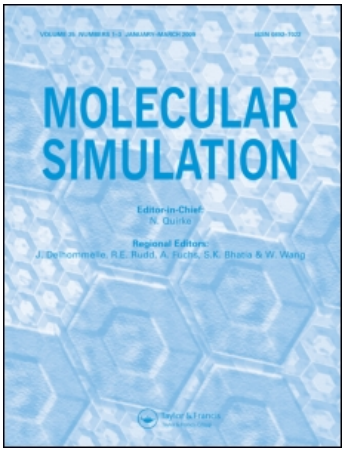

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Optimisation of OPLS-UA force-field parameters for protein systems using protein data bank

Yoshitake Sakae[a]; Yuko Okamoto[a]
[a] Department of Physics, School of Science, Nagoya University, Nagoya, Aichi, Japan

Online publication date: 10 December 2010

To cite this Article Sakae, Yoshitake and Okamoto, Yuko(2010) 'Optimisation of OPLS-UA force-field parameters for protein systems using protein data bank', Molecular Simulation, 36: 14, 1148 — 1156
To link to this Article: DOI: 10.1080/08927022.2010.507247
URL: http://dx.doi.org/10.1080/08927022.2010.507247

Molecular Simulation
Vol. 36, No. 14, December 2010, 1148–1156

Taylor & Francis
Taylor & Francis Group

Optimisation of OPLS–UA force-field parameters for protein systems using protein data bank

Yoshitake Sakae and Yuko Okamoto*

Department of Physics, School of Science, Nagoya University, Nagoya, Aichi 4648602, Japan

(*Received 18 December 2009; final version received 5 July 2010*)

We have previously proposed a method for refining force-field parameters of protein systems, which consists of minimising the summation of the square of the force acting on each atom in the proteins with the structures from the protein data bank (PDB). The results showed that the modified force-field parameters for all-atom model gave structures more consistent with the experimental implications than the original force fields. In this work, we applied this method and a new method to the OPLS–UA force field. In the new method, we perform a minimisation of the average of the root-mean-square deviation of various protein structures from the native structure. We selected some torsion-energy parameters for this optimisation, and 100 molecules from the PDB were used. The results imply that the new force-field parameters gave structures of two peptides more consistent with the experimental implications for the secondary structure-forming tendencies than the original OPLS–UA force field.

Keywords: force field; protein; folding simulation; optimisation; secondary structure

1. Introduction

Biomolecular systems are now often studied by computational methods. Especially, simulations of protein folding, protein–protein interactions and receptor-ligand docking are performed using potential energy functions with force-field parameters. Some potential energy terms and force-field parameters have been proposed and developed. For example, AMBER [1–3], CHARMM [4], OPLS [5,6], GROMOS [7] and ECEPP [8] are used for many studies of molecular simulations. Usually, these force-field parameters are determined based on experimental results for small molecules and theoretical results using quantum chemistry calculations and molecular dynamics (MD) simulations. However, the simulations using different force-field parameters will give different results. Therefore, we have performed detailed comparisons of three versions of AMBER (parm94 [1], parm96 [2] and parm99 [3]), CHARMM [4], OPLS–AA/L [6] and GROMOS [7] by generalised ensemble simulations [9] of two small peptides in an explicit solvent [10,11]. From the results of these comparisons, we saw that these force fields showed clearly different behaviours of peptides, especially, about secondary structure-forming tendencies. The folding simulations of the two peptides with implicit solvent model also showed the same secondary structure-forming tendencies [12–14].

These results imply that it is necessary to refine and improve the existing force-field parameters. Therefore, in order to yield the secondary structure-forming tendencies that agree with experimental implications, some methods of force-field refinement have been proposed. In a force field, the potential energy is usually composed of the bond-stretching term, the bond-bending term, the torsion-energy term and the non-bonded energy term. Of these energy terms, the torsion-energy term is the most problematic. For instance, the parm94, parm96 and parm99 versions of AMBER differ mainly in the backbone torsion-energy parameters. The main changes from OPLS–AA to OPLS–AA/L can also be found in the torsion-energy term [6]. The proposed methods of force-field refinement mainly concentrate on the torsion-energy terms. These modifications of the torsion energy are usually based on quantum chemistry calculations [15–19]. It was also proposed to set the backbone torsion-energy term simply to zero [20].

We have also proposed a force-field refinement method in the previous works [12–14]. This method consists of minimising the sum of the square of the force acting on each atom in the proteins with the structures from the protein data bank (PDB). We selected the partial charge and backbone torsion-energy parameters for this optimisation and refined the three versions of AMBER (parm94, parm96 and parm99), CHARMM 27 and OPLS–AA [12–14].

In this article, we introduce another optimisation method, which minimises the average of the root-mean-square deviations of various protein structures from the native structures. We applied this method and our previous method [12–14] to some torsion-energy parameters of the OPLS–UA force field.

*Corresponding author. Email: okamoto@phys.nagoya-u.ac.jp

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927022.2010.507247
http://www.informaworld.com

In Section 2, the details of the methodology are given. In Section 3, the results of applications of the optimisation method to the OPLS–UA force field are presented. Section 4 is devoted to conclusions.

2. Methods

2.1 Previous optimisation method (Method 1)

The force fields for protein systems such as AMBER, CHARMM and OPLS use essentially the same functional forms for the potential energy E_{conf} except for minor differences. The commonly used conformational potential energy E_{conf} is given by

$$E_{\text{conf}} = E_{\text{BL}} + E_{\text{BA}} + E_{\text{torsion}} + E_{\text{nonbond}}. \quad (1)$$

Here, E_{BL}, E_{BA}, E_{torsion} and E_{nonbond} represent the bond-stretching term, the bond-bending term, the torsion energy term and the non-bonded energy term, respectively. The torsion energy term is usually given by

$$E_{\text{torsion}} = \sum_{\text{dihedral angle } \Phi} \sum_{n} \frac{V_n}{2}[1 + \cos(n\Phi - \gamma_n)], \quad (2)$$

where the first summation is taken over all dihedral angles Φ (both in the backbone and in the side chains), n is the number of waves, γ_n is the phase and V_n is the Fourier coefficient.

Equation (1) represents a standard set of the potential energy terms. As mentioned above, there are minor differences in the energy functions among different force fields. For instance, Urey–Bradley term is used in CHARMM and OPLS, but not in AMBER. In our method, we try to optimise a certain set of parameters in the existing force fields. Therefore, if the original force field has non-standard terms, then the optimised one also has them.

Our optimisation method for these force-field parameters is now described [12–14]. We first retrieve N native structures (one structure per protein) from PDB. We try to choose proteins from different amino acid sequence homology as much as possible. If the force-field parameters are of ideal values, then all the chosen native structures are stable without any force acting on each atom in the molecules. Hence, we expect

$$F = 0, \quad (3)$$

where

$$F = \sum_{m=1}^{N} \frac{1}{N_m} \sum_{i_m=1}^{N_m} \left| \vec{f}_{i_m} \right|^2 \quad (4)$$

and

$$\vec{f}_{i_m} = -\frac{\partial E_{\text{tot}}^{\{m\}}}{\partial \vec{x}_{i_m}}. \quad (5)$$

Here, N_m is the total number of atoms in molecule m, $E_{\text{tot}}^{\{m\}}$ is the total potential energy for molecule m, and $\vec{f}_i$ is the force acting on atom i. In reality, $F \neq 0$, and because $F \geq 0$, we can optimise the force-field parameters by minimising F with respect to these parameters. In practice, we perform a simulation in the force-field parameter space for this minimisation.

Proteins are usually in aqueous solution. However, the more the total number of proteins *(N)* is, the better the force-field parameter optimisations are expected to be, and hence we want to minimise our efforts in the calculations of the solvent effects. Here, we employ the generalised-Born/surface area (GB/SA) terms for the solvent contributions [21,22]. Hence, we use in Equation (5) (we suppress the label m for each molecule)

$$E_{\text{tot}} = E_{\text{conf}} + E_{\text{solv}}, \quad (6)$$

where

$$E_{\text{solv}} = E_{\text{GB}} + E_{\text{SA}}, \quad (7)$$

$$E_{\text{GB}} = -166\left(1 - \frac{1}{\epsilon_s}\right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-Dij}}}, \quad (8)$$

$$E_{\text{SA}} = \sum_k \sigma_k A_k. \quad (9)$$

Namely, in the GB/SA model, the total solvation free energy in Equation (7) is given by the sum of a solute–solvent electrostatic polarisation term, a solvent–solvent cavity term and a solute–solvent van der Waals term. A solute–solvent electrostatic polarisation term can be calculated by the generalised Born Equation (8), where $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$, α_i is the so-called Born radius of atom i, $D_{ij} = r_{ij}^2/(2\alpha_{ij})^2$ and ϵ_s is the dielectric constant of bulk water (we take $\epsilon_s = 78.3$). A solvent–solvent cavity term and a solute–solvent van der Waals term can be approximated by the term that is proportional to the solvent accessible surface area in Equation (9). Here, A_k is the total solvent-accessible surface area of atoms of type k, and σ_k is an empirically determined proportionality constant [21,22].

At first, we try to obtain as many structures as possible from PDB. The number is limited by the computer power that we have available in our laboratory. We want to choose proteins with different sizes (numbers of amino acids), different folds and different homology classes as much as possible. We also want to use only those with high experimental resolutions. Note that only atomic coordinates of proteins are extracted from PDB (and coordinates from other molecules such as crystal water are neglected).

If we use data from X-ray experiments, hydrogen atoms are missing, and thus we have to add hydrogen coordinates. Many protein simulation software packages

provide with routines that add hydrogen atoms to the PDB coordinates, and one can use one of such routines.

We now have N protein coordinates ready, but usually such 'raw data' result in very high total potential energy, and strong forces will be acting on some of the atoms in the molecules. This is because the hydrogen coordinates that we added as above are not based on experimental results and have rather large uncertainties. The coordinates of heavy atoms from PDB also have experimental errors. We take the position that we leave the coordinates of heavy atoms as they are in PDB as much as possible, and adjust the hydrogen coordinates to reduce this mismatch. This is why we want to include as many PDB data as possible with high experimental resolutions (so that the effects of experimental errors in PDB may be minimal). We thus minimise the total potential energy $E_{\rm tot} = E_{\rm conf} + E_{\rm solv} + E_{\rm constr}$ with respect to the coordinates for each protein conformation, where $E_{\rm constr}$ is the constraint energy term that is imposed on the heavy atoms in PDB (it is referred to as the 'predefined constraints'):

$$E_{\rm constr} = \sum_{\rm heavy\,atom} K_x(\vec{x} - \vec{x}_0)^2. \tag{10}$$

Here, K_x is the force constant of the restriction, and $\vec{x}_0$ are the original coordinate vectors of heavy atoms in PDB. Because we are searching for the nearest local-minimum states, usual minimisation routines such as the conjugate-gradient method and Newton–Raphson method can be employed here. As one can see from Equation (10), the coordinates of hydrogen atoms will be mainly adjusted, but unnatural heavy-atom coordinates will also be modified. We perform this minimisation for all N protein structures separately and obtain N refined structures.

Given N set of 'ideal' reference coordinates, we now optimise the first set of force-field parameters. In Equation (1), we have five classes of force-field parameters as mentioned above. Namely, the force-field parameters are those in the bond-stretching term (K_l and $l_{\rm eq}$), those in the bond-bending term (K_θ and $\theta_{\rm eq}$), those in the torsion term (V_n and γ_n), those in the Lennard-Jones term (A_{ij} and B_{ij}), and those in the electrostatic term (q_i). Because they are of very different nature, we believe that it is better to optimise these classes of force-field parameters separately. We note also that if we optimise all the parameters simultaneously, the null result (all the parameter values equal to zero) is a solution to Equation (3). This is the main reason why we optimise each class of parameters separately.

For each set of force-field parameters, the optimisation is carried out by minimising F in Equation (4) with respect to these parameters. Here, $E_{\rm tot}$ in Equation (5) is given by Equation (6). For this purpose, usual minimisation routines such as the conjugate-gradient method are not adequate, because we need a global optimisation. One should employ more powerful methods such as simulated annealing [23] and generalised-ensemble algorithms [9]. We perform this minimisation simulation in the above parameter space to obtain the parameter values that give the global minimum of F.

These processes are repeated until the optimised force-field parameters converge.

2.2 New optimisation method (Method 2)

We now describe our new method for optimising the force-field parameters. We use 100 proteins which are the same proteins from the PDB as those that we used in Method 1. If the force-field parameters are of ideal values, we expect that all the chosen native structures minimised by the ideal force field do not change after minimisations. Namely, we believe that force-field parameters are better, if they have lower deviations obtained by minimisations of protein structures. Hence, we expect

$$R = 0, \tag{11}$$

where

$$R = \frac{\sum_{i=1}^{n} {\rm RMSD}_i}{n}. \tag{12}$$

Here, ${\rm RMSD}_i$ is the root-mean-square deviation between the native structure of protein i and the corresponding minimised structure using trial force-field parameters. In reality, $R \neq 0$, and because $R \geq 0$, we expect that we can optimise the force-field parameters by minimising R with respect to these force-field parameters. In practice, we perform a simulation in the force-field parameter space for this minimisation. Namely, in the previous method (which we refer to as Method 1), we minimise F in Equation (4), and in the present method (which we refer to as Method 2), we minimise R in Equation (12) instead.

3. Results and discussion

3.1 Applications of the optimisation methods

We now present the results of the applications of our new optimisation method (Method 2) as well as the previous method (Method 1) of force-field parameters.

At first, we chose 100 PDB files with resolution 2.0 Å or better, with sequence similarity of amino acid 30.0% or lower and with less than 200 residues (the average number of residues is 117.0) from PDB–REPRDB [24]. Next, we refine these selected 100 structures. Generally, data from X-ray experiments do not have hydrogen atoms. Therefore, we have to add hydrogen coordinates. Many protein simulation software packages provide with routines that add hydrogen atoms to the PDB coordinates. We used the TINKER program package [25]. We thus minimise the

total potential energy $E_{\text{total}} = E_{\text{conf}} + E_{\text{solv}} + E_{\text{constr}}$ with respect to the coordinates for each protein conformation, where E_{constr} is the constraint energy term in Equation (10). Here, K_x is the force constant of the restriction and $\vec{x}_0$ are the original coordinate vectors of heavy atoms in PDB. As one can see from Equation (10), the coordinates of hydrogen atoms will be mainly adjusted, but unnatural heavy-atom coordinates will also be modified. We performed this minimisation for all the 100 protein structures separately and obtained 100 refined structures.

We focused on the parameters of torsion energy term, which we believe to be an important force-field term that influences the backbone conformational preferences such as α-helix structure and β-sheet structure. For example, AMBER parm94 [1] and AMBER parm96 [2] have very different behaviours about the secondary structure-forming tendencies, although these force fields differ only in the backbone torsion energy terms for rotations of the backbone ϕ and ψ angles. Recently, new force-field parameters of the backbone torsion-energy term about ϕ and ψ angles have been developed, which are, e.g. AMBER ff99SB [26], AMBER ff03 [27] and CHARMM 22/CMAP [28].

The force field that we optimised is the OPLS–UA [29]. The torsion-energy term E_{torsion} (Φ) for this force field is given by Equation (2). We performed the force-field parameter optimisations that correspond to the following torsion angles by Methods 1 and/or 2:

(1) N—C_α—C_β—C_γ and C—C_α—C_β—C_γ (χ_1) by Method 2;
(2) C—N—C_α—C (ϕ), N—C_α—C—N (ψ), C—N—C_α—C_β and N—C—C_α—C_β by Methods 1 and 2;
(3) C—N—C_α—C_β by Method 2;
(4) N—C_α—C—N by Method 2; and
(5) C_α—C_β—C_γ—C_δ (χ_2 of Glu) by Methods 1 and 2.

Here, we also optimised the force-field parameters of χ_2 of Glu. The reason is given below.

In Method 1, the minimisations of F in Equation (4) by the Monte Carlo (MC) simulated annealing simulations of the torsion-energy parameters with 10,000 MC steps were performed 10 times. In order to make a better force field, we have to optimise many force-field parameters. However, we ignored the uncertainty of improper torsion-energy parameters with this optimisation, because we wanted to focus on the torsion-energy parameters, and Method 1 is very sensitive for the energy of dihedral angles. For example, one of the results of the simulations of Method 1 above is shown in Figure 1.

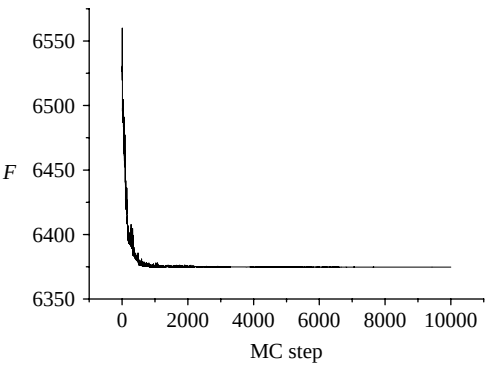

Figure 1. Time series of MC simulated annealing simulations in force-field parameter space of torsion-energy for OPLS–UA. The ordinate is the value of F in Equation (4).

In Method 2, the lowest R value was selected from about 10–30 optimisation runs with different initial conditions. In order to calculate R, the minimisations of 100 proteins were performed using these new parameter sets. In Table 1, all the optimised torsion-energy parameters are listed. As one can see in Table 1, the original parameters of OPLS–UA force field for the optimisation are almost zero.

In comparison with Method 1, Method 2 can optimise force-field parameters appropriately even if there are some errors in PDB structures. However, the computational cost of Method 2 is much larger than that of Method 1. Therefore, we could not apply Method 2 to the global optimisation in the force-field parameter space. The force-field parameters of the backbone-torsion angles need the global optimisation, because we consider that these parameters are the most problematic. Thus, at first, we

Table 1. Original and optimised torsion-energy parameters of OPLS–UA.

	$V_1/2$			$V_2/2$			$V_3/2$		
	Orig.	Opt.	γ_1	Orig.	Opt.	γ_2	Orig.	Opt.	γ_3
N—C_α—C_β—C_γ (χ_1)							0.500 or 1.000	1.950	0.0
C—C_α—C_β—C_γ (χ_1)							0.500 or 1.000	1.950	0.0
C—N—C_α—C (ϕ)	0.000	−0.662	0.0	0.000	0.277	π	0.000	−0.050	0.0
N—C_α—C—N (ψ)	0.000	0.974	0.0	0.000	0.576	π	0.000	−0.083	0.0
C—N—C_α—C_β	0.000	0.811	0.0	0.000	0.328	π	0.000	0.155	0.0
N—C—C_α—C_β	0.000	0.215	0.0	0.000	0.036	π	0.000	0.015	0.0
C_α—C_β—C_γ—C_δ (χ_2 of Glu)	0.000	0.565	0.0	0.000	0.177	π	2.000	−0.025	0.0

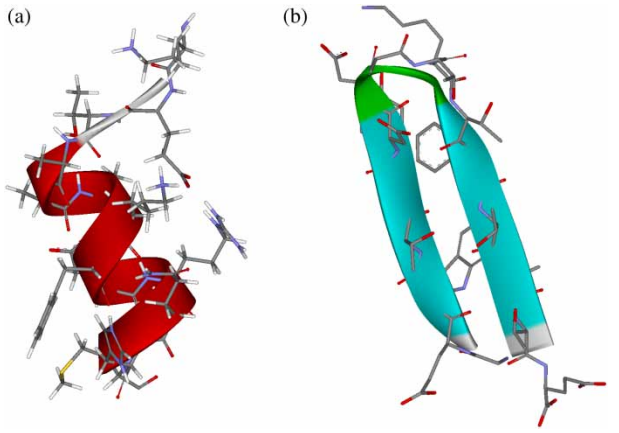

Figure 2. Structures of C-peptide (a) and G-peptide (b) obtained from the experimental results. The figures were created with DS Visualizer v1.5 [38].

performed the global optimisation of the backbone-torsion parameters by using Method 1. After that, Method 2 was applied only on the local region of the parameter space, which was identified as relevant by Method 1.

3.2 Folding simulations of two peptides

In order to test the validity of the force-field parameters obtained by our optimisation methods, we performed folding simulations using two peptides, namely, C-peptide of ribonuclease A and the C-terminal fragment of the Bl domain of streptococcal protein G, which is sometimes referred to as G-peptide [30]. The C-peptide has 13 residues and its amino acid sequence is Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-Gln-His-Met. This peptide has been extensively studied by experiments and is known to form an α-helix structure, as shown in Figure 2(a) [31,32]. Because the charges at peptide termini are known to affect helix stability [31,32], we blocked the termini by a neutral COCH3-group and a neutral-NH2 group. The G-peptide has 16 residues and its amino acid sequence is Gly-Glu-Trp-Thr-Tyr-Asp-Asp-Ala-Thr-Lys-Thr-Phe-Thr-Val-Thr-Glu. The termini were also blocked by a neutral COCH3-group and a neutral-NH2 group. This peptide is known to form a β-hairpin structure by experiments, as shown in Figure 2(b) [30,33,34].

Only Glu amino acid appears twice in each of the two peptides. Therefore, we consider that Glu amino acid is the most important, and the χ_2 parameters were optimised for this amino acid. (Of course, we expect that it becomes a better force field if the remaining force-field parameters of other amino acids are also optimised.)

For the folding simulations, we used the replica-exchange MD (REMD) method [35]. REMD is one of the generalised-ensemble simulation algorithms and has high conformational sampling efficiency by allowing configurations to heat up and cool down while maintaining proper Boltzmann distributions. We used the TINKER program package [25] modified by us for the folding simulations. The unit time step was set to 1.0 fs. Each simulation was carried out for 10 ns (hence, it consisted of 10,000,000 MD steps) with 16 replicas. The temperature during MD simulations was controlled by Nosé–Hoover method [36]. For each replica, the temperature was distributed exponentially: 700, 662, 625, 591, 558, 528, 499, 471, 446, 421, 398, 376, 355, 336, 317 and 300 K. As for solvent effects, we used the GB/SA model [21,22] included in the TINKER program package [25]. These folding simulations were repeated 10 times with different sets of randomly generated initial velocities.

In Figure 3, the helicity and strandness of C-peptide, which were obtained with the original OPLS–UA and its optimised force field are shown. These values are the averages of the 10 REMD simulations at 300 K. In comparison with the helicity of the original OPLS–UA, the helicity of the optimised force field increases at the amino acid sequence between 6 and 12. For the strandness, both the original and optimised OPLS–UA force fields are almost zero.

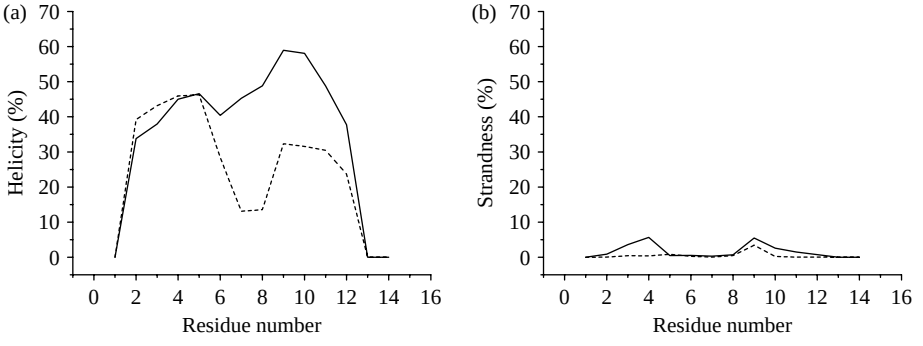

Figure 3. Helicity (a) and strandness (b) of C-peptide as functions of the residue number. These values are the average of the 10 independent REMD [35] simulations at 300 K. Normal and dotted lines stand for the optimised and original OPLS–UA force fields, respectively.

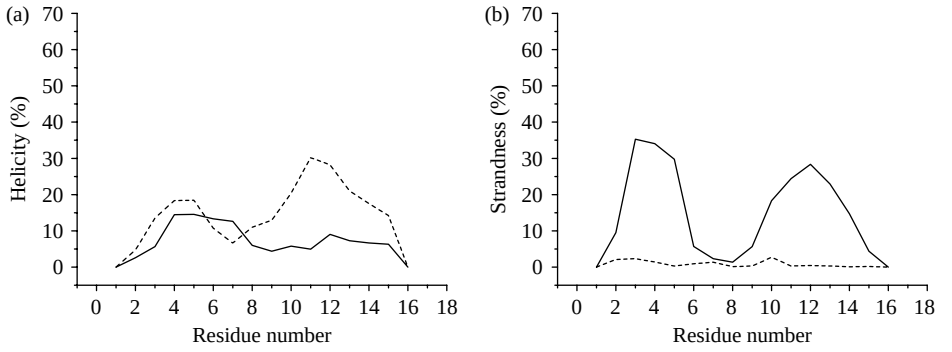

Figure 4. Helicity (a) and strandness (b) of G-peptide as functions of the residue number. These values are the average of the 10 independent REMD [35] simulations at 300 K. Normal and dotted lines stand for the optimised and original OPLS–UA force fields, respectively.

In Figure 4, the helicity and strandness of G-peptide at the original OPLS–UA and its optimised force field are shown. In comparison with the helicity of the original OPLS–UA, the helicity of the optimised force field decreases at the area of amino acid sequence between 8 and 15, and in comparison with the strandness of the original OPLS–UA, the strandness of the optimised force field clearly increases at the two areas of amino acid sequences 2–6 and 9–15. We checked the secondary-structure formations using the DSSP program [37], which is based on the formations of the intra-backbone hydrogen bonds. Strandness means that there are β-bridge or extended strand in the corresponding amino acid. In the experimental results, there is a turn region around residues 7–10, and there are five intra-backbone hydrogen bond pairs, namely, between residue pairs 2–15, 3–14, 4–13,

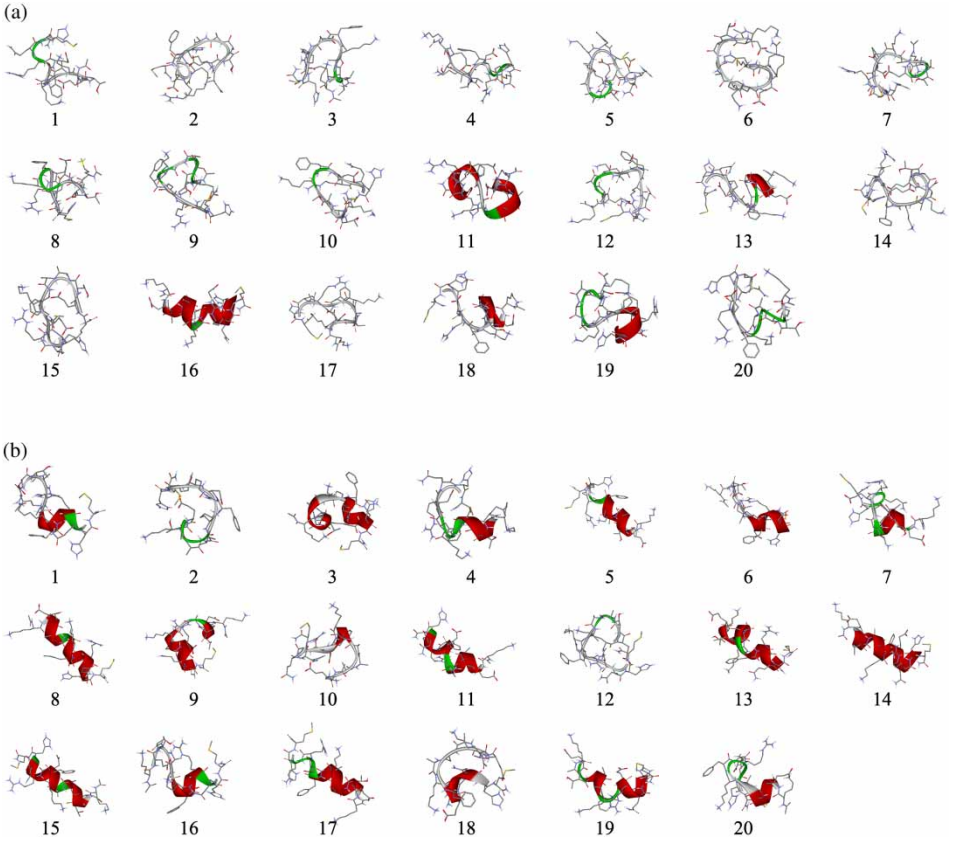

Figure 5. Twenty lowest-energy conformations of C-peptide obtained from 10 sets of REMD [35] simulation runs. (a) and (b) are the results of the original and optimised OPLS–UA force field, respectively. The conformations are ordered in the increasing order of energy for each case. The figures were created with DS Visualizer v1.5 [38].

5–12 and 6–11 in G-peptide. In Figure 4(b), the strandness decreases in the region around 7–8 residues in agreement with the experiments.

These results show that the optimised force field favours helix structures more than the original OPLS–UA in the case of C-peptide and favours β structures more than the original OPLS–UA in the case of G-peptide. We see that these secondary structure-forming tendencies of the optimised force field are better than those of the original OPLS–UA, because these results are consistent with the native structures of the two peptides.

In Figures 5 and 6, we show the 20 lowest-energy conformations of C-peptide and G-peptide obtained by the REMD simulations in the case of the original and optimised OPLS–UA force fields, respectively. In Figure 5(a), five conformations (Nos. 11, 13, 16, 18 and 19) have α-helix structures for the original OPLS–UA in the case of C-peptide. In Figure 5(b), 18 conformations (all conformations except for Nos. 2 and 12) have α-helix structures for the optimised OPLS–UA in the case of C-peptide. From these results, we can see that the optimised OPLS–UA force field favours α-helix structure more than the original OPLS–UA force field, in the case of C-peptide. In Figure 6(a), 11 conformations have α-helix structures for the original OPLS–UA, in the case of G-peptide. In Figure 6(b), seven conformations have α-helix structures, and eight conformations have β-hairpin structures for the optimised OPLS–UA, in the case of G-peptide. In Figure 6(b), two conformations (Nos. 3 and 16) out of the eight β-hairpin conformations have the right hydrogen bond formations that are inferred by the experiments. Namely, conformation No. 3 has three native-like hydrogen bonds between residue pairs 3–14, 4–13 and 5–12, and conformation No.16 has two native-like hydrogen bonds between residue pairs 3–14 and 4–13. These results for G-peptide show that the optimised OPLS–UA force field does not favour α-helix structure and clearly favours β-hairpin structure more than the original OPLS–UA force field.

These secondary structure-forming tendencies of the optimised OPLS–UA force field for two peptides agree with experimental implications in comparison with those of the original OPLS–UA force field. Therefore, our optimisation methods succeeded in enhancing the accuracy of the OPLS–UA force field.

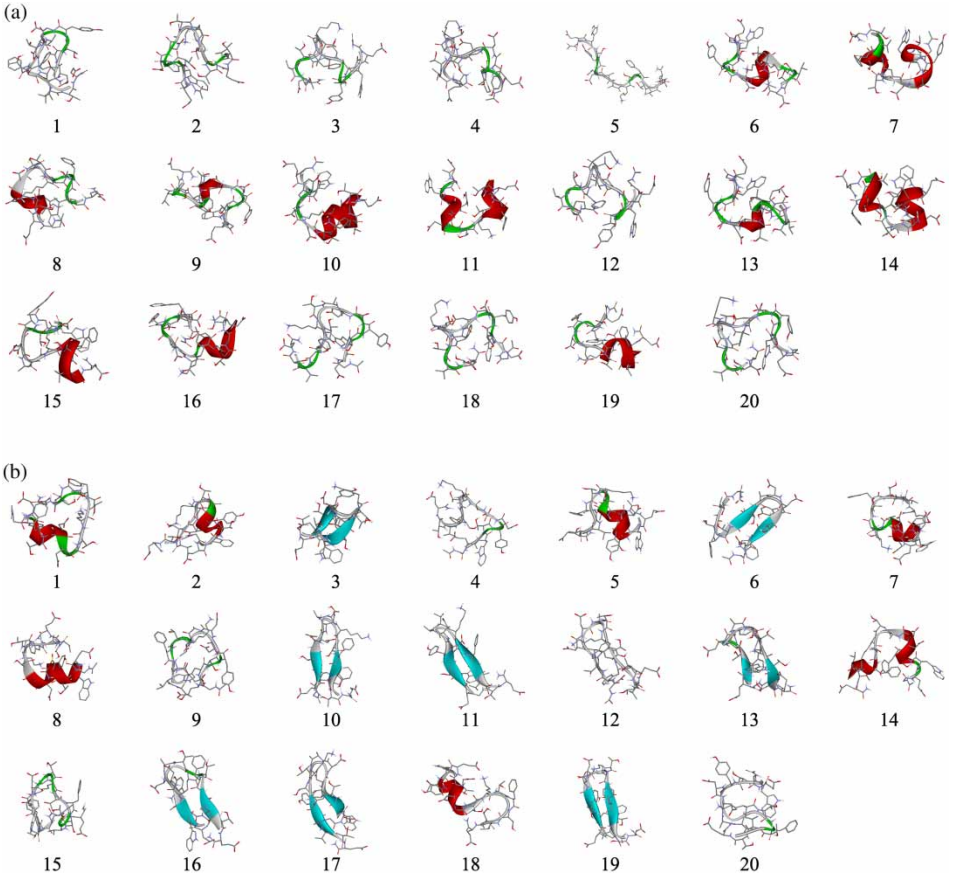

Figure 6. Twenty lowest-energy conformations of G-peptide obtained from 10 sets of REMD [35] simulation runs. (a) and (b) are the results of the original and optimised OPLS–UA force field, respectively. The conformations are ordered in the increasing order of energy for each case. The figures were created with DS Visualizer v1.5 [38].

4. Conclusions

In this work, we proposed a new optimisation method of force-field parameters and optimised some torsion-energy parameters of the OPLS–UA force field by using the method and our previous method. These two methods can optimise force-field parameters using only PDB structures. We applied these optimisation methods using 100 protein molecules from the PDB. We then performed folding simulations of α-helical and β-hairpin peptides. We found that the helicity of the optimised force field increases for α-helical peptide and that the strandness of the optimised force field increases for β-hairpin peptide in comparison with those of the original OPLS–UA. The results, therefore, showed that the optimised force-field parameters gave structures more consistent with the experimental implications than the original OPLS–UA force field. We have shown that we can indeed improve the force-field parameters by using our optimisation procedures.

Acknowledgements

The computations were performed on Super NOVA of Miki Laboratory at Doshisha University and the computers at the Research Center for Computational Science, Institute for Molecular Science. This work was supported, in part, by the Grants-in-Aid for the Academic Frontier Project, 'Intelligent Information Science', for Scientific Research on Innovative Areas ('Fluctuations and Biological Functions'), and for the Next Generation Super Computing Project, Nanoscience Program from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

[1] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, J. Kenneth, M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*, J. Am. Chem. Soc. 117 (1995), pp. 5179–5197.

[2] P.A. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot, and A. Pohorille, Vol. 3, ESCOM, Dordrecht, 1997, pp. 83–96.

[3] J. Wang, P. Cieplak, and P.A. Kollman, *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* J. Comput. Chem. 21 (2000), pp. 1049–1074.

[4] A.D. MacKerell, Jr., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, I, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins*, J. Phys. Chem. B 102 (1998), pp. 3586–3616.

[5] W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*, J. Am. Chem. Soc. 118 (1996), pp. 11225–11236.

[6] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, and W.L. Jorgensen, *Evaluation and reparametrization of the OPLS–AA force field for proteins via comparison with accurate quantum chemical calculations on peptides*, J. Phys. Chem. B 105 (2001), pp. 6474–6487.

[7] W.F. Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott, and I.G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag AG an der ETH Zürich, Zürich, 1996.

[8] G. Némethy, K.D. Gibson, K.A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. Scheraga, *Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides*, J. Phys. Chem. 96 (1992), pp. 6472–6484.

[9] A. Mitsutake, Y. Sugita, and Y. Okamoto, *Generalized-ensemble algorithms for molecular simulations of biopolymers*, Biopolym. (Pept. Sci.) 60 (2001), pp. 96–123.

[10] T. Yoda, Y. Sugita, and Y. Okamoto, *Comparisons of force fields for proteins by generalized-ensemble simulations*, Chem. Phys. Lett. 386 (2004), pp. 460–467.

[11] T. Yoda, Y. Sugita, and Y. Okamoto, *Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations*, Chem. Phys. 307 (2004), pp. 269–283.

[12] Y. Sakae and Y. Okamoto, *Optimization of protein force-field parameters with the protein data bank*, Chem. Phys. Lett. 382 (2003), pp. 626–636.

[13] Y. Sakae and Y. Okamoto, *Protein force-field parameters optimized with the protein data bank. I. force-field opti-mizations*, J. Theo. Comput. Chem. 3 (2004), pp. 339–358.

[14] Y. Sakae and Y. Okamoto, *Protein force-field parameters optimized with the protein data bank. II. Comparisons of force fields by folding simulations of short peptides*, J. Theo. Comput. Chem. 3 (2004), pp. 359–378.

[15] C. Simmerling, B. Strockbine, and A.E. Roitberg, *All-atom structure prediction and folding simulations of a stable protein*, J. Am. Chem. Soc. 124 (2002), pp. 11258–11259.

[16] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*, J. Comput. Chem. 24 (2003), pp. 1999–2012.

[17] M. Iwaoka and S. Tomoda, *The SAAP force field. A simple approach to a new all-atom protein force field by using single amino acid potential (SAAP) functions in various solvents*, J. Comput. Chem. 24 (2003), pp. 1192–1200.

[18] A.D. MacKerell, M. Feig, and C.L. Brooks III, *Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations*, J. Comp. Chem. 25(11) (2004), pp. 1400–1415.

[19] N. Kamiya, Y. Watanabe, S. Ono, and J. Higo, *AMBER-based hybrid force field for conformational sampling of polypeptides*, Chem. Phys. Lett. 401 (2005), pp. 312–317.

[20] A.E. Garcia and K.Y. Sanbonmatsu, *α-Helical stabilization by side chain shielding of backbone hydrogen bonds*, Proc. Natl. Acad. USA 99 (2002), pp. 2782–2787.

[21] W.C. Still, A. Tempczyk, R.C. Hawley, and T. Hendrickson, *Semianalytical treatment of solvation for molecular mechanics and dynamics*, J. Am. Chem. Soc. 112 (1990), pp. 6127–6129.

[22] D. Qiu, P.S. Shenkin, F.P. Hollinger, and W.C. Still, *The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate born radii*, J. Phys. Chem. A 101 (1990), pp. 3005–3014.

[23] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, *Optimization by simulated annealing*, Science 220 (1983), pp. 671–680.

[24] T. Noguchi, K. Onizuka, Y. Akiyama, and M. Saito, *PDB-REPRDB: A Database of Representative Protein Chains in PDB (Protein Data Bank), in Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology*, AAAI press, Menlo Park, CA, 1997.

[25] *TINKER program package*, software available at http://dasher.wustl.edu/tinker/

[26] V. Hornak, A. Abel, R. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple amber force fields and*

development of improved protein backbone parameters, Proteins 65 (2006), pp. 712–725.
[27] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, and T. Lee, *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*, J. Comput. Chem. 24 (2003), pp. 1999–2012.
[28] A. MacKerell, Jr., M. Feig, and C.L. Brooks III, *Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations*, J. Comput. Chem. 25 (2004), pp. 1400–1415.
[29] W.L. Jorgensen and J. Tirado-Rives, *The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin*, J. Am. Chem. Soc. 110 (1988), pp. 1657–1666.
[30] S. Honda, N. Kobayashi, and E. Munekata, *Thermodynamics of a –hairpin structure: Evidence for cooperative formation of folding nucleus*, J. Mol. Biol. 295 (2000), pp. 269–278.
[31] K.R. Shoemaker, P.S. Kim, D.N. Brems, S. Marqusee, E.J. York, I.M. Chaiken, J.M. Stewart, and R.L. Baldwin, *Nature of the charged-group effect on the stability of the C-peptide helix*, Proc. Natl. Acad. Sci. U.S.A. 82 (1985), pp. 2349–2353.
[32] J.J. Osterhout, Jr., R.L. Baldwin, E.J. York, J.M. Stewart, H.J. Dyson, and P.E. Wright, *1H NMR studies of the solution conformations of an analogue of the C-peptide of ribonuclease A*, Biochemistry 28 (1989), pp. 7059–7064.
[33] F.J. Blanco, G. Rivas, and L. Serrano, *A short linear peptide that folds into a native stable bold beta-hairpin in aqueous solution*, Nat. Struct. Biol. 1 (1994), pp. 584–590.
[34] N. Kobayashi, S. Honda, H. Yoshii, H. Uedaira, and E. Munekata, *Complement assembly of two fragments of the streptococcal protein G B1 domain in aqueous solution*, FEBS Lett. 366 (1995), pp. 99–103.
[35] Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chem. Phys. Lett. 314 (1999), pp. 141–151.
[36] W.G. Hoover, *Canonical dynamics: Equilibrium phase-space distributions*, Phys. Rev. A 31 (1985), pp. 1695–1697.
[37] W. Kabsch and C. Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features*, Biopolymers 22 (1983), pp. 2577–2637.
[38] *Accelrys Discovery Studio Visualizer*, software available at http://www.accelrys.com/