### STAP ARTICLE

Linear-scaling first-principles molecular dynamics of complex biological systems with the Conquest code

To cite this article: Takao Otsuka et al 2016 Jpn. J. Appl. Phys. 55 1102B1

View the article online for updates and enhancements.

# The Japan Society of Applied Physics

### You may also like

- First-principles adaptive-boost accelerated molecular dynamics simulation with effective boost potential construction methods: a study of Li diffusion in Si crystal Masahiro Yamamoto, Akio Ishii, Shuhei Shinzato et al.
- Force probe simulations using an adaptive resolution scheme
  Marco Oestereich, Jürgen Gauss and Gregor Diezemann
- Prediction of solute diffusivity in Al assisted by first-principles molecular dynamics
  O M Løvvik, E Sagvolden and Y J Li

This content was downloaded from IP address 3.144.77.71 on 28/04/2024 at 06:46

Nanoarchitectonics for manipulation of atom, molecule, and materials

# Linear-scaling first-principles molecular dynamics of complex biological systems with the CONQUEST code

Takao Otsuka<sup>1</sup>, Makoto Taiji<sup>1</sup>, David R. Bowler<sup>2,3,4</sup>, and Tsuyoshi Miyazaki<sup>2,3,5</sup>

<sup>1</sup>Quantitative Biology Center (QBiC), RIKEN, Suita, Osaka 565-0874, Japan

<sup>2</sup>International Center for Materials Nanoarchitechtonics (MANA), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki 305-0044, Japan <sup>3</sup>Department of Physics and Astronomy, University College London (UCL), London WC1E 6BT, U.K.

<sup>4</sup>London Centre for Nanotechnology, University College London (UCL), London WC1H 0AH, U.K.

<sup>5</sup>Division of Bio-organometallics, Research Institute for Science and Technology, Tokyo University of Science, Noda, Chiba 278-8510, Japan

Received June 29, 2016; accepted July 29, 2016; published online September 21, 2016

The recent progress of linear-scaling or  $\mathcal{O}(N)$  methods in density functional theory (DFT) is remarkable. In this paper, we show that all-atom molecular dynamics simulations of complex biological systems based on DFT are now possible using our linear-scaling DFT code CONQUEST. We first overview the calculation methods used in CONQUEST and explain the method introduced recently to realise efficient and robust first-principles molecular dynamics (FPMD) with  $\mathcal{O}(N)$  DFT. Then, we show that we can perform reliable all-atom FPMD simulations of a hydrated DNA model containing about 3400 atoms. We also report that the velocity scaling method is both reliable and useful for controlling the temperature of the FPMD simulation of this system. From these results, we conclude that reliable FPMD simulations of complex biological systems are now possible with CONQUEST. © 2016 The Japan Society of Applied Physics

#### 1. Introduction

Molecular simulation technology is now commonly used to explore biological phenomena of biomolecular systems. It helps us understand the mechanisms of various biological phenomena, including enzyme reactions, photoexcitations, molecular interactions, and so on.<sup>1)</sup> Although most molecular simulations of biological systems use parameterised interatomic potentials, the reliability of such empirical potentials in various environments or for some phenomena is sometimes doubtful. Thus, it is important that we should be able to perform molecular simulations based on quantum mechanics. However, quantum simulations, such as first-principles (FP) simulations based on density functional theory (DFT), are usually very expensive, especially for large systems. As is well known, the CPU time of a conventional DFT calculation is proportional to the cube of the number of atoms N in the simulation cell. It is very difficult and expensive to treat systems containing more than 1000 atoms within DFT. To reduce this demanding cost, the quantum mechanics and molecular mechanics (QM/MM) hybrid method or its molecular dynamics version (QM/MM-MD) is often used for molecular simulations on biological systems. However, it is usually impossible to remove the effect of the artificial boundary between the two regions introduced in the hybrid calculations. There are increasing demands for all-atom DFT simulations on complex biological systems.

In this respect, the recent advances in a computational technique for large-scale DFT calculation called the linear scaling or  $\mathcal{O}(N)$  method, whose calculation cost is only proportional to *N*, are encouraging.<sup>2)</sup> We have been developing our own linear-scaling DFT code called CoNQUEST<sup>3)</sup> and have recently demonstrated that we can treat million-atom systems with DFT using the code.<sup>4,5)</sup> We have also recently introduced a method to realise efficient and reliable first-principles molecular dynamics (FPMD) on large systems, by combining the  $\mathcal{O}(N)$  DFT and extended Lagrangian Born–Oppenheimer molecular dynamics (XL-BOMD) methods.<sup>6)</sup> We investigated the requirements for calculations with accurate  $\mathcal{O}(N)$  FPMD simulations and actually performed the FPMD simulation of a very large crystalline silicon system containing

32,768 atoms.<sup>7)</sup> We expect that we can also employ this technique for large and complex biological systems.

In this paper, we overview the calculation methods used in CONQUEST and explain the combined method for efficient and accurate FPMD simulations. Then, we show that we can perform reliable all-atom FPMD simulations of a test DNA model, a DNA decamer hydrated with a large number of water molecules, consisting of about 3400 atoms. We demonstrate that the FPMD simulations on the hydrated DNA system are robust and accurate.

## 2. Linear-scaling first-principles molecular dynamics method

#### 2.1 Linear-scaling DFT code CONQUEST

In this subsection, we first overview the computational methods and recent progress of CONQUEST.

In CONQUEST, we use the Kohn-Sham density matrix defined as

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_{n} f_{n} \Psi_{n}(\boldsymbol{r}) \Psi_{n}^{*}(\boldsymbol{r}'), \qquad (1)$$

where  $\Psi_n(\mathbf{r})$  is the eigenfunction (Kohn–Sham orbital) of the Kohn–Sham Hamiltonian for the band index *n* and  $f_n$  is its occupation number.<sup>8–10)</sup> The total energy based on DFT can be calculated from the density matrix using the pseudopotential method and a standard exchange–correlation functional such as the local density approximation (LDA) or a generalised gradient approximation (GGA). Note that an efficient technique for calculating the exact exchange term has recently been introduced<sup>11)</sup> to the code, and thus hybrid functionals are also now available.

In CONQUEST, we represent the density matrix by localised orbitals called "support functions",  $\phi_{i\alpha}(\mathbf{r})$ , with the matrix elements  $K_{i\alpha,j\beta}$ , which are the coefficients of the density matrix expressed in this nonorthogonal basis of support functions.

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{i\alpha,j\beta} \phi_{i\alpha}(\mathbf{r}) K_{i\alpha,j\beta} \phi_{j\beta}(\mathbf{r}').$$
(2)

The support function  $\phi_{i\alpha}(r)$  for the orbital  $\alpha$  is centred on the atom *i* and is nonzero only inside the "support region".



STAP ARTICLE



**Fig. 1.** Total energy of a hydrated DNA system, whose structure is shown in Fig. 2, as a function of  $R_L$ , the cutoff range of the auxiliary density matrix *L*. The total energy obtained by the non-self-consistent technique (NonSCF) is also presented.

The support functions themselves are represented in terms of basis functions, and two types of basis sets are available in CONQUEST: B-splines on regular grids<sup>12)</sup> and numerical pseudo-atomic orbitals (PAOs).<sup>13-15)</sup> When B-splines are used, we can systematically improve the accuracy of the basis set by reducing the grid spacing and can reach the planewave accuracy. On the other hand, we can employ efficient calculations with a reasonable accuracy by using PAOs as basis sets. Even with PAO basis sets, we can improve the accuracy by increasing the number of basis functions, but the computational cost usually increases very rapidly. We have recently introduced a method to treat such accurate but large PAO basis sets efficiently.<sup>16-18)</sup> With this method, called the multisite support function (MSSF) method, we can perform accurate calculations without increasing the CPU time significantly.

The matrix  $K_{i\alpha,j\beta}$  is obtained either by the conventional diagonalization method, or by a linear-scaling [or  $\mathcal{O}(N)$ ] method. In the case of  $\mathcal{O}(N)$  calculations, CONQUEST uses the density matrix minimisation (DMM) method proposed by Li et al.<sup>19</sup> In this method, we express the matrix *K* following McWeeny's purification transformation,<sup>20</sup>

$$K = 3LSL - 2LSLSL, \tag{3}$$

to impose weak idempotency on the density matrix. Here, the matrix L is called the auxiliary density matrix and S $(S_{i\alpha,j\beta} = \langle \phi_{i\alpha} | \phi_{j\beta} \rangle)$  is the overlap matrix between the support functions. To achieve the  $\mathcal{O}(N)$  behaviour using the locality of the density matrix, we introduce a spatial cut-off  $R_L$  on the *L*-matrix,  $L_{i\alpha,j\beta} = 0$  for  $|\mathbf{R}_i - \mathbf{R}_j| > R_L$ , where  $\mathbf{R}_i$  are the atomic positions. Then, we calculate the matrix elements  $L_{i\alpha,i\beta}$ , which minimise the DFT total energy using numerical optimisation methods, such as the residual minimisation method.<sup>21,22)</sup> One of the advantages of the DMM method is that it satisfies the variational principle and we can monitor the accuracy of the  $\mathcal{O}(N)$  calculations by checking the  $R_L$ dependence of the total energy. Figure 1 shows the  $R_L$ dependence of the total energy for a DNA system, which is the target of the MD study shown in the next section. Here, the total energy using the Harris-Foulkes functional obtained by non-self-consistent (NSC) technique<sup>23,24)</sup> is also presented together with the DFT total energy using the self-consistentfield (SCF) charge density. This graph shows that the total energy converges as the cutoff applied to the L matrix is increased, regardless of self-consistency. We can also see that the result with the NSC technique converges faster than the SCF result. This is probably due to the fact that the electronic structure obtained by the NSC technique usually has a larger energy gap and is more localised than that obtained using the SCF. We also note that tests on smaller systems (dry DNA with NSC) show that it converges to the exact diagonalisation result.<sup>25)</sup>

Another strong point of CONQUEST is its excellent efficiency on massively parallel computers. Since CONQUEST uses the locality of the electronic structure, it also has an advantage in parallelisation. We have recently reported its parallel efficiency on the K computer and showed that it has almost ideal parallel efficiency even when we use more than 200,000 cores.<sup>5,26)</sup> It was also demonstrated that we can now treat million-atom systems using CONQUEST on such large-scale parallel computers. Using this ability of the code, the code has been used for structure relaxations of the nanoscale systems of semiconductor surfaces.<sup>27,28)</sup>

#### 2.2 Molecular dynamics with the CONQUEST code

Even though we can now calculate the total energy and atomic forces<sup>29,30)</sup> of very large systems using  $\mathcal{O}(N)$  DFT method, this does not guarantee that stable, efficient, and accurate FPMD simulations are also practically possible. There are two methods widely used in conventional FPMD simulations: Car-Parrinello MD (CPMD) and Born-Oppenheimer MD (BOMD) methods. To realise  $\mathcal{O}(N)$  FPMD simulations, we adopt the BOMD method to eliminate the ambiguity of fictitious mass, which is used in CPMD simulations. Another advantage of the BOMD method is we can use a larger time step than the CPMD method. However, note that the stability or accuracy of BOMD simulations strongly depends on the accuracy of the calculated forces. We usually use an iterative method to calculate the ground state of the electronic structure even in conventional methods. It is well known that we can have an unphysical energy drift, if the electronic structure is not well converged and the calculated forces are not sufficiently accurate. This problem is closely related to the time reversibility of the optimised electronic structure. To solve this problem, Niklasson and coworkers have recently proposed a new method called the XL-BOMD method. With this method, the time reversibility of the electronic structure is maintained and the stability of the BOMD simulations is greatly improved.<sup>6,31,32)</sup>

Recently, we have combined this XL-BOMD scheme with the DMM method and demonstrated that the combined method enables us to perform efficient and reliable FPMD simulations with the  $\mathcal{O}(N)$  method.<sup>7)</sup> The Lagrangian in the XL-BOMD scheme  $\mathcal{L}^{XBO}$  is defined as follows, using the Lagrangian in the usual BOMD method  $\mathcal{L}^{BO}$ ,

$$\mathcal{L}^{\text{XBO}}(X, \dot{X}, \boldsymbol{R}, \dot{\boldsymbol{R}}) = \mathcal{L}^{\text{BO}}(\boldsymbol{R}, \dot{\boldsymbol{R}}) + \frac{\mu}{2} \operatorname{Tr}[\dot{X}^2] - \frac{\mu \omega^2}{2} \operatorname{Tr}[(LS - X)^2], \quad (4)$$

where the matrix X is the sparse matrix introduced to prepare the initial guess of the L matrix at each MD step.  $\mu$  is the fictitious electronic mass and  $\omega$  is the curvature of the electronic harmonic potential. As in the original XL-BOMD method, if we take the limit  $\mu \rightarrow 0$ ,  $\mathcal{L}^{XBO}$  becomes  $\mathcal{L}^{BO}$  and we have two equations of motion for nuclear positions and X. If we apply the time-reversible Verlet scheme to calculate *X* using the equation of motion, we have

$$X(t + \delta t) = 2X(t) - X(t - \delta t) + \delta t^2 \omega^2 (L(t)S(t) - X(t)),$$
(5)

which shows that X(t) is time-reversible and evolves in a harmonic potential centred around the ground-state L(t)S(t). We can expect that a good initial guess for the *L*-matrix, which will obey time reversal symmetry, can be calculated by multiplying *X* and  $S^{-1}$  (in CONQUEST, the sparse approximate inverse *S* is computed using Hotelling's method<sup>33</sup>). As a result, the optimised *L* matrix also satisfies time reversibility and the trajectories of FPMD simulations become stable and accurate. In practice, for the numerical propagation of the matrix *X*, we use an equation of motion with a dissipative term to maintain the numerical stability of the matrix X.<sup>34</sup>)

In our previous study, we clarified the effects of control parameters used in the DMM method in the FPMD simulations. We found that, even when the total energy is not fully converged, MD trajectories are almost the same as those in more accurate MD simulations. We also demonstrated that reliable MD simulations can be actually performed on a 32,768-atom crystalline silicon system using 1024 CPUs (8192 cores) of the K computer. Since we already showed that the parallel efficiency of CONQUEST is ideal even when using more than 200,000 cores, we can conclude that FPMD simulations on million-atom systems are now feasible using a large supercomputer such as the K computer.

## 3. All-atom FPMD simulations on a hydrated DNA system with the CONQUEST code

Although we already demonstrated the practical ability of the combined (DMM+XL-BOMD) method, the examples of FPMD simulations using CONQUEST have been limited to simple systems so far, such as crystalline silicon and bulk water. In this section, we present another example of MD simulations on a more complex system, a hydrated DNA system, whose structure is shown in Fig. 2. The system, which was studied in our previous work,<sup>25)</sup> consists of 10 DNA base pairs [d(CCATTAATGG)2 in PDB ID: 1WQZ] of 634 atoms, 9 Mg atoms being counter ions, and 932 H<sub>2</sub>O molecules, with a total of 3,439 atoms. The initial structure is prepared by classical MD simulation using AMBER9 with the force fields of PARM99<sup>35)</sup> for the DNA atoms and TIP3P for water molecules. In classical MD simulations, the system is equilibrated with a constant pressure and the structure at the last step is adopted as the initial structure of the FPMD simulation. Figure 3 shows the energy profile of the FPMD simulation in the microcanonical case (NVE simulation). In this O(N) FPMD simulation, a periodic boundary condition, a single-zeta with polarization (SZP) basis set, the Perdew-Burke-Ernzerhof (PBE)<sup>36)</sup> exchange-correlation functional, the NSC technique with the Harris-Foulkes energy functional, the cutoff range of 16 bohr for the L matrix, and the numerical integration grid cutoff of 75 Ha are used. Note that the SZP basis set used in the present MD simulations tends to show a slightly larger error than the DZP basis set. For example, the mean absolute error of the bond lengths in the adenine molecule is 0.05 Å using the SZP basis set, while it is 0.02 Å using the DZP basis set. However, we believe that the qualitative aspects reported in this paper are



**Fig. 2.** (Color online) Structure of a DNA decamer (PDB ID:1WQZ) hydrated with 934 water molecules, consisting of about 3400 atoms.



**Fig. 3.** (Color online) Energy profile in the FPMD simulation of the hydrated DNA system (NVE simulation). Total energy (green), potential energy (blue), and kinetic energy (red) are shown.

not affected by the choice of the basis set. As a dissipation term in the equation of motion for the X matrix, we consider the terms up to the order of 5. The time step is 0.5 fs and the initial temperature for the atomic velocity is 300 K.

The most important point we can see from Fig. 3 is that the total energy, which is the sum of the potential energy (DFT total energy) and kinetic energy of nuclei, is constant during the simulation. This means that the present method is reliable also for this complex system. We can also see that the potential and the kinetic energies of the system both fluctuate relatively markedly in the early stage (up to about 50 MD steps). This probably shows a rapid response to the differences of chemical bonds between classical force fields and the CONQUEST calculation with the present conditions. However, after about 100 MD steps, these two energies show much slower changes. We do not clearly understand why we have observed such a behaviour, but we expect that using the initial structure given by classical MD simulations would help to reduce the simulation time for the equilibration of FPMD simulations.

Next, we investigate the temperature-controlled FPMD simulations by the velocity scaling method. Figure 4 shows the energy profiles of the FPMD simulation of the same



**Fig. 4.** (Color online) Energy profile of the FPMD simulation on the hydrated DNA system (NVT simulation) with the velocity scaling method at (a) 300 and (b) 600 K; total energy (green), potential energy (blue), and kinetic energy (red) calculated from the atomic velocities before the correction.

hydrated DNA system at the temperatures of (a) 300 and (b) 600 K, respectively. In this method, we simply rescale the velocity at each MD step to make the average of the kinetic energy the same as the given temperature, and use the corrected velocity in the equation of motion for nuclei. In Fig. 4, the profiles of the kinetic energy calculated with the velocities before the correction are plotted, together with the potential and total energies.

As can be seen in Fig. 4, although the fluctuation of the kinetic energy defined by the velocities before the correction is large in the early stages (0-50 MD steps), the change in kinetic energy becomes very small after 50 MD steps. In both cases, i.e., at 300 and 600 K, the kinetic energy is close to the correct temperature after around 120 steps, and the profile of the total energy becomes flat. This rapid convergence should be useful for controlling the temperature in FPMD simulations. We expect that it will also be possible to perform stable microcanonical (NVE) MD simulations around the given temperature after we employ the velocity scaling method for a short time. This may be a useful technique for FPMD simulations of similar hydrated biological systems. We have also recently implemented another method to control the temperature, the Nose-Hoover-chain method. The stability of this method with the DMM+XL-BOMD scheme will be reported elsewhere. We believe that these techniques will contribute to the realisation of many efficient, reliable, and accurate FPMD simulations of complex biological systems in the near future.

#### 4. Summary

The linear-scaling or  $\mathcal{O}(N)$  code CONQUEST has the ability to treat million-atom systems based on DFT. In this paper,

we first gave an overview of the methods used in the code and introduced recent progress, especially the newly introduced method that combines the DMM and XL-BOMD methods to realise accurate and efficient FPMD simulations with the O(N) method.

Then, we demonstrated that the method can also be applied to a complex biological system, a hydrated DNA system, containing about 3400 atoms. We showed that the total energy is conserved accurately in the microcanonical simulations when the combined method (DMM+XL-BOMD) is applied. Furthermore, we found that the velocity scaling method is useful for controlling the temperature of the FPMD simulation of this system. From these results, we can conclude that reliable FPMD simulations of complex biological systems can be performed with CONQUEST.

#### Acknowledgments

This work is partly supported by the JSPS KAKENHI projects (Grant Numbers 22790122, 26610120, and 26246021). We also acknowledge the Strategic Programs for Innovative Research (SPIRE) of MEXT and the Computational Materials Science Initiative (CMSI), Japan. The calculations were performed in part on the NIMS material numerical simulator and the COMA (PACS-IX) system of the University of Tsukuba. The authors also acknowledge Professor M. J. Gillan (University College London) for useful discussion and Dr. M. Arita (Tokyo University of Science) for his contribution to the implementation of the DMM+XL-BOMD method into CONQUEST.

- For example, Novel prize of chemistry 2013, Martin Karplus, Michael Levitt, Arieh Warshel, "for the development of multiscale models for complex chemical systems".
- 2) D. R. Bowler and T. Miyazaki, Rep. Prog. Phys. 75, 036503 (2012).
- 3) Web [http://www.linear-scaling.org/].
- D. R. Bowler and T. Miyazaki, J. Phys.: Condens. Matter 22, 074207 (2010).
- M. Arita, S. Arapan, D. R. Bowler, and T. Miyazaki, J. Adv. Simulation Sci. Eng. 1, 87 (2014).
- 6) A. M. N. Niklasson, Phys. Rev. Lett. 100, 123004 (2008).
- M. Arita, D. R. Bowler, and T. Miyazaki, J. Chem. Theory Comput. 10, 5419 (2014).
- 8) E. Hernández and M. J. Gillan, Phys. Rev. B 51, 10157 (1995).
- D. R. Bowler, T. Miyazaki, and M. J. Gillan, J. Phys.: Condens. Matter 14, 2781 (2002).
- D. R. Bowler, R. Choudhury, M. J. Gillan, and T. Miyazaki, Phys. Status Solidi B 243, 989 (2006).
- 11) L. A. Truflandier, T. Miyazaki, and D. R. Bowler, arXiv:1112.5989v2.
- 12) E. Hernández, M. J. Gillan, and C. M. Goringe, Phys. Rev. B 55, 13485 (1997).
- 13) O. F. Sankey and D. J. Niklewski, Phys. Rev. B 40, 3979 (1989).
- 14) J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter 14, 2745 (2002).
- 15) A. S. Torralba, M. Todorović, V. Brázdová, R. Choudhury, T. Miyazaki, M. J. Gillan, and D. R. Bowler, J. Phys.: Condens. Matter 20, 294206 (2008).
- 16) M. J. Rayson and P. R. Briddon, Phys. Rev. B 80, 205104 (2009).
- 17) A. Nakata, D. R. Bowler, and T. Miyazaki, Phys. Chem. Chem. Phys. 17, 31427 (2015).
- 18) A. Nakata, D. R. Bowler, and T. Miyazaki, J. Chem. Theory Comput. 10, 4813 (2014).
- 19) X. P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B 47, 10891 (1993).
- 20) R. McWeeny, Rev. Mod. Phys. 32, 335 (1960).
- 21) D. R. Bowler and M. J. Gillan, Comput. Phys. Commun. 120, 95 (1999).
- 22) D. R. Bowler and M. J. Gillan, Chem. Phys. Lett. 325, 473 (2000).
- 23) J. Harris, Phys. Rev. B 31, 1770 (1985).
- 24) W. M. C. Foulkes and R. Haydock, Phys. Rev. B 39, 12520 (1989).

- 25) T. Otsuka, T. Miyazaki, T. Ohno, D. R. Bowler, and M. J. Gillan, J. Phys.: Condens. Matter 20, 294201 (2008).
- 26) T. Miyazaki, NIMS NOW Int. 11 [9], 6 (2013).
- 27) T. Miyazaki, D. R. Bowler, R. Choudhury, and M. J. Gillan, Phys. Rev. B 76, 115327 (2007).
- 28) T. Miyazaki, D. R. Bowler, M. J. Gillan, and T. Ohno, J. Phys. Soc. Jpn. 77, 123706 (2008).
- 29) T. Miyazaki, D. R. Bowler, R. Choudhury, and M. J. Gillan, J. Chem. Phys. 121, 6186 (2004).
- 30) A. S. Torralba, D. R. Bowler, T. Miyazaki, and M. J. Gillan, J. Chem. Theory Comput. 5, 1499 (2009).
- 31) A. M. N. Niklasson, C. J. Tymczak, and M. Challacombe, Phys. Rev. Lett.

**97**, 123001 (2006).

- 32) P. Souvatzis and A. M. N. Niklasson, J. Chem. Phys. 140, 044117 (2014).
- 33) W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (Cambridge University Press, Cambridge, U.K., 1992) 2nd ed., p. 49.
- 34) A. M. N. Niklasson, P. Steneteg, A. Odell, N. Bock, M. Challacombe, C. J. Tymczak, E. Holmström, G. Zheng, and V. Weber, J. Chem. Phys. 130, 214109 (2009).
- 35) J. Wang, O. Cieplak, and P. A. Kollman, J. Comput. Chem. 21, 1049 (2000).
- 36) J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).



Takao Otsuka is a researcher of RIKEN Quantitative Biology Center (QBiC). His research field is theoretical and computational chemistry, and its application. His current interest is computational molecular design by large-scale electronic structure and first-principles molecular dynamics calculations.



Makoto Taiji is a deputy director of RIKEN Quantitative Biology Center (QBiC), and a team leader of Processor Research Team, RIKEN Advanced Institute of Computational Sciences. His research field is High-Performance Computing and its application in life sciences and drug discovery.



**David Bowler** is Professor of Physics at UCL, PI in the London Centre for Nanotechnology and PI in the MANA WPI at NIMS, Japan. He is the co-leader of the CONQUEST linear scaling DFT project, and collaborates extensively with experiment on nanostructures and growth on semiconductor surfaces. He tweets as @MillionAtomMan and blogs at www.atomisticsimulations.org.



**Tsuyoshi Miyazaki** is a MANA Principal Investigator in the International Center for Nanoarchitectonics (MANA) at National Research Institute for Materials Science (NIMS), and a group leader of First-principles Simulation Group in Nano-Theory Field of MANA. He is the co-leader of the CONQUEST project and has been working on various materials, such as organic solids, semiconductor surfaces, nano-structured materials, and biological systems. He is a head editor of Journal of

the Physical Society of Japan (JPSJ) and a visiting Professor of Research Institute for Science and Technology at Tokyo University of Science.