ELSEVIER

A further implementation of the rotational symmetry boundary conditions for calculations of $P4_32_12$ symmetry crystals

Shigetaka Yoneda

School of Science, Kitasato University, Sagamihara-Shi, Kanagawa-Ken 228, Japan

One of the most accurate styles of protein simulation is to calculate proteins in crystalline environment without neglect of long-range interactions. The long-range interactions can be accelerated by various methods. However, as a unit cell of a protein crystal is a large molecular assembly, its simulation is still unpractical without high-speed computers. Thus this article is addressed to the reduction of calculational volumes for protein crystal simulation by a further implementation of the rotational symmetry boundary condition method. For protein crystals in P4_32_12 *symmetry, a computational cell and related tables were developed. A 120-ps molecular dynamics simulation was performed for a* P4_32_12 *symmetry crystal of glycogen phosphorylase* b *under rotational symmetry boundary conditions. The computational cell was one-eighth of the unit cell in volume, and less than about one-fourth of the conventional periodic boundary box. Generation of neighbor atom pair lists was greatly accelerated, and thus the simulation was practical even with a personal computer.*

Keywords: rotational symmetry boundary condition, molecular dynamics simulation, protein crystal, protein assembly, glycogen phosphorylase b

INTRODUCTION

Simulation in the actual crystallographic environment without neglect of long-range electrostatic interactions is one of the most accurate computational approaches to reproduce the experimental structures and dynamics of biological macromolecules. In optimal situations, better than 1-Å accuracy can be achieved, as reported for *Streptomyces griseus* protease A in a molecular dynamics (MD) simulation with $P4_2$ crystallographic symmetry and 25-Å cutoff.[1] The high accuracy is, however, accompanied by a great increase in the central processing unit (CPU) time, for calculations of long-range interactions and by a large unit cell volume necessary for the crystallographic environment. As various methods of accelerating calculations of long-range interactions have been developed, the CPU time for long-range interactions can be much reduced.[2–5] However, the large volume of a unit cell, which is usually an assembly of several copies of a protein, is still an obstacle for rapid calculations. Thus this article addresses acceleration of simulation in the crystallographic environment by reducing the volume to be calculated.

The acceleration is based on the rotational symmetry boundary condition (RSBC) proposed by Cagin et al.[6] The method was implemented by Yoneda et al. in an MD simulation program, APRICOT,[7] for calculations of an icosahedrally symmetrical rhinovirus capsid.[8] In their implementation, the generation of neighbor atom pair lists (neighbor lists) was accelerated by about 100 times, thus making the MD simulation of capsids practical even with readily available, inexpensive computers.

For the second implementation of the RSBC, we developed the necessary tables and definitions for protein crystal simulation. Among the 230 crystallographic space groups, tetragonal $P4_32_12$ symmetry is selected as the target of the development, as many X-ray structures of protein crystals in $P4_32_12$ have been determined. Practical calculations are performed for a crystal of glycogen phosphorylase *b* (GP*b*). GP*b*, an enzyme with a molecular weight of 98 000, catalyzes the intracellular degradation of glycogen into glucose 1-phosphate. The control mechanism of GP*b* by phosphorylation in signal transduction processes has been extensively studied. Moreover, GP*b* is an allosteric enzyme active in a symmetric dimer form. Thus, GP*b* is an appropriate target for MD studies with symmetry consideration.

Color Plate for this article is on page 260.

Address reprint requests to: Dr. S. Yoneda, School of Science, Kitasato University, 1-15-1 Kitasato, Sagamihara-Shi, Kanagawa-Ken 228, Japan.
Received 4 July 1997; accepted 10 September 1997.

655 Avenue of the Americas, New York, NY 10010

1093-3263/97/$17.00
PII S1093-3263(97)00081-8

Before the presentation of this study, we briefly review the principle of the RSBC, and we also note that a more detailed formulation was presented previously.[8] In principle, calculation of a symmetric molecular assembly can be greatly accelerated, if only a single constituent molecule is calculated instead of the entire assembly. The intramolecular interactions in the constituent molecule are calculable by conventional energy calculation methods. The intermolecular interactions are also easily calculable, if the nonbonded interaction is truncated at a reasonable cutoff distance R_c and the neighbor list is already generated. For example, the conventional R_c of about 10 Å, if adopted, is much smaller than the size of the proteins constituting a virus capsid. Thus the number of interactions (atom pairs) from the neighbor molecules is much smaller than the number of intramolecular interactions. However, this straightforward approach carries a large computational load for the generation of neighbor lists, because it is difficult to determine which neighbor molecules make contact with the central molecule, as the molecular shape is complicated and varies with the simulation. All the numerous neighbor molecules are candidates to contribute intermolecular interactions, and thus the atomic distance calculation for neighbor list generation is time-consuming. The neighbor molecules can be restricted by the RSBC. In the RSBC, a symmetric assembly is divided into spatial frames termed *computational cells* (CCs). The CCs are equivalent to "asymmetric units" in crystallographic terminology, and should be simple in shape. Calculations are performed for the molecular structure within a CC. As the shape of the CC is fixed, it is possible to predetermine which CCs are in contact with the central CC. For the icosahedral rhinovirus capsid, a CC makes contacts with only eight neighbor CCs. Thus, the atomic distance calculations for the neighbor list generation are much reduced. However, the molecular structure within a CC is a mixture of fragments of different molecules. Thus it is necessary to track the molecular numbers of the atoms. For this purpose, a combination of two numbers, denoted k_p, is used, where k is the sequential atom number defined in the molecule and the subscript p is the molecular number. Another notation, k^c, where the superscript c is the CC number, is also used. As the positions of the molecules uniquely correspond to the symmetry operators, the molecular numbers follow the rules of group theory. The symbol $\oplus$ is used for multiplication of two molecular numbers, and also for multiplication of two CC numbers. The molecular numbers are trackable with Eq. (1):

$$\mathrm{pr}(k^c) \equiv c \oplus [-\mathrm{cl}(k_0)] \qquad (1)$$

where the functions pr() and cl() present the molecular number and the CC number, respectively. The symbol $\equiv$ represents equality in the group. Further, if Eq. (2) is satisfied, atoms i^0 and j^d are in the same molecule:

$$d \equiv [-\mathrm{cl}(i_0)] \oplus \mathrm{cl}(j_0) \qquad (2)$$

Equations (1) and (2) are the protomer number equations; the biochemical term *protomer* is used instead of *molecule* for the sake of precision. In addition, the border residue flags, brf(a,k), were introduced for RSBC calculations.[8] If a residue *a* is near boundary k, then brf(a,k) is operative; otherwise brf(a,k) is inoperative. The criterion for the proximity is known as the border cutoff. The border cutoff is set at longer than R_c. With brf(a,K), numerous pairs of atoms can be omitted from the distance calculations for the neighbor list generation, and thus calculations in the RSBC are further accelerated.

DEVELOPMENT OF TABLES AND DEFINITIONS FOR $P4_32_12$ SYMMETRY

Regular quadrangular cylinders, shown in Figure 1, are the CCs adopted for the RSBC calculations. The CCs were numbered from -1 to 25, as shown in Figure 2. The central CC was numbered 0. As the CCs numbered 5, 11, 17, and 23 do not make direct contact with the central CC, they were not considered in the calculations. All the necessary operators of rotation and translation were prepared in the APRICOT program. The multiplication table concerning the product $\oplus$ was also prepared to apply the protomer number equations. The boundaries of the CCs were defined as shown in Figure 3. The contact relationship of the boundaries of the central CC and the neighbor CCs was summarized as shown in Table 1. Table 1 and Figure 3 are necessary for the border residue flag method.

PRACTICAL CALCULATIONS

Calculational procedure

An MD simulation of a crystal of GP*b* under rotational symmetry boundary conditions was performed to estimate the degree of acceleration and to confirm the integrity of the developed program. The initial coordinates were taken from the X-ray analysis of the T state GP*b*–AMP complex from rabbit skeletal muscle,[9] deposited as 8GPB in the Protein Data Bank.[10] As the size of the unit cell is $128.5 \times 128.5 \times 116.3 Å^3$, the CC size is $64.25 \times 64.25 \times 58.15 Å^3$. Thus, considering that CCs are placed in staggered fashion in the z direction (see Figure 2), if $R^c < 58.15/2$Å (i.e., 29 Å) is used, inter-CC interactions are contributed only by the neighbor CCs, and thus the RSBC method is applicable.

First, a minimization with positional constraints was performed for the X-ray coordinates *in vacuo.* The AMBER energy parameters were used.[11] The atomic charges for the cofactor, pyrodoxamine 5-phosphate in the enol imine form, were determined considering the quantum chemical study.[12] The empty space of the CCs not occupied by the minimized GP*b* structure was filled with water molecules from the AMBER wat216.dat file. The water molecules making short contact with GP*b* residues were excluded. The short contacts were calculated in the RSBC. The total number of atoms including those of water was 16 465. A constraint minimization of GP*b* and water in the RSBC was performed. After the minimization, MD simulation in the RSBC was started. The time step used was 2 fs with the SHAKE algorithm.[13] The neighbor list was updated at every 30-step (0.06-ps) interval. A residue-based cutoff of 14 Å and a border cutoff of 19 Å were adopted. An initial 5 ps of dynamics was performed with positional constraints to the protein atoms. The constraints decreased stepwise with time and no constraints were added after 5 ps. The temperature was gradually increased and was kept at 300 K after 10 ps. A series of nine simulations of length 10–25 ps was calculated, starting from the coordinates and velocities of the preceding simulations. The simulations were calculated with a personal computer with LINUX and a Pen-

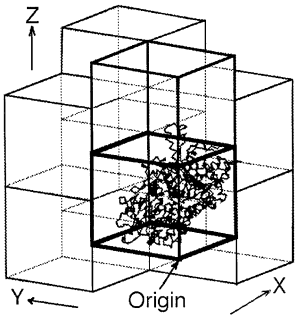

Figure 1. Eight asymmetric units in $P4_32_12$ *symmetry. The assembly of those eight frames is not of box shape, but fills the entire crystal by parallel translation, and thus is a unit cell. The asymmetric units are used for the CC for the RSBC calculations. The CC numbers are defined in Figure 2. The* C_α *carbons of glycogen phosphorylase* b *are drawn within the CC numbered 0. The directions of the axes adopted for the calculations are denoted as x, y, and z.*

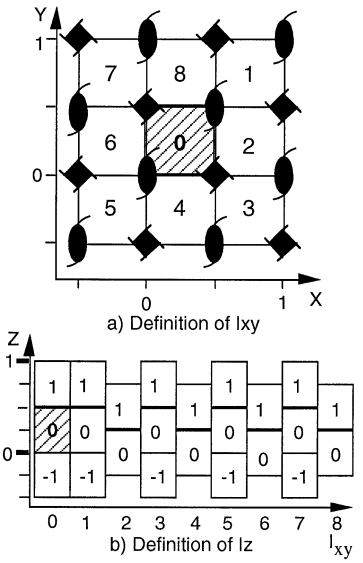

Figure 2. CC number defined as $3I_{xy} + I_z$*, where* I_{xy} *is determined by the xy position of the CC (a), and* I_z *by the z position (b). The unit of the length is the unit cell length in each direction. The 23 CCs are related with the rotations and parallel translations for* $P4_32_12$*. Not all the rotation axes are shown.*

tium (200 MHz) and older workstation The length of the entire simulation was 120 ps.

Calculational results

The final structure of the 120-ps MD simulation is shown in Color Plate 1.[14] Figure 4 shows the time profile of the root mean square (r.m.s.) deviations. A plateau is shown after 70 ps in Figure 4, indicating the equilibration of the simulation for global structural properties of GP*b*. The component analysis of energy at 120 ps is shown in Table 2. Not only the nonbonded energy but also the energies of bonds, angles, and dihedrals are decomposed into the inter-CC and intra-CC interactions in the APRICOT program. The electrostatic and Lennard–Jones components of the inter-CC interactions are about one-tenth those of the intra-CC interactions, respectively.

The CPU time for a 15-ps simulation was about 93 h when the Pentium personal computer was used. Thus if the entire 120-ps simulation is calculated with the personal computer, the total CPU time is 1 month (31 days). The CPU time consumption was itemized with a short simulation. The CPU time required to generate a neighbor list was 210.8 s. As the neighbor list was regenerated at every 30-step interval, the averaged CPU time of the generation is 7.0 s for each step. The CPU time to calculate nonbonded interaction was 35.3 s. The CPU time of miscellaneous calculations for integration and for the energies of bonds, angles, and dihedrals was 1.1 s. The total of 43.4 s consumed for each step is approximately consistent with 1 month for the 120-ps simulation.

DISCUSSION

The CCs used in the RSBC method are the conceptual frames similar to the periodic boxes used in the periodic boundary method. However, the CCs can be smaller than the periodic boxes. For example, the size of the CC used in this study is $64.25 \times 64.25 \times 58.15 = 240\,047 Å^3$ for the GP*b* molecule, which spans a space of $80 \times 82 \times 81 Å^3$. Thus many residues of GP*b* protrude from the CC as shown in Color Plate 1. However, the calculations were performed for the molecular structure within the CC using the RSBC. On the other hand, in calculations for proteins under periodic boundary conditions, it is usual to assume that the proteins are isolated under infinitely dilute conditions. Thus the periodic boundary calculations are impossible if the box is smaller than the space spanned by the protein. For example, the box for periodic boundary simulations of a GP*b* must be larger than $96 \times 98 \times 97 = 912\,576 Å^3$, which is the size of the space spanned by the GP*b* added to the cutoff distance R_c of 14 Å and a surplus 2 Å for surety for the protein fluctuations. The volume is about 3.8 times larger than the volume of the CC of the RSBC. Any boxes smaller than this would not exclude direct interactions between the image GP*b* molecules, and thus the calculations would be erroneous. If the same criterion of the water addition in this study is used, the periodic box includes a total of 74 887 atoms, about 4.5 times more than the 16 465 atoms included in the RSBC calculations. For estimation from a minor calculation, generation of a neighbor list in the periodic box consumed 6 383 s and other calculations consumed 201 s. Thus 6383/30 + 201 = 414s was consumed for each averaged step, and an impractical length of 288 days will be necessary for a 120-ps simulation with the personal computer. Furthermore, as GP*b* is active in the dimer form, dimer simulations seem more probable under periodic boundary conditions. As the GP*b* dimer is $86 \times 86 \times 121 Å^3$, the necessary size of the periodic box is $102 \times 102 \times 137 Å^3$ and the number of the atoms would be 115 931. Thus additional CPU time would be necessary for the dimer simulation. In most ideal simulations of a unit cell under periodic boundary conditions, the dilute solution conditions are not assumed, and thus the box size is equal to the unit cell size, which is eight times larger than a CC. The unit cell size of the GP*b* crystal, $128.5 \times 128.5 \times 116.3 Å^3$, is large for present practical simulations.

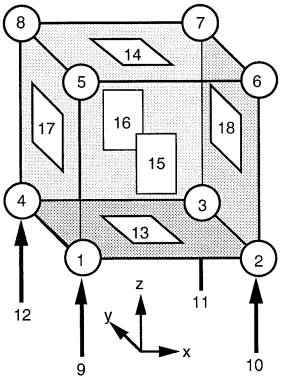

Figure 3. Eighteen boundaries used for the border residue flag method. Eight boundaries are vertices, four are edges, and six are planes.

Table 1. Contacting boundaries[a]

Boundaries of surrounding cells	Boundaries of central cell
13p of (0, −1)	13p
14p of (0, 1)	14p
7v of (1, −1)	3v
11e of (1, 0)	11e
3v of (1, 1)	7v
17pl of (2, 0)	18pl
15pl of (2, 1)	18pu
6v of (3, −1)	2v
12e of (3, 0)	10e
2v of (3, 1)	6v
18pu of (4, 0)	15pl
16pu of (4, 1)	15pu
5v of (5, −1)	1v
9e of (5, 0)	9e
1v of (5, 1)	5v
18pl of (6, 0)	17pl
16pl of (6, 1)	17pu
8v of (7, −1)	4v
10e of (7, 0)	12e
4v of (7, 1)	8v
17pu of (8, 0)	16pl
15pu of (8, 1)	16pu

[a] The boundaries in the same columns are in contact. The numbers in parentheses are I_{xy} and I_z, defined in Figure 2. The letters v, e, p, pu, and pl, after the boundary numbers, indicate the vertex, edge, plane, upper part of the plane, and lower part of the plane, respectively.

The r.m.s. deviations were equilibrated after 70 ps, as shown in Figure 4. The equilibrated value of about 2.5 Å is a little larger than those from the previous studies.[1,15] However, the molecular weights of the proteins in the previous studies are less than a few ten thousandths, much smaller than that of GP*b*. Furthermore the cutoff lengths used were longer. Thus the same results are not expected.

The CC shape selected is not unique, and other selections of CC shape are possible. The most important factor for the selection is smoothness and compactness, which make a long cutoff distance possible in the RSBC. The definition of the CC and the related tables in this study are a good selection, as an R_c of 29 Å is available.

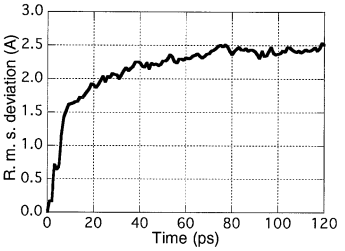

Figure 4. Time profile of the r.m.s. deviations from the 0-ps structure. All heavy atoms with an experimental B factor of less than 20 Å2 were included in the calculation.

Symmetry constraints may obscure asymmetric properties, but they are a conventional tool to study matters related to symmetry. In RSBC calculations, intermolecular interactions in crystallographic symmetry are correctly considered by using a small CC. Furthermore, the calculations can be greatly accelerated in the RSBC. The RSBC method is much faster and appropriate for studies of protein crystals, especially when problems related to intermolecular interactions are addressed. Additional implementation of various methods of calculating long-range interactions[2–5] will further accelerate the calculations. Thus, although the development of programs for RSBC calculations is not straightforward, further implementations of the RSBC to other types of crystallographic symmetry will be helpful for the studies of symmetrical macromolecular assemblies.

CONCLUSIONS

The RSBC method was implemented in the MD simulation program, APRICOT, for calculations in $P4_32_12$ symmetry. The necessary tables and definitions were developed. As the calculational volume was reduced and the neighbor list generation was greatly accelerated, the CPU time of the 120-ps MD simulation of the GP*b* crystal in the RSBC was only 1 month

Table 2. Component analysis of energy at 120 ps

Component	Total[a]	Intra-CC[a]	Inter-CC[a]
Electrostatic	−69 285.8	−61 430.9	−7 854.9
Lennard–Jones	−5 880.3	−5 317.2	−563.1
H bond	−962.1	−876.9	−85.2
Bond	2 364.5	2 346.7	17.8
Angle	3 413.4	3 360.6	52.8
Dihedral	2 152.9	2 110.0	42.9
1–4 LJ	1 883.7	1 831.4	52.3
1–4 el	24 801.2	23 612.8	1 188.4
Improper dihedral	759.4	734.9	24.5
Kinetic	14 803.7		

[a] Energy in kcal/mol.

with a personal computer. The CC was one-eighth the unit cell in size, and less than about one-fourth of the conventional periodic boundary box. The CPU time was about one-tenth of those under the conventional periodic boundary simulation of a molecule of the same protein. The MD simulation of GP*b* under the RSBC was equilibrated after 70 ps. Computational study of macromolecular assemblies with consideration of intermolecular interaction is becoming more practical as a result of improving computing machinery and this theoretical development.

REFERENCES

1 Kitson, D.H., Avbelj, F., Moult, J., Nguyen, D.T., Mertz, J.E., Hadzi, D., and Hagler, A.T. On achieving better than 1-Å accuracy in a simulation of a large protein: *Streptomyces griseus* protease A. *Proc. Natl. Acad. Sci. U.S.A.* 1993, **90,** 8920–8924

2 Mathiowetz, A.M., Jain, A., Karasawa, N., and Goddard, W.A., III. Protein simulations using techniques suitable for very large systems: The cell multipole method for nonbond interactions and the Newton–Euler inverse mass operator method for internal coordinate dynamics. *Proteins* 1994, **20,** 227–247

3 Saito, M. Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. *J. Chem. Phys.* 1994, **101,** 4055–4061

4 York, D.M., Wlodawer, A., Pedersen, L.G., and Darden, T.A. Atomic-level accuracy in simulations of large protein crystals. *Proc. Natl. Acad. Sci. U.S.A.* 1994, **91,** 8715–8718

5 Cheatham, T.E., III, Miller, J.L., Fox, T., Darden, T.A., and Kollman, P.A. Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* 1995, **117,** 4193–4194

6 Cagin, T., Holder, M., and Pettitt, B.M. A method for modeling icosahedral virions: Rotational symmetry boundary conditions. *J. Comput. Chem.* 1991, **12,** 627–634

7 Yoneda, S., and Umeyama, H. Free energy perturbation calculations on multiple mutation bases. *J. Chem. Phys.* 1992, **97,** 6730–6736

8 Yoneda, S., Kitazawa, M., and Umeyama, H. Molecular dynamics simulation of a rhinovirus capsid under rotational symmetry boundary conditions. *J. Comput. Chem.* 1996, **17,** 191–203

9 Barford, D., Hu, S.-H., and Johnson, L.N. Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. *J. Mol. Biol.* 1991, **218,** 233–260

10 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–542

11 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, Jr., S., and Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 1984, **106,** 765–784

12 Wampler, J.E. Electrostatic potential derived charges for enzyme cofactors: Methods, correlations, and scaling for organic cofactors. *J. Chem. Inf. Comput. Sci.* 1995, **35,** 617–632

13 van Gunsteren, W.F., and Berendsen, H.J.C. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 1977, **34,** 1311–1327

14 Sayle, R.A., and Milner-White, E.J. RASMOL: Biomolecuar graphics for all. *Trends Biochem. Sci.* 1995, **20,** 374–376

15 York, D.M., Darden, T.A., Pedersen, L.G., and Anderson, M.W. Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution. *Biochemistry* 1993, **32,** 1443–1453