A Heuristic Molecular-Dynamics Approach for the Prediction of a Molecular Crystal Structure

Nobuo Tajima, Takayuki Tanaka, Tomoko Arikawa,† Takashi Sakurai,†† Shoichi Teramae,†† and Tsuneo Hirano*,†

Department of Applied Chemistry, Faculty of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113

†Department of Chemistry, Faculty of Science, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112

††Toray Systems Center, 1-8-1 Mihama, Urayasu, Chiba 279

(Received August 12, 1994)

A general method has been developed to predict the structures of molecular crystals without any constraints on the space symmetry nor on the lattice constants. A constant-pressure molecular dynamics (MD) technique, treating the lattice constants as variables, is employed in the search of candidate structures ("shaking" process). A heuristic MD search is proposed to reduce redundant sampling in the shaking process. The structures of the thus sampled species are further optimized in geometry by the steepest descent method ("quenching" process). This general method was coded in a program named MDCP (Molecular Dynamics for Crystal Packing), and was applied to the prediction of the crystal structures of CO_2, benzene, and pyrimidine molecules, as test examples. For CO_2 and benzene crystals, the experimentally observed structures were well reproduced through our method. For pyrimidine, however, a slight difficulty was observed under the atom–atom potentials employed here, although near-miss structures were frequently encountered.

Prediction of the crystal structure of a given new molecule has been an important challenge in the field of applied crystallography, such as organic superconductors and nonlinear optical materials. The *a priori* prediction of the crystal packing for an assembly of proposed molecules can suggest the potential possibility of the crystal as a superconductor or a nonlinear optical material even before starting their syntheses. Namely, when the Se–Se overlap is large, for example, in an organic charge-transfer complex, the crystal may behave as a superconductor; when an inversion center is absent in the crystal of an organic molecule showing a high 2nd-order hyperpolarizability, the crystal may behave as a nonlinear optical material.

For the prediction of the structure of a molecular crystal, a criterion for selecting the most probable structure among many different calculated structures should be discussed first. Conventionally, the predicted structures are compared in terms of the potential energies.[1] A more sound criterion based on thermochemistry is the free energy, which can be calculated for each proposed crystal structure through a partition function derived under a harmonic approximation.[2] In both schemes, the structure at the global minimum, i.e. the lowest potential energy or free-energy structure, is considered as the most probable crystal structure. Thus, locating the global minimum on the potential surface is an important step for predicting a crystal structure.

The steepest-descent method and the Newton–Raphson method are often used for minimizing the potential energy. Since the potential energy surface of a molecular assembly has many energy minima, and since these methods can only locate the nearest minimum-energy point around the initial structures, the key point for locating the global minimum in these methods is how to efficiently select a possible initial guess. Thus, previous studies have been aimed at obtaining a good initial guess for the crystal structure.[3—5] The coordinate space to be considered for a molecular crystal, which consists of the coordinates of molecules and the lattice constants, is very large in its dimension. The space symmetry and the lattice constants for the molecular system are often assumed to reduce the number of independent coordinates. Successful methods to be run under the preassumed lattice constants and/or space symmetries have been proposed.[4,5] However, these assumptions cannot be made for the prediction of the packing structure of an entirely new molecule which has not been synthesized yet, because the preassumed lattice constants and spatial symmetry are derived from empirical guesses or experimental information. Thus, the method to be developed should be based only on information concerning the properties obtainable either theoretically and/or experimentally for a given molecule.

For a general prediction of a molecular crystal structure without any assumption concerning the space symmetry and lattice constants, we make use of a lattice-variable constant-pressure molecular dynamics (MD) method. Note that MD has been employed here not to obtain an equilibrium configuration of a molecular assembly, but to efficiently sample the possible initial-guess structures for energy minimization. For this purpose we have coded a program for molecular crystals using constant-pressure MD treating the lattice constants as variables under no spatial symmetry constraints. We assume only a periodic boundary condition. In this study, the efficiency and capability of our algorithm were examined through trial predictions of several known crystal structures.

Algorithm for the Search of Crystal Structures

The basic idea in searching for the possible packing structure of a molecular crystal is to let a small number of molecules find themselves stable packing while moving around in a flexible unit cell under the periodic boundary condition. To adapt to various crystal structures, we chose mostly 8 (sometimes 4) as the number of molecules in a unit cell, since 84.6% of the organic homomolecular crystals contain 1, 2, 4 or 8 molecules in the unit cells according to a statistical study by Belsky and Zorkii.[6] We also assumed that the molecules do not change their conformation (rigid-body approximation). Thus, our search is performed with a unit cell composed of 8 rigid molecules under a periodic boundary condition. It follows that independent geometrical variables in our search are 24 (=3×8) coordinates for the centers of mass and 24 (=3×8) Euler angles for the orientations of the molecules, and 6 lattice constants (a, b, c, α, β, γ). Thus, the potential energy surface becomes a surface of 54 dimensions, and a point on the potential energy surface represents each crystal-packing structure.

Since we have no *a priori* knowledge concerning the potential surface for a particular molecular system, a useful way to search for the global minimum is to wander about over the potential energy surface under the molecular-dynamics constraint. Since the lattice constants should be treated as variables, we adopted a constant-pressure MD[7] of the type proposed by Nose and Klein[8] for searching for the minima on the potential surface. Thus, we formulated constant pressure MD equations for a molecular crystal system comprizing a small number of rigid molecules in a unit cell (see Appendix).

When the MD simulation is started at an appropriate temperature ("shaking" process), the system may relax from the given initial structure and wander about on the potential energy surface. One of the low-energy structures picked up from the thus calculated trajectories on the potential energy surface would be a possible candidate packing structure to be optimized so as to go down to the global minimum by the succeeding "quenching" process.

For the initial structures, the centers of mass of 8 molecules are placed so as to form 2×2×2 blocks of a simple cubic lattice with the orientation of each molecule being set randomly. The volume of the initial MD cell is set to be large enough to allow the 8 molecules to change their orientations easily. The initial velocities for geometrical variables are set to be proportional to the forces acting in the initial structures. These artificial initial settings do not mean any predetermined constraint, since the lattice constants as well as the spatial configuration of the molecules in the lattice are changed by the MD process, and the artifact imposed at the initial stage becomes rapidly relaxed in the first few hundreds steps.

During the MD process, we sample a possible crystal-packing structure by picking up a point having a low potential energy from the trajectory. Note that a point on the MD trajectory at a certain temperature seldom goes down to the very bottom of a basin on the potential surface, since the molecular system moves around with the kinetic energy corresponding to a given temperature. Since the locations of the potential minima are not known before starting MD walking, any points on the trajectory which show lower potential energies than previous points could be considered to represent a better packing structure, and, hence, should be sampled. In order to avoid redundant samplings, we set up a selection rule of sampling on the basis of the resemblance between the structure S of the current time-step and the structure S' of the most recent sample stored in a temporary buffer. The resemblance of two structures is evaluated in terms of the standard deviation d_{pair} for the relative distances of all atom pairs found in the unit cell.

$$d_{\mathrm{pair}} = \left\{ \sum_{i<j} (r_{ij} - r_{ij}')^2 / N_{\mathrm{pair}} \right\}^{1/2}, \qquad (1)$$

where r_{ij} and r_{ij}' denote the distances between atoms i and j in S and S', respectively, and N_{pair} denotes the number of atom pairs. When d_{pair} is greater than a given threshold $d_{\max}$, S' is accepted as a valid sample and S is stored in the temporary buffer as the most recent sample. On the other hand, when d_{pair} is smaller than the threshold $d_{\max}$, S' in the temporary buffer will be updated by the present sample S only when the energy of S is lower than that of S'. In this way, we obtain one sample point every time we visit a supposed-to-be different basin on an unfamiliar potential surface. While sorting in energy, we accumulate the thus sampled points up to a certain total number.

After obtaining sufficient sampling points, energy minimization is performed for each sampled structure in order to bring it down to the bottom of the corresponding potential-energy minimum by removing kinetic energy at every time-step of the MD process (a steep-

est-descent method). We call this process "quenching". We then identify the independent minimum by classifying each minimum-energy point according to its energy, a set of lattice constants, and space symmetry. For a unique definition of the lattice constants, each MD cell obtained through quenching should be reduced to a primitive unit cell, which is the minimal unit of the periodic structure. The primitive cell need further be transformed so that $c<a<b$ and $\alpha,\beta>90°$. The space group can be determined by examining possible symmetry operations on the thus obtained reduced structure. Numerically, a candidate space group can be confirmed by checking the standard deviation σ for the distances between the corresponding atoms in the reduced structure before and after the symmetry operation is applied.

By repeating the procedures described above, we can obtain, hopefully a small number of, different minimum-energy points that are independent of each other. Among them, the structure showing the lowest potential energy could be the global minimum structure, and should be the one to be observed as a real molecular crystal. The potential functions used in the MD processes in the shaking and quenching stages are inaccurate in most cases; we could say, instead, that one of the structures having low potential energy should be proposed as a candidate for a real crystal structure.

The entire flow of our procedure for finding the global minimum is as follows:

1) The geometry and atomic charges of a given molecule are determined, for example, by *ab initio* molecular orbital calculations.
2) The appropriate potential functions and parameters are adopted.
3) Eight molecules (sometimes four) are set in a large unit cell, and the initial velocities are calculated according to the forces acting in the initial structure.
4) A constant-temperature and constant-pressure MD simulation regarding the lattice constants as variables is performed at certain temperature ("shaking"). Low-energy structures are sampled while avoiding redundancy, and are stored after sorting by energy.
5) Steps 3 and 4 are repeated until sufficient number of samples are obtained.
6) The potential energies of the sampled structures are minimized in order to identify the potential minima to which each sample should belong to ("quenching").
7) The quenched structures are classified into groups with respect to their potential energies.
8) The lattice constants and the space symmetry are checked for each classified group.

We would like to call the entire process a "heuristic MD search."

Program MDCP

For the calculations in steps 3)—6) mentioned above, a program called MDCP[9] (Molecular Dynamics for Crystal Packing) was coded in FORTRAN. In this program the intermolecular interactions are expressed in the form of an atom–atom potential. For van der Waals interactions, three types of potential functions, i.e. exp-6, Lennard-Jones (6,12), and Morse type functions, are installed. These short-range forces are summed over all the atom pairs within a given cutoff radius. For Coulombic interactions, Ewald's sum method[10] is employed in order to accelerate their convergence. A numerical integration for the equations of motion is performed using a predictor-corrector algorithm.[11] For energy minimization, the steepest descent cycles are applied by using a temperature-scaling routine.

Results and Discussion

In order to demonstrate the validity and efficiency of our method, it has been applied to predicting the crystal structures of CO_2, benzene, and pyrimidine molecules, for which the crystal structures are already known experimentally. Since the potential surface to be used in our method is entirely dependent on the potential functions and their parameters, our goal of demonstration here is how efficiently to reproduce the crystal structure determined by the global minimum of the potential surface. The experimentally observed crystal structure should correctly be reproduced, when the employed potential functions and parameters are appropriate.

1. CO_2. Table 1 gives the potential parameters, cut-off radius r_{cut}, Ewald's convergence parameter α, and geometry of the CO_2 molecule employed in our calculations. The potential parameters and geometry were taken from those proposed by Williams et al.[12] for lattice energy calculations. The external pressure P_{ext} was kept at 1 atm.

First, we examined the quality of potential surface calculated from the potential functions employed. We performed a "quenching" process starting from the experimentally observed crystal structure in order to search for the structure at the nearest potential-energy minimum point (S_{min}) defined by the given potential functions and parameters. This energy-minimum point should be the goal of our MDCP search. Table 2 gives the lattice constants, potential energy, and space symmetry for the thus determined structure S_{min}, which are in good agreement with those of experimentally observed values.

Next, we ran the heuristic MD search. Eight CO_2 molecules were placed in the MD cell. The initial length of the MD cell edg L_{init} was set to 10 Å, which gave a volume about three times larger than that of the experimentally observed structure. The centers of mass of the molecules were set to form a simple cubic lattice. Since the CO_2 molecule has an inversion center, the initial forces acting on the centers of mass are zero. Thus, if the initial velocities are set to be proportional to the initial accelerations, the centers of mass will not move. To avoid this situation, we set the initial velocities in a random fashion.

Table 1. Potential Parameters and Geometry for the CO_2 Molecule System,[a] Adopted from Those Reported by Williams et al.[12]

- Potential function and parameters[12]

$$E_{ij}=B_{ij}\exp(-C_{ij}r_{ij})-A_{ij}r_{ij}^{-6}+1389.4q_iq_jr_{ij}^{-1}$$

$A_{CC}=2.41\times10^3$,	$B_{CC}=3.67\times10^5$,	$C_{CC}=3.60$,	$q_C=0.820$,
$A_{OO}=1.13\times10^3$,	$B_{OO}=2.30\times10^5$,	$C_{OO}=3.96$,	$q_O=-0.410$,
$A_{CO}=1.65\times10^3$,	$B_{CO}=2.91\times10^5$,	$C_{CO}=3.78$.	

- Cut-off distance $r_{cut}=14$ Å
- Ewald's convergence parameter $\alpha=0.221$ Å^{-1}
- Geometry of CO_2[12,13] $r_{CO}=1.12$ Å

a) r_{ij} is the interatomic distance between atoms i and j, and q_i is the partial charge on the atom i. Units: E_{ij} in kJ mol^{-1}, r_{ij} in Å, and q_i in e (elementary charge).

Table 2. Lattice Constants, Potential Energies and Space Symmetries of the Predicted CO_2 Crystals

	Z[a]	a/Å	b/Å	c/Å	α/°	β/°	γ/°	P.E.[b] / kJ mol^{-1}	σ[c] / 10^{-3}Å	Symmetry
S_{obs}[d]	4	5.54	5.54	5.54	90	90	90			$Pa3$
$S_{Williams}$[e]	4	5.66	5.66	5.66	90	90	90			$Pa3$
S_{min}[f]	4	5.66	5.66	5.66	90.0	90.0	90.0	−27.40	0.22	$Pa3$
C-1	8	8.00	8.00	5.66	90.0	90.0	90.0	−27.40		
	(4)[g]	(5.66)	(5.66)	(5.66)	(90.0)	(90.0)	(90.0)		0.04	$Pa3$
C-2	8	6.89	7.78	7.26	90.0	113.0	90.0	−26.71		
	(4)[g]	(6.89)	(7.26)	(3.89)	(90.0)	(90.0)	(113.0)		0.02	$P2_1/a$
C-3	8	7.73	6.92	7.73	90.0	119.9	90.0	−26.69		
	(4)[g]	(6.67)	(6.92)	(3.87)	(90.0)	(90.1)	(90.0)		1.03	$P2_1ab$

a) Number of molecules in a unit cell. b) Potential energy for the structure after "quenching". c) Standard deviation in geometry between the predicted packing and that obtained by symmetry operation. d) Experimentally observed structure.[13] e) Calculated values by Williams et al.[12] f) The structure obtained by "quenching" from the experimentally observed crystal structure. g) Values in parentheses denote those of the reduced unit cell.

Starting with these initial settings, the shaking process was performed for 5000 time-steps with a time-step Δt of 2 fs at a temperature T of 200 K. The hypothetical mass W (see Appendix and Refs. 7 and 8) for the lattice was empirically determined to be 150 u (unified atomic mass unit). A trace of the potential energy in the shaking process is shown in Fig. 1, where the sampling points (denoted by circles) and the minimum points obtained by quenching the sampling points (denoted by crosses) are also shown. As shown in Fig. 1, the system relaxed rapidly at the initial MD stage, and then moved around the rims of the basins on the potential energy surface. The system seems to stay longer in the potential basin of the lower energy. Thus, the employed MD has been proved to work well in wandering about concentratingly on regions of the low potential-energy minima. A set of the initial setting and shaking process was repeated four times. As a result, 98 structures in total were sampled with the sampling threshold $d_{max}=0.5$ Å.

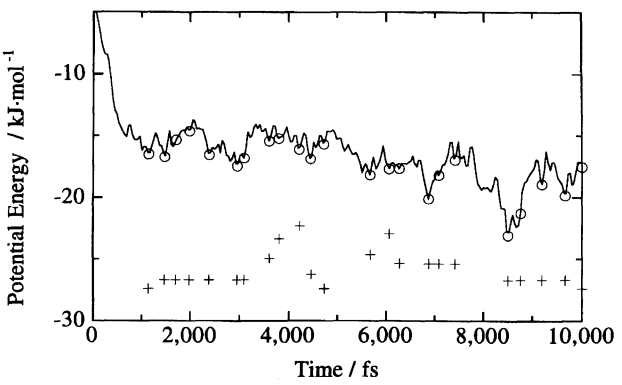

Fig. 1. The trace of the potential energy of the eight CO_2 molecule system. Open circles indicate sampling points and crosses show the potential energies for the samples quenched from each sampling point.

A histogram of the potential energy for each quenched structure derived from 98 sampling points is shown in Fig. 2. According to the potential energy, the quenched structures can be classified into groups in energy: {−27.40 (×21)}, {−26.71 (×9)}, {−26.69 (×11)}, {−26.61 (×7)}, {−26.25 (×11)}, ··· kJ mol^{-1}, where the numbers of occurrence are shown in parentheses. Examining the structure of each group after the unit cell reduction, we found that the structures belonging to one group are all the same. The lattice constants and space symmetries for the three lowest energy groups, i.e. **C-1**, **C-2**, and **C-3**, are shown in Table 2, and their stereoviews are shown in Fig. 3. The structure of **C-1** is the most energetically stable among all of the quenched structures. This structure has the same lattice con-

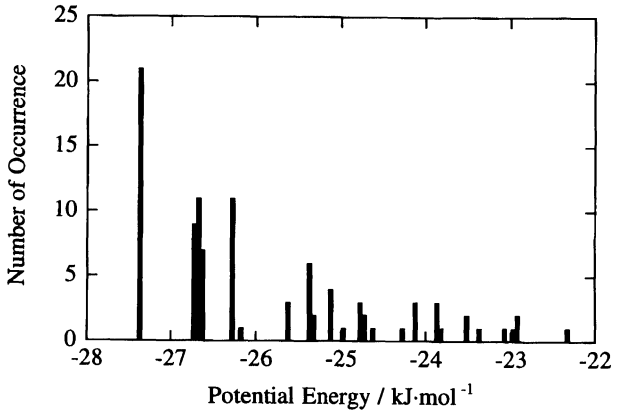

Fig. 2. Histogram of the potential energies of the 98 quenched samples of the eight CO_2 molecule system.

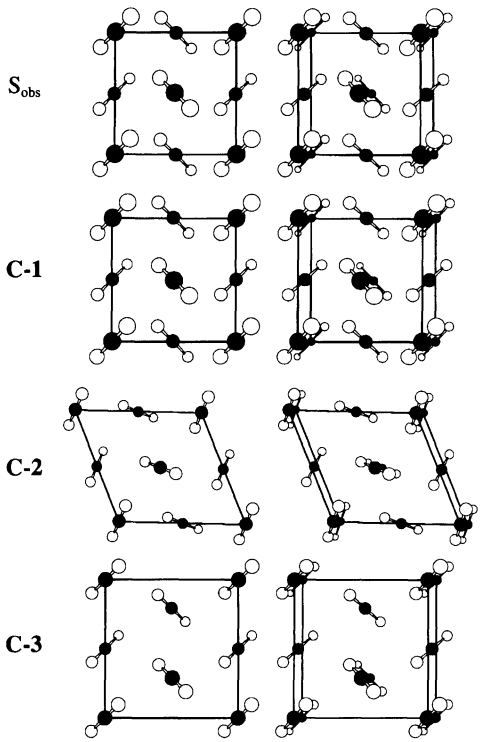

Fig. 3. Stereoviews of the experimentally observed and theoretically predicted crystal structures of CO_2 molecule (● and ○ denote carbon and oxygen atoms, respectively). See Table 2 for the structures S_{obs}, **C-1**, **C-2**, and **C-3**.

stants as those of S_{min}, and the space symmetry is also found to be the same *Pa*3. This means that MDCP has been successful in locating the global minimum on the potential surface defined by the Williams potentials,[12] and that the Williams potentials are sufficiently good for this particular case. In this case, since the Williams potentials are of good quality, we can also reproduce the experimentally observed crystal structure. Thus, our method was proven to be valid in finding the crystal structure of CO_2.

2. Benzene. The geometry of the benzene molecule and the potential parameters employed (Hall and Williams, 1975)[14] are listed in Table 3. The validity of these potential parameters was checked by quenching the experimentally observed crystal structures under $P_{ext}=1$ atm. For a benzene crystal, two polymorphs are reported: one is observed at ordinary pressure and shows *Pbca* symmetry ($S_{obs\text{-}I}$);[15] the other is observed at 25000 atm and shows $P2_1/c$ ($S_{obs\text{-}II}$).[16] The minimum-energy structures, $S_{min\text{-}I}$ and $S_{min\text{-}II}$, obtained by the quenching of $S_{obs\text{-}I}$ and $S_{obs\text{-}II}$, respectively, were very similar to the corresponding observed structures (see Table 4).

The parameters for shaking were set as follows: the number of molecules in a unit cell $Z=8$, $d_{max}=0.7$ Å, $T=200$ K, $P_{ext}=1$ atm, $L_{init}=12$ Å, $\Delta t=2$ fs, and $W=150$ u. Since the benzene molecules also have centers of symmetry, as is the case of CO_2, the initial velocities for the centers of mass of benzene molecules were set in a random fashion. Shaking for 5000 time-steps was repeated four times, and 40 sampled structures in total were obtained. After quenching, these samples were able to be classified into several groups, as shown in Fig. 4. Two of the quenched structures of these groups are listed in Table 4 and their stereoviews are shown in Fig. 5.

The **B-1** structure has a potential energy of -54.54 kJ mol^{-1}, which was the lowest among those of all the quenched structures. The potential energy, the lattice constants, and the predicted space symmetry (*Pbca*) were almost the same as those of $S_{min\text{-}I}$, as shown in Table 4. Thus, for the system of the benzene molecules as well as for that of CO_2 molecules, our method successfully located a crystal structure having the correct space symmetry.

The **B-2** and **B-3** structures have potential energies of -54.34 and -54.06 kJ mol^{-1} and show *Pbca* and $P2_1/n$ symmetries, respectively. They were encountered only once each. They represent independent energy minima, the structures of which are different from **B-1** and **B-4**.

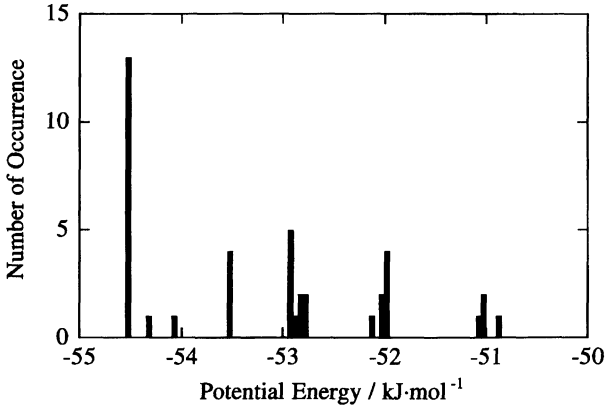

Fig. 4. Histogram of the potential energies of the 40 quenched samples of the eight benzene molecule system.

Table 3. Potential Parameters and Geometry for the Benzene Molecule System,[a] Adopted from Those Reported by Hall and Williams[14]

• Potential function and parameters[14]			
$E_{ij}=B_{ij}\exp(-C_{ij}r_{ij})-A_{ij}r_{ij}^{-6}+1389.4q_iq_jr_{ij}^{-1}$			
$A_{CC}=1.88\times10^3$,	$B_{CC}=2.99\times10^5$,	$C_{CC}=3.60$,	$q_C=-0.179$,
$A_{HH}=1.68\times10^2$,	$B_{HH}=1.20\times10^4$,	$C_{HH}=3.74$,	$q_H=0.179$,
$A_{CH}=5.62\times10^2$,	$B_{CH}=5.99\times10^4$,	$C_{CH}=3.67$.	
• Cut-off distance		$r_{cut}=14$ Å	
• Ewald's convergence parameter		$\alpha=0.355$ Å^{-1}	
• Geometry of benzene[14]		$r_{CC}=1.397$ Å, $r_{CH}=1.027$ Å	

a) r_{ij} is the interatomic distance between atoms i and j, and q_i is the partial charge on the atom i. Units: E_{ij} in kJ mol^{-1}, r_{ij} in Å, and q_i in e (elementary charge).

Table 4. Lattice Constants, Potential Energies and Space Symmetries of the Predicted Benzene Crystals

	Z[a]	a/Å	b/Å	c/Å	α/°	β/°	γ/°	P.E.[b] / kJ mol^{-1}	σ[c] / 10^{-3}Å	Symmetry
$S_{obs\text{-}I}$[d]	4	7.39	9.42	6.81	90	90	90			*Pbca*
$S_{obs\text{-}II}$[e]	2	5.42	5.38	7.53	90	110	90			$P2_1/c$
$S_{Williams\text{-}I}$[f]	4	7.43	9.29	6.93	90	90	90	−54.2		*Pbca*
$S_{Williams\text{-}II}$[f]	4	5.69	5.52	7.95	90	111.8	90	−53.1		$P2_1/c$
$S_{min\text{-}I}$[g]	4	7.43	9.31	6.94	90.0	90.0	90.0	−54.54	0.00	*Pbca*
$S_{min\text{-}II}$[g]	2	5.65	5.55	7.92	90.0	108.1	90.0	−53.54	0.00	$P2_1/c$
B-1	8	10.20	11.56	10.20	114.2	94.5	113.8	−54.54		
	(4)[h]	(7.49)	(9.26)	(6.93)	(90.0)	(90.0)	(90.4)		7.00	*Pbca*
B-4	8	9.68	11.27	9.68	104.6	110.2	75.3	−53.55		
	(2)[h]	(5.63)	(5.54)	(7.94)	(90.0)	(108.0)	(90.0)		0.26	$P2_1/c$

a) Number of molecules in a unit cell. b) Potential energy for the structure after "quenching". c) Standard deviation in geometry between the predicted packing and that obtained by symmetry operation. d) Experimentally observed structure at ordinary pressure.[15] e) Experimentally observed structure at 25000 atm.[16] f) Calculated values by Hall and Williams.[14] g) The structure obtained by "quenching" from the corresponding experimentally observed crystal structure. h) Values in parentheses denote those of the reduced unit cell.

The **B-4** structure has a potential energy of −53.55 kJ mol^{-1}, which belongs to the group of the forth lowest potential energy. This structure was reduced into a monoclinic unit cell; the consequent structure was similar to those of $S_{obs\text{-}II}$ as shown in Fig. 5. Both of them belong to the same space group, $P2_1/c$. Thus, a polymorphous structure at high pressure was successfully located as the fourth lowest energy structure at an external pressure of 1 atm.

3. Pyrimidine. For the geometry of the pyrimidine molecule and potential parameters, those of Williams and Cox[17] (listed in Table 5) were adopted. Here, additional interaction sites for lone-pair electrons are placed outward from the nitrogen atoms (Fig. 6).[17] The parameters for shaking were set as follows: $Z=8$, $d_{max}=1.0$ Å, $T=300$ K, $P_{ext}=1$ atm, $L_{init}=12$ Å, $\Delta t=5$ fs and $W=150$ u. Since many local minima were observed for this molecule in preliminary calculations, we considered that a wide search of the potential surface would be necessary. Thus, the shaking temperature for pyrimidine was set to be 300 K, higher by 100 K than that for CO_2 and benzene, and the number of the sets of initial setting and shaking cycle was increased to 20. From these shaking cycles, 475 sample structures in total were obtained.

The distribution of the potential-energy values for the quenched structures obtained from 475 samples is shown in Fig. 7. These quenched structures seemed to represent independent local minima, and were unable to be classified in terms of their potential energies. The presence of such numerous independent minima arises from the low symmetry of the pyrimidine molecule. For instance, different rotation angles of one of the eight pyrimidine molecules in its molecular plane give different packing structures, resulting in a small, but meaningful, potential-energy difference.

Among the structures quenched from the 475 samples, the structure corresponding to S_{min} was not found. The lowest potential energy found for these quenched structures was higher by 0.85 kJ mol^{-1} than the energy for S_{min}. The packing information for three quenched samples selected from the lowest potential energy side, i.e. **P-1**, **P-2**, and **P-3**, is listed in Table 6, and their stereoviews are shown in Fig. 8. In these structures, as is the case for the experimentally observed crystal structure, the molecules pack with their molecular planes in parallel and the centers of mass of the four molecules are in a parallelogram. The difference from the experimentally observed packing structure is in the rotational angles of a few molecules in their molecular planes,

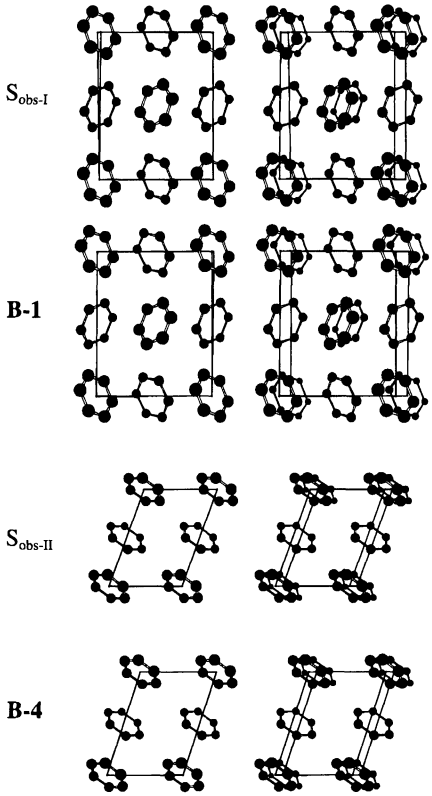

Fig. 5. Stereoviews of the experimentally observed and theoretically predicted crystal structures of benzene molecule (● denotes carbon atom). See Table 4 for the structures $S_{obs\text{-}I}$, $S_{obs\text{-}II}$, **B-1**, and **B-4**.

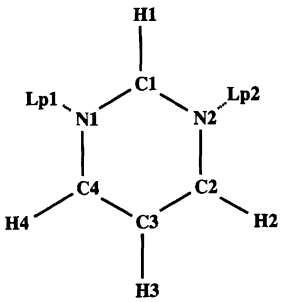

Fig. 6. Numbering of atoms in a pyrimidine molecule, to be used with Table 5. Lp1 and Lp2 are placed to represent lone pair electrons.

which results in the differences in the lattice constants and space symmetries. This type of packing form was frequently observed in those quenched samples having energies under -54 kJ mol^{-1}. Take **P-1** for example, the packing structure was almost in complete agreement with S_{min}, except for the directions of molecules 3 and 4 in their molecular planes. When these odd molecules were rotated properly in their molecular planes followed by quenching, the thus revised **P-1** structure coincided with that of S_{min} and the potential energy was lowered by 0.85 kJ mol^{-1} from that for the original **P-1** itself.

In this example, the difference in energy due to the difference in the rotational angles in the molecular plane was so small under the atom–atom potentials employed here that difficulty arose in predicting the correct packing structure. For such cases having many local minima of negligible energy difference, assistance of human interaction at the very last stage will be necessary to correct the oddly disposed molecules.

Table 5. Potential Parameters and Geometry for the Pyrimidine Molecule System,[a] Adopted from Those Reported by Williams and Cox[17]

- Potential function and parameters[17]

$$E_{ij}=B_{ij}\exp(-C_{ij}r_{ij})-A_{ij}r_{ij}^{-6}+1389.4q_iq_jr_{ij}^{-1}$$

$A_{CC}=2.44\times10^3$, $B_{CC}=3.70\times10^5$, $C_{CC}=3.60$,
$A_{NN}=1.38\times10^3$, $B_{NN}=2.55\times10^5$, $C_{NN}=3.78$,
$A_{HH}=1.36\times10^2$, $B_{HH}=1.20\times10^4$, $C_{HH}=3.74$,
$A_{CN}=1.83\times10^3$, $B_{CN}=3.07\times10^5$, $C_{CN}=3.69$,
$A_{CH}=5.77\times10^2$, $B_{CH}=6.65\times10^4$, $C_{CH}=3.67$,
$A_{NH}=4.34\times10^2$, $B_{NH}=5.52\times10^4$, $C_{NH}=3.76$.

- Cut-off distance $r_{cut}=14$ Å
- Ewald's convergence parameter $\alpha=0.266$ Å^{-1}
- Cartesian coordinates and charges for pyrimidine[b]

	X	Y	Z	q
N1	−1.192	0.690	0.000	1.067
N2	1.192	0.682	−0.002	1.067
C1	0.000	1.284	0.003	−0.079
C2	1.181	−0.657	0.000	−0.025
C3	−0.000	−1.382	0.003	−0.269
C4	−1.179	−0.646	−0.002	−0.025
H1	−0.002	2.311	0.002	0.154
H2	2.074	−1.164	0.000	0.141
H3	0.000	−2.408	0.003	0.166
H4	−2.075	−1.149	−0.003	0.141
Lp1[c]	−1.410	0.814	0.000	−1.169
Lp2[c]	1.411	0.805	0.000	−1.169

a) r_{ij} is the interatomic distance between atoms i and j, and q_i is the partial charge on the atom i. Units: E_{ij} in kJ mol^{-1}, r_{ij} and Cartesian coordinates in Å, and q_i in e (elementary charge). b) See Fig. 6 for the atom numbering. Williams and Cox[17] adopted the geometry reported in Ref. 18. c) Lp1 and Lp2 denote the pseudo-atoms representing lone pairs.

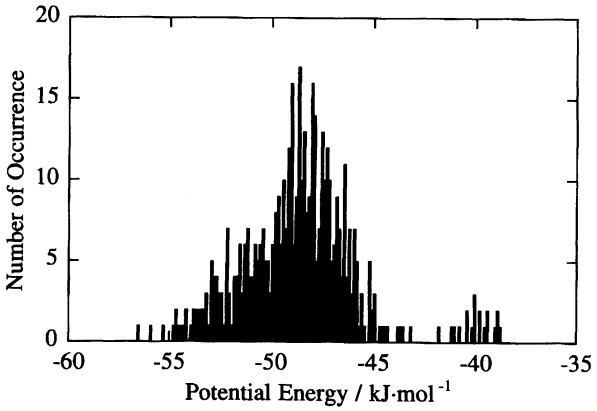

Fig. 7. Histogram of the potential energies of the 475 quenched samples of the eight pyrimidine molecule system.

Table 6. Lattice Constants, Potential Energies and Space Symmetries of the Predicted Pyrimidine Crystals

	Z[a]	a/Å	b/Å	c/Å	α/°	β/°	γ/°	P.E.[b] / kJ mol^{-1}	σ[c] / 10^{-3}Å	Symmetry
S_{obs}[d]	4	11.56	9.46	3.69	90	90	90			$Pna2_1$
$S_{Williams}$[e]	4	11.56	9.72	3.64	90	90	90	−57.3		$Pna2_1$
S_{min}[f]	4	11.56	9.72	3.65	90.0	90.0	90.0	−57.36	9.32	$Pna2_1$
P-1[g]	8	10.34	11.66	7.26	90.1	109.9	110.6	−56.51		$P1$
P-2[g]	8	10.28	11.78	7.24	90.1	110.6	92.3	−55.90		$P1$
P-3[g]	8	10.52	11.24	7.04	95.2	93.0	90.1	−55.30		$P1$

a) Number of molecules in a unit cell. b) Potential energy for the structure after "quenching". c) Standard deviation in geometry between the predicted packing and that obtained by symmetry operation. d) Experimentally observed structure.[18] e) Calculated values by Williams and Cox.[17] f) The structure obtained by "quenching" from the observed crystal structure. g) Unable to be reduced to smaller unit cell.

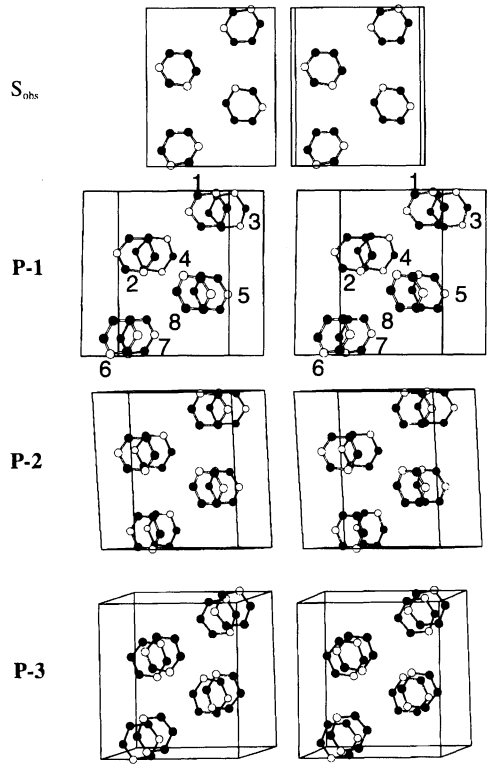

Fig. 8. Stereoviews of the experimentally observed and theoretically predicted crystal structures of pyrimidine molecule (● and ○ denote carbon and nitrogen atoms, respectively). See Table 6 for the structures S_{obs}, **P-1**, **P-2**, and **P-3**.

Conclusion

We have proposed a method to search for the structures of molecular crystals without using any information concerning the space symmetry and lattice constants. Molecules were treated as rigid bodies and periodic boundary condition was assumed. An algorithm for searching the global minimum has been developed by using a lattice-variable constant-pressure MD technique.

For high-symmetry molecules, such as CO_2 and benzene, our method was proved to be successful in finding a potential-energy minimum corresponding to the experimentally observed crystal structure. The total number of time-steps in the present MD calculations for generating possible packing structures was 20000, which is an acceptable number from a practical viewpoint. However, for a low-symmetry molecule, i.e. pyrimidine, even such MD calculations involving as many as 100000 time-steps ended with a near miss in finding the true global minimum. The usefulness of human interaction at the very last stage is suggested in this case.

This research was supported by the Ministry of Education, Science and Culture (Grant-in-Aid for Scientific Research on Priority Areas: Molecular Crystal). The authors are grateful to Professor Yuji Ohashi, Tokyo Institute of Technology, for his stimulative discussion.

Appendix

The following formulation for the MD, which treats the lattice constants as well as the coordinates of the interacting molecules as variables, is based on studies by Parrinello and Rahman[19] as well as Nose and Klein.[8]

The Cartesian coordinates $\boldsymbol{r}_i=(x_i,y_i,z_i)^{\mathrm{T}}$ (the superscript T, hereafter, denotes the transverse of the vector or matrix) and the fractional coordinates in a unit cell, $\boldsymbol{s}_i=(\xi_i,\eta_i,\zeta_i)^{\mathrm{T}}$, for the i-th atom are related through the lattice vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ as

$$\begin{aligned}\boldsymbol{r}_i&=\xi_i\boldsymbol{a}+\eta_i\boldsymbol{b}+\zeta_i\boldsymbol{c}\\&=\boldsymbol{B}\boldsymbol{s}_i,\end{aligned}\tag{1}$$

where the lattice matrix $\boldsymbol{B}$ is defined as

$$\boldsymbol{B}=(\boldsymbol{a}\ \boldsymbol{b}\ \boldsymbol{c}).\tag{2}$$

The Lagrangian L for a system consisting of atom i with a mass of m_i can be written as[19]

$$\begin{aligned}L&=\frac{1}{2}\sum_i m_i\dot{\boldsymbol{r}}_i^{\mathrm{T}}\dot{\boldsymbol{r}}_i-\frac{1}{2}\sum_i\sum_{j\neq i}\phi_{ij}(r_{ij})-P_{\mathrm{ext}}\det\boldsymbol{B}\\&=\frac{1}{2}\sum_i m_i\left\{\dot{\boldsymbol{s}}_i^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\dot{\boldsymbol{s}}_i+2\boldsymbol{s}_i^{\mathrm{T}}\dot{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}\dot{\boldsymbol{s}}_i+\boldsymbol{s}_i^{\mathrm{T}}\dot{\boldsymbol{B}}^{\mathrm{T}}\dot{\boldsymbol{B}}\boldsymbol{s}_i\right\}\\&\quad-\frac{1}{2}\sum_i\sum_{j\neq i}\phi_{ij}(r_{ij})-P_{\mathrm{ext}}\det\boldsymbol{B}.\end{aligned}\tag{3}$$

Here, $\phi_{ij}(r_{ij})$ is the potential energy between the i-th and j-th atoms separated by the distance r_{ij}, and P_{ext} is the external pressure. When the cross term, $\sum_i 2m_i \boldsymbol{s}_i^{\mathrm{T}} \dot{\boldsymbol{B}}^{\mathrm{T}} \boldsymbol{B} \dot{\boldsymbol{s}}_i$, is neglected and the kinetic-energy term, $\frac{1}{2}\sum_i m_i \boldsymbol{s}_i^{\mathrm{T}} \dot{\boldsymbol{B}}^{\mathrm{T}} \dot{\boldsymbol{B}} \boldsymbol{s}_i$, is replaced by $\frac{1}{2} W \mathrm{tr}(\dot{\boldsymbol{B}}^{\mathrm{T}} \dot{\boldsymbol{B}})$,[19] i.e. kinetic-energy term of the lattice vectors having a hypothetical mass W, the thus-approximated Lagrangian becomes

$$L = \frac{1}{2}\sum_i m_i \dot{\boldsymbol{s}}_i^{\mathrm{T}} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{B} \dot{\boldsymbol{s}}_i - \frac{1}{2}\sum_i \sum_{j \neq i} \phi_{ij}(r_{ij}) + \frac{1}{2} W \mathrm{tr}(\dot{\boldsymbol{B}}^{\mathrm{T}} \dot{\boldsymbol{B}}) - P_{\text{ext}} \det \boldsymbol{B}. \quad (4)$$

Thus, the Lagrange equations of motion $\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0$ derived for each atom in a unit cell and for each component of matrix $\boldsymbol{B}$ become

$$\ddot{\boldsymbol{s}}_i = -m_i^{-1} \sum_{j \neq i} (\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B})^{-1} \frac{\partial \phi_{ij}}{\partial \boldsymbol{s}_i} - (\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B})^{-1} (\dot{\boldsymbol{B}}^{\mathrm{T}} \boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}} \dot{\boldsymbol{B}}) \dot{\boldsymbol{s}}_i, \quad (5)$$

$$\ddot{B}_{\alpha\beta} = W^{-1} \left\{ \sum_i m_i (\dot{\boldsymbol{s}}_i)_\beta (\boldsymbol{B} \dot{\boldsymbol{s}}_i)_\alpha - \sum_i \sum_{j>i} \frac{\partial \phi_{ij}}{\partial B_{\alpha\beta}} - P_{\text{ext}} \frac{\partial}{\partial B_{\alpha\beta}} (\det \boldsymbol{B}) \right\}. \quad (6)$$

For a rigid-molecule system under periodic boundary condition, the Lagrangian in Eq. 4 becomes,

$$L = \frac{1}{2}\sum_i M_i \dot{\boldsymbol{S}}_i^{\mathrm{T}} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{B} \dot{\boldsymbol{S}}_i + \frac{1}{2} W \mathrm{tr}(\dot{\boldsymbol{B}}^{\mathrm{T}} \dot{\boldsymbol{B}}) + \frac{1}{2}\sum_i \boldsymbol{\omega}_i^{\mathrm{T}} \boldsymbol{I}_i \boldsymbol{\omega}_i - \frac{1}{2}\sum_{\boldsymbol{n}} \sum_i \sum_j (1 - \delta_{i,j}\delta_{\boldsymbol{n},\boldsymbol{0}}) \sum_{k \in i} \sum_{l \in j} \phi_{ij}^{kl}(r_{ijn}^{kl}) - P_{\text{ext}} \det \boldsymbol{B}, \quad (7)$$

where

$$\boldsymbol{r}_{ijn}^{kl} = \boldsymbol{r}_i^k - \boldsymbol{r}_j^l + \boldsymbol{B}\boldsymbol{n} \quad (8)$$

$$= \boldsymbol{B}(\boldsymbol{S}_{ij} + \boldsymbol{n}) + \boldsymbol{d}_{ij}^{kl}, \quad (9)$$

$$\boldsymbol{r}_i^k = \boldsymbol{B}\boldsymbol{S}_i + \boldsymbol{d}_i^k, \quad (10)$$

and M_i is the mass, $\boldsymbol{\omega}_i$ the angular velocity vector, $\boldsymbol{I}_i$ the inertia tensor, and $\boldsymbol{S}_i$ the fractional coordinates for the center of mass of the i-th molecule, respectively, $\boldsymbol{r}_i^k$ the Cartesian coordinates of the k-th atom belonging to the i-th molecule, $\phi_{ij}^{kl}(r_{ijn}^{kl})$ the potential energy between the k-th atom of the i-th molecule and the l-th atom of the j-th molecule separated by r_{ijn}^{kl}, and $\boldsymbol{d}_i^k$ the relative vector from $\boldsymbol{B}\boldsymbol{S}_i$ to $\boldsymbol{r}_i^k$, and $\boldsymbol{d}_{ij}^{kl} = \boldsymbol{d}_i^k - \boldsymbol{d}_j^l$. Each component of $\boldsymbol{n}$ is an integer for the repeated symmetry translations, and δ is Kronecker's δ. Thus, the following equations of translational motion, rotational motion and lattice vectors for the rigid-molecule system are coded in the program MDCP:[9]

$$\ddot{\boldsymbol{S}}_i = -M_i^{-1} \sum_{\boldsymbol{n}} \sum_{j \neq i} \sum_{k \in i} \sum_{l \in j} (\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B})^{-1} \frac{\partial \phi_{ij}^{kl}}{\partial \boldsymbol{S}_i} - (\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B})^{-1} (\dot{\boldsymbol{B}}^{\mathrm{T}} \boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}} \dot{\boldsymbol{B}}) \dot{\boldsymbol{S}}_i, \quad (11)$$

$$\dot{\boldsymbol{J}}_i = \boldsymbol{N}_i = \sum_{k \in i} \boldsymbol{d}_i^k \times \left\{ -\sum_{\boldsymbol{n}} \sum_j (1 - \delta_{i,j}\delta_{\boldsymbol{n},\boldsymbol{0}}) \sum_{l \in j} \frac{\partial \phi_{ij}^{kl}}{\partial \boldsymbol{r}_i^k} \right\}, \quad (12)$$

$$\ddot{B}_{\alpha\beta} = W^{-1} \left\{ \sum_i M_i (\boldsymbol{B}\dot{\boldsymbol{S}}_i)_\alpha (\dot{\boldsymbol{S}}_i)_\beta - \frac{1}{2}\sum_{\boldsymbol{n}} \sum_i \sum_j (1 - \delta_{i,j}\delta_{\boldsymbol{n},\boldsymbol{0}}) \sum_{k \in i} \sum_{l \in j} \frac{\partial \phi_{ij}^{kl}}{\partial B_{\alpha\beta}} - P_{\text{ext}} \frac{\partial}{\partial B_{\alpha\beta}} (\det \boldsymbol{B}) \right\}. \quad (13)$$

Here $\boldsymbol{J}_i$ is the angular momentum and $\boldsymbol{N}_i$ the torque vector of the i-th molecule. The derivatives of ϕ_{ij}^{kl} (r_{ijn}^{kl}) in Eqs. 11, 12, and 13 read

$$(\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B})^{-1} \frac{\partial \phi_{ij}^{kl}}{\partial \boldsymbol{S}_i} = \frac{\partial \phi_{ij}^{kl}}{\partial r_{ijn}^{kl}} \frac{1}{r_{ijn}^{kl}} (\boldsymbol{S}_{ij} + \boldsymbol{n} + \boldsymbol{B}^{-1} \boldsymbol{d}_{ij}^{kl}), \quad (14)$$

$$\frac{\partial \phi_{ij}^{kl}}{\partial \boldsymbol{r}_i^k} = \frac{\partial \phi_{ij}^{kl}}{\partial r_{ijn}^{kl}} \frac{1}{r_{ijn}^{kl}} \boldsymbol{r}_{ijn}^{kl}, \quad (15)$$

and

$$\frac{\partial \phi_{ij}^{kl}}{\partial B_{\alpha\beta}} = \frac{\partial \phi_{ij}^{kl}}{\partial r_{ijn}^{kl}} \frac{1}{r_{ijn}^{kl}} (\boldsymbol{S}_{ij} + \boldsymbol{n})_\beta (\boldsymbol{r}_{ijn}^{kl})_\alpha. \quad (16)$$

Equation 13 implies that the $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ vectors fluctuate around a certain equilibrium point in balance between the internal and external pressures.

References

1) For example: J. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, **111**, 8576 (1989).

2) A. J. Pertsin and A. I. Kitaigorodsky, "The Atom–Atom Potential Method," Springer-Verlag, Berlin (1987), pp. 283—337.

3) H. R. Karfunkel and R. J. Gdanitz, *J. Comput. Chem.*, **13**, 1171 (1992).

4) A. Gavezzotti, *J. Am. Chem. Soc.*, **113**, 4622 (1991).

5) J. R. Holden, Z. Du, and H. L. Ammon, *J. Comput. Chem.*, **14**, 422 (1993).

6) V. K. Belsky and P. M. Zorkii, *Acta Crystallogr., Sect. A*, **A33**, 1004 (1977).

7) N. Tajima and T. Hirano, "Reactivity in Molecular Crystal," ed by Y. Ohashi, Kodansha, Tokyo (1993), pp. 9—23.

8) S. Nose and M. L. Klein, *J. Chem. Phys.*, **78**, 6928 (1983).

9) MDCP: N. Tajima, T. Tanaka, T. Arikawa, T. Sakurai, S. Teramae, and T. Hirano, "JCPE (Japan Chemistry Program Exchange), Program #064," 1992.

10) P. P. Ewald, *Ann. Physik.*, **64**, 253 (1921).

11) A. Rahman, *Phys. Rev., Sect. A, Second Ser.*, **136A**, 405 (1964).

12) S. R. Cox, L. Y. Hsu, and D. E. Williams, *Acta Crystallogr., Sect. A*, **A37**, 293 (1981).

13) W. H. Keesom and J. W. L. Kohler, *Physica (Utrecht)*, **1**, 655 (1934).

14) D. Hall and D. E. Williams, *Acta Crystallogr., Sect. A*, **A31**, 56 (1975).

15) G. E. Bacon, N. A. Curry, and S. A. Wilson, *Proc. R. Soc. London, Ser. A*, **279**, 98 (1964).

16) G. J. Piermarini, A. D. Mighell, C. E. Weir, and S. Block, *Science*, **165**, 1250 (1969).

17) D. E. Williams and S. R. Cox, *Acta Crystallogr., Sect. B*, **B40**, 404 (1984).

18) S. Furberg, J. Grogaard, and B. Smedsrud, *Acta Chem. Scand., Ser. B*, **33**, 715 (1979).

19) M. Parrinello and A. Rahman, *J. Appl. Phys.*, **52**, 7182 (1981).