·5 |
| NAO | 120·7±14·4 | 124·0±1·9 | 124·6±3·1 | 124·6±0·9 |
| CAO | 119·1±12·6 | 118·6±1·8 | 119·3±5·0 | 119·0±1·2 |
| CAN | 116·8±11·0 | 115·3±2·7 | 114·6±3·2 | 114·7±1·4 |
| CAC | 111·2±10·5 | 106·0±2·3 | 104·5±0·9 | 102·5±0·3 |
| NBN | 119·2±15·0 | 117·2±2·4 | 114·8±5·8 | 115·5±3·6 |
| CBN | 118·0±0·0 | 130·4±0·0 | 131·6±3·4 | 131·3±1·3 |
| BBO | 121·8±9·5 | 120·0±1·0 | 119·9±3·0 | 120·2±0·4 |
| BBN | 114·0±14·4 | 112·8±7·7 | 108·4±5·3 | 109·0±4·3 |
| BBC | 122·7±11·7 | 123·1±3·5 | 121·9±3·5 | 121·4±2·4 |
| BBB | 118·8±10·8 | 118·9±6·0 | 119·5±5·2 | 119·6±4·0 |
| csc | 112·9±0·0 | 99·2±0·0 | 98·7±4·8 | 98·9±0·2 |
| ssc | 103·5±4·5 | 105·9±1·9 | — | — |

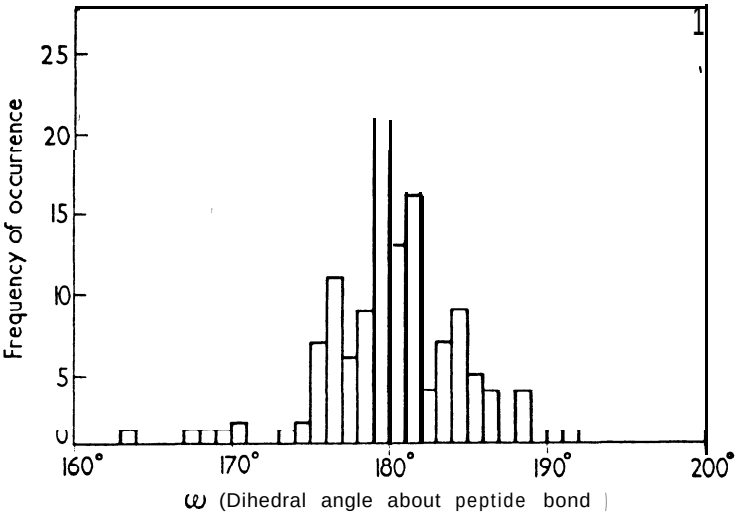


FIG. 6. Distribution of the peptide-bond dihedral angles of lysozyme after refinement.

tively, while the final values are 0.03 A and 8, respectively. Earlier treatments assumed that the peptide group is planar; in the present method, a twofold periodic function with a barrier height of 20 kcal./mole is used, and this gives a spread of peptide bond dihedral angles (see Fig. 5).

The final root mean square deviation of the refined lysozyme co-ordinates relative to the rough initial co-ordinates is 0.22 A. In myoglobin the root meansquare deviation is 0.086 A, which is in agreement with the uncertainty introduced by using refined co-ordinates published to 0.1 A.

The calculation described above uses a constraint (the last term in equation (1)) to compensate for the inadequacies of the long-range part of the force-field and to ensure convergence to a reasonable conformation close to that observed experimentally. Using a suitable constraining force, rough measured co-ordinates can be refined without taking account of any non-bonded or hydrogen-bond interactions.

An attempt was next made to find the stable conformation in a more realistic force-field including non-bonded and hydrogen-bond forces and eliminating the constraint. A preliminary set of non-bonded and hydrogen-bond energy function parameters, describing the equilibrium of a folded protein in an aqueous medium, was selected by the criterion that the refined conformation should converge rapidly to a stereochemically reasonable conformation close to the experimental conformation. The initial values of these parameters were selected using the usual accepted values of van der Waals contacts, hydrogen-bond lengths and interaction energies. Hydrogen atoms are not included explicitly in this refinement but the energy functions are designed to take account of the effect of these atoms in a simple way. The non-bonded and hydrogen-bond interactions are calculated between atoms not covalently linked to each other or to a common atom. using a modified Lennard-Jones potential function

$$E_{ij} = \epsilon \{ (R_0/R_{ij})^{12} - 2\lambda(R_0/R_{ij})^6 \}$$

where $R_{ij}$ is the distance between atoms $i$ & j, and $\lambda = 1$ if the interaction function has a minimum and $\lambda = 0$ if it does not. The chosen values of the parameters $\epsilon$, $R_0$ and $\lambda$ are given in Table 8 and the shapes of some of the functions are shown in Figure 6. The symbol V (Table 8) is used to designate an OH group which is both a hydrogen-bond donor and acceptor. The potentials are strongly attractive between non-polar, hydrophobic groups and those polar groups which can form hydrogen bonds. Other interactions between polar groups unable to form hydrogen bonds, and between polar and non-polar groups, are made entirely repulsive. These interactions will form a hydrophobic core and make polar groups point out into the medium; this is observed in all protein structures determined by X-ray crystallography. These functions differ significantly from those used by Scott & Scheraga (1966) but are suitable for a folded protein, without hydrogen atoms, in a polar solvent.

Myoglobin has been re-examined in greater detail, using these new non-bonded and hydrogen-bond potential functions. The final average values and standard deviations of bond lengths and bond angles are similar to those presented in Tables 6 and 7. The final root mean square deviation from the inital co-ordinates is 0.15 Å. This is greater than the value obtained in the fist refinement described above (0.086 Å) because atoms move to satisfy the requirements of the long-range part of the force-field. Although hydrogen bonds are calculated using a spherically symmetrical interaction between **acceptor and donor** atoms, all the hydrogen bonds of the myoglobin molecule

TABLE 8

*Non-bonded and hydrogen bond parameters* $E = \epsilon\{ (R_0/R)^{12} - 2\lambda (R_0/R)^6\}$

| Interacting atoms | $\epsilon$ (kcal./mole) | $R_0$ (Å) | $\lambda$ |
|---|---|---|---|
| O...O | 0·233 | 3·04 | 0 |
| V...O | 2·5 | 2·8 | 1 |
| V...V | 0·233 | 3·04 | 0 |
| N...O | 4·5 | 2·8 | 1 |
| N...V | 2·5 | 2·8 | 1 |
| N...N | 0·205 | 3·3 | 0 |
| C...O | 0·165 | 3·22 | 0 |
| C...V | 0·165 | 3·22 | 0 |
| C...N | 0·155 | 3·25 | 0 |
| C...C | 2·0 | 3·8 | 1 |
| A...O | 0·15 | 3·1 | 0 |
| A...V | 0·15 | 3·1 | 1 |
| A...N | 0·15 | 3·18 | 0 |
| A...C | 1·0 | 3·7 | 1 |
| A...A | 1·0 | 3·4 | 0 |
| s...o | 0·1 | 3·4 | 1 |
| s...v | 0·5 | 2·8 | 1 |
| S...N | 0·5 | 2·8 | 1 |
| s...c | 0·1 | 3·8 | 1 |
| S...A | 0·1 | 3·65 | 1 |
| s...s | 0·1 | 3·6 | 0 |

Atom types A and B (see Table 1) use the same non-bond parameters.

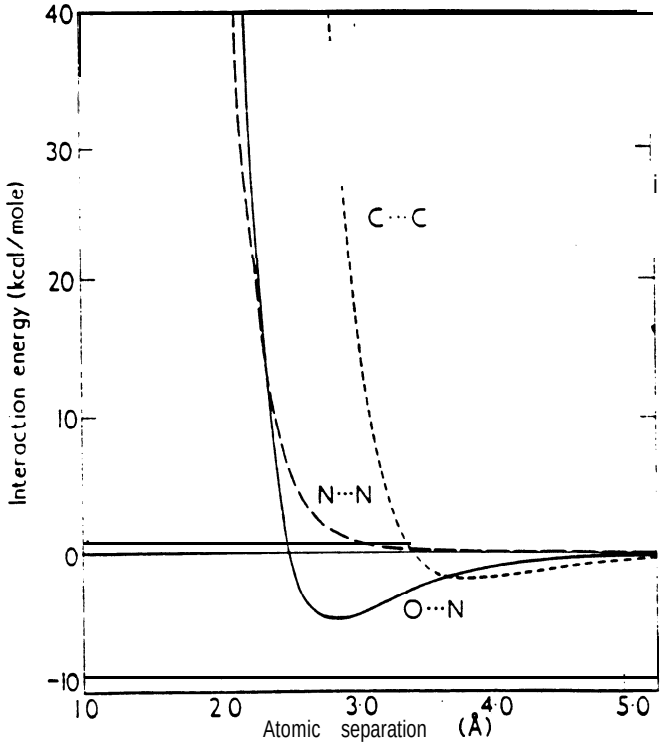


FIG. 6. Shape of some of the non-bonded potentials used.

are maintained with lengths between 2*65 and 3.0 Å and angles in the acceptable ranges. All non-bonded contact distances between atoms not forming a hydrogen bond are greater than the minimum van der Waals contact distances (Ramachandran & Sasisekharan, 1968).

These results indicate that the empirical energy expression used in these calculations has a minimum energy conformation close to the observed conformation of myoglobin. The energy function parameters used in the present study are preliminary and subject to further improvement, and indeed it should be possible to derive reliable parameters from the conformations of proteins determined by X-ray crystallographv. These potential functions may be useful in examining changes in known protein conformations caused by substrate binding or ammo acid substitution.

Fifty steepest descent iterations in the refinement of lysozyme (964 atoms), without non-bonded forces, required 18 minutes of computing time on *a* Golem computer (access time, 2 μsec ; multiply time, 12 μsec). Myoglobin, a larger molecule, required proportionately more computer time. The inclusion of non-bonded interactions, to a maximum interatomic distance of 6 A, caused a tenfold increase in computing time per iteration.

One of us (M.L.) was supported by a Royal Society Fellowship while in Israel.

## REFERENCES

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

**Bixon, M. & Lifson,** S. **(1967). Tetrahedron, 23, 769.**

Diamond, R. (1966). *Acta Cryst. 21, 253.*

**Kendrew,** J. C., Dickerson, R. **E., Strandberg, B. E., Hart, R. G., Davies.** D. **R., Phillips,** D. *C.* **& Shore, V.** *C. (1960). Nature,* **185, 422.**

**Pearson,** J. D. **(1969).** *Computer Journal,* **12, 171.**

**Perutz, M.** F., **Muirhead,** H., Cox, **J. M.** & **Goaman, L. C. G. (1968).** *Nature,* **219,** 131.

**Ramachandran,** G. N. & **Sasisekharan,** V. **(1968). In** *Advances in Protein Chemistry,* **23.** New York : Academic Press.

Scott, R. **A.** & Scheraga, H. A. ( **1966).** *J. Chem. Phys.* **45,** 2091.