FEATURE ARTICLE

Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method

Dmitri G. Fedorov* and Kazuo Kitaura

Research Institute for Computational Sciences (RICS), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki, Japan 305-8568

Received: March 1, 2007; In Final Form: April 11, 2007

Following the brief review of the modern fragment-based methods and other approaches to perform quantum-mechanical calculations of large systems, the theoretical development of the fragment molecular orbital method (FMO) is covered in detail, with the emphasis on the physical properties, which can be computed with FMO. The FMO-based polarizable continuum model (PCM) for treating the solvent effects in large systems and the pair interaction energy decomposition analysis (PIEDA) are described in some detail, and a range of applications of FMO to biological studies is introduced. The factors determining the relative stability of polypeptide conformers (α-helix, β-turn, and extended form) are elucidated using FMO/PCM and PIEDA, and the interactions in the Trp-cage miniprotein construct (PDB: 1L2Y) are analyzed using PIEDA.

1. Introduction

The power of quantum chemistry enables very accurate calculations of small systems, and with appropriate methods one can achieve the desired goal of either the chemical or even the spectroscopic accuracy. The applications of these methods to large systems, however, are hindered by the steep growth of the computational cost with the system size. Density functional theory (DFT) calculations of an RNA piece,[1] cytochrome *c*,[2] insulin,[3] and its hexamer[4] have been reported, and a small protein has been optimized with restricted Hartree−Fock (RHF).[5] The semiempiric methods,[6−8] while retaining the quantum-mechanical (QM) description of the whole system, succeeded in reducing the calculation load by relying on experiment-based parameters. In the hybrid quantum mechanics/molecular mechanics (QM/MM) approach[9,10] and ONIOM,[11−13] one uses force fields for the larger part of the system, having to perform the more expensive QM calculations only for the smaller remaining part. In the effective fragment potential (EFP) method[14,15] the QM description of the system is successfully modeled with the properly designed potential functions.

An efficient alternative to either the full ab initio or some MM compromise treatment (QM/MM, ONIOM) lies in the fragment-based methods, which form an actively developed field of research. The basic idea shared by the fragment methods is to divide the system into pieces (fragments) and to obtain the total properties from those of fragments and their conglomerates. Beginning with the proposal of the self-consistent group scheme combining electrons into groups,[16] there has been a great amount of research devoted to reducing the computational cost by grouping some parts of the system into larger units.

Among the more modern work, the divide-and-conquer (DC) approach[17] to DFT calculations was introduced, projecting the total Hamiltonian upon fragments, retaining, however, the long-range Coulomb interaction from the whole system. There have been a number of more purely DC approaches introduced, where the fragments are treated completely independently, such as the molecular tailoring (MTA)[18−20] and several other approaches.[21−23] It is also conceivable to account for the environment partially, which is done in the molecular fractionation with conjugate caps (MFCC) approach[24−40] and the integrated multicenter molecular orbital method[41,42] by including a number of surrounding atoms in the fragment calculations.

Alternatively, one can use the idea of locality in the form of localized orbitals. The elongation method developed originally to describe polymers[43−53] was later extended to other system types such as polypeptides; however, it suffers from ambiguities when the system is not linear. Excitation energies of large clusters can be computed with the symmetry adapted cluster configuration interaction (SAC-CI) method.[54] The incremental correlation method[55,56] has been used in describing the electron correlation in crystals[57] and polymers,[58−60] relying on the full calculations of the uncorrelated wave function, and thus it seems to be better classified to the category of the local correlation methods. Some overlap, at least conceptually, with the fragment methods can be seen in the grouping of orbitals for the functional groups[60] and the many-body expansion used to compute the correlation energy.

The fragment molecular orbital method (FMO) was introduced by Kitaura and co-workers.[61−63] The distinctive feature of FMO is the inclusion of the electrostatic field from the whole system in each individual fragment calculation, and in using the systematic many-body expansion. Thus, the underlying principles are different from the typical divide-and-conquer approaches, which lead us to suggest[63a] the *e pluribus unum* category for the methods, where the influence of the whole system is retained in the subsystem calculations and the total

* Corresponding author. E-mail: d.g.fedorov@aist.go.jp.

10.1021/jp0716740 CCC: $37.00
Published on Web 05/19/2007

Dmitri G. Fedorov received his M.S. in quantum chemistry from Saint Petersburg State University in Russia in 1993. He was awarded the Ph.D. in physical chemistry at Iowa State University in 1999, working under the guidance of Mark S. Gordon. He spent two years at the University of Tokyo as a JSPS postdoctoral research fellow, working with Kimihiko Hirao and then moved to the National Institute of Advanced Industrial Science and Technology in Japan in 2002, where he is currently employed as a research scientist. His research interests include the relativistic effects in chemistry as well as the quantum-chemical method development for describing large systems, such as proteins.

Kazuo Kitaura received his Ph.D. in quantum chemistry in 1976 from Osaka City University. He is currently Professor at Kyoto University and Principal Research Scientist at the National Institute of Advanced Industrial Science and Technology (AIST). His primary research interests are the development of quantum-chemical methods for large molecules and their applications to structure, properties, and reactions of biomolecules.

properties are computed. Several methods closely related to FMO have been suggested, where either the dipole field is employed to describe the polarization[64] or the electrostatic field is treated with fixed partial charges,[65] corresponding to the Mulliken charge approximation[66] in FMO.

The adjustable density matrix assembler (ADMA) approach[67−71] uses the additive fuzzy density fragmentation principle to obtain the fragment densities (optionally,[69] in the field of the external point charges), from which the total electron density is constructed and used in the conventional RHF Fock matrix construction for the whole system. The total energy has been shown to be in good agreement with ab initio values.[68,69,71]

2. Fragment Molecular Orbital Method

The emphasis in FMO is laid on accuracy and the ability to obtain the physical properties of fragments and the interaction between them. The notion of groups of atoms retaining to the large part their properties when they become part of a larger system is very familiar to chemists: functional groups, benzene rings, or amino acid residues are the building blocks on which the chemical way of thinking is based. When the groups of atoms are described not as balls and sticks, but as electron density distributions, care has to be taken to fragment the system in the least disturbing way, that is, so that the electron density within each fragment be as localized as possible. The process of fragmentation has to be performed on the basis of the chemical knowledge and not with mechanical means: it is not difficult to imagine that benzene rings are to be kept intact.

Supposing that fragments are defined, the next step is to compute their electron density distribution. In the full quantum-mechanical calculation, each fragment is immersed in the Coulomb field due to the remaining part of the system (environment), to which the exchange interaction with other fragments is added and the electron density is fully relaxed. In FMO, one adds the environmental Coulomb field (long-ranged) and neglects the corresponding exchange and charge-transfer interactions (short-ranged). Consequently, the fragment (monomer) densities are converged self-consistently giving the fully polarized fragments and their respective energies. In the second step, one obtains the quantum-mechanical interaction between pairs of fragments (dimers), which is accomplished by performing dimer calculations in the Coulomb field due to the remaining fragments.

FMO is variational at the monomer level because of the self-consistent convergence of monomer densities, and the dimer correction can be thought of as a perturbation, so that overall FMO is not fully variational. The fragment wave functions are not orthogonal to each other, which leads to some complications in the diagrammatic representation of FMO (vide infra).

Naturally, the question of fractioning covalent bonds arises. Various methods solved this issue differently, and in a large number of cases simple hydrogen atom capping is employed. In FMO, bonds are fractioned electrostatically, which means that we divide the electrons along a covalent bond that becomes the fragment border (in the ratio 2:0 for the two fragments concerned), and do not append hydrogen caps. Because the Coulomb field is added to each fragment from all other fragments, the dangling bonds are effectively saturated by such a field and no other capping is necessary, which has the advantage of eliminating unphysical caps and well reproducing the density distribution of the whole system. A very detailed explanation of the fragmentation can be found in ref 63a.

The fragmentation of polypeptides is performed at C_α atoms. Typically, two residues are assigned to one fragment if the total energetics is of interest, and for the pair interaction analysis it is convenient to assign one residue per fragment. For good accuracy, water clusters are usually divided into two molecules per fragment. The total properties in FMO slightly depend upon the particular way of dividing a system into fragments, just as the ab initio properties depend upon the basis set, and thus in quoting the results it is necessary to specify the fragmentation details. The PDB data for proteins can be easily fragmented using FMOutil software.[72] More detailed explanation of the fragmentation issues can be found in ref 63a.

2.1. Outline of Methodology. The FMO expression for the total energy is given by

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J) \qquad (1)$$

where monomer (E_I) and dimer (E_{IJ}) energies are obtained from the corresponding calculations of N fragments (monomers) and

their pairs (dimers) in the external Coulomb field due to the remaining monomers (including contributions from both nuclei and electron density). Other properties such as the dipole moment can be defined with a similar expression, and the energy gradient for the whole system is obtained[73] by differentiating eq 1. The spatial separation between fragments is efficiently made use of[66,74] by (a) the Mulliken atomic orbital population and charge approximations of the Coulomb field and (b) the electrostatic approximation to the RHF calculations of far separated dimers.

The two-body expansion can be made more accurate by proceeding to higher terms: the addition of three-body corrections[75] nearly fully recovers the ab initio properties. FMO has been shown to describe many types of systems with high accuracy already at the two-body level (FMO2). For instance, at the production level of approximations, the error in the total energy of the β-strand of polyalanine with 40 residues relative to ab initio was 0.13 kcal/mol (6-31G* basis set).[76] The error for systems with a significant charge transfer and polarization is more substantial: for the Trp-cage miniprotein construct[77] (PDB code: 1L2Y, 304 atoms), and the 6-31G* basis set with diffuse functions on carboxyl groups, denoted by 6-31G(+)G* it was[76] 0.40 kcal/mol, and the total energetics can be refined at the three-body level (FMO3), in which case the error decreases to 0.052 kcal/mol. An alternative way to compute the three-body corrections was also suggested,[78] providing the diagrammatic representation of FMO as a perturbation theory of the interaction of fragments with non-orthogonal wave functions. An efficient implementation of the ideas in this approach may prove very useful for practical applications.

The original RHF formulation of FMO was combined with density functional theory (DFT),[79−81] second-order Møller-Plesset perturbation theory (MP2),[76,82−84] multiconfiguration self-consistent field (MCSCF),[85] and coupled cluster (CC).[86] Each method has its own merits: MP2 describes well the dispersion interaction, which is of great importance to most biological systems, DFT is an inexpensive way to a good description of equilibrium properties, MCSCF excels in the proper treatment of states with near degeneracies, such as transition metals or transition states of chemical reactions, and CC in its CCSD(T) variety is often regarded as the golden standard of the chemical accuracy (for systems without near degeneracies). The FMO-MP2 error in the total correlation energy relative to ab initio for the Trp-cage protein (6-31(+)-G*) was 2.1 and 0.16 kcal/mol for the two and three-body expansions, respectively.[76]

The accuracy of the correlation energy in FMO compares quite well with the local orbital correlation methods. The percentage of the CCSD(T) correlation energy recovery in water clusters with 3−8 molecules and the cc-pVDZ basis sets varied from 99.956 to 99.998 already at the two-body level. The local correlation methods[56,87−89] appear to work in the range of the 95−99% correlation energy recovery, which is acceptable for small molecules but may be insufficient to describe the dispersion in large systems like proteins.

The other interesting feature of MCSCF is its applicability to excited states. The FMO-based MCSCF was applied to the study of singlet (ground) and triplet (excited) states of phenol solvated in explicit water,[85] reproducing the ab initio MCSCF vertical excitation energies with the 0.007 eV error (for phenol in 64 water molecules, 6-31G*). It is generally thought that MCSCF delivers good geometries and an exciting application of FMO-MCSCF might be to study biological systems with near degeneracies (such as protein and enzyme systems involving transition metals).

Frequently, not all parts of the system have the same physical nature, that is, not all of them can be properly described by the same wave function. In addition, some part of the system may be more important than the rest, such as the active center of a chemical reaction. In these cases the need to describe fragments differently arises, and to satisfy it, the multilayer FMO method (MFMO) was introduced.[90]

In MFMO, all fragments are assigned to layers, and each layer may have its own wave function and basis set. It should be noted that the Coulomb field due to the whole system is added to all layers, so that both low and high layers are immersed in the full system (high layers are computed in the Coulomb field from their own (high layer) and the other (low layer) fragments. The main difference to the original (unilayer) case is that in MFMO, the pair interactions are computed at the lowest layer level of the two fragments between which the interaction is considered. The high level calculation can thus be limited to a small part of the system, which may be especially attractive if pair calculations are difficult to perform.

In the limiting case, the high layer has only one fragment, in which case all dimer calculations are performed at the lower level. This type of MFMO has been successfully applied[85] to the MCSCF studies of the ground and excited states and vertical excitations, the latter being described with the 0.009 eV error ($PhOH + (H_2O)_{64}$, 6-31G*) vs ab initio MCSCF. The decarboxylation reaction catalyzed by β-cyclodextrin was studied,[90] where the reaction center was treated with DFT/6-31G* and the rest (cyclodextrin) with RHF/3-21G*. The error in the activation barrier relative to the full DFT/6-31G* calculation was 1.0 kcal/mol while the computational time was reduced by a factor of 36. Another very interesting application of MFMO is to configuration interaction (CI) methods. Mochizuki et al. developed the FMO-based CI with single excitations (CIS)[91] and CIS with the perturbative correction for double excitations CIS(D).[92]

The computational scaling of FMO with respect to the basis set increase while fixing the number of atoms is determined by the scaling of individual ab initio monomer and dimer calculations and is thus like the corresponding ab initio methods. The scaling with respect to the system size with the fixed basis set is nearly linear, as has been shown on numerous occasions.[82,85,86] The means to enforce nearly linear scaling[86] are in the efficient use of the spatial separation, where carefully established approximations permit a drastic reduction of computational cost at nearly no loss of accuracy.

2.2. Calculation of Properties. The ability to obtain the properties of individual fragments, the pair corrections and the total properties are an attractive feature of FMO. As pointed out above, in FMO many properties can be computed following the expansion in eq 1. In particular, the accuracy of the dipole moments was studied for a number of wave functions[75,80,82,85] and Mochizuki et al.[93] reported FMO-based dynamic polarizability calculations.

The molecular orbital shapes and phases provide valuable insights into the nature of the chemical bonding. Inadomi et al.[94] implemented the approach of computing molecular orbitals and orbital energies using FMO, where the Fock matrix for the whole system is computed from the monomer and dimer densities and diagonalized once yielding the molecular orbitals and their energies for the whole system. Sekino et al.[95] also compared the FMO and full ab initio molecular orbitals.

TABLE 1: FMO-RHF Accuracy (rmsd) in Reproducing the ab Initio RHF Optimized Geometries

molecule	basis set	heavy (Å)[a]	backbone (Å)[b]	bond length (Å)[c]	bond angle (deg)[d]	ϕ (deg)[e]	ψ (deg)[e]	ω (deg)[e]
1UAO	3-21(+)G	0.097	0.041	0.0022	0.24	0.63	0.84	1.16
EMP1	3-21(+)G	0.095	0.073	0.0019	0.35	1.73	2.32	1.24
1L2Y	3-21G	0.198	0.157	0.0048	0.60	6.63	4.43	1.43

[a] Cartesian coordinates of all heavy atoms. [b] Cartesian coordinates of all backbone atoms. [c] All covalent bond lengths. [d] All covalent bond angles. [e] Peptide dihedral angles (see, e.g., ref 101 for their definition).

The biochemical applications pose a serious challenge to theoretical chemistry, as not only very large systems have to be computed, but the number of such calculations is necessarily large. Geometry optimizations typically take the number of steps similar to the number of atoms, so their cost for a 10 and 1000 atom molecule differs roughly by a factor of $100 \times 100 = 10000$ even assuming the linear scaling of single point energy calculations.

In addition, one has to face the issue of the configurational sampling dealing with finite temperature corrections, which frequently make a substantial contribution to the energetics of large systems. The solution is to perform molecular dynamics simulations. The straightforward applications of FMO to such studies[96−99] were possible only for small systems and have only recently been brought into the production level.[100] A satisfactory solution to the sampling problem (possibly done with methods other than FMO) is of high importance, and the applicability of FMO to real life biological processes is hindered by this issue.

To perform geometry optimizations of large systems, a number of methods based on fragmentation, semiempiric approaches, etc. has been proposed.[6−8,12,13,19,25] A comparison of the FMO-RHF and RHF optimized structures was conducted[101] for the α-helix, β-strand, and the extended form of the capped 10-residue polyalanine, a synthesized polypeptide chignolin (PDB: 1UAO), the Trp-cage protein (1L2Y), an agonist polypeptide of the erythropoietin receptor protein (EMP1, extracted from the complex in 1EBP), and met-enkephalin monomer and dimer.

The root mean square deviations (rmsd) from ab initio optimized structures of polypeptides with 138−304 atoms are summarized in Table 1. The FMO optimized bond lengths and angles agree with ab initio values within 0.005 Å and 0.6°, respectively. The overall rmsd for all atoms excluding hydrogen was about 0.1−0.2 Å. The flexible peptide dihedral angles ϕ, ψ, and ω are accurately reproduced in FMO optimizations, which is illustrated in Figure 1 for the Trp-cage protein, and the FMO-RHF and RHF optimized structures are shown in Figure 2, where it can be seen that they nearly coincide.

The effect of solvation upon the structure optimization was also considered by adding TIP3P water molecules to the FMO description of polypeptides. The rmsd between the optimized solvated and gas-phase structures of chignolin and the Trp-cage protein were 1.10 and 1.52 (Å), respectively. The difference between the gas-phase and solvated structures is due to the formation of salt bridges and hydrogen bonds, the distortion of their neighborhood and the extent to which the latter is spread. The effect of the distortion propagation is considerable for random coil but is largely blocked by the rigid conformations of α-helices and strong hydrogen bonds. The FMO structures of the solvated polypeptides showed good agreement with experiment, but a numeric accuracy criterion was difficult to establish, as the experimental structures of chignolin and the Trp-cage protein came from NMR and thus included many conformations (which suggests that geometry optimization may be viewed as the starting point to study dynamics and include the configuration sampling).

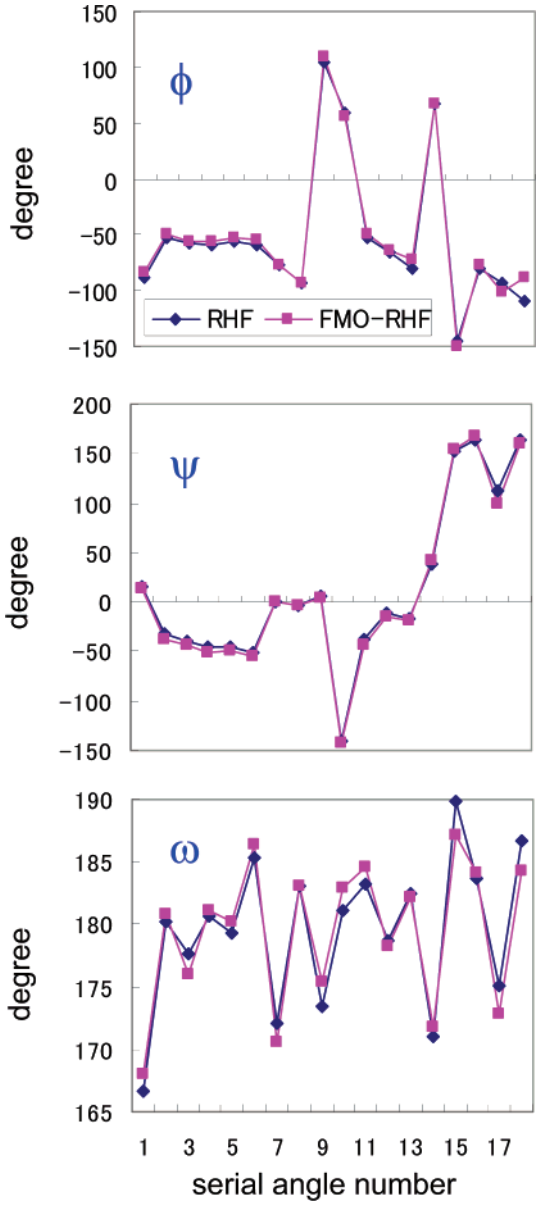

Figure 1. Comparison of the 3-21G optimized geometry parameters (the three peptide dihedral angles ϕ, ψ, and ω) of the Trp-Cage protein (1L2Y), FMO-RHF (blue diamonds) vs RHF (magenta squares). The angles are numbered from the N to C-terminus.

One of the important biological applications of theoretical methods is to drug design.[102,103] Drug discovery using grid technology[63c] takes advantage of FMO calculations to compute the docking energy. In the visualized cluster analysis of the protein−ligand interaction (VISCANA),[104] FMO is the main tool to compute the interaction energy between amino acid residues of a protein and a ligand, which is consequently visualized permitting an easy ligand comparison by the graphic representation of the interaction patterns.

Among other interesting and practically important properties developed with FMO, one can name the study of the isotope

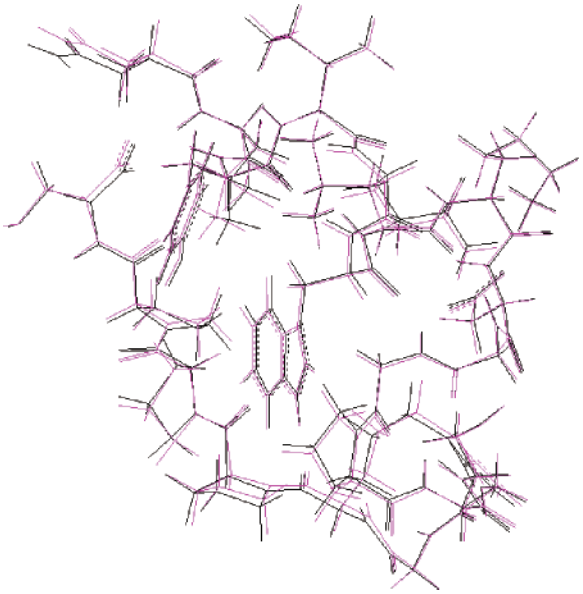

Figure 2. Overlay of the FMO-RHF (black) and RHF (magenta) optimized structures of the Trp-Cage protein (3-21G).

effect by Ishimoto et al.[105] By treating hydrogen atom nuclei quantum mechanically, it was possible to study the effect of the deuterium substitution in polypeptides and observe the weakening of deuterized hydrogen bonds due to their elongation (the Ubbelohde effect). The isotope substitution is not infrequently done experimentally, for the purpose of identifying the structural features, and this approach is useful to study the isotope effect in proteins.

In the FMO method, there is no limitation to the type of atom to describe; to extend the applicability to the systems containing heavy atoms, Ishikawa et al.[106] used model core potentials (MCP) and studied the Pt-containing DNA complex, and Hg^{2+} solvated in up to 256 explicit water molecules. FMO was also applied to Si-containing systems by Ishikawa et al.,[107] who demonstrated the good accuracy in describing systems fragmented along Si−Si and Si−O bonds.

The FMO code is freely distributed in two program packages, GAMESS[108−110] and ABINIT-MP.[63b,97,111] The former code was parallelized with the generalized distributed data interface (GDDI).[112] In GDDI, all computer nodes are divided into groups, and individual monomer and dimer calculations are performed on such small groups, whereby many difficulties in parallelizing ab initio QM methods on a large number of CPUs are greatly reduced, emphasizing, however, the problem of properly balancing the workload between groups. The latter problem is efficiently solved by the dynamic load balancing, and the GDDI-parallelized FMO code scales well on massively parallel computers. On 128 PC nodes connected by FastEthernet, the scaling of a large waters clusters $(H_2O)_{1024}$ and a protein complex (2036 atoms) was[112] 90 and 83%, respectively (90% scaling corresponds to about a $0.90 \times 128 = 115$-fold speed-up on 128 CPUs).

2.3. Solvation. Nearly all biological phenomena occur in aqueous solution. Although some processes may be well described by gas-phase calculations (e.g., excitations in inner protein regions, membrane proteins, etc.), in general one needs to take the solvent into account. The means to do that may be divided into two large groups: explicit solvent molecules and some averaged models. The former approach may rely on the corpuscular (explicit solvent) or potential (e.g., EFP[113]) representation of water molecules, but it faces the necessity to do the configurational sampling of water molecules. Although potentially the explicit treatment of solvent is more realistic and accurate, it is also much more expensive. It may be noted that in the absence of the solute, water molecules build hydrogen-bonding network among themselves, and when the solute is added, some water−water hydrogen bonds are replaced by typically the same number of solute−solvent hydrogen bonds. Thus the net effect is the very small difference between the two types of hydrogen bonds, and the entropy loss due to the reduced configuration space of water.[114]

The continuum approach to solvation avoids many of the above problems, as it accounts for the configurational sampling in some averaged way. Although it is sometimes argued that it does not properly describe the solvent−solute hydrogen bonding, as pointed out above, the net effect of solvation is the cancellation of pairs of hydrogen bonds rather than just the formation of those between solvent and solute. An alternative statistically averaged approach of the reference interaction site model (RISM) was also suggested,[115] and the Poisson−Boltzmann (PB) model for the electrostatics was used in its linearized form with the EFP method[116] or complemented by other interaction terms from the solvent accessible surface area (PB/SA).[117]

The polarizable continuum model (PCM)[118] is a model of solvation very well established for small molecules. It was interfaced[119] with FMO providing the means for treating the solvent effects while describing the large solute molecules quantum-mechanically with FMO. In PCM, the solute molecule is put in a polarizable cavity (representing the solvent), which is made of the union of atomic spheres, with the overlapping pieces removed. The atomic radii used for the cavity construction are parameters, frequently empirical. For small molecules, the united atom set of atomic radii[120] gained some popularity frequently attaining the desired level of accuracy in reproducing the experimental solvation free energies within 1 kcal/mol, although much larger errors are also found for molecules outside the fitting set. Recently, some progress has been achieved in enabling larger PCM calculations: Li et al.[121] interfaced it with QM/MM, Scalmani et al.[122] with MM, Barone et al.[123] with ONIOM, and Mei et al.[37] with MFCC.

The solvation free energy in FMO/PCM is divided into the same components, as in PCM:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{corr} + \Delta G_{cav} + \Delta G_{disp} + \Delta G_{rep} \quad (2)$$

The first term ΔG_{elec} gives the electronic energy difference between the total energies (usually RHF) for the solvated and gas-phase electronic states (this terms includes the electrostatic solute−solvent interaction). The second term ΔG_{corr} is the intramolecular solute correlation energy (e.g., MP2) difference between the solvated and gas-phase states. The third term ΔG_{cav} is the cavitation energy, parametrized to experiment to account for the energy required to form the cavity, which includes the solvent entropy loss due to the reduced configurational space. The remaining two terms ΔG_{disp} and ΔG_{rep} are the solute−solvent dispersion and repulsion energy, respectively, which roughly correspond to the correlated and uncorrelated energy contributions to the solvent−solute interactions (the latter term excludes the electrostatic and charge-transfer interaction).

FMO/PCM differs from the full PCM in the $\Delta G_{elec} + \Delta G_{corr}$ term, which is obtained from FMO rather than ab initio calculations (i.e., FMO-MP2); all other terms are exactly the same. Several levels of calculations were proposed, depending on the truncation of the many-body expansion of the solute electron density, which is used to determine the induced apparent

TABLE 2: Total Relative Stabilities ΔE_{tot}^{PCM} (kcal/mol) of the α Helices vs the β-Strands of the *n*-Residue Polyalanine with One Middle Residue Mutated to Glutamate, Computed with PCM/6-31G(+)G* and Divided into the Gas-Phase RHF (ΔE_{RHF}), MP2 Correlation[a] (ΔE_{corr}), and Solvation ($\Delta\Delta G_{solv}$) Energy Differences (e.g., $\Delta E_{RHF} = \Delta E_{RHF}^{helix} - \Delta E_{RHF}^{strand}$)

n	ΔE_{RHF}	ΔE_{corr}	$\Delta\Delta G_{solv}$	ΔE_{tot}^{PCM}
10	−16.7	−37.7	25.7	−28.6
20	−48.6	−82.8	69.4	−61.8
40	−133.4	−170.2	179.9	−123.7

[a] Obtained with FMO-MP2, two residues per fragment.

TABLE 3: Contributions (kcal/mol) to the FMO2-RHF/PCM[1(2)] Solvation Energy ΔG_{solv}, Decomposed into the Electronic ΔG_{elec}, Cavitation ΔG_{cav}, Dispersion ΔG_{disp}, and Repulsion ΔG_{rep} Energies (6-31(+)G*)

system[a]	ΔG_{elec}	ΔG_{cav}	ΔG_{disp}	ΔG_{rep}	ΔG_{solv}
1L2Y	−330.0	252.0	−167.1	40.4	−204.7
1IO5	−2249.9	1574.6	−622.1	147.8	−1149.7
2CGA	−2978.7	2935.6	−976.7	243.3	−776.6

[a] Divided into two residues per fragment.

surface charges (ASC) on the cavity. The most practical approach was found to be FMO/PCM[1(2)], in which the ASCs are determined self-consistently using the one-body expansion of the solute density, followed by a single ASC calculation with the two-body density. The electronic state of the solute is finally computed in the electrostatic field of the ASCs thus calculated.

To establish the FMO/PCM accuracy in comparison to full PCM, numerous test calculations were performed on the Trp-cage protein and 10-, 20-, and 40-residue polyalanines, including the mutants, in which one middle residue was changed into positively charged arginine and negatively charged glutamate. At the FMO/PCM[1(2)] level, the error in the total solvation energies was at most 0.8 kcal/mol for the systems with 112−412 atoms.

The relative stability of the solvated α-helices and β-strands of mutated polyalanine was studied,[119] and the representative data for the glutamate mutant are given in Table 2. It was found that although the mutant residue had a large effect upon the solvation energy (especially neutral vs charged), the relative stability (α-helix vs β-strand) is quite insensible to the change. The solvated α-helices are more stable than the β-strands due to the large stabilization from the electronic contributions and the intramolecular solute dispersion energy, whereas the solvation energy is more exothermic for the β-strands. The details of the individual contributors to the solvation energy (eq 2) can be found in ref 119. The obtained FMO/PCM value of the α-helix polyalanine being more stable than the β-strand by about −3.0 kcal/(mol·residue) can be compared with the Amber/PCM value[122] of about −5.8 kcal/(mol·residue).

The computed FMO/PCM[1(2)] solvation energy of the Trp-cage protein (1L2Y), Hen Egg Lysozyme (1IO5), and Bovine chymotrypsinogen chain A (2CGA) is summarized in Table 3. It can be seen that the very endothermic cavitation energy is compensated by a similar in magnitude electronic contribution (which includes the electrostatic solute−solvent interactions), and the solute−solvent dispersion + repulsion terms drive the solvation energy to be exothermic. The present absence of the experimental data (to the best of our knowledge) prohibits the direct comparison with experiment, and thus the issue of the best choice of the atomic radii for the cavity remains unsettled for polypeptides.

2.4. Pair Interaction Analysis. In the standard ab initio quantum-mechanical calculations, partial properties are rather limited to quantities like atomic charges, although some interesting techniques are being developed for ab initio density partitioning as well, so far limited to very small systems.[124] To elucidate the components of the interaction energy, Kitaura and Morokuma[125] proposed the energy decomposition analysis (EDA), and a number of researchers contributed to its further development[126−131] or suggested other schemes.[132−138]

To define the pair interactions in FMO, eq 1 is rewritten as follows

$$E = \sum_{I}^{N} E_I' + \sum_{I>J}^{N} (E_{IJ}' - E_I' - E_J') + \sum_{I>J}^{N} Tr(\Delta \mathbf{D}^{IJ}\mathbf{V}^{IJ}) \quad (3)$$

where E_I' and E_{IJ}' are the internal energies of monomers and dimers, respectively (both are obtained from the E_I and E_{IJ} energies in eq 1 by subtracting the electrostatic interaction with the external potential, e.g., $E_{IJ}' = E_{IJ} - Tr(\mathbf{D}^{IJ}\mathbf{V}^{IJ})$). $\Delta\mathbf{D}^{IJ}$is the density matrix difference of dimer *IJ* and the sum of monomer *I* and *J* electron densities and $\mathbf{V}^{IJ}$ is the electrostatic potential due to the external fragments acting upon dimer *IJ*. For those dimers, where the interfragment distance is large, the third term in eq 3 is very small and the second term can be approximated[66] by the electrostatic interaction between fragments *I* and *J*.

The pair interaction energy of a pair of fragments is given by

$$\Delta E_{IJ}^{int} = (E_{IJ}' - E_I' - E_J') + \mathrm{Tr}(\Delta \mathbf{D}^{IJ}\mathbf{V}^{IJ}) \quad (4)$$

The first term gives the amount of the pair interaction between the fragments polarized by the environment, and the second term is the interaction of the relaxed density (of dimer vs two monomers) with the external Coulomb field.

It is of interest to decompose the total interaction value of ΔE_{IJ}^{int} into contributions. In the configuration analysis for fragment interaction (CAFI),[139] the polarization (PL) and charge-transfer (CT) components were extracted, by performing the appropriate CIS calculations within each dimer. In addition to the total PL and CT values, it is possible to analyze the individual orbital contributions, which may be of special interest to some applications.

The pair interaction energy decomposition analysis (PIEDA)[140] was proposed as the FMO-based EDA, in which the bulk values of ΔE_{IJ}^{int} are decomposed into the same components as in EDA.

$$\Delta E_{IJ}^{int} = \Delta E_{IJ}^{ES} + \Delta E_{IJ}^{EX} + \Delta E_{IJ}^{CT+mix} + \Delta E_{IJ}^{DI} \quad (5)$$

The pair interaction energy is thus divided into the electrostatic (ES), exchange−repulsion (EX), charge-transfer plus higher order mixed terms (CT+mix), and dispersion (DI) contributions (DI absent in the original EDA is frequently added to it in the straightforward way). The notion of the polarization appears as the electrostatic stabilizing interaction between fragments (monomers) mutually destabilized in the system relative to their free state; the polarization component is thus obtained from monomer energies E_I' and a fraction of ΔE_{IJ}^{ES}. From eq 5 the coupling terms can be further extracted, such as the polarization−exchange, polarization−dispersion, polarization−charge transfer, and the many-body polarization terms.

For molecular clusters the free state is naturally available as the standalone molecules. With the fragmentation of covalent bonds, the definition of the free state is somewhat arbitrary. In the PIEDA scheme, such a state was defined for fragments with minimally possible caps, which for C−C bonds results in methyl

TABLE 4: PIEDA/6-31G* Contributions (kcal/mol) to the Stabilities of the α-Helix and *β*-Turn, Each Relative[a] to the Extended Form of the Capped $(ALA)_{10}$

	α-helix	*β*-turn
ΔE^{BB}	66.2	25.1
$\Delta\Delta E^{ES}$	−78.3	−43.3
$\Delta\Delta E^{EX}$	16.3	18.7
$\Delta\Delta E^{CT+mix}$	−16.8	−9.2
$\Delta\Delta E^{DI}$	−22.4	−17.3
ΔE^{tot}	−35.0	−26.0

[a] E.g., $\Delta\Delta E^{ES} = \Delta E^{ES,helix} - \Delta E^{ES,extended}$ and $\Delta E^{ES,helix} = \sum_{R\neq 0}^{I>J}\Delta E_{IJ}^{ES}$ is the sum of all unconnected dimer ES contributions for the α-helix.

caps. The issues related to the choice of the free state affect specifically the polarization energy, and the extraction of the coupling terms; however, the pair interaction energies between the unconnected fragments (which are not linked by a fractioned covalent bond) in eq 5 bears no such ambiguity.

A convenient way to consider the interactions is to divide the two dimer sums in eq 3 into the connected and unconnected dimer contributions. The sum of the former with the monomer energies is called the backbone energy, to which the latter terms add the nonbonding interaction of the molecular cluster type.

$$E = E^{BB} + \Delta E^{U,int}$$

$$E^{BB} = \sum_{I}^{N} E_I' + \sum_{\substack{I>J \\ R_{IJ}=0}}^{N} \Delta E_{IJ}^{int}$$

$$\Delta E^{U,int} = \sum_{\substack{I>J \\ R_{IJ}\neq 0}}^{N} \Delta E_{IJ}^{int} \qquad (6)$$

where the distance between two fragments R_{IJ} is zero if they are connected. The backbone energy E^{BB} thus represents the stability of the polarized chain of fragments without any other interactions (except for the joint points holding the chain together, the latter type is included in E^{BB}).

In contrast to EDA, PIEDA can be applied not only to molecular clusters, but also to covalently bound systems. The former type, which can be described by both methods was used for accuracy tests. The comparison is convenient to perform for the polarized state of free molecules,[140] in which case the components of the two are most closely related, as follows from the theoretical considerations. In $(H_2O)_8$, 6-31G*, the errors in the exchange-repulsion and charge transfer were 0.76 and 2.3 (kcal/mol), or 1.0 and 6.9 (%), respectively (the total interaction in PIEDA vs EDA had the error of 1.2 kcal/mol, or about 1.6%). The polarization and electrostatic components in PIEDA are exactly the same as in EDA (the former because of the self-consistent monomer relaxation in FMO and the latter because ES is pair additive and is perfectly described at the dimer level).

An application of PIEDA (MP2/6-31G*) is illustrated in Table 4, where we analyzed the reason of the greater stability of α-helices vs *β*-turns, comparing each of the two isomers of 10-residue polyalanine to the extended form. The backbone stress is much larger in the more distorted α-helix (by 41.1 kcal/mol), and it is compensated by the larger interfragment interaction energy for the unconnected pairs. In the α-helix, the electrostatic and charge-transfer energies are lower by 35.0 and 7.6 (kcal/mol), respectively, and the exchange repulsion is reduced by 2.4 kcal/mol. This is mostly due to the more numerous hydrogen bonds in the α-helix, although the individual hydrogen bonds are stronger in the *β*-turn, where they are enhanced by the strongly attractive dipole−dipole interaction. Finally, the dispersion is also stronger in the α-helix (by 5.1 kcal/mol), probably due to its more compact (three-dimensional) conformation. The net result is that the α-helix is more stable by 9.0 kcal/mol, or 0.9 kcal/(mol·residue).

TABLE 5: Interaction Analysis of the Trp-Cage Protein (1L2Y), Performed with PIEDA/6-31(+)G* (Charge Transfer $\Delta q_{I\to J}^{CT}$ in Atomic Units, All Energies in kcal/mol

pair	$\Delta q_{I\to J}^{CT}$	ΔE_{IJ}^{ES}	ΔE_{IJ}^{EX}	ΔE_{IJ}^{CT+mix}	ΔE_{IJ}^{DI}	ΔE_{IJ}^{int}
Arg-16, Asp-9[a]	−0.003	−70.8	2.7	−1.4	−1.6	−71.1
Lys-8, Gln-5[b]	0.060	−41.3	16.1	−5.9	−8.3	−39.4
Pro-17, Trp-6[c]	−0.040	−15.8	8.1	−3.3	−3.8	−14.8
Lys-8, Trp-6[a,d]	0.021	−9.3	2.1	−1.0	−3.0	−11.2
Asp-9, Trp-6[a]	0.028	13.3	0.6	−1.5	−2.0	10.3
Trp-6, Tyr-3[c]	0.029	−5.8	7.1	−2.9	−7.5	−9.1
Pro-18, Trp-6[e]	0.008	−2.6	6.2	−2.6	−7.6	−6.5
Pro-12, Trp-6[f]	0.006	−4.5	3.6	1.6	−5.3	−4.6
Pro-19, Trp-6[f]	0.009	−2.5	2.2	1.1	−3.6	−2.9
Gly-11, Trp-6[f]	0.012	−1.4	3.5	−1.3	−2.8	−2.1

[a] Electrostatic (charge−charge and charge-dipole) interaction. [b] Two hydrogen bonds. [c] Hydrogen bond. [d] Mostly dispersion (CH···O, sometimes referred to as a hydrogen bond; see ref 153 for the theoretical analysis of this type of bonding between a protein and ligands). In this case the fairly large amount of charge transfer is evoked by the charged residue (Lys-8). [e] Distorted hydrogen bond. [f] CH···π interaction (dispersion).

Next, we applied PIEDA to the study of the interactions in the Trp-Cage miniprotein (gas phase), as a model of a protein. The experimental geometry available from the PDB database was amended by the addition of hydrogen atoms (whose position was optimized with the Amber96 force field in Hyperchem) and the removal of crystal water. The Trp-Cage protein has 20 residues and all pair interactions are too numerous to discuss and only some (obtained with MP2/6-31(+)G*) are included in Table 5.

It should be born in mind that the fragments in FMO are slightly shifted relative to the conventional residues, and the discussion below refers, strictly speaking, to fragments. The reason for the difference is that the common residue definition relies on the semantic matter of convenience of choosing the basic unit, whereas in FMO the criterion is the divide at such places as to avoid the electron density delocalization, and the conventional place across the peptide bond involves some considerable electron delocalization.

The main components holding this miniprotein construct in the native conformation are (a) very strong electrostatic attraction between the oppositely charged Arg-16 and Asp-9 (−71.1 kcal/mol in gas phase; the short distance between them permits a fairly weak solvent screening, which according to some estimates would reduce the interaction to about one-half, much less than the typical factor of $1/\epsilon \approx 1/80$ for charges surrounded by solvent), and (b) a number of hydrogen bonds, of which a particularly strong example is by the two such bonds between Lys-8 and Gln-5 further enhanced by the charge-dipole interaction (the total pair interaction is −39.4 kcal/mol, the large part of which comes from the electrostatic energy).

Trp-6 can be seen to be the key binder in the Trp-Cage protein, as it interacts strongly with a very large number of other residues. The schematic structure of the protein and the pair interactions of Trp-6 with other residue fragments are shown in Figure 3. The total amount of the connected pair interactions involving Trp-6 is −46.9 kcal/mol, which is divided into the ES, EX, CT+mix, and DI contributions as −29.6, +37.5, −12.5, and −42.3 kcal/mol, respectively. Trp-6 is involved in a number of hydrogen bonds with Tyr-3, Pro-17, and Pro-18, and due to

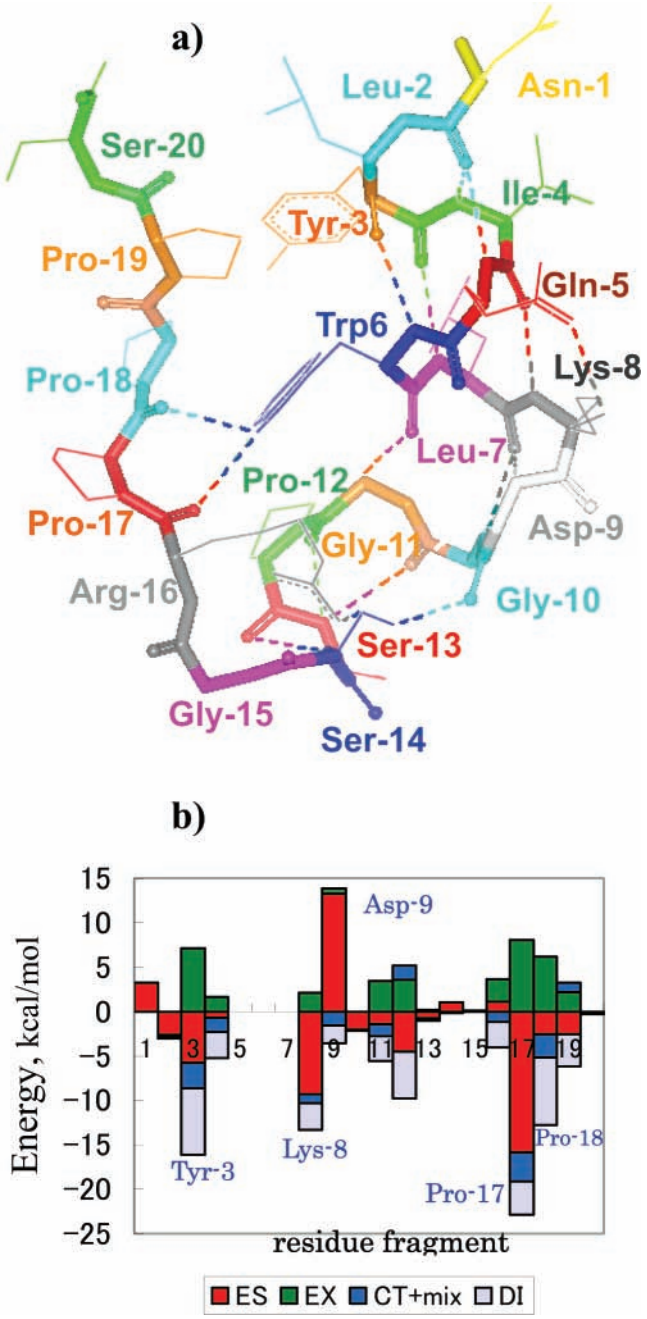

Figure 3. (a) Schematic diagram of the Trp-Cage protein and (b) the PIEDA/6-31G* results for the interaction of TRP-6 with other unconnected residue fragments. The pair interaction energy is divided into the electrostatic (ES), exchange−repulsion (EX), charge-transfer plus higher order mixed terms (CT+mix), and dispersion (DI) contributions.

its long side chain with the indole ring it is a source of the large attractive dispersion (hydrophobic) interaction.

Lys-8 forms a bond with Trp-6, which is mostly electrostatic with a large dispersion contribution (the C−H···O bond between them is sometimes assigned as a hydrogen bond). Another important stabilization factor is Arg-16, which has the total interaction energy with other fragments of −38.6 kcal/mol (excluding the interaction between oppositely charged pairs, largely weakened in solution). It helps to reduce the tension at the turning point by the strongly attractive electrostatic interaction (−31.5 kcal/mol). We also note that an experimental NMR structure was used in the above demonstration of PIEDA, and the interaction details may change if the structure is refined with optimization.

2.5. Applications. One of the first barriers that quantum-mechanical methods must overcome to be applied to biochemical studies is the ability to treat large systems. The FMO method has been routinely used with systems of the globular protein size (several thousands atoms), and Ikegami et al.[141] succeeded in performing an FMO-RHF/6-31G* calculation of the photosynthetic reaction center of *Rhodopseudomonas viridis* consisting of 20 581 atoms and 164 442 basis functions, which was accomplished in 72.5 h on 600 2.0 GHz Opteron CPUs on the AIST Super Cluster.

The protein−ligand binding is one of the most important application fields of theoretical methods.[142] Fukuzawa et al.[143−145] applied FMO calculations to the binding of human estrogen receptor and its ligands. The correlation coefficient between the calculated FMO-RHF/STO-3G binding energies of 11 ligands and the experimental relative binding affinities was 0.837, whereas the value obtained with CHARMM was 0.035 (no correlation).[144] By combining the CAFI/6-31G* analysis with the dispersion treatment by MP2/6-31G*, Fukuzawa et al. performed the quantitative analysis of the pair interactions for the binding of 17β-estradiol to estrogen receptor.[145]

Nemoto et al.[146] used the FMO-RHF/3-21G calculations to analyze the protein−ligand interactions, involving pheromone binding protein (BmPBP) that occurs in silkworm moth *Bombyx mori*, captures and carries airborne pheromone molecules to a pheromone receptor. By dividing the ligand (bombykol) into four fragments, it was possible to determine the importance of the ligand pieces to the overall binding.

Sugiki et al.[147] calculated the interaction between the catabolite activator protein and cyclic adenosine monophosphate (AMP) using FMO-DFT/STO-3G and the PBE0 functional, and found that the electrostatic interaction plays the key role in the binding. The functions of key residues in the ligand-binding pocket of vitamin D receptor were analyzed at the FMO-RHF/6-31G* level by Yamagishi et al.,[148] who discussed the effectiveness of the three pharmacophore hydroxyl groups.

The interactions between the complex of the cyclic AMP receptor protein (CRP) and cyclic AMP bound to DNA were studied by Fukuzawa et al.[149] using FMO-MP2/6-31G, who found that although the nucleotide pair and CRP-DNA interactions are mostly electrostatic in nature and are described similarly to AMBER94, other interactions (base-CRP) were quite different between FMO and MM, and it was suggested that the classical approach is not enough to describe that type of interaction. Ito et al.[150] investigated the complex of liganded retinoid X receptor with steroid receptor coactivating factor-1 coactivator at the FMO-MP2/6-31G level, finding that the interaction of the latter with helix 12 is the main cause for the stabilization of the coactivator binding.

Sawada et al.[151] applied FMO-RHF/STO-3G to determine the reason of the avian influenza A virus hemagglutinin binding stronger to avian rather than human receptor, which is related to the issue of the outbreaks of avian and human influenza. It was concluded that the reason for the stronger binding and thus to the weaker human virulence lies in the single amino acid mutation and the difference it evokes in the interaction energies. In the consequent work[152] the importance of the protein bulkiness and complexity was examined at the FMO-RHF/STO-3G for the same protein-receptor complex, and it was found that truncated protein model results in a significant underestimate of the binding.

Nakanishi et al.[153] studied the molecular recognition mechanism of FK506 Binding Protein (FKBP), using FMO-MP2/6-31G*, complemented by the solvation energies obtained with

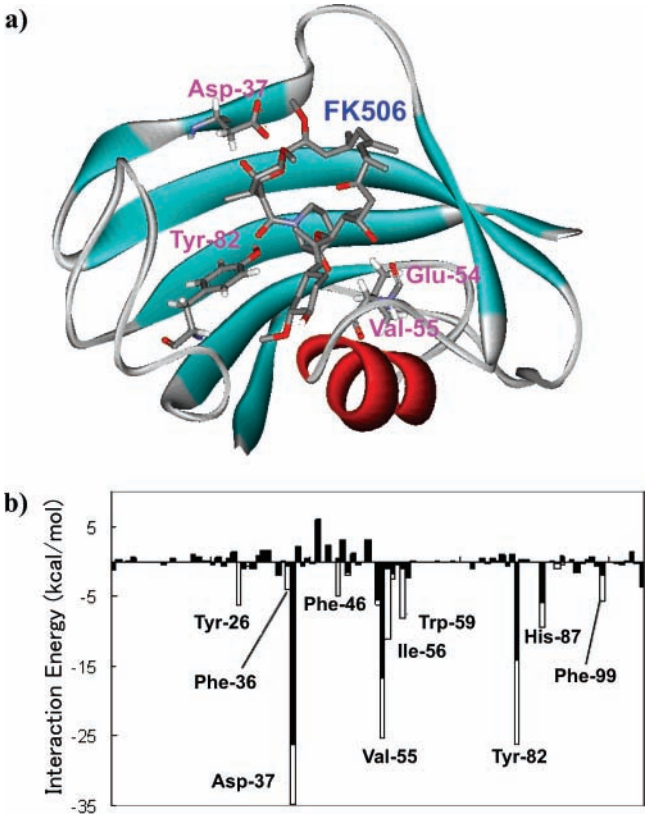

Figure 4. (a) Structure of the ligand−protein complex (FK506) used in the binding energy calculations. The important residues in the protein are shown with sticks. (b) Ligand−protein pair interaction energies (FK506). The FMO-RHF/6-31G* values (filled bars) are complimented by the FMO-MP2/6-31G* correlation contributions (empty bars).

PB/SA. The structure of the protein complex with the FK506 ligand and the pair interaction analysis are shown in Figure 4. The FKBP binding study stressed the importance of the electron correlation and solvation to theoretically predict the binding energies, which were similar to the experimental values with deviations of several kcal/mol, and the order of the binding strength of the four ligands was not reproduced. The detailed protein−ligand binding analysis provided the detailed information about hydrogen bonding, nonpolar interaction and the effect of the protein environment upon the residue-ligand pair interactions.

Ishida et al.[154] applied the multilayer method of FMO-RHF:MP2/6-31(+)G* to study the chemical reaction of Claisen rearrangement of chorismate to prephenate, catalyzed by *Bacillus subtilis* chorismate mutase. By comparing the wild type (Arg90) and the two lysine and citrulline mutants, they found that the catalytic behavior of the latter mutant is quite different from the former two cases, due to the substrate destabilization in the surrounding electrostatic field. The transition state stabilization was largely determined by the role of Arg90, which polarizes the substrate to gain the maximum electrostatic stabilization and controls the overall relative stability through the collective hydrogen-bonding network.

Komeiji et al.[155] used explicit solvent to model the solvation effects of proteins[133] with FMO-MP2/6-31G* (treating water quantum-mechanically) and suggested that the change in the protein electronic structure due to the charge transfer to the solvent and the solvent-induced polarization results in the general destabilization of the protein internal energy and, in particular, in the weakening of the intramolecular interactions, which can have significant implications in the ligand binding control and drug design.

The multilayer method of FMO-RHF:CIS/6-31G was applied[91] to the vertical excitations in the photoactive yellow protein. The lack of the electron correlation in the CIS method was compensated by the perturbative treatment of doubly excited configurations in CIS(D) in the consequent work,[156] where FMO-RHF:CIS(D)/6-31G* was used to compute the vertical and relaxed excitations in the red fluorescent protein and an excellent agreement with experiment (within 0.1 eV) was obtained. The molecular dynamics study of the n-π^* excitation in formaldehyde[100] performed at the FMO-RHF:CIS(D)/6-31G* level succeeded in reproducing the experimental value within 0.01 eV. Finally, in several studies, FMO was used as a secondary tool to elucidate some interaction details in biological systems.[157,158]

3. Summary and Future Outlook

Among the large number of fragment-based methods, only a few have been supported by a continuous method development and a series of applications validating their usefulness to practical problems. The fragment molecular orbital method has been developed by taking advantage of the standard apparatus of quantum chemistry, with the interface to a number of wave function types, along with other necessary tools such as the solvation model. The meticulous accuracy evaluation in comparison to ab initio methods performed for each wave function is expected to solicit interest of researchers who look for a reliable method to conduct practical studies of large systems.

The general trend in the fragment methods has been to increase the complexity and the accuracy, largely by increasing the inclusion of the environment (the rest of the system beyond the fragment treated explicitly). It can be expected that the efficient findings to accomplish that will be interfaced and reused in other methods, gradually decreasing the difference between them. With the advent of the multiple core CPUs parallel computations are becoming routine for the end users, and the high efficiency of the fragment methods in utilizing massively parallel computers is an advantage.

The applications of FMO clearly show the continuous increase in the basis set quality and the treatment of the electron correlation. A number of problems remains to be solved, such as the difficulties in the diffuse function description, an appropriate set of PCM atomic radii for polypeptides, and the very important treatment of the configurational sampling. Although it can hardly be expected that ab initio based methods can without fail reproduce the experimental values of binding and reaction energetics for large biological systems with the 1 kcal/mol accuracy (which is often the difference between several ligands in the protein−ligand binding), the aid of the detailed interaction analysis in understanding the binding mechanism may be of considerable interest in designing new drugs and suggesting mutations of enzymes to improve their efficacy. After all, the goal of theoretical studies is not so much in the mere reproduction of the experimental values, but in the understanding of the driving principles and the design of novel agents for practical problems.

Acknowledgment. This work was partially supported by the Next Generation Supercomputing Project, Nanoscience Program (MEXT, Japan), a Grant-in-Aid for Scientific Research (JSPS, Japan), and the CREST project (JST, Japan).

References and Notes

(1) Scuseria, G. E. *J. Phys. Chem. A.* **1999**, *103*, 4782.
(2) Sato, F.; Yoshihiro, T.; Era, M.; Kashiwagi, H. *Chem. Phys. Lett.* **2001**, *341*, 645.
(3) Inaba, T.; Tahara, S.; Nisikawa, N.; Kashiwagi, H.; Sato, F. *J. Comput. Chem.* **2005**, *26*, 987.
(4) Inaba, T.; Sato, F. *J. Comput. Chem.* **2007**, *28*, 984.
(5) Alsenoy, C. V.; Yu, C. H.; Peeters, A.; Martin, J. M. L.; Schäfer, L. *J. Phys. Chem. A* **1998**, *102*, 2246.
(6) Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. *Int. J. Quantum Chem.* **2004**, *97*, 747.
(7) Wada, M.; Sakurai, M. A. *J. Comput. Chem.* **2005**, *26*, 160.
(8) Yu, N.; Yennawar, H. P.; Merz, K. M., Jr. *Acta Crystallogr. Sect. D* **2005**, *61*, 322.
(9) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
(10) Friesner, R. A.; Guallar, Y. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
(11) Svensson, M.; Humbel, S.; Froese, R. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
(12) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.
(13) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. *J. Chem. Phys.* **2006**, *124*, 024301.
(14) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.
(15) Minikis, R. M.; Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2001**, *105*, 3829.
(16) Klessinger, M.; McWeeny, R. *J. Chem. Phys.* **1965**, *42*, 3343.
(17) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(18) Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.
(19) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.
(20) Deshmukh, M. M.; Gadre, S. R.; Bartolotti, L. J. *J. Phys. Chem. A* **2006**, *110*, 12519.
(21) Collins, M. A.; Deev, V. *J. Chem. Phys.* **2006**, *125*, 104104.
(22) Das, G. P.; Yeates, A. T.; Dudis, D. S. *Int. J. Quantum Chem.* **2003**, *92*, 22.
(23) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.
(24) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
(25) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039.
(26) Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. *Chem. Phys. Lett.* **2004**, *394*, 293.
(27) Chen, X. H.; Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 839.
(28) Zhang, D. W.; Xiang, Y.; Gao, A. M.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 1145.
(29) Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, 120, 11386; **2006**, *125*, 044903.
(30) Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Comput. Chem.* **2004**, *25*, 1431.
(31) Chen, X.; Zhang, Y.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 184105.
(32) Zhang, D. W.; Zhang, J. Z. H. *Int. J. Quantum Chem.* **2005**, *103*, 246.
(33) Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Phys. Chem. A* **2005**, *109*, 2.
(34) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7251.
(35) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112.
(36) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 031103; **2006**, *124*, 184703.
(37) Mei, Y.; Ji, C.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *125*, 094906.
(38) Mei, Y.; Wu, E. L.; Han, K. L.; Zhang, J. Z. H. *Int. J. Quantum Chem.* **2006**, *106*, 1267.
(39) Mei, Y.; He, X.; Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *Proteins-Struct. Funct. Bioinform.* **2005**, *59*, 489.
(40) He, X.; Mei, Y.; Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *Proteins-Struct. Funct. Bioinform.* **2005**, *61*, 423.
(41) Morita, S.; Sakai, S. *J. Comput. Chem.* **2001**, *22*, 1107.
(42) Sakai, S.; Morita, S. *J. Phys. Chem. A* **2005**, *109*, 8424.
(43) Imamura, A.; Aoki, Y.; Maekawa, K. *J. Chem. Phys.* **1991**, *95*, 5419.
(44) Aoki, Y.; Suhai, S.; Imamura, A. *Int. J. Quantum Chem.* **1994**, *52*, 267.
(45) Aoki, Y.; Suhai, S.; Imamura, A. *J. Chem. Phys.* **1994**, *101*, 10808.
(46) Mitani, M.; Aoki, Y.; Imamura, A. *Int. J. Quantum Chem.* **1995**, *54*, 167; **1997**, *64*, 301.
(47) Rather, G.; Aoki, Y.; Imamura, A. *Int. J. Quantum Chem.* **1999**, *74*, 35.
(48) Gu, F. L.; Aoki, Y.; Imamura, A.; Bishop, D. M.; Kirtman, B. *Mol. Phys.* **2003**, *101*, 1487.
(49) Gu, F. L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. *J. Chem. Phys.* **2004**, *121*, 10385.
(50) Korchowiec, J.; Gu, F. L.; Imamura, A.; Kirtman, B.; Aoki, Y. *Int. J. Quantum Chem.* **2005**, *102*, 785.
(51) Korchowiec, J.; Gu, F. L.; Aoki, Y. *Int. J. Quantum Chem.* **2005**, *105*, 875.
(52) Makowski, M.; Korchowiec, J.; Gu, F. L.; Aoki, Y. *J. Comput. Chem.* **2006**, *27*, 1603.
(53) Orimoto, Y.; Aoki, Y. *J. Polym. Sci. B: Polym. Phys.* **2006**, *44*, 119.
(54) Nakatsuji, H.; Miyahara, T.; Fukuda, R. *J. Chem. Phys.* **2007**, *126*, 084104.
(55) Stoll, H. *Phys. Rev. B.* **1992**, *46*, 6700.
(56) Paulus, B. *Phys. Rep.* **2006**, *428*, 1.
(57) Rościszewski, K.; Paulus, B.; Fulde, P.; Stoll, H. *Phys. Rev. B.* **2000**, *62*, 5482.
(58) Yu, M.; Kalvoda, S.; Dolg, M. *Chem. Phys.* **1997**, *224*, 121.
(59) Buth, C.; Paulus, B. *Chem. Phys. Lett.* **2004**, *398*, 44.
(60) Willnauer, C.; Birkenheuer, U. *J. Chem. Phys.* **2004**, *120*, 11910.
(61) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
(62) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
(63) (a) Fedorov, D. G.; Kitaura, K. In *Modern Methods for Theoretical Physical Chemistry and Biopolymers*; Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier: Amsterdam, 2006; pp 3−38. (b) Nakano, T.; Mochizuki, Y.; Fukuzawa, K.; Amari, S.; Tanaka, S. In *Modern Methods for Theoretical Physical Chemistry and Biopolymers*; Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier: Amsterdam, 2006; pp 39−52. (c) Goto, H.; Obata, S.; Kamakura, T.; Nakayama, N.; Sato, M.; Nakajima, Y.; Nagashima, U.; Watanabe, T.; Inadomi, Y.; Ito, M.; Nishikawa, T.; Nakano, T.; Nilsson, L.; Tanaka, S.; Fukuzawa, K.; Inagaki, Y.; Hamada, M.; Chuman, H. In *Modern Methods for Theoretical Physical Chemistry and Biopolymers*; Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier: Amsterdam, 2006; pp 227−248.
(64) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255.
(65) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theor. Comput.* **2007**, *3*, 46.
(66) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.
(67) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2002**, *106*, 11791.
(68) Exner, T. E.; Mezey, P. G. *J. Comput. Chem.* **2003**, *24*, 1980.
(69) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2004**, *108*, 4301.
(70) Szekeres, Z.; Exner, T. E.; Mezey, P. G. *Int. J. Quant. Chem.* **2005**, *104*, 847.
(71) Exner, T. E.; Mezey, P. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061.
(72) *FMOutil*, version 2.1; http://staff.aist.go.jp/d.g.fedorov/fmo/fmoutil.html.
(73) Kitaura, K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.* **2001**, *336*, 163.
(74) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2006**, *433*, 182.
(75) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
(76) Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. *J. Comp. Chem.* **2007**, *28*, 1476.
(77) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nature Struct. Biol.* **2002**, *9*, 425.
(78) Yasuda, K.; Yamaki, D. *J. Chem. Phys.* **2006**, *125*, 154101.
(79) Sugiki, S.-I.; Kurita, N.; Sengoku, Y.; Sekino, H. *Chem. Phys. Lett.* **2003**, *382*, 611.
(80) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129.
(81) Shimodo, Y.; Morihashi, K.; Nakano, T. *J. Mol. Struct. (THEOCHEM)* **2006**, *770*, 163.
(82) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.
(83) Mochizuki, Y.; Koikegami, S.; Nakano, T.; Amari, S.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *396*, 473.
(84) Mochizuki, Y.; Nakano, T.; Koikegami, S.; Tanimori, S.; Abe, Y.; Nagashima, U.; Kitaura, K. *Theor. Chem. Acc.* **2004**, *112*, 442.
(85) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *122*, 054108.
(86) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
(87) Pulay, P. *Chem. Phys. Lett.* **1983**, *100*, 151.
(88) Schütz, M.; Hetzer, G.; Stoll, H.; Werner, H.-J. *J. Chem. Phys.* **1999**, *111*, 5691.
(89) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592.
(90) Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638.
(91) Mochizuki, Y.; Koikegami, S.; Amari, S.; Segawa, K.; Kitaura, K.; Nakano, T. *Chem. Phys. Lett.* **2005**, *406*, 283.
(92) Mochizuki, Y.; Tanaka, K.; Yamashita, K.; Ishikawa, T.; Nakano, T.; Amari, S.; Segawa, K.; Murase, T.; Tokiwa, H.; Sakurai, M. *Theor. Chem. Acc.* **2007**, *117*, 541.
(93) Mochizuki, Y.; Ishikawa, T.; Tanaka, K.; Tokiwa, H.; Nakano, T.; Tanaka, S. *Chem. Phys. Lett.* **2006**, *418*, 418.

(94) Inadomi, Y.; Nakano, T.; Kitaura, K.; Nagashima, U. *Chem. Phys. Lett.* **2002**, *364*, 139.
(95) Sekino, H.; Sengoku, Y.; Sugiki. S.-I.; Kurita, N. *Chem. Phys. Lett.* **2003**, *378*, 589.
(96) Komeiji, Y.; Nakano, T.; Fukuzawa, K.; Ueno, Y.; Inadomi, Y.; Nemoto, T.; Uebayasi, M.; Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2003**, *372*, 342.
(97) Komeiji, Y.; Inadomi, Y.; Nakano, T. *Comput. Biol. Chem.* **2004**, *28*, 155.
(98) Ishimoto, T.; Tokiwa, H.; Teramae, H.; Nagashima, U. *Chem. Phys. Lett.* **2004**, *387*, 460.
(99) Ishimoto, T.; Tokiwa, H.; Teramae, H.; Nagashima, U. *J. Chem. Phys.* **2005**, *122*, 094905.
(100) Mochizuki, Y.; Komeiji, Y.; Ishikawa, T.; Nakano, T.; Yamataka, H. *Chem. Phys. Lett.* **2007**, *437*, 66.
(101) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A.* **2007**, *111*, 2722.
(102) Peters, M. B.; Raha, K.; Merz, K. M., Jr. *Curr. Opin. Drug Disc. Dev.* **2006**, *9*, 370.
(103) Cavalli, A.; Carloni, P.; Recanatini, M. *Chem. Rev.* **2006**, *106*, 3497.
(104) Amari, S.; Aizawa, M.; Zhang, J.; Fukuzawa, K.; Mochizuki, Y.; Iwasawa, Y.; Nakata, K.; Chuman, H.; Nakano, T. *J. Chem. Inf. Comput. Sci.* **2006**, *46*, 221.
(105) Ishimoto, T.; Tachikawa, M.; Nagashima, U. *J. Chem. Phys.* **2006**, *124*, 014112.
(106) Ishikawa, T.; Mochizuki, Y.; Nakano, T.; Amari, S.; Mori, H.; Honda, H.; Fujita, T.; Tokiwa, H.; Tanaka, S.; Komeiji, Y.; Fukuzawa, K.; Tanaka, K.; Miyoshi, E. *Chem. Phys. Lett.* **2006**, *427*, 159.
(107) Ishikawa, T.; Mochizuki, Y.; Imamura, K.; Nakano, T.; Mori, H.; Tokiwa, H.; Tanaka, K.; Miyoshi, E.; Tanaka, S. *Chem. Phys. Lett.* **2006**, *430*, 361.
(108) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347.
(109) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: New York, 2005; pp 1167−1189.
(110) *GAMESS*, http://www.msg.ameslab.gov/.
(111) *ABINIT-MP*, http://moldb.nihs.go.jp/abinitmp/.
(112) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.* **2004**, *25*, 872.
(113) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.
(114) Finkelstein, A. V.; Ptitsyn, O. B. *Protein Physics: A Course of Lectures*; Academic Press: Amsterdam, 2002.
(115) Ten-no, S.; Hirata, F.; Kato, S. *Chem. Phys. Lett.* **1993**, *214*, 391.
(116) Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106*, 3486.
(117) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
(118) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
(119) Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. *J. Comput. Chem.* **2006**, *27*, 976.
(120) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.
(121) Li, H.; Pomelli, C. S.; Jensen, J. H. *Theor. Chem. Acc.* **2003**, *109*, 71.
(122) Scalmani, G.; Barone, V.; Kudin, K. N.; Pomelli, C. S.; Scuseria, G. E.; Frisch, M. J. *Theor. Chem. Acc.* **2004**, *111*, 90.
(123) Barone, V.; Improta, R.; Rega, N. *Theor. Chem. Acc.* **2004**, *111*, 237.
(124) Pendás, A. M.; Blanco, M. A.; Francisco, E. *J. Comput. Chem.* **2007**, *28*, 161.
(125) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325.
(126) Sokalski, W. A.; Roszak, S.; Hariharan, P. C.; Kaufman, J. *J. Int. J. Quantum Chem.* **1983**, *23*, 847.
(127) Cammi, R.; Bonaccorsi, R.; Tomasi, J. *Theor. Chim. Acta* **1985**, *68*, 271.
(128) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15.
(129) Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *100*, 14316.
(130) Curutchet, C.; Bofill, J. M.; Hernández, B.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2003**, *24*, 1263.
(131) Kawamura, Y.; Nakai, H. *J. Comput. Chem.* **2004**, *25*, 1882.
(132) van der Vaart, A.; Merz, K. M., Jr. *J. Phys. Chem. A* **1999**, *103*, 3321.
(133) van der Vaart, A.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1999**, *121*, 9182.
(134) Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530.
(135) Titmuss, S. J.; Cummins, P. L.; Rendell, A. P.; Bliznyuk, A. A.; Gready, J. E. *J. Comput. Chem.* **2002**, *23*, 1314.
(136) Glendening, E. D. *J. Phys. Chem. A* **2005**, *109*, 11936.
(137) Raha, K.; van der Vaart, A. J.; Riley, K. E.; Peters, M. B.; Westerhoff, L. M.; Kim, H.; Merz, K. M.; Jr. *J. Am. Chem. Soc.* **2005**, *127*, 6583.
(138) Ababou, A.; van der Wart, A.; Gogonea, V.; Merz, K. M., Jr. *Biophys. Chem.* **2007**, *125*, 221.
(139) Mochizuki, Y.; Fukuzawa, K.; Kato, A.; Tanaka, S.; Kitaura, K.; Nakano, T. *Chem. Phys. Lett.* **2005**, *410*, 247.
(140) Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.* **2007**, *28*, 222.
(141) Ikegami, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Inadomi, Y.; Umeda, H.; Yokokawa, M.; Sekiguchi, S. *Proc. Supercomputing 2005;* IEEE Computer Society: Seattle, 2005.
(142) Raha, K.; Merz, K. M., Jr. *J. Med. Chem.* **2005**, *48*, 4558.
(143) Fukuzawa, K.; Kitaura, K.; Nakata, K.; Kaminuma, T.; Nakano, T. *Pure Appl. Chem.* **2003**, *75*, 2405.
(144) Fukuzawa, K.; Kitaura, K.; Uebayasi, M.; Nakata, K.; Kaminuma, T.; Nakano, T. *J. Comput. Chem.* **2005**, *26*, 1.
(145) Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. *J. Phys. Chem. B* **2006**, *110*, 16102.
(146) Nemoto, T.; Fedorov, D. G.; Uebayasi, M.; Kanazawa, K.; Komeiji, Y.; Kitaura, K. *Comput. Biol. Chem.* **2005**, *29*, 434.
(147) Sugiki, S.-I.; Matsuoka, M.; Usuki, R.; Sengoku, Y.; Kurita, N.; Sekino, H.; Tanaka, S. *J. Theor. Comput. Chem.* **2005**, *4*, 183.
(148) Yamagishi, K.; Yamamoto, K.; Yamada, S.; Tokiwa, H. *Chem. Phys. Lett.* **2006**, *420*, 465.
(149) Fukuzawa, K.; Komeiji, Y.; Mochizuki, Y.; Kato, A.; Nakano, T.; Tanaka, S. *J. Comput. Chem.* **2006**, *27*, 948.
(150) Ito, M.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. *J. Phys. Chem. B* **2007**, *111*, 3525.
(151) Sawada, T.; Hashimoto, T.; Nakano, H.; Suzuki, T.; Ishida, H.; Kiso, M. *Biochem. Biophys. Res. Commun.* **2006**, *351*, 40.
(152) Sawada, T.; Hashimoto, T.; Nakano, H.; Suzuki, T.; Suzuki, Y.; Kawaoka, Y.; Ishida, H.; Kiso, M. *Biochem. Biophys. Res. Commun.* **2007**, *355*, 6.
(153) Nakanishi, I.; Fedorov, D. G.; Kitaura, K. *Proteins-Struct. Funct. Bioinform.*, in press (DOI: 10.1002/prot.21389).
(154) Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. B* **2006**, *110*, 1457.
(155) Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.*, in press (DOI: 10.1002/jcc.20686).
(156) Mochizuki, Y.; Nakano, T.; Amari, S.; Ishikawa, T.; Tanaka, K.; Sakurai, M.; Tanaka, S. *Chem. Phys. Lett.* **2007**, *433*, 360.
(157) Yamamoto, K.; Abe, D.; Yoshimoto, N.; Choi, M.; Yamagishi, K.; Tokiwa, H.; Shimizu, M.; Makishima, M.; Yamada, S. *J. Med. Chem.* **2006**, *49*, 1313.
(158) Tsukamoto, K.; Shimizu, H.; Ishida, T.; Akiyama, Y.; Nukina, N. *J. Mol. Struct. (THEOCHEM)* **2006**, *778*, 85.