Structural relaxation and binding energy calculations of FK506 binding protein complexes using the large-scale DFT code CONQUEST

T Otsuka[1], N Okimoto[1], M Taiji[1], DR Bowler[2] and T Miyazaki[3]

[1]Quantitative Biology Center (QBiC), RIKEN, 1-6-5 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyoho, 650-0047, Japan

[2]Department of Physics & Astronomy, University College London, Gower St, London, WC1E 6BT, UK

[3]Computational Materials Science Center, National Institute for Materials Science (NIMS), 1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047, Japan

E-mail: totsuka@riken.jp

Abstract. We have performed structural relaxation and calculated binding energy of protein-ligand complex systems, FK506 binding protein (FKBP) and some ligand molecules, using a large-scale density functional theory (DFT) code CONQUEST. Detailed comparison of the calculated binding energies of FKBP with various ligand molecules is reported including the effects of the full geometry relaxation.

1. Introduction

In the field of drug discovery, computational methods are now commonly used to evaluate the binding affinities of drug molecules to proteins. Such computational methods are expected to improve the efficiency of drug design in the near future [1, 2]. Recently, *first-principles* (or *ab initio*) electronic structure approaches are also used in this field in order to calculate protein-ligand interactions accurately [3]. In the structure-based drug design (SBDD), however, (accurate) experimental atomic positions are not available in most cases. In general, less reliable experimental structures and/or partially relaxed structures by QM/MM or MM method are usually used in such studies. It is a great challenge for theorists to clarify the effects on the binding energy by relaxing all atomic positions solely by QM.

In the present study, we report fully relaxed structure and binding energy of the complex systems consisting of FK506 binding protein (FKBP) and a ligand molecule, using our density functional theory (DFT) code CONQUEST. The code has a high parallel efficiency and enables us to perform DFT studies on very large systems. The systems we have calculated in this work contain about 1700 atoms. Experimental structures of FKBP and various ligands can be obtained from the PDB codes: 1FKF, 1FKG, and 1FKI. To our knowledge, there have been no theoretical reports of the fully relaxed structures of these systems by first-principles DFT calculations. We believe it is important to collect such basic information, for example, to analyze the experimental or theoretical structural data for a protein-ligand system. The CONQUEST code can be used as a linear scaling code [4], and so it will be possible to model these systems in aqueous solution; the present work is a preparation also for such future studies.

2. Computational Methods

In this study, all calculations are performed with our own DFT code CONQUEST [5]. Details of the methods used in the code are explained in our previous papers [6]. Here we describe briefly some key points. In CONQUEST, we use the Kohn-Sham density matrix defined as

$$\rho(r, r') = \sum_n f_n \psi_n(r) \psi_n^*(r'),$$

where $\psi_n(r)$ is the Kohn-Sham orbital for band index n, and f_n is its occupation number between 0 and 1. The DFT total energy of system can be calculated from density matrix, using the pseudopotential technique and the standard exchange-correlation functionals such as the local density approximation (LDA) or the generalized gradient approximation (GGA). The density matrix is represented by localized orbitals called "support functions",

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

where $\phi_{i\alpha}(r)$ are the support function that are non-zero only inside "support region" centered on the atoms i, and α runs over the support functions on a given atom. The matrix $K_{i\alpha, j\beta}$ is the density matrix expressed by non-orthogonal basis of support functions. In the CONQUEST code, two types of basis function are prepared for the support functions, one is B-splines on regular grids [7] and the other is numerical pseudo atomic orbitals (PAOs)[8]. In order to achieve the density matrix using the support functions, CONQUEST can employ two methods: conventional diagonalization or order-N methods. Although we only use the former method in this work, one of the advantage of using the code is we can treat extremely large systems, up to million atom systems [9], if we use the latter method.

In this work, we consider the FK506 binding protein (FKBP) complexes, which have been well studied both experimentally [10] and theoretically [11-13]. There is also our previous related work [14]. The systems consist of 107 residues for FKBP and small ligands. Experimental structures of FKBP and various ligands can be obtained from the PDB codes, 1FKF, 1FKG, and 1FKI, summarized in table 1. The initial (that is, experimental) structures as the reference structures are prepared by the following procedure: some missing hydrogen atoms are added to the PDB structures, and only hydrogen atoms are relaxed by using the classical force field of parm99. The fully relaxed structures are obtained by using our large-scale DFT code CONQUEST with double-zeta with polarization (DZP) basis set, Perdew–Burke–Ernzerhof (PBE) functional [15] for GGA, the diagonalization with non-self-consistent calculations, and the numerical integration grid cutoff of 100 Ha. The binding energy between a FKBP and the ligand molecule is simply evaluated by

$$\Delta E_{binding} = E_{complex} - \left(E_{protein} + E_{ligand} \right).$$

Here we calculate the total energy of protein, $E_{protein}$, using the structure of the protein in the complex system. In this study, we use only PBE functional in our DFT calculations for the evaluation of binding energy. The PBE functional reproduces the structure and the interaction energy of hydrogen bonds of biological systems (for example, see ref. [16,17]), but in some cases its lack of the description of long-range dispersion interactions [18,19] may cause a serious problem. We are preparing for DFT-D2 [18] and vdW-DF [19] scheme, and the results of dispersion effects using those methods will be presented in a future publication. Entropic contributions are also neglected in our DFT calculations.

Table 1. PDB code of FKBP complex systems, the name of ligand, and the total number of atoms.

PDB code	Ligand	Number of atoms[a]
1FKF	L20	1792
1FKG	LG8	1734
1FKI	L13	1736

[a] The total number of atoms of FKBP is 1666.

3. Results and Discussion

In the protein and ligand complex system, the binding structure of the ligand molecule is significantly affected by the side chains of the amino acids surrounding the ligand. In order to collect the structural information of the ligand in such an environment, we first investigate the fully relaxed structures of FKBP-ligand systems. Figures 1, 2, and 3 show the structure of FKBP-L20, FKBP-LG8, and FKBP-L13 molecules before (blue line) and after (red line) the structure relaxation.
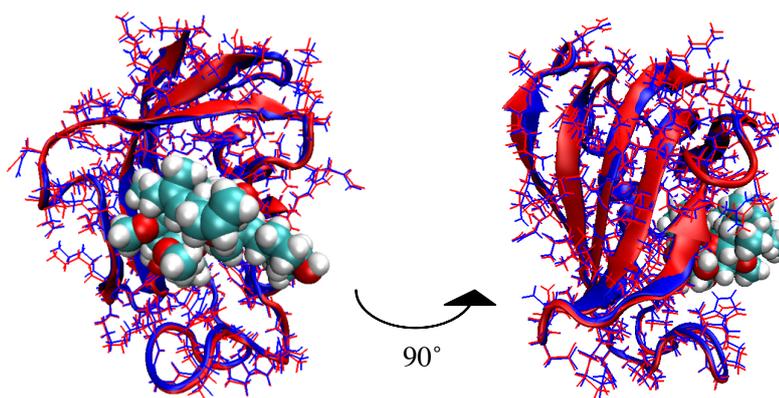

**Figure 1.** Structure of FKBP-L20 complex; before (blue) and after (red) fully structure relaxation.
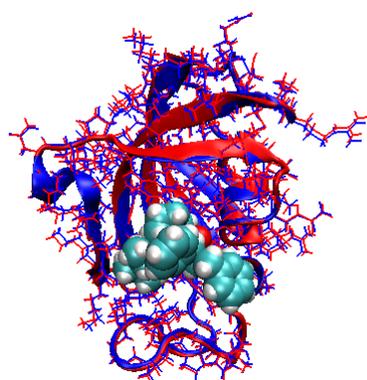
Figure 2. Structure of FKBP-LG8 complex; before (blue) and after (red) fully structure relaxation.
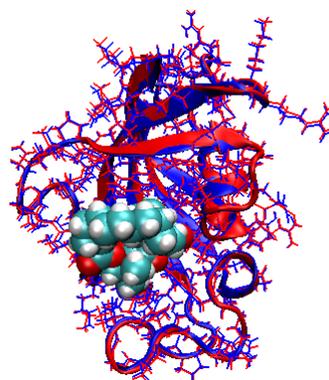
Figure 3. Structure of FKBP-L13 complex; before (blue) and after (red) fully structure relaxation.

Table 2 shows the root mean square deviation (RMSD) of the relaxed structure from the initial structure. In table 2, the RMSD of the backbone is much smaller than that of the whole protein including the residues. From table 2 and these figures (1, 2 and 3), we can see that the change of the backbone structure of the protein is small, while the structure of the side chains are largely changed by the full relaxation.

Table 2. Root mean square deviation (RMSD) of fully relaxed structure from the experimental one.

System	$RMSD_{backbone}$	$RMSD_{residue}$
FKBP-L20	0.27125	0.45675
FKBP-LG8	0.24112	0.38802
FKBP-L13	0.30690	0.45652

Next we compare the binding energies between FKBP and the ligand molecules. Table 3 shows the binding energies with or without the structure relaxation of FKBP-ligand complex systems. Note that the experimental values shown here are the binding affinities obtained by Holt et al [10]. In the present study, we regard the values as the approximate binding free energies. Table 3 shows that the order of the calculated binding energies is the same as the experimental one, for both cases using the initial or relaxed structures. The calculated binding energies are slightly overestimated in the present PBE calculations. It also shows that the binding energies calculated by using the relaxed structures are larger than those obtained by the initial structures. We can conclude that the ligand molecule can bind to the residues in FKBP more tightly in the fully relaxed structure. For the overestimation in the calculated binding energy, we should consider the hydration effects, entropic contribution in the future.

Table 3. Calculated biding energy of FKBP-ligand complex systems.

System	Experimental[a]	Initial structure	Relaxed structure
FKBP-L20	-12.8	-18.95	-26.62
FKBP-LG8	-10.9	-17.25	-24.37
FKBP-L13	-9.5	-14.02	-21.88

[a] Experimental values indicate the experimental binding affinities as referred in [10]

4. Conclusion

In this study, we have performed the structural relaxation of FKBP-ligand complexes using the large-scale DFT code CONQUEST. We have calculated the binding energies of FKBP and various ligands and found that the energy of the relaxed structure is stronger than that obtained by using the experimental structure. We have also analyzed the fully relaxed structure of the FKBP complexes in detail and have found that the structure of the side chains of the amino acids surrounding the ligand is largely changed. From these results, it seems that the structure of the entire protein, not just the area around the ligand, should be modelled when performing a study on SBDD. The results in this work assist in promoting the study of the "relaxed" SBDD. In order to clarify these aspects theoretically, very large-scale electronic structure calculations on FKBP-ligand complex systems in aqueous solution are necessary. Linear scaling DFT codes such as CONQUEST form the tools to perform these studies and the results will be presented in a future publication.

References

[1] Gilson M K, Given J A, Bush B and McCammon J A 1997 *Biophys. J.* **72**.1047-1069
[2] Gilson M K, and Zhou H 2007 *Annu. Rev. Biophys. Biomol. Struct.* **36** 21-42
[3] Kitaura K, Ikeo E, Asada T, Nakano T and Uebayasi M 1999 *Chem. Phys. Lett.* **313** 701–706
[4] Bowler D R and Miyazaki T 2012 *Rep. Prog. Phys.* **75** 036503
[5] See our web page; http://www.order-n.org/
[6] Bowler D R, Miyazaki T and Gillan M J 2002 *J. Phys.: Condens. Matter* **14** 2781; Bowler D R, Choudhury R, Gillan M J and Miyazaki, T 2006 P*hys. Stat. Sol.* B **243** 989
[7] Hernandez E, Gillan M J, and Goringe C M 1997 *Phys. Rev. B* **55** 13485
[8] Torralba A S, Todorovic M, Brazdova V, Choudhury R, Miyazaki T, Gillan M J and Bowler D R 2008 *J. Phys.: Condens. Matter* **20** 294206
[9] Bowler D R and Miyazaki T 2010 *J. Phys.: Condens. Matter* **22** 074207
[10] Holt D A, Luengo J I, Yamashita D S, Oh H J, Konialian A L, Yen H K, Rozamus L W, Brandt M, Bossard M J, Levy M A, Eggleston D S, Liang J, Schultz L W, Stout T J and Clardy J 1993 *J. Am. Chem. Soc.* **115** 9925-9938
[11] Fujitani H, Tanida Y and Matsuura A 2009 *Phys. Rev. E* **79** 021914
[12] Nakanishi I, Fedorov D G and Kitaura K 2007 *Proteins* **68** 145-158
[13] Xu Y and Wang R 2006 *Proteins* **64** 1058-1068
[14] Watanabe H, Tanaka S, Okimoto N, Hasegawa A, Taiji M, Tanida Y, Mitsui T, Katsuyama M and Fujitani H 2010 *CBI Journal* **10** 32-45
[15] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865, Errata: 1997 Phys Rev. Lett. **78** 1396
[16] Otsuka T, Miyazaki T, Ohno T, Bowler D R and Gillan M J 2008 *J. Phys. Condens. Matter* **20** 294201
[17] Sugiki S, Matsuoka M, Usuki R, Sengoku Y, Kurita N, Sekino H and Tanaka S 2005 *J. Theor. Comput. Chem.* **4** 183
[18] Grimme S 2006 *J. Comput. Chem.* **27** 1787-1799
[19] Dion M, Rydberg H, Schröder E, Langreth D C and Lundqvist B I 2004 *Phys. Rev. Lett.* **92** 246401