

position. This strategy was employed for the straightforward structure solution of 3-chloro-1,3,4-triphenylazetid-2-one (AZET) (Colens, Declercq, Germain, Putzeys & Van Meerssche, 1974) with two Cl atoms at arbitrary starting positions (for example, both placed at the origin). For such a search, a considerable amount of computing time may be saved by specifying a proper value for the allowed intermolecular distance (e.g. for metal complexes); otherwise the distance tests could be rather ineffective. It may also be useful to calculate CFOM without TPRSUM (a *PATSEE* option) in order to give increased weight to TFOM, which is very reliable in such cases.

The values for the figures of merit show that RFOM, TFOM and  $R_E$  together are strongly indicative of the correct solution. TPRSUM is often only a local maximum but it enables the rapid location of the search fragment. All test examples (not only the five discussed here) confirm that *PATSEE* is reliable and widely applicable. In terms of computing times, it is also competitive with direct methods; under favourable circumstances (see SUOA and MUNICH1) it can even prove more economical. In any case, *PATSEE* offers a powerful alternative if chemical information is available.

We thank the Fonds der Chemischen Industrie for a Liebig-Stipendium (to EE) and the Deutsche Forschungsgemeinschaft for financial support. One of us (EE) is very much indebted to Professor Jack D.

Dunitz for his enthusiastic interest and hospitality during a three-months stay in Zürich.

#### References

- BLOW, D. M. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 229-238. Copenhagen: Munksgaard.
- BRAUN, P. B., HORNSTRA, J. & LEENHOUTS, J. I. (1969). *Philips Res. Rep.* **42**, 85-118.
- COLENS, A., DECLERCQ, J.-P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119-122.
- DOESBURG, H. M. & BEURSKENS, P. T. (1983). *Acta Cryst.* **A39**, 368-376.
- EGERT, E. (1983). *Acta Cryst.* **A39**, 936-940.
- EGERT, E., CRUSE, W. B. T. & KENNARD, O. (1983). *Acta Cryst.* **C39**, 95-99.
- HANSON, J. C., SIEKER, L. C. & JENSEN, L. H. (1973). *Acta Cryst.* **B29**, 797-808.
- HIRSHFELD, F. L. (1968). *Acta Cryst.* **A24**, 301-311.
- HOEKSTRA, A., VOS, A., BRAUN, P. B. & HORNSTRA, J. (1975). *Acta Cryst.* **B31**, 1708-1715.
- HORNSTRA, J. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 103-109. Copenhagen: Munksgaard.
- LINDNER, H. J. (1974). *Tetrahedron*, **30**, 1127-1132.
- NORDMAN, C. E. (1966). *Trans. Am. Crystallogr. Assoc.* **2**, 29-38.
- NORDMAN, C. E. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 5.01-5.13. Bangalore: The Indian Academy of Sciences.
- NORDMAN, C. E. & SCHILLING, J. W. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 110-114. Copenhagen: Munksgaard.
- OLIVER, J. D. & STRICKLAND, L. C. (1984). *Acta Cryst.* **C40**, 820-824.
- SHELDRIK, G. M. (1985). In preparation.
- SZEIMIES-SEEBACH, U., HARNISCH, J., SZEIMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J.-P. (1978). *Angew. Chem. Int. Ed. Engl.* **17**, 848-850.
- TOLLIN, P., MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **20**, 404-407.

*Acta Cryst.* (1985). **A41**, 268-273

## Statistical Mechanics Approach to the Structure Determination of a Crystal

BY S. V. SEMENOVSKAYA

*Institute of Organo-Element Compounds, Academy of Sciences of the USSR, Moscow, USSR*

AND K. A. KHACHATURYAN AND A. G. KHACHATURYAN

*Institute of Crystallography, Academy of Sciences, Moscow, USSR*

(Received 22 June 1984; accepted 7 December 1984)

### Abstract

The previously formulated new approach to the structure analysis of a crystal based on the profound analogy between the problem of determination of thermodynamic equilibrium in statistical mechanics and the optimization problem for a function of many variables [Khachaturyan, Semenovskaya & Vainshtein (1979). *Sov. Phys. Crystallogr.* **24**, 519-524; (1981). *Acta Cryst.* **A37**, 742-754] is developed. In

this approach, a crystal structure is determined by the equilibrium low-temperature state of a model non-ideal gas composed of the atoms within a crystal unit cell, the unit cell and the  $R$  factor being regarded as a vessel and an interatomic interaction Hamiltonian, respectively. In contrast to the above cited papers, the low-temperature equilibrium state is found by means of the Monte Carlo sampling scheme usually utilized in statistical mechanics applications. The main advantage of such a treatment is that the

system reaches the equilibrium state avoiding all 'traps', metastable (local) minima of the free energy, and thus provides automatic determination of the crystal structure. The method is successfully tested on the known L-proline structure containing 32 non-hydrogen atoms per unit cell. The proposed statistical mechanics approach can also be applied with minor modifications to optimization problems in the multi-dimensional space of many variables.

### 1. Introduction

As is known, all input quantitative information used in crystal structure determination is stored in the intensity array of diffraction reflections. The crystal structure determination is thus reduced to finding the spatial atomic distribution within a unit cell that would provide coincidence between calculated and observed intensities of diffraction reflections of the array. The linear  $R$  factor,

$$R = \frac{\sum_{\mathbf{H}} a(\mathbf{H}) \left| |F(\mathbf{H})| - |F(\mathbf{H})_{\text{obs}}| \right|}{\sum_{\mathbf{H}} a(\mathbf{H}) |F(\mathbf{H})_{\text{obs}}|}, \quad (1)$$

may be considered for example as a quantitative measure of this coincidence. In (1),

$$F(\mathbf{H}) = \sum_j f_j(\mathbf{H}) \exp(-i2\pi\mathbf{H}\mathbf{r}_j) \quad (2)$$

is the calculated structure amplitude related to a reciprocal-lattice vector  $\mathbf{H}$ ;  $\mathbf{r}_j$  is the vector indicating the position of the  $j$ th atom of an atomic configuration in a unit cell; index  $j$  enumerates all atoms entering the unit cell;  $|F(\mathbf{H})_{\text{obs}}| = [I(\mathbf{H})_{\text{obs}}]^{1/2}$ ,  $I(\mathbf{H})_{\text{obs}}$  is the observed intensity of the reflection related to  $\mathbf{H}$ ;  $a(\mathbf{H})$  are arbitrary positive constants; summation in (1) is made over all reciprocal-lattice vectors  $\mathbf{H}$  labelling the reflections of the intensity array.

It follows from (1) that in the case of high-precision intensity measurements the minimum value of the  $R$  factor tends to zero at the atomic configuration corresponding to the correct structure. Therefore, the crystal structure determination may be reduced to finding the global minimum of the  $R$  factor in the multi-dimensional space of many variables (coordinates of all atoms in a unit cell). Similar problems arise in various fields of research. Their solution faces serious difficulties that increase dramatically with increasing number of variables. The main difficulty is the many local minima in which the system gets trapped during the minimization procedure. Being trapped in a local minimum results in determination of an incorrect configuration.

These difficulties can be overcome in a new statistical mechanics approach to the structure determination problem (Khachaturyan, Semenovskaya & Vainshtein, 1979, 1981), which is based on the idea that the minimization problem can be reformulated

in terms of statistical mechanics whose well developed theoretical concepts could be fruitfully applied to the new area of research. This idea was realized by utilizing the model of a thermodynamic system whose mathematical characteristics are analogous to the characteristics of the minimization problem for the function  $R$ .

The statistical mechanics approach has two advantages. The first is that statistical mechanics is created to deal with systems described by numerous degrees of freedom. Therefore any statistical mechanics model in the crystal structure determination problem does not impose too strict limitations on the number of atoms per unit cell. The second is that a proper statistical mechanics formalism necessarily involves thermal fluctuations. This is of especial importance since escapes from local free-energy minima occur only because of fluctuations, *viz* because of fluctuation-induced uphill 'diffusion'.

Since statistical mechanics studies the most probable states of a system, the most probable configurations at low temperature should be close to the ground-state configuration providing the absolute minimum of the model 'Hamiltonian',  $R$ . This is why the determination of the equilibrium state at low temperature enables us to solve the minimization problem for the  $R$  factor and thus to solve the structure determination problem.

Below we shall use the same model system as in the previous papers (Khachaturyan, Semenovskaya & Vainshtein, 1979, 1981), in which atoms occupying an independent part of a unit cell are regarded as a non-ideal gas, the independent part of the unit cell as a vessel, the  $R$  factor (1) as a Hamiltonian simulating interaction in the non-ideal gas. The temperature  $T$ , the partition sum  $Z$  and the free energy  $\Phi$  are introduced according to the conventional equations

$$Z = \sum \exp(-R/T) \quad \text{and} \quad \Phi = -T \ln Z, \quad (3)$$

where the summation is carried out over all atomic configurations in the independent part of a unit cell. The equilibrium atomic distribution in this system at low temperature yields the structure, slightly disordered, however, by 'thermal' fluctuations (this disordering is a model analogue of the real temperature disordering described by the Debye-Waller factor).

Khachaturyan, Semenovskaya & Vainshtein (1979, 1981) found the low-temperature equilibrium state related to the true crystal structure by means of numerical solution of Onsager kinetic equations. Below we shall propose to utilize the more convenient Monte Carlo method (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953) formulated to solve the statistical mechanics problems in liquids, alloys, magnetic materials and so on. The main advantage of the Monte Carlo method is that it enables us to find the equilibrium stable state corresponding to

the absolute minimum of the free energy at a given temperature whereas the kinetic equation method sometimes yields a metastable state corresponding to an incorrect structure.

## 2. Monte Carlo simulation of thermodynamic equilibrium

A Monte Carlo sampling scheme similar to that used by Metropolis, Rosenbluth, Rosenbluth, Teller & Teller (1953) and others to compute the desired thermodynamic functions in statistical mechanics is chosen for the crystal structure determination. The idea of the Monte Carlo method consists in simulation of random walks of atoms by means of constant stepwise transition probabilities for elementary 'diffusion' events, which generate the succession of atomic configurations forming a Markov chain. The probabilities of elementary atomic transitions are chosen so that mean values over the segment of the Markov chain, which starts from a certain atomic configuration remote from the beginning of the chain, would tend to the mean values over the petite canonical ensemble with the given model Hamiltonian  $R$ .

The brief account of the Monte Carlo algorithm formulated for the crystal structure determination problem is as follows.

Let  $N$  specific atoms be contained in the independent part of a unit cell. To simplify the calculation procedure the atoms are assumed to migrate only over  $N_0$  sites of a fine rectangular grid inscribed in the independent part of the unit cell. These  $N$  atoms are tested by picking them up one by one in a random sequence.\*

Let one of  $N$  atoms be chosen. This atom may migrate to one of the six nearest-neighbour sites of the grid. The random number generator selects a nearest-neighbour site to which the atom may migrate. The question whether or not the atom will move to the selected site is answered as follows.

1. If the selected nearest-neighbour site is already occupied by an atom, the migration is not made and the previous configuration is taken to be the new configuration.

2. If the selected nearest site is vacant, then the change in  $R$  factor,  $\Delta R$ , resulting from the migration to the nearest site is computed. Two outcomes are then possible:

(a) if  $\Delta R \leq 0$ , the migration to the tested nearest site is made and the resultant configuration is assumed to be the new one;

(b) if  $\Delta R > 0$ , the migration may be realized with the probability  $\exp(-\Delta R/T)$ : it occurs if the random number generator gives a number  $\xi$  taken from a uniform distribution on the interval  $(0, 1)$ , which

satisfies the inequality  $\xi < \exp(-\Delta R/T)$ ; and it does not occur if  $\xi \geq \exp(-\Delta R/T)$ . In both cases the resultant configuration is taken to be the new one.

The chosen fundamental stepwise transition probabilities between the nearest sites of the grid produce a Markov chain of atomic configurations. The stationary segment of the Markov chain sufficiently remote from its starting point consists of atomic configurations (states) that are distributed according to the petite canonical ensemble (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953; Fosdick, 1959). In other words, if the 'time',  $\tau$ , enumerating the successive atomic configurations of the chain is introduced and the stationary part of the Markov chain starts from the 'time'  $\tau = t_0$ , the average of any function calculated over configurations of the stationary segment of the chain in the 'time' interval  $(t_0, t_0 + t)$  tends to the thermodynamic average at  $t \rightarrow \infty$ . For example, the average structure amplitude over the 'time' interval  $(t_0, t_0 + t)$  is

$$\langle F(\mathbf{H}, t) \rangle = \frac{1}{t} \sum_{\tau=t_0}^{\tau=t_0+t} F(\mathbf{H}, \tau), \quad (4)$$

where  $F(\mathbf{H}, \tau)$  is the structure amplitude (2) related to the  $\tau$ th configuration of the Markov chain. At  $t \rightarrow \infty$ ,  $\langle F(\mathbf{H}, t) \rangle \rightarrow \langle F(\mathbf{H}) \rangle_T$ , where  $\langle F(\mathbf{H}) \rangle_T$  is the thermodynamic average structure amplitude. The corresponding average density is

$$\langle \rho(\mathbf{r}) \rangle_T = \sum_{\mathbf{H}} \langle F(\mathbf{H}) \rangle_T \exp(i2\pi\mathbf{H}\mathbf{r}). \quad (5)$$

At low temperature where the system is close to the ground state (the state with the minimum  $R$  factor), its average amplitude  $\langle F(\mathbf{H}) \rangle_T$  and the average density (5) tend to the structure amplitudes  $F(\mathbf{H})_{\text{obs}}$  and density of the structure wanted, respectively. Therefore, at low temperatures and  $t \rightarrow \infty$  the phases  $\varphi(\mathbf{H}, t)$  of the average amplitudes  $\langle F(\mathbf{H}, t) \rangle$  tend to the phases  $\varphi_0(\mathbf{H})$  of the structure amplitudes  $F(\mathbf{H})_{\text{obs}}$  of the structure wanted.

On the contrary, a temperature increase results in disordering of the structure because of 'thermal' fluctuations and thus leads to deviation of the calculated phases  $\varphi(\mathbf{H}, t)$  from their true values,  $\varphi_0(\mathbf{H})$ . Since the measure of disordering is deviation of the calculated average structure amplitudes  $\langle F(\mathbf{H}, t) \rangle$  from their true values  $F(\mathbf{H})_{\text{obs}}$ , a quantitative criterion of reliability of the calculated phases  $\varphi(\mathbf{H}, t)$  is the ratio

$$\begin{aligned} \eta(\mathbf{H}, t) &= |\langle F(\mathbf{H}, t) \rangle| / |F(\mathbf{H})_{\text{obs}}| \\ &= |\langle F(\mathbf{H}, t) \rangle| / [I(\mathbf{H})_{\text{obs}}]^{1/2}. \end{aligned}$$

At  $t \rightarrow \infty$ ,  $\eta(\mathbf{H}, t) \rightarrow \eta_T(\mathbf{H})$ , where  $\eta_T(\mathbf{H})$  is the thermodynamic average. The ratio  $\eta_T(\mathbf{H})$  plays the part of the long-range parameter of the reflection  $\mathbf{H}$  since  $\eta_T(\mathbf{H}) = 0$  at the completely disordered state when  $\langle \rho(\mathbf{r}) \rangle_T = \text{constant}$  and thus  $\langle F(\mathbf{H}) \rangle_T = 0$ . At  $T \rightarrow 0$

\* The final result does not depend on the order of testing  $N$  atoms (Fosdick, 1959).

when the system approaches the completely ordered state corresponding to the true structure,  $\langle F(\mathbf{H}) \rangle_T \rightarrow F(\mathbf{H})_{\text{obs}}$  and thus  $\eta_T(\mathbf{H}) \rightarrow 1$ . Therefore, the closer  $\eta_T(\mathbf{H})$  is to unity, the more reliable is the determination of the structure amplitude  $\langle F(\mathbf{H}) \rangle_T$  and its phase  $\varphi(\mathbf{H})$ .

The following conclusion can thus be made: the temperature  $T$  cannot be chosen too high because it would result in considerable deviation of the calculated phases  $\varphi(\mathbf{H}, t)$  from their true values  $\varphi_0(\mathbf{H})$  caused by 'temperature disordering'; the temperature  $T$  cannot however be made too low because it would result in 'temperature freezing', *i.e.* 'thermoactivated' migration of atoms would cease.

### 3. Testing the Monte Carlo algorithm in the case of L-proline

Potentialities of the Monte Carlo method in structure analysis are tested below for the same L-proline structure on which the method of kinetic equations (Khachaturyan, Semenovskaya & Vainshtein, 1979, 1981) has already been tested. The structure of L-proline,  $\text{C}_4\text{H}_7(\text{NH})\text{COOH}$ , is orthorhombic with  $a = 11.55$ ,  $b = 9.02$ ,  $c = 5.20$  Å and belongs to space group  $P2_12_12_1$  (Kayushina & Vainshtein, 1966). The rectangular grid with  $N_0 = 15 \times 15 \times 15$  sites was inscribed in an independent part of the unit cell of volume  $\frac{1}{2}a \times \frac{1}{2}b \times c$ . Its sites are occupied by eight atoms: five C atoms, two O atoms and one N atom.\* Random walk of these atoms generating the Markov chain was realized by means of the stochastic procedure described in § 2.

The calculations were carried out under so-called toroidal boundary conditions, which are imposed by the crystallographic symmetry: exit of any atom from the independent part of the unit cell through one of its six faces automatically implies that this atom returns to the same independent part of the unit cell through another face in accordance with the multiplication rule for a point in the  $P2_12_12_1$  space group:

$$\begin{aligned} (x, y, z) &\rightarrow \left(\frac{1}{2} - x, -y, \frac{1}{2} + z\right) \rightarrow \left(\frac{1}{2} + x, \frac{1}{2} - y, -z\right) \\ &\rightarrow (-x, \frac{1}{2} + y, \frac{1}{2} - z), \end{aligned}$$

where  $x = r_x/a$ ,  $y = r_y/b$ ,  $z = r_z/c$  are dimensionless coordinates of the unit-cell point  $\mathbf{r} = (r_x, r_y, r_z)$ . Within the independent part of the unit cell

$$0 \leq x < \frac{1}{2}, 0 \leq y < \frac{1}{2}, 0 \leq z < 1.$$

The coordinates of  $N_0 = 15 \times 15 \times 15$  sites of the grid inscribed in the independent part of the unit cell are

$$x_m = \frac{1}{30}(m - \frac{1}{2}), y_n = \frac{1}{30}(n - \frac{1}{2}), z_p = \frac{1}{15}(p - \frac{1}{2}), \quad (6)$$

where  $m, n, p$  are integer coordinates of the site:

$$1 \leq m \leq 15, 1 \leq n \leq 15, 1 \leq p \leq 15.$$

\* Low scattering hydrogen atoms are not taken into account.

The  $R$  factor was chosen in the form

$$R = A^{-1} \left\{ \sum_{\mathbf{H}}' b(\mathbf{H}) |F(\mathbf{H}) - [I(\mathbf{H})_{\text{obs}}]^{1/2} \exp[i\varphi_0(\mathbf{H})]| + \sum_{\mathbf{H}}'' a(\mathbf{H}) ||F(\mathbf{H})| - [I(\mathbf{H})_{\text{obs}}]^{1/2} \right\}, \quad (7)$$

where the normalization constant  $A$  is

$$A = \sum_{\mathbf{H}}' b(\mathbf{H}) [I(\mathbf{H})_{\text{obs}}]^{1/2} + \sum_{\mathbf{H}}'' a(\mathbf{H}) [I(\mathbf{H})_{\text{obs}}]^{1/2},$$

$a(\mathbf{H}) = 1$ ,  $b(\mathbf{H}) = 2$ . The summation  $\sum_{\mathbf{H}}' (\dots)$  was carried out over four reflections, 210, 201, 720, 032. The phases  $\varphi_0(\mathbf{H})$  for them were chosen to be  $\pi$ ,  $-\frac{1}{2}\pi$ ,  $-\frac{1}{2}\pi$ ,  $-\frac{1}{2}\pi$ , respectively. Such a choice fixes the origin of the structure and its enantiomorphic form. The summation  $\sum_{\mathbf{H}}'' (\dots)$  was taken over the remaining reflections.

The 'observed' array of intensities  $\{I(\mathbf{H})_{\text{obs}}\}$  was calculated from the known atomic coordinates (Kayushina & Vainshtein, 1966). These coordinates expressed in terms of  $m, n, p$  [see (6)] are

$$\begin{aligned} \text{C}(1) &(1.37, 8.54, 2.46); \text{C}(2) (2.15, 9.35, 6.72); \\ \text{C}(3) &(5.09, 12.68, 7.25); \text{C}(4) (8.27, 9.65, 6.69); \\ \text{C}(5) &(7.28, 5.39, 8.85); \text{O}(1) (2.21, 4.82, 0.96); \\ \text{O}(2) &(14.81, 4.58, 0.51); \text{N} (3.47, 4.94, 8.34). \end{aligned} \quad (8)$$

The coordinates  $m, n, p$  in (8) are not integers since the real locations of the atoms do not coincide with sites of the grid (6). The assumption that the atoms are located on the sites of the grid and thus are shifted from their real positions is equivalent to introducing a certain artificial 'experimental' error in 'observed' intensities. The error is of the order of 10%. The structure amplitudes in the space group  $P2_12_12_1$  were calculated from the equations

$$\begin{aligned} \text{Re } F(\mathbf{H}) &= \sum_{j=1}^{j=8} f_j(\mathbf{H}) \cos 2\pi[hx_j - (h-k)/4] \\ &\quad \times \cos 2\pi[ky_j - (k-l)/4] \\ &\quad \times \cos 2\pi[lz_j - (l-h)/4] \\ \text{Im } F(\mathbf{H}) &= - \sum_{j=1}^{j=8} f_j(\mathbf{H}) \sin 2\pi[hx_j - (h-k)/4] \\ &\quad \times \sin 2\pi[ky_j - (k-l)/4] \\ &\quad \times \sin 2\pi[lz_j - (l-h)/4], \end{aligned} \quad (9)$$

where  $f_j$  is the atomic factor of the  $j$ th atom,  $h, k, l$  are Miller indexes of the reciprocal-lattice vector  $\mathbf{H}$ . The average structure amplitude is determined by (4). The corresponding phases  $\varphi(\mathbf{H}, t)$  were calculated from the equation

$$\varphi(\mathbf{H}, t) = \arctan(\text{Im} \langle F(\mathbf{H}, t) \rangle / \text{Re} \langle F(\mathbf{H}, t) \rangle). \quad (10)$$

The calculations were carried out at  $T = 0.0055$ . The

Table 1. Summary of averaging data

TIME = 704000							JUMPS = 66823						
<i>h</i>	<i>k</i>	<i>l</i>	$I(\mathbf{H})_{\text{obs}}$	$\eta(\mathbf{H}, t)$	$\varphi_0(\mathbf{H})$	$\varphi(\mathbf{H}, t)$	<i>h</i>	<i>k</i>	<i>l</i>	$I(\mathbf{H})_{\text{obs}}$	$\eta(\mathbf{H}, t)$	$\varphi_0(\mathbf{H})$	$\varphi(\mathbf{H}, t)$
0	0	2	0.73E 04	0.91E 00	0.0	0.0	0	2	0	0.45E 04	0.95E 00	3.14	3.14
2	1	1	0.30E 04	0.96E 00	1.45	1.45	2	1	0	0.33E 04	0.10E 01	3.14	-3.14
0	4	0	0.16E 04	0.97E 00	3.14	3.14	0	4	2	0.16E 04	0.84E 00	3.14	3.14
4	0	0	0.13E 04	0.87E 00	0.0	0.0	5	1	0	0.56E 03	0.85E 00	1.57	1.57
4	2	0	0.87E 03	0.67E 00	3.14	3.14	4	1	0	0.87E 03	0.52E 00	3.14	3.14
2	0	1	0.13E 04	0.98E 00	-1.57	-1.57	7	2	0	0.11E 04	0.10E 01	-1.57	-1.57
2	0	0	0.19E 04	0.74E 00	0.0	0.0	0	3	2	0.85E 03	0.91E 00	-1.57	-1.57
3	4	0	0.61E 03	0.69E 00	1.57	1.57	2	1	2	0.68E 03	0.82E 00	-2.95	-3.00
7	2	1	0.88E 03	0.61E 00	-3.13	3.08	3	3	0	0.46E 03	0.34E 00	-1.57	-1.57
3	2	1	0.59E 03	0.67E 00	-2.45	-2.63	5	2	0	0.40E 03	0.59E 00	-1.57	-1.57
5	5	0	0.99E 03	0.67E 00	-1.57	-1.57	0	1	1	0.51E 03	0.92E 00	-1.57	-1.57
1	0	3	0.42E 03	0.53E 00	1.57	1.57	6	1	1	0.56E 03	0.58E 00	1.58	1.44
1	0	1	0.74E 03	0.68E 00	1.57	1.57	5	1	2	0.48E 03	0.73E 00	2.01	1.74
3	4	1	0.57E 02	0.30E 00	1.17	0.77	0	1	2	0.58E 03	0.87E 00	1.57	1.57
3	3	2	0.34E 03	0.42E 00	-1.66	-1.67	2	2	2	0.29E 03	0.86E 00	-1.68	-1.73
2	5	1	0.44E 03	0.86E 00	-1.49	-1.50	0	2	1	0.45E 03	0.79E 00	0.0	-0.00
0	2	2	0.35E 03	0.89E 00	3.14	3.14	6	0	0	0.28E 03	0.64E 00	3.14	3.14
0	4	1	0.23E 03	0.63E 00	3.14	3.14	4	3	0	0.37E 03	0.64E 00	0.0	0.0
5	2	2	0.31E 03	0.52E 00	-1.84	-1.77	6	3	0	0.35E 03	0.44E 00	0.0	0.00
7	2	2	0.77E 03	0.94E 00	-1.27	-1.43	0	2	3	0.58E 03	0.71E 00	0.0	0.0
6	2	0	0.27E 03	0.71E 00	0.0	0.00	2	4	2	0.28E 03	0.72E 00	2.94	2.69
5	3	2	0.35E 03	0.69E 00	-1.08	-1.49	5	1	1	0.28E 03	0.69E 00	-1.97	-2.38
1	3	0	0.18E 03	0.73E 00	-1.57	-1.57	2	2	0	0.30E 03	0.80E 00	3.14	3.14
1	7	0	0.79E 03	0.64E 00	1.57	1.57	3	2	0	0.22E 03	0.56E 00	-1.57	-1.57
5	3	1	0.29E 03	0.82E 00	0.89	0.60	1	2	1	0.36E 03	0.63E 00	-2.08	-1.77
1	6	0	0.21E 03	0.63E 00	-1.57	-1.57	2	3	1	0.13E 03	0.90E 00	2.09	2.25
2	3	0	0.12E 03	0.78E 00	3.14	3.14	3	2	3	0.33E 03	0.56E 00	-2.18	-2.48
6	1	0	0.22E 03	0.13E 01	3.14	3.14	5	3	0	0.12E 03	0.93E 00	-1.57	-1.57
2	3	3	0.32E 03	0.73E 00	2.01	2.21	0	3	4	0.99E 03	0.77E 00	-1.57	-1.57

RM = 0.13033

8	9	7	2	9	3	2	10	7	3	5	9
5	13	8	15	5	15	2	5	1	7	5	9

latter temperature was chosen as a compromise between two contradictory requirements, to perform computations at the lowest possible temperature and to maintain a sufficient number of successful trials for atomic migration (at  $T = 0.0055$  atomic migration occurs in about 10% of all trials).\*

The random distribution of eight atoms,  $5C + 2O + N$ , produced by the random number generator was used as the initial configuration of the Markov chain at  $\tau = 0$ . The averaging procedure was started from  $\tau = t_0 = 2.24 \times 10^6$ . The averaging interval,  $t$ , was chosen to be  $t = 7.04 \times 10^5$ . The averaging was taken over configurations separated by the 'time' interval  $\Delta t = 32$  to avoid correlation between successive configurations.

The summary of averaging data is presented in Table 1. The columns of the table are reflection indexes  $h, k, l$ , corresponding to the observed intensities  $I(\mathbf{H})_{\text{obs}}$ , 'long-range-order parameters'  $\eta(\mathbf{H}, t)$ , true phases,  $\varphi_0(\mathbf{H})$ , and calculated phases,  $\varphi(\mathbf{H}, t)$ , at  $t = 7.04 \times 10^5$ .

Comparison of the phases  $\varphi_0(\mathbf{H})$  and  $\varphi(\mathbf{H}, t)$  in the fourth and fifth columns shows that all phases

\* The temperature  $T$  has to be of the same order of magnitude as the lowest difference between values  $\Delta R$  from the spectrum  $\{\Delta R\}$ .

$\varphi(\mathbf{H}, t)$  are correct: the maximum deviation  $\sim 23^\circ$  is observed for three reflections, 341, 532, 511.

There is another interesting fact that deserves discussion. During the computer simulation of the Markov chain, the comparison of the  $R$  factors of all configurations of the chain was carried out and the coordinates of all atoms forming the minimum  $R$  factor configuration were 'memorized'. The minimum  $R$  factor attained along the entire Markov chain was  $R_{\text{min}} = 0.13033$ . The coordinates of the configuration with the minimum  $R$  factor are presented at the bottom of Table 1 in terms of integers  $m, n, p$  [see (6)]. The coordinates in Table 1 are arrayed in the same sequence as in (8): the upper line yields the  $(m, n, p)$  coordinates of atoms C(1), C(2), C(3), C(4); the bottom line yields the coordinates of C(5), O(1), O(2), N. Comparison of the atomic coordinates  $(m, n, p)$  displayed in Table 1 with the correct coordinates (8) shows that

(i) the positions of all eight atoms are correctly determined within the accuracy of a spacing of the grid with  $N_0 = 15 \times 15 \times 15$  sites (deviation of each coordinate is less than unity);

(ii) distribution of the atoms over their positions differs from the correct one. The difference is associated with the only faulty interchange between the N

and C(5) atoms [ $N \rightleftharpoons C(5)$ ]. The other interchanges involve atoms of the same kind, *i.e.* result in mere renumbering of the C and O atoms. 'Mixing up' the N and C(5) atoms in the crystal structure determination process is quite natural because it is caused by the small 'diffraction distinction' between them. Refining the kinds of atoms can be easily done either in the framework of the Monte Carlo scheme (increasing  $N_0$  and decreasing  $T$ ) or by means of the conventional procedure when heights of electron density peaks calculated through the phases  $\varphi(\mathbf{H}, t)$  are compared.

Table 1 displays results of calculations in the framework of the simplest scheme: random initial conditions, the grid with  $N_0 = 15 \times 15 \times 15$  and constant temperature  $T = 0.0055$ . It was, however, shown that the proposed method yields the same correct crystal structure regardless of initial conditions and the averaging interval ( $t_0, t_0 + t$ ) of the stationary part of a Markov chain.

It was shown that at low resolution when  $N_0 = 10 \times 10 \times 10$  (spacing of the grid of the order of  $0.5 \text{ \AA}$  is close to the atomic radius) accurate crystal structure determination turns out to be impossible. This is not too surprising because in this case the atomic configuration related to the  $R$ -factor minimum differs considerably from the correct one. What is surprising is that even in this unfavourable case the calculations at  $T = 0.007$  and  $T = 0.0055$  yield correct phases for about 80% of all reflections entering the  $R$  factor. Furthermore, the calculated phases of all reflections with  $\eta(\mathbf{H}, t) > 0.4$  are correct, *i.e.* the value of  $\eta(\mathbf{H}, t)$  may be regarded as a reliability criterion.

As for higher resolution, when  $N_0 = 15 \times 15 \times 15$  (spacing of the grid of the order of  $0.3 \text{ \AA}$  is close to half of the atomic radius), the crystal structure determination was attained from any initial conditions.

The fastest way to find the structure is, however, stepwise 'cooling' similar to that employed earlier (Khachaturyan, Semenovskaya & Vainshtein, 1979, 1981). The L-proline structure was determined by means of the following scheme. The Markov chain was generated at  $T = 0.007$  on the grid with  $N_0 = 10 \times 10 \times 10$  with  $t = 6.4 \times 10^5$ . The minimum  $R$ -factor configuration attained was utilized as the initial one for further calculation at  $T = 0.0055$  on the  $15 \times 15 \times 15$  grid with  $t = 1.36 \times 10^6$ . The minimum  $R$ -factor configuration obtained also solves the structure determination problem within the accuracy of a  $15 \times 15 \times 15$  grid spacing.

Of course, the L-proline structure with 32 atoms per unit cell and eight atoms per independent part

of the unit cell is not too complex by modern standards. It is described nevertheless by 24 independent degrees of freedom. This is considerably higher than the number of degrees of freedom acceptable for a non-local method (Gelfand, Vul, Ginzburg & Fedorov, 1966), the only operative method based on the  $R$ -factor minimization. In contrast to this non-local minimization method, the proposed statistical mechanics algorithm seems to be free from constraints associated with the number of atoms in the unit cell. Furthermore, the full advantages of the Monte Carlo method are usually revealed in the cases of systems with many variables. Numerous successful calculations of thermodynamical characteristics of non-ideal gas models operating with about  $10^2$ – $10^3$  atoms per cyclic volume (see, for example, Binder, 1979; Wood, 1968) seem to confirm this conclusion, because the non-ideal gas models differ in no way from the statistical mechanics model considered in this paper. We hope that the direct proof of this statement will be obtained in our calculations for structures with more atoms per unit cell, which are under way now.

In conclusion, the following point should be especially emphasized. The statistical mechanics approach may be efficient not only in the case of the structure analysis; it can be an efficient method of solving various deterministic minimization problems with many variables. But the most exciting advantage of the statistical mechanics approach is that it may be absolutely irreplaceable in dealing with optimization problems for big systems composed of elements with probabilistic interaction between them. Economics and biological models may serve as good examples of such systems.

#### References

- BINDER, K. (1979). Editor. *Monte Carlo Methods in Statistical Physics*. Berlin: Springer-Verlag.
- FOSDICK, L. D. (1959). *Phys. Rev.* **116**, 565–573.
- GELFAND, E. B., VUL, E. B., GINZBURG, S. L. & FEDOROV, JU. G. (1966). *The Method of Ravines in Problems of X-ray Analysis*. Moscow: Nauka. (In Russian.)
- KAYUSHINA, R. L. & VAINSHTEIN, B. K. (1966). *Sov. Phys. Crystallogr.* **10**, 698–706.
- KHACHATURYAN, A. G., SEMENOVSKAYA, S. V. & VAINSHTEIN, B. K. (1979). *Sov. Phys. Crystallogr.* **24**, 519–524.
- KHACHATURYAN, A. G., SEMENOVSKAYA, S. V. & VAINSHTEIN, B. K. (1981). *Acta Cryst.* **A37**, 742–754.
- METROPOLIS, N. A., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953). *J. Chem. Phys.* **21**, 1087–1099.
- WOOD, W. W. (1968). In *Physics of Simple Liquids*, edited by H. N. V. TEMPERLEY, J. S. ROWLINSON & G. S. RUSHBROOK. Amsterdam: North-Holland.