The Fragment Molecular Orbital Method for Geometry Optimizations of Polypeptides and Proteins

Dmitri G. Fedorov,* Toyokazu Ishida, Masami Uebayasi, and Kazuo Kitaura

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

Received: October 29, 2006; In Final Form: January 30, 2007

The fragment molecular orbital method (FMO) has been used with a large number of wave functions for single-point calculations, and its high accuracy in comparison to ab initio methods has been well established. We have developed the analytic derivative of the electrostatic interaction between far separated fragments and performed a number of restricted Hartree−Fock (RHF) geometry optimizations using FMO and ab initio methods. In particular, the α-helix, β-turn, and extended conformers of a 10-residue polyalanine were studied and the good FMO accuracy was established (the rms deviations for the former two forms were about 0.2 Å and for the latter structure about 0.001 Å). Met-enkephalin dimer was used as a model for the polypeptide binding and computed at the 3-21G and 6-31G* levels with a similar accuracy achieved; the error in the binding energy predictions (FMO vs ab initio) was 1−3 kcal/mol. Chignolin (PDB: 1uao) and an agonist polypeptide of the erythropoietin receptor protein (emp1) were optimized at the 3-21(+)G level, with the rms deviation from ab initio of about 0.2 Å, or 0.5° in terms of bond angles. The effect of solvation on the structure optimization was studied in chignolin and the Trp-cage miniprotein construct (PDB:1l2y), by describing water with TIP3P. The computed structures in gas phase and solution are compared to each other and experiment.

1. Introduction

The quantum-chemical methods have demonstrated their high accuracy and generality in describing molecular systems from the first principles. However, the traditional approaches require computational resources that scale very steeply with the system size. Consequently, the applications of ab initio methods, even to single-point calculations of systems of biological size, are scarce. The calculations of cytochrome C and insulin were reported using PROTEINDF.[1,2] Geometry optimizations pose a far greater challenge, as they typically require hundreds of single-point calculations to optimize flexible biological systems such as proteins. Among the first applications, Alsenoy et al.[3] optimized a small protein with restricted Hartree−Fock (RHF) and the 4-21G basis set. More progress has been achieved in treating large systems with semiempirical methods,[4−6] the ONIOM scheme,[7,8] the quantum mechanics/molecular mechanics (QM/MM) approach,[9,10] and the effective fragment potential theory.[11] The incremental correlation method has been successful in treating the electron correlation in the periodic systems such as crystals and polymers.[12,13]

Fragment-based methods are a fast-developing field of research[14] whereby the system is divided into pieces commonly called fragments and the total properties are obtained from those of fragments and their conglomerates. To name just a few, the molecular tailoring approach,[15,16] the molecular fractionation with conjugate caps,[17,18] and the elongation approach[19,20] use different strategies in describing the effect of the environment upon the electronic structure of fragments; several FMO-like energy expansions have also been suggested.[21,22] However, the ability to perform geometry optimizations remains rather limited.[23,24]

The fragment molecular orbital (FMO) method originally proposed by Kitaura et al.[25] has been shown[26] to closely reproduce ab initio properties, including the energy, energy gradient, and dipole moment. The original RHF implementation has been generalized into a variety of common wave function types: density functional theory,[27] second-order Møller−Plesset perturbation theory,[28] multiconfiguration self-consistent field,[29] and coupled cluster,[30] all of which can be combined into a multilayer scheme.[31] Nearly analytic RHF gradients have been developed by Kitaura et al.[32]

The polarizable continuum model has been interfaced with the FMO method,[33] providing means for treating solvent effects. The efficiency in handling large molecules has been demonstrated in the all-electron FMO study of a system with more than 20 000 atoms.[34] The recently proposed pair interaction energy decomposition analysis (PIEDA)[35] demonstrated the high importance of the dispersion interaction in polypeptides and water clusters. The usefulness of the FMO method for treating large systems has been exemplified in a number of applications to biological systems.[36−41]

The organization of this paper is as follows. After a brief FMO introduction, an important improvement in the computation of the FMO gradient is described. For the purpose of comparing the FMO and ab initio optimizations, we use RHF wave function throughout. Next, a careful analysis of the accuracy in reproducing the optimized geometries for polypeptides and small proteins was conducted in comparison to the ab initio structures. In particular, we studied the α-helix (denoted by α-ala10), β-turn (β-ala10), and extended (e-ala10) conformers of 10-residue polyalanine using the 3-21G and 6-31G* basis sets. The basis set effects upon geometry, as well as the accuracy in reproducing the relative energetics, are discussed. Dimer of met-enkephalin was used as a model of the polypeptide

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

10.1021/jp0671042 CCC: $37.00
Published on Web 03/16/2007

TABLE 1: Charges *Q* and the Number of Residues N_R, Atoms N_A, and Basis Functions N_{BF} for All Systems

molecule	*Q*	N_R	N_A	N_{BF}
MeCO-$(Ala)_{10}$-NHMe	0	10	112	609[a], 939[c]
Met-enkephalin				
monomer	0	5	75	446[b], 686[d]
dimer	0	10	150	892[b], 1372[d]
chignolin (1uao)	−2	10	138	851[b]
emp1 (1ebp)	2	16	254	1445[b]
1l2y	1	20	304	1686[a]

[a] 3-21G. [b] 3-21(+)G. [c] 6-31G*. [d] 6-31(+)G*.

interaction, and the 3-21(+)G and 6-31(+)G* geometry optimization results are discussed ((+) denotes that diffuse functions were added to all carboxyl groups). Chignolin (PDB: 1uao) and an agonist polypeptide of erythropoietin receptor protein (emp1) were optimized at the 3-21(+)G level.

Finally, we considered the importance of the solvent effects upon geometry by performing geometry optimizations of chignolin and Trp-cage miniprotein construct (PDB: 1l2y) in gas phase as well as in the first solvation shell, modeled by molecular mechanics combined with FMO.[42] The solvated structures were compared to experiment and a reasonable agreement was achieved.

2. Methodology

2.1. Theory. In the two-body expansion of the FMO method that we use for geometry optimizations, the total energy *E* is given by:

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

where the 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. The rigorous Coulomb field (including contributions from both nuclei and electron density) allows proper description of the density distribution in large systems, eliminating the need to cap dangling bonds found in almost all other fragment-based methods.

The FMO calculation begins by making an initial guess of the electron density for each fragment and converging the monomer densities in the external Coulomb field added to the Fock matrix for each fragment. The monomer calculations change the total Coulomb field and they are repeated self-consistently. Consequently, all pairs of fragments are computed, close pairs as RHF in the external Coulomb field and remote ones as the electrostatic (ES) interaction between nuclei and electron densities in the two fragments (the ES approximation). The original gradient work by Kitaura et al.[32] did not introduce the derivative of the latter dimer energy type, so it was implemented in this work. Note that the number of dimers for which SCF is performed is in general linear[14] (proportional to the number of fragments) and all other dimers are computed in the ES approximation (their number being quadratic). Development of the gradient for the latter terms leads to a very significant reduction of the computational cost.

Equation 1 can be rewritten as follows,

$$E = \sum_{I}^{N} E'_I + \sum_{I>J}^{N} (E'_{IJ} - E'_I - E'_J) + \sum_{I>J}^{N} Tr(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \tag{2}$$

where E'_I and E'_{IJ} are the internal energies of monomers and dimers, respectively (with the electrostatic energy subtracted from the total energy). $\Delta \mathbf{D}^{IJ}$ is the density matrix difference[14] of dimer *IJ* and the sum of monomer *I* and *J* electron densities. $\mathbf{V}^{IJ}$ is the electrostatic field due to the external fragments acting upon dimer *IJ*. For those dimers, where the interfragment distance is large, the third term in eq 2 vanishes and the second term can be approximated by the ES interaction[43] between fragments *I* and *J*.

The contribution of dimer *IJ* to the total energy is given in the ES approximation by:

$$E'_{IJ} - E'_I - E'_J = Tr(\mathbf{D}^I \mathbf{u}^{I(J)}) + Tr(\mathbf{D}^J \mathbf{u}^{J(I)}) + \sum_{\mu\nu\in I}^{M_I} \sum_{\rho\sigma\in J}^{M_J} D^I_{\mu\nu} D^J_{\rho\sigma} (\mu\nu|\rho\sigma) + \Delta E^{NR}_{IJ} \tag{3}$$

where $\mathbf{u}^{K(L)}$are the one-electron (nucleus-electron) Coulomb potentials exerted by fragment *L* unto fragment *K*; $\mathbf{D}^I$ is the electron density for monomer *I*; M_I and M_J are the number of basis functions in fragments *I* and *J*, respectively. The nuclear repulsion contribution is given by:

$$\Delta E^{NR}_{IJ} = \sum_{\alpha\in I} \sum_{\beta\in J} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}$$

where α and β number atoms with nuclear charges Z_α and coordinates $\mathbf{R}_\alpha$.

The ES approximation is used if the distance R_{IJ} between fragments *I* and *J* is larger than the threshold R^{ES}. It is computationally efficient to combine all ES dimer derivatives into one sum, thus one obtains from summing all similar contributions given in eq 3,

$$\Delta E_{ES} = \sum_{\substack{I>J \\ R_{IJ}>R^{ES}}} (E'_{IJ} - E'_I - E'_J) \tag{4}$$

By taking the analytic derivative of eq 4 with respect to some nuclear coordinate *a* of atom α, one obtains (derivatives are denoted by superscript *a*) the combined contribution to the total energy gradient due to all ES dimers ΔE^a_{ES}:

$$\Delta E^a_{ES} = \Delta E^a_{AO} + \Delta E^a_{MO} + \Delta E^a_{NR}$$

$$\Delta E^a_{INT} = -\sum_{\substack{I \\ \alpha\in I}} \sum_{\mu\nu\in I} D^I_{\mu\nu} \left\{ \left\langle \mu^a \left| \sum_{\substack{\beta\in J \\ R_{IJ}>R^{ES}}} \frac{Z_\beta}{|\mathbf{r} - \mathbf{R}_\beta|} \right| \nu \right\rangle + \left\langle \mu \left| \sum_{\substack{\beta\in J \\ R_{IJ}>R^{ES}}} \frac{Z_\beta}{|\mathbf{r} - \mathbf{R}_\beta|} \right| \nu^a \right\rangle \right\} - \sum_{I} \sum_{\mu\nu\in I} D^I_{\mu\nu} \left\langle \mu \left| \sum_{\substack{\beta\in J \\ R_{IJ}>R^{ES}}} \frac{\partial}{\partial a} \left(\frac{Z_\beta}{|\mathbf{r} - \mathbf{R}_\beta|} \right) \right| \nu \right\rangle + \sum_{\substack{I \\ \alpha\in I}} \sum_{\mu\nu\in I} D^I_{\mu\nu} \left\{ \sum_{\substack{J\neq I \\ R_{IJ}>R^{ES}}} \sum_{\rho\sigma\in J} D^J_{\rho\sigma} [(\mu^a\nu|\rho\sigma) + (\mu\nu^a|\rho\sigma)] \right\}$$

$$\Delta E^a_{MO} = -2 \sum_{\substack{I \\ \alpha\in I}} \sum_{\mu\nu\in I} W_{\mu\nu}{}^I \frac{\partial S_{\mu\nu}{}^I}{\partial a}$$

$$\Delta E^a_{NR} = -\sum_{\substack{I \\ \alpha\in I}} \sum_{\substack{\beta\in J \\ R_{IJ}>R^{ES}}} \frac{Z_\alpha Z_\beta (R_a - R_b)}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|^3} \tag{5}$$

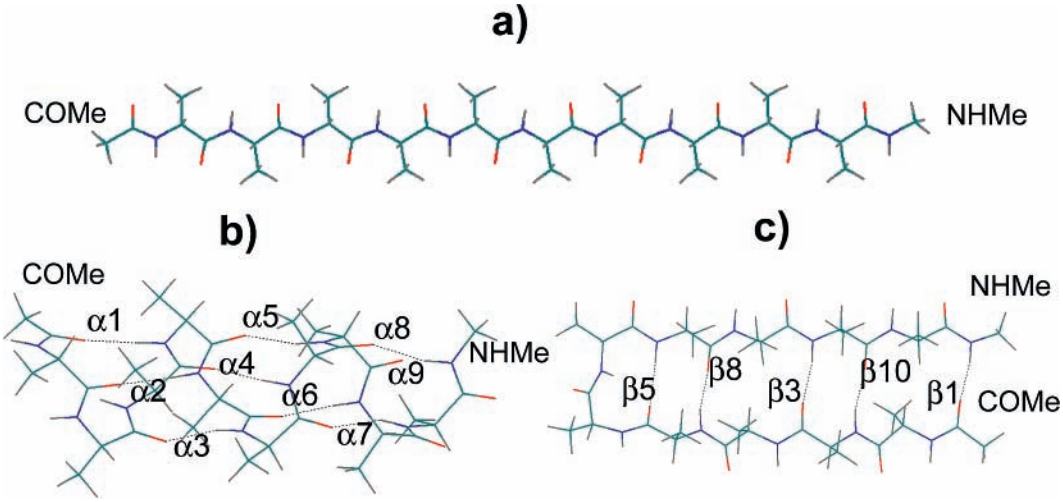

Figure 1. Model geometries of the capped alanine conformers: (a) extended, (b) α-helix, and (c) β-turn. αi and βi designate the proton acceptor atoms of hydrogen bonds in the α-helix and β-turn conformers, respectively.

TABLE 2: RMSD between FMO and ab initio Optimized Geometrical Parameters of MeCO-(Ala)$_{10}$-NHMe

basis set	conformer	all (Å)[a]	bond length (Å)[b]	bond angle (deg)[c]	ϕ (deg)[d]	ψ (deg)[e]	ω (deg)[f]
3-21G	extended	0.0003	0.0004	0.035	0.04	0.03	0.02
	α-helix	0.113	0.0020	0.297	2.35	2.83	1.07
	β-turn	0.173	0.0017	0.431	3.03	2.79	1.25
6-31G*	extended	0.0015	0.0006	0.051	0.10	0.05	0.07
	α-helix	0.198	0.0019	0.272	2.80	4.12	1.40
	β-turn	0.203	0.0037	0.331	2.68	3.12	1.11

[a] All Cartesian coordinates, including hydrogen atoms. [b] All covalent bond lengths are included. [c] All covalent bond angles are included. [d] Dihedral angle of C′(i − 1)−N(i)−C$_\alpha$(i)−C′(i) (i numbers residues). [e] Dihedral angle of N(i)−C$_\alpha$(i)−C′(i)−N(i + 1). [f] Dihedral angle of C$_\alpha$(i)−C′(i)−N(i + 1)−C$_\alpha$(i + 1).

TABLE 3: Comparison of the Optimized Geometric Parameters for the One- (m = 1) and Two-Residue (m = 2) Fragmentation (3-21G) of MeCO-(Ala)$_{10}$-NHMe

conformer	m	all (Å)[a]	bond length (Å)	bond angle (deg)	ϕ (deg)	ψ (deg)	ω (deg)
extended	1	0.123	0.0015	0.35	1.24	0.83	0.63
	2	0.000	0.0004	0.035	0.037	0.034	0.017
α-helix	1	0.325	0.0032	0.51	6.23	7.78	2.21
	2	0.113	0.0020	0.30	2.35	2.83	1.07
β-turn	1	0.311	0.0031	0.43	4.75	3.03	1.59
	2	0.173	0.0017	0.43	3.03	2.79	1.25

[a] See Table 2 for the definition of all quantities.

In the last sum, b stands for the (x,y,z) coordinate of atom β corresponding to (x,y,z) of a in α. The three gradient contributions $\Delta E^a_{\mathrm{INT}}$, ΔE^a_{MO}, and ΔE^a_{NR} are the integral, molecular orbital coefficient, and nuclear repulsion derivative terms, respectively.

$$\mathbf{W}^I = {}^1/_4\mathbf{D}^I\tilde{\mathrm{V}}^I\mathbf{D}^{I\dagger}$$

$$\tilde{\mathrm{V}}^I_{\mu\nu} = \sum_{\substack{J\neq I \\ R_{IJ}>R^{\mathrm{ES}}}} [u^{I(J)}_{\mu\nu} + \sum_{\rho\sigma\in J} D^J_{\rho\sigma}(\mu\nu|\rho\sigma)] \qquad (6)$$

ΔE^a_{ES} in eq 5 gives the gradient correction due to far separated dimers $R_{IJ} > R^{\mathrm{ES}}$. It is added to the dimer gradients for which SCF is performed ($R_{IJ} \leq R^{\mathrm{ES}}$), yielding the energy gradient ∇E for the whole system. $\mathbf{W}^I$ in eq 5 is the electrostatic potential energy weighted density matrix defined in eq 6.

2.2. Computational Details. Geometry optimizations were performed using the default geometry optimizer in GAMESS[44] based on the numeric updates of the Hessian, which are accumulated from ∇E, producing a new geometry at each optimization step. All degrees of freedom were optimized and the calculations were performed on several PC clusters.

The following systems were studied with the two-body FMO2-RHF method and ab initio RHF: the α-helix, β-turn, and the extended conformers of polyalanine, with the 3-21G and 6-31G* basis sets. Both monomer and dimer of met-enkephalin were treated with 3-21(+)G and 6-31(+)G*; chignolin (PDB: 1uao) and an agonist polypeptide of erythropoietin receptor protein (emp1) were computed at the 3-21(+)G level. The sizes of all systems are summarized in Table 1.

Chignolin and Trp-cage miniprotein construct (PDB: 1l2y) were optimized in gas phase as well as in the first solvation shell using 3-21(+)G for 1uao and 3-21G for 1l2y. In the case of the solvated calculations, the proteins were treated with FMO2-RHF and water was described by molecular mechanics (MM). The combination of FMO and MM was implemented in the IMOMM fashion[45] with the Tinker[46] program linked to GAMESS. In MM calculations, the protein and water were described by the Amber94[47] and TIP3P,[48] respectively.

Specifically, the FMO/MM calculations proceeded as follows. The FMO gradient computation of the polypeptide systems (chignolin and 1l2y) were performed in gas phase as usual, followed by the MM calculation of the solvated system whereby the total gradient was complemented by the polypeptide−solvent

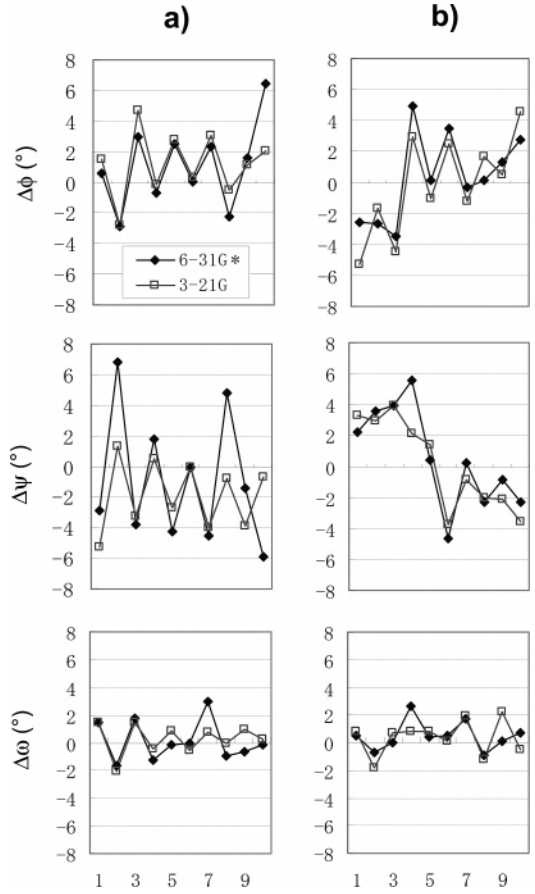

Figure 2. Deviations of dihedral angles ($\Delta\phi$, $\Delta\psi$, and $\Delta\omega$) of FMO optimized geometries with the 6-31G* and 3-21G basis sets relative to the corresponding ab initio values: (a) α-helix and (b) β-turn. The horizontal axis indicates the serial number of alanine residues.

TABLE 4: Comparison of N(H)···O Hydrogen Bond Lengths (Å) between the FMO and ab initio Optimized Geometries in the α-Helix and β-Turn of MeCO-(Ala)$_{10}$-NHMe

	3-21G		6-31G*	
hydrogen bond[a]	FMO	ab initio	FMO	ab initio
α1	3.137	2.987	3.181	3.186
α2	2.987	2.979	3.364	3.242
α3	3.026	3.015	3.218	3.248
α4	3.002	2.994	3.277	3.221
α5	2.982	2.961	3.403[b]	3.524[b]
α6	3.015	3.109	4.370[b]	4.421[b]
α7	3.178	3.200	3.343	3.330
α8	3.036	3.048	3.319	3.239
α9	3.064	3.064	3.190	3.255
β1	2.895	2.910	3.037	3.061
β3	2.974	2.957	3.157	3.141
β5	3.293	3.287	3.757[b]	3.745[b]
β8	2.802	2.816	3.038	3.013
β10	2.939	2.927	3.167	3.099

[a] See Figure 1. [b] The N···O interatomic distance is too long for a hydrogen bond, but it is given for the comparison of the two basis sets.

and solvent−solvent contributions obtained from molecular mechanics. The combined gradient was used in geometry optimization, producing a new set of coordinates for the polypeptide part, used iteratively to repeat the above steps until convergence. No link atoms were used as there was no covalent bond on the boundary of FMO and MM. The only difference between the regular IMOMM and FMO/MM calculations was the computation of the gradient for the quantum region (polypeptide).

Fragmentation of polypeptides was performed at C_α atoms. Typically, we assigned two residues per fragment, except 1l2y, where one residue per fragment division was done (only small Gly residues were appended to their neighbors). In addition, we also probed the magnitude of the fragmentation effects by comparing the one and two residue per fragment division for polyalanines. All FMO input files including the automatic fragmentation were generated with FMOutil.[49]

In all calculations, spherical atomic basis functions were used (5d). The geometry optimization convergence threshold was 1×10^{-4} hartree/bohr, which means that the maximum gradient component with respect to a single coordinate should fall below 1×10^{-4} and the rms gradient should be smaller than $^1/_3 \times 10^{-4}$. In the IMOMM calculations, the optimization threshold was 5×10^{-4} hartree/bohr and 0.01 kcal/mol·Å^{-1}, for molecular orbital and MM calculations, respectively.

We used the following approximations[43] in the FMO polyalanine calculations: RESPAP = 1.0 (the Mulliken atomic populations were used to reduce the cost of the two-electron contributions in the external electrostatic potentials, if the interfragment separation exceeded this value), RESPPC = 2.0 (the Mulliken charges were used instead of the two-electron contributions in the external electrostatic potentials if the interfragment separation exceeded this value), and RESDIM = 2.0 (the ES approximation described above to approximate the total energy of far separated dimers); in all other systems, to improve the accuracy, we did not use the RESPAP approximation.

The initial polyalanine structures (MeCO-(Ala)$_{10}$-NHMe) were prepared with Hyperchem, met-enkephalin dimer was taken from the Cambridge Crystallographic Data (ID: FABJIB), and all other systems were taken from PDB (experimental geometries). For 1uao and 1l2y, the PDB data came from NMR containing many structures, and we used model 1. For emp1, an X-ray structure was available (PDB: 1ebp), which was protonated assuming the standard ionized state using Hyperchem.

To model the first solvation shell, water was added with Hyperchem as a periodic box, and the water molecules within 8 Å from the protein were extracted. Consequently, the protein structure was frozen and the position of the water molecules optimized. From this, we took all water molecules within 5 Å from the protein surface. Thus constructed first solvation shells consisted of 177 and 293 water molecules for 1uao and 1l2y, respectively.

The parallelization of the ES dimer derivatives was accomplished with the generalized distributed data interface (GDDI)[50] based on the same work load distribution as used for the ES dimer energy. The additional cost of computing the ES dimer gradients is negligible compared to other much more time-consuming parts of an FMO calculation (that is, SCF of monomers and dimers, including the computation of their external electrostatic potential). As an example, the energy gradient for the 10-residue polyalanine with the 6-31G* basis set took 6.8% more than the corresponding energy calculation.

3. Results and Discussion

3.1. Polyalanine. The structures of polyalanine conformers are visualized in Figure 1. In Table 2, we summarized the reduced mean-square deviations (rmsd) between the FMO and ab initio Cartesian coordinates, bond lengths, bond angles, and polypeptide dihedral angles ϕ, ψ, and ω for the three conformers

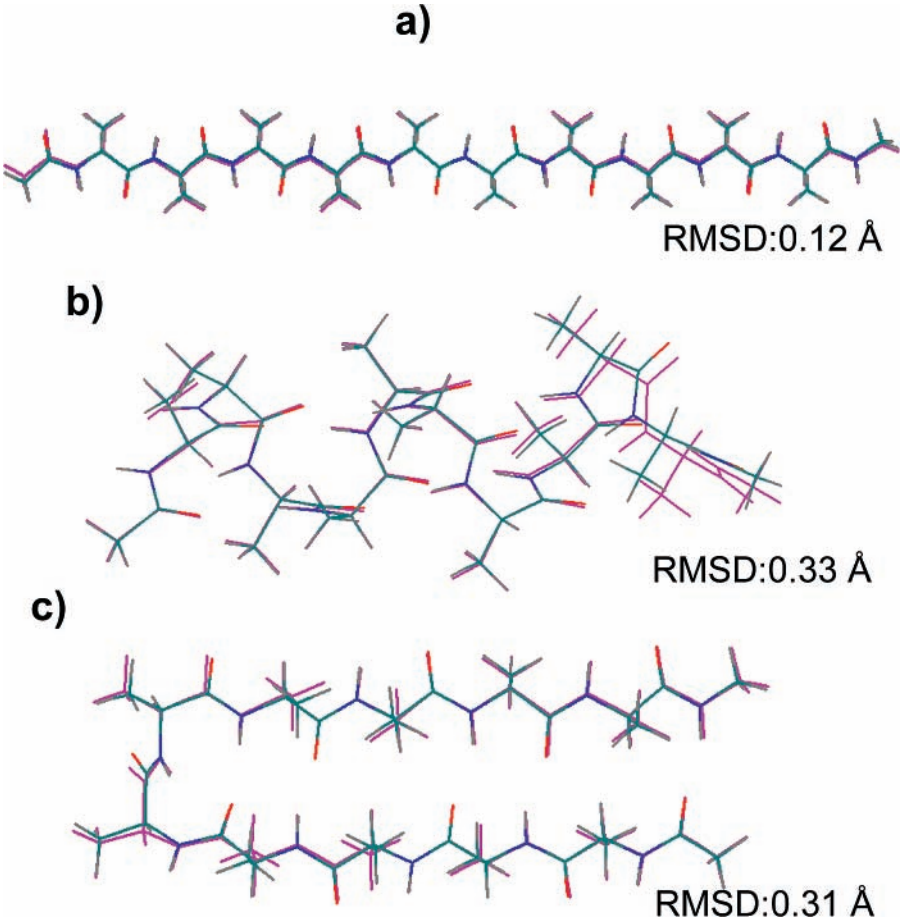

Figure 3. Superposition of one residue/fragment partition FMO (colored by chemical elements as gray (H), green (C), blue (N), and red (O)) and ab initio (violet) optimized geometries (both at the RHF/3-21G level): (a) extended, (b) α-helix, and (c) β-turn. The rmsd is calculated from the Cartesian coordinates of all atoms (including hydrogens). For the α-helix, the superposition of structures is done with the best fit excluding NHMe and the three neighbor C-terminus residues to show that the overall deviation is largely local to these fragments.

TABLE 5: FMO and the ab initio Total Energies (au) of MeCO-(Ala)$_{10}$-NHMe at the Corresponding Optimized Geometries[a]

basis set	conformer	FMO2	FMO3	ab initio
3-21G	extended	−2690.509496	−2690.509539	−2690.509541
	α-helix	−2690.571221 (−38.7)	−2690.564465 (−34.5)	−2690.566724 (−35.9)
	β-turn	−2690.550923 (−26.0)	−2690.555551 (−28.9)	−2690.557882 (−30.3)
6-31G*	extended	−2705.537661	−2705.537736	−2705.537745
	α-helix	−2705.561143 (−14.7)	−2705.558627 (−13.1)	−2705.560242 (−14.1)
	β-turn	−2705.556104 (−11.6)	−2705.557719 (−12.5)	−2705.559355 (−13.6)

[a] The energy relative to the extended conformer is given in parentheses in kcal/mol.

of polyalanine: the α-helix, β-turn, and extended form. In computing the rmsd values, we aligned the two structures under comparison minimizing the mean-square deviation. The details of the deviations between FMO and ab initio structural parameters are also plotted in Figure 2. It can be seen (Table 2) that, for the latter conformer, the errors are nearly zero. Thus, the fragmentation scheme in FMO by itself (the fact that covalent bonds are divided) does not introduce much error and the FMO structures nearly perfectly coincide with ab initio ones. The other two conformers have the errors about 1 order of magnitude larger. The origin of the FMO error in general was elucidated earlier,[51] and it is mainly the three-body coupling between pairs of hydrogen bonds that is responsible for it.

One can observe that the overall agreement between FMO and ab initio is good: in terms of the rmsd values, at worst about 0.1−0.2 Å total, for bond lengths 0.001−0.004 Å, and for bond angles 0.1−0.4°. Rotations along the polypeptide angles (ϕ, ψ) correspond to very small energy differences with about 1 kcal/mol rotational barriers, and the observed error of 1°−4° is not relevant in practice, as the thermal energy at room temperature brings enough energy to allow nearly free rotation about these angles.[52] Also, the error for these two angles is concerted in the sense that the whole structure is slightly distorted relative to ab initio in a coherent fashion, which can be seen from the small rmsd values. The other dihedral angle ω, which is more rigid and corresponds to the preferential orientation of polypeptides as the *trans* conformations, had the error of about 1°−1.5°.

The relative stability of the polypeptide conformers and proteins is largely driven by the fairly weak noncovalent interactions, most importantly, by the hydrogen bond network. The hydrogen bond distances for the α-helix and β-turn are listed in Table 4. In FMO, pair interactions (corresponding to hydrogen bonds) are described by dimer calculations at the ab initio level, and the error mostly comes from the three-body and higher-order coupling of interactions involving charge transfer. In a network of hydrogen bonds, the interactions are

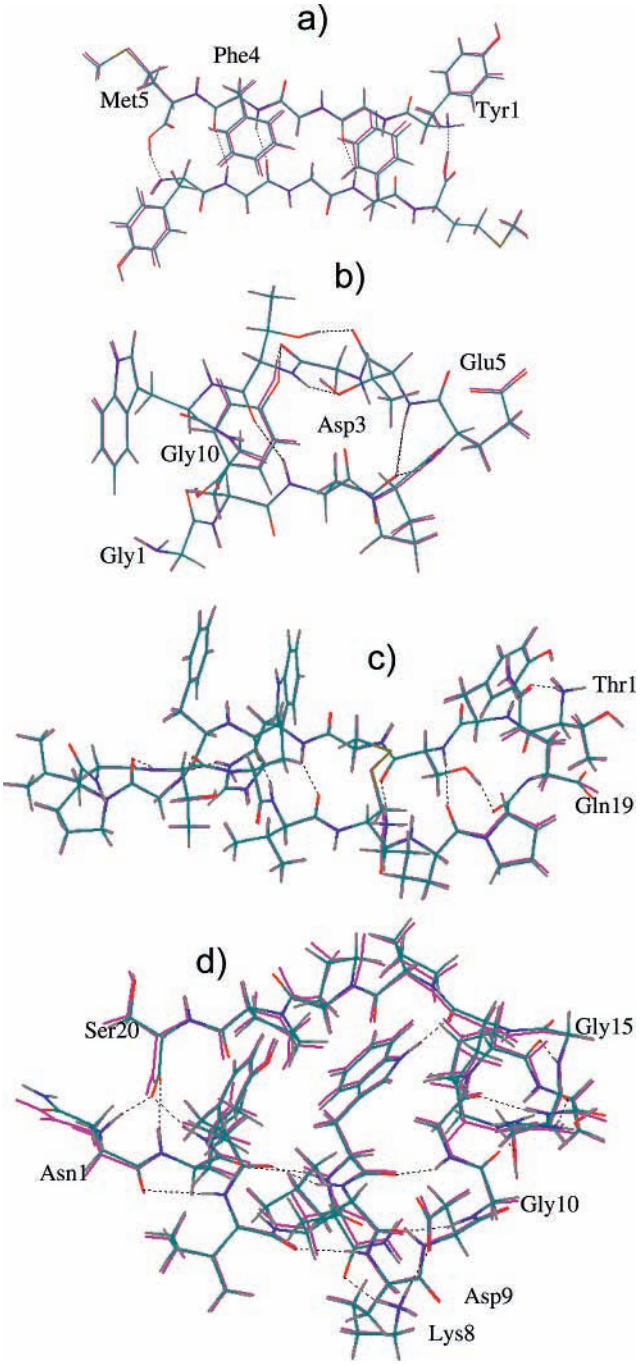

Figure 4. Superposition of the FMO (colored by chemical elements as gray (H), green (C), blue (N), and red (O)) and ab initio (all violet) optimized geometries of (a) Met-enkephalin dimer, (b) emp1, (c) 1uao, and (d) 1l2y. The 3-21(+)G basis set is used for (a−c), and 3-21G is used for (d). Dotted lines show hydrogen bonds.

complicated and no pronounced trend in the hydrogen bond length could be observed (Table 4). In particular, both over- and underestimation of the bond lengths are seen. One can also comment that the energy surface with respect to a small noncovalent bond length variation (about 0.1 Å) is flat and the total energy is not sensitive to such changes. The overall discrepancy between FMO and ab initio is small, and the individual hydrogen bond differences are not random but reflect some small coherent change of the whole network.

In Table 5, we listed the relative stabilities of conformers, as obtained from FMO and ab initio calculations using their optimized structures. The errors in the relative stabilities do not exceed 3.7 and 2.0 kcal/mol for the 3-21G and 6-31G* basis sets, respectively. The larger error of the former basis can be expected, as it was also observed in the previous accuracy study.[26] By using the FMO2 optimized structures and performing single-point three-body FMO3 calculations,[26] the errors are decreased to 1.5 and 1.0 kcal/mol for the 3-21G and 6-31G* basis sets, respectively. Thus, a reasonable compromise between the accuracy and computational efficiency is in using FMO2 for geometry optimizations and refining the final energetics by FMO3.

3.2. Fragmentation Effects. It is known that doubling the fragment size increases the FMO accuracy (the error in the total energy relative to ab initio is typically halved).[14] To investigate the effect of the fragment size upon the geometry optimizations, we compared the 3-21G structures of polyalanines, optimized for the one and two residue per fragment division, denoted by FMO2/1 and FMO2/2, respectively. The superimposed structures for FMO2/1 and ab initio are shown in Figure 3. In the discussion below, the two residue per fragment data (FMO2/2) from Table 2 were complemented by the corresponding set of FMO2/1 data and tabulated in Table 3.

For the extended conformer, the rms deviations for the FMO2/1 and FMO2/2 structures relative to ab initio are 0.123 and 0.000 Å, respectively. It is known from the earlier accuracy tests that extended conformers are described by FMO2/2 nearly exactly,[26] whereas the error for FMO2/1 is fairly large. The rmsd values for the covalent bonds and angles are 0.0015 Å and 0.35° for FMO2/1, whereas for FMO2/2 they are 0.0004 Å and 0.035°, respectively. The dihedral angle ϕ, ψ, and ω deviations are 1.24, 0.83, and 0.63 (deg) for FMO2/1, which can be compared to the corresponding FMO2/2 values of 0.037, 0.034, and 0.017 (deg). While a drastic improvement can be achieved by doubling the fragment size, the structure accuracy of FMO2/1 is good, which can be seen from the fact that the ab initio RHF energy computed at the FMO2/1 and RHF optimized geometries differs only by 0.8 kcal/mol.

In the α-helix, the overall rmsd values are 0.325 and 0.113 (Å) for FMO2/1 and FMO2/2, respectively. For FMO2/1, the rmsd of the covalent bonds and angles are 0.0032 Å and 0.51°, which is about 50% more than in FMO2/2 (0.0020 Å and 0.30°). The 3-fold difference in the rmsd values comes from the dihedral angles ϕ, ψ, and ω, whose deviations are 6.23, 7.78, and 2.21 (deg) for FMO2/1, and 2.35, 2.83, and 1.07 (deg) for FMO2/2. The main factor that is responsible for the larger error of FMO2/1 is the C-terminus and three adjacent residues. The FMO2/1 deviations of the residue-specific angles (ϕ, ψ, ω) are: Ala8 (3.73, −10.57, 4.47), Ala9 (4.42, −11.08, 3.89), and Ala10 (14.54, −13.52, 1.35), all in degrees. If one excludes these fragments and calculates the rmsd for all remaining atoms, the value of 0.102 Å is obtained, which is essentially the same as in FMO2/2. Therefore, the FMO2/1 and FMO2/2 structures of the α-helix are essentially the same, with the exception of the C-terminus, the reason for the latter deviation being the flat energy surface, as the terminus is not as tightly bound by hydrogen bonds as the rest of the helix. We conclude that the accuracy of the FMO2/1 optimization is reasonably good. In proteins, the termini of most α-helices are not free, so that the problem is expected to be insignificant. The difference of the ab initio RHF energy computed at the FMO2/1 and RHF optimized structures was 3.4 kcal/mol.

For the β-turn, the overall rmsd values are 0.311 and 0.173 (Å) for FMO2/1 and FMO2/2, respectively. The rmsd of the covalent bonds and angles are 0.0031 Å and 0.43° in FMO2/1, which can be compared to FMO2/2 (0.0017 Å and 0.43°). The rmsd of the dihedral angles (ϕ, ψ, ω) are (4.75, 3.03, 1.59) and

TABLE 6: RMSD between FMO and ab initio Optimized Geometrical Parameters of Polypeptides

molecule	basis set	heavy (Å)[a]	backbone (Å)[b]	bond length (Å)[c]	bond angle (deg)[d]	ϕ (deg)	ψ (deg)	ω (deg)
enk-mon[e]	3-21(+)G	0.001	0.001	0.0012	0.05	0.03	0.06	0.03
enk-mon	6-31(+)G*	0.161	0.104	0.0081	0.19	3.36	1.91	1.46
enk-dim[f]	3-21(+)G	0.167	0.080	0.0014	0.26	2.24	1.20	1.28
enk-dim	6-31(+)G*	0.195	0.075	0.0105	0.07	3.84	4.16	2.88
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] Met-enkephalin monomer. [f] Met-enkephalin dimer.

TABLE 7: FMO and ab initio Total (E, au) and Binding Energies (ΔE, kcal/mol) of Met-enkephalin Dimer

	basis set	FMO2[a]	FMO3[a]	ab initio
E (monomer)	3-21(+)G	−2224.573695 (0.00)	−2224.573694 (0.00)	−2224.573691
E (dimer)	3-21(+)G	−4449.260742 (-2.89)	−4449.253738 (1.51)	−4449.256144
ΔE[b]	3-21(+)G	−71.13 (−2.89)	−66.74 (1.51)	−68.25
E (monomer)	6-31(+)G*	−2236.720017 (0.16)	−2236.720014 (0.16)	−2236.720267
E (dimer)	6-31(+)G*	−4473.485026 (-0.68)	−4473.482129 (1.14)	−4473.483950
ΔE[b]	6-31(+)G*	−28.23 (−0.99)	−26.42(0.82)	−27.24

[a] The deviation of FMO from the corresponding ab initio values is given in parentheses in kcal/mol. FMOn denotes the n-body FMO expansion. [b] $\Delta E = E$ (dimer) $- 2E$ (monomer)

TABLE 8: Comparison between the FMO and ab initio Total Energies (au) at Their Respective Geometries

molecule	FMO2[a]	ab initio
1uao	−3779.029339 (−0.49)	−3779.028553
emp1	−6875.929363 (−0.01)	−6875.929349
1l2y	−7398.823548 (1.48)	−7398.825906

[a] The difference (FMO − ab initio) is given in parentheses in kcal/mol. All FMO energies are for a two residue per fragment division; the energy of 1l2y with one residue per fragment is −7398.835068.

(3.03, 2.79, 1.25) for FMO2/1 and FMO2/2, respectively (all in degrees). They are very similar, except for ϕ, which has a somewhat larger deviation for FMO2/1. The source of this is the pair of residues at the turning point (Ala4 and Ala6), for which the deviation of FMO2/1 is −8.5 and 10.2 (deg), respectively. With these two residues excluded, the rmsd of ϕ in FMO2/1 becomes 2.5°, which is similar to FMO2/2. Thus, we observe that the β-turn structures are nearly the same for FMO2/1 and FMO2/2, with the exception of the somewhat larger error in ϕ for the former, local to the turning point. The difference of the ab initio RHF energy computed at the FMO2/1 and RHF optimized structures was 2.2 kcal/mol. Summarizing, we conclude that the one residue per fragment division is reasonable for geometry optimizations and the accuracy can be systematically improved by doubling the fragment size.

3.3. Polypeptide Interaction in Met-enkephalin Dimer. Met-enkephalin dimer consists of two monomers bound head-to-tail to each other. Geometry optimizations of this originally zwitterion dimer in gas phase for both ab initio and FMO resulted in a proton transfer from the N-terminus (Thy1) to the C-terminus (Met5). Therefore, we optimized monomer and dimer as neutral species at both termini.

The summary of the FMO and ab initio structure deviations is provided in Table 6, and the superposition of the two structures is given in Figure 4a. The initial structure of met-enkephalin monomer was taken from the dimer geometry, and its optimization yielded the extended form. As can be seen, all rmsd values (total, bond length, bond angle, and polypeptide angles) for the extended form of met-enkephalin monomer are very similar to those of e-ala10. The corresponding deviations in the case of 6-31G* are larger, which is due to the flatter potential energy surface corresponding to the polypeptide dihedral angle rotation. Note that the gradient accuracy[26] of FMO vs ab initio is quite similar for the two basis sets, with 6-31G* having about 20−30% larger error, so the difference observed here does not come from the gradient deviation alone. The very small energy difference of 0.16 kcal/mol between the FMO and ab initio total energies also supports this assertion.

The two antiparallel monomer chains form dimer with the resultant structure of a β-sheet (Figure 4a). The deviations of the dimer structural parameters (Table 6) are very similar to those observed in β-ala10 (slightly larger). In contrast to monomer, dimer has a fairly rigid structure not featuring a flat energy surface along many angle rotations, thus the basis set dependence of the error is small. The polypeptide angle deviation of several degrees is similar to polyalanine and does not present a problem, as the overall structure shows good accuracy: the total rmsd of 0.167 and 0.195 Å, for 3-21(+)G and 6-31(+)G*, respectively. Overall, we conclude that the FMO optimization results are in a good agreement with ab initio.

An accurate evaluation of the binding energies requires both accurate structures and reliable energy calculations. Next, we demonstrate that FMO possesses sufficient accuracy in this regard. The total and binding energies obtained with ab initio and FMO-based RHF are summarized in Table 7, where the energies are obtained at the geometries optimized with each method (the three-body FMO predictions (FMO3) are obtained with the FMO2 structures). It can be seen that the small error in the binding energies comes from the small inaccuracy in the total energy of dimer; the deviation of the FMO2 binding energies from ab initio is −2.9 and −1.0 (kcal/mol) for 3-21(+)G and 6-31(+)G*, respectively. It may be noted that the general trend of the 3-21G basis set to have a somewhat larger error compared to 6-31G* was observed earlier as well;[26] nevertheless, the basis set dependence of the binding energy error is small.

The higher-level FMO calculations (FMO3) for the two basis sets reduced the errors in the binding energy from 2.9 to 1.5 kcal/mol (in the worst case). Because this change is small, we conclude that FMO2 can reliably predict the binding energies (that is, it can reproduce the corresponding ab initio values). The basis set dependence of the binding energies is very large in this system; the ab initio values are −68.3 and −27.2 (kcal/mol) for 3-21(+)G and 6-31(+)G*, respectively. The former basis set largely overestimates the binding energies and the use

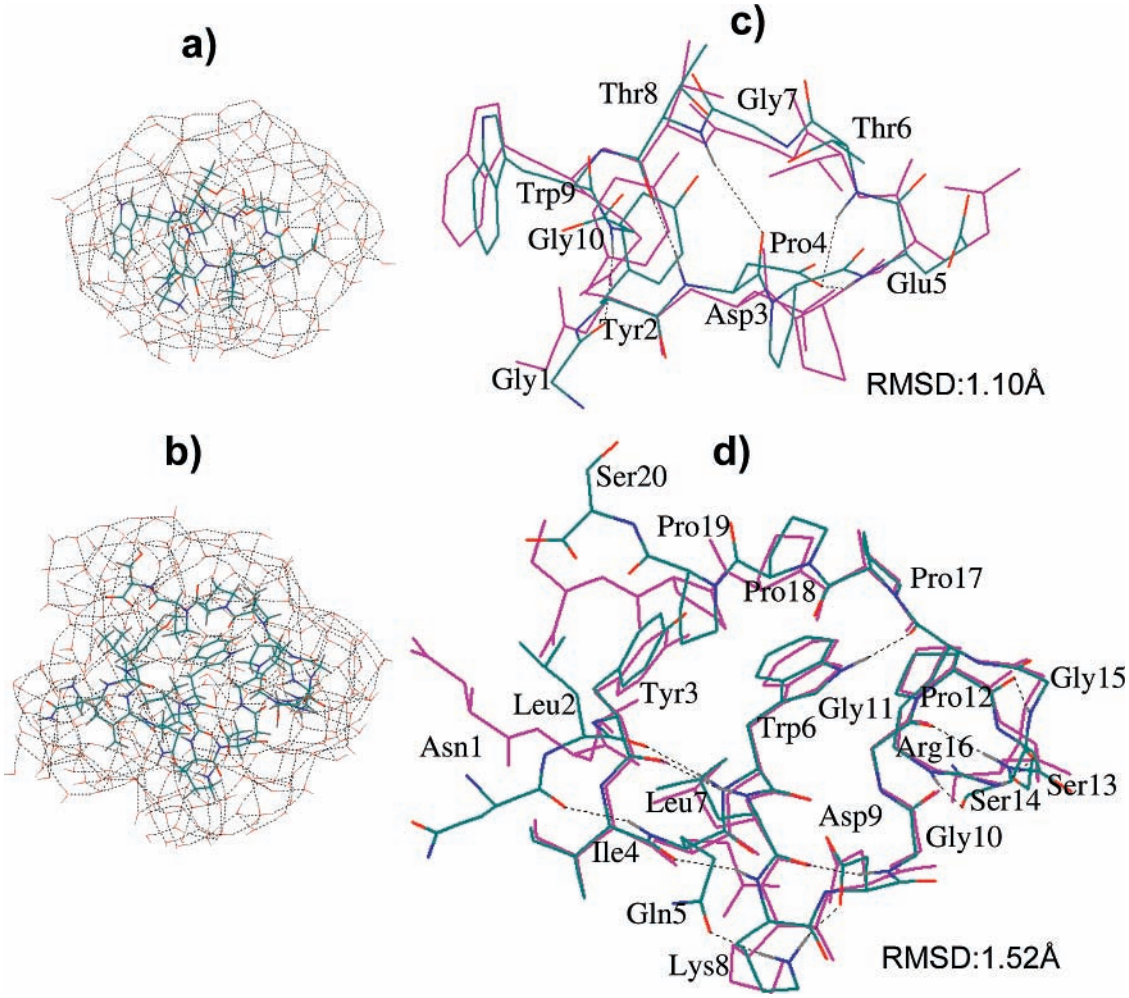

Figure 5. Hydrated polypeptide models of (a) 1uao and (b) 1l2y. The superposition of heavy atoms in the FMO optimized geometries in gas phase (violet) and in the hydrated cluster (colored by chemical elements as green (C), blue (N), and red (O)) are shown for (c) 1uao and (d) 1l2y. The rmsd values between the gas phase and solvated structures are obtained for the Cartesian coordinates of all heavy atoms.

of an at least double-ζ with polarization basis set (6-31G*) seems necessary in practical applications.

3.4. Other Polypeptide Structures. Here we used the 3-21(+)G basis set with the diffuse functions added to the carboxyl groups (COO^-), except for 1l2y, in which case, SCF convergence problems prevented such addition and we thus used 3-21G. It should be noted that frequently the lack of diffuse functions on carboxyl groups leads to the unphysical elongation of the adjacent C−C bonds (up to 1.6 Å, in our experience). However, if carboxyl groups are bound to other groups with salt bridges or hydrogen bonds, such a problem does not occur. In 1l2y, there are two carboxyl groups in Ser20 (C-terminus) and Asp9. During geometry optimizations, salt bridges were formed between the former group and the amino group in the N-terminus as well as between the latter group and Lys8. As a result, the C−C bond lengths in the two carboxyl groups were 1.524 and 1.512 Å (ab initio) or 1.543 and 1.517 Å (FMO). Thus we observe that the absence of diffuse functions in 1l2y did not lead to problems in this regard.

In 1uao (Figure 4), for both ab initio and FMO, a salt bridge was formed between the protonated amino group (NH_3^+) of the N-terminus (the cap of Gly1) and the carboxyl group of the C-ternimus (the cap of Gly10). Consequently, the gas-phase optimization resulted in a proton transfer between the two residues, which happens frequently in such cases. A similar problem was described above for met-enkephalin dimer (proton transfer between two monomers). In our experience, proton transfer occurs when a pair of COO^- and NH_3^+ groups is able to arrange itself so that NH···O forms a nearly straight line. If, due to some steric hindrance this is not possible, then proton transfer does not occur. In this system (1uao), there was no such hindrance and a proton transfer took place. The other two carboxyl groups in Asp3 and Glu5 had no partner to form salt bridges, and no proton transfer occurred. It should be pointed out that the main goal of comparing the FMO and ab initio structures can be successfully achieved here, as this proton transfer took place in both cases. As related below, no proton transfer occurred in the solvated model.

As can be seen in Figure 4b, emp1 has a structure of a β-turn, with Thr1 (N-terminus) and Gln16 (C-terminus) being close to each other. However, the amino group of the former does not form a salt bridge with the carboxyl group of the latter, and instead, the two residues are bound by a hydrogen bond between the amino group of Thr1 and Oϵ1 of Gln16. l12y includes two salt bridges between Asn1 (N-terminus) and Ser20 (C-terminus), as well as between Lys8 and Asp9. However, in cases of both emp1 and 1l2y, due to the structural hindrance, NH···O cannot be arranged linearly, and no proton-transfer takes place.

The comparison of the FMO and ab initio structural parameters is given in Table 6 as well as in Figure 4a−d. The rmsd of the covalent bond lengths in FMO and ab initio is 0.0022, 0.0019, and 0.0048 (Å), for 1uao, emp1, and 1l2y, respectively. The corresponding rmsd of the bond angles is 0.24, 0.35, and 0.60 (deg). The difference between the FMO and ab initio values

is thus small: several milliångstrom or fractions of a degree. The dihedral angle rmsd (ϕ, ψ) is (0.63°,0.84°), (1.73°, 2.32°), and (6.63°,4.43°) for 1uao, emp1, and 1l2y, respectively. As pointed out above, this error of several degrees comes from the several kcal/mol error in the FMO total energy, coupled with the flat energy surface (with respect to the rotations about these angles). The larger error in these angles does not appear to lead to practical problems, and the overall all heavy atom rmsd values show good accuracy: 0.097, 0.095, and 0.198 (Å), for 1uao, emp1, and 1l2y, respectively.

Both 1uao and emp1 are distorted β-turns so that the error behavior resembles that for β-ala10. As shown in Table 2, the β-turn rmsd for all atoms is 0.173 Å, and for heavy atoms it is 0.125 Å. The structure of emp1 is only slightly distorted from the ideal β-turn geometry, its rmsd value being very similar to that of β-ala10. In contrast to emp1, the rmsd values for 1uao (except for ω) are smaller than those for β-ala10. The reason behind the larger distortion of 1uao from the ideal β-turn geometry is thought to be the lack of the hydrogen bonds holding the main chain together (Figure 4c). In the 20-residue protein 1l2y (Figure 4d), there are two α-helices (from Asn1 to Asp9 and from Gly10 to Gly15), with a large relative helical fraction. Thus, the rmsd values for 1l2y resemble those for α-ala10 (slightly larger).

It should be noted that 1l2y has twice as many residues and nearly thrice as many atoms as ala10, and the overall rms deviations from ab initio for 1l2y and α,β-ala10 are essentially the same, despite the fact that 1l2y was computed with one residue per fragment partition, whereas the polyalanines were computed with two residues per fragment division. According to our experience, this is a general trend of large protein structure optimizations to be better described by FMO compared to small test polypeptides. The reason for that is the native structure of globular proteins, which is typically quite rigid and held together by fairly strong hydrogen bonds, whereas the linear chains of polypeptides considered here are very flexible.

The accuracy of the FMO total energies (relative to ab initio) is summarized in Table 8. For 1uao, emp1, and 1l2y, the errors are −0.49, −0.01, and 1.48 (kcal/mol) for the two-body FMO method with two residues per fragment partition. The quality of the absolute total energies is high, and one can reasonably expect satisfactory performance with even better accuracy for the energy differences.

3.5. Solvation Effects upon Structure. Generally, in gas phase, the charged residues on the surface tend to form salt bridges, and the structure optimization favors the formation of intramolecular hydrogen bonds. In solution, both types of bonding are to a large extent substituted by solute−solvent interactions. The solvated structures of 1uao and 1l2y, including the explicit solvent, are depicted in parts a and b of Figure 5, respectively. This difference results in a distortion between gas phase and solvated structures, including both main and side chains. In addition, the gas-phase optimization often promotes the proton transfer between basic and acidic residues.

The superimposed gas phase and solvated structures of 1uao are shown in Figure 5c, and the dihedral angles (ϕ, ψ) are given in Figure 6a. As discussed above, despite the fact that the N- and C-termini were largely separated in the initial structure, during the gas-phase geometry optimization, a salt bridge between them was formed, followed by a proton transfer. In contrast, in solution, water molecules enter between the two termini and no salt bridge is formed. As a result, the gas phase and solvated structures have a considerably different separation between the two termini, and side chains adjust accordingly.

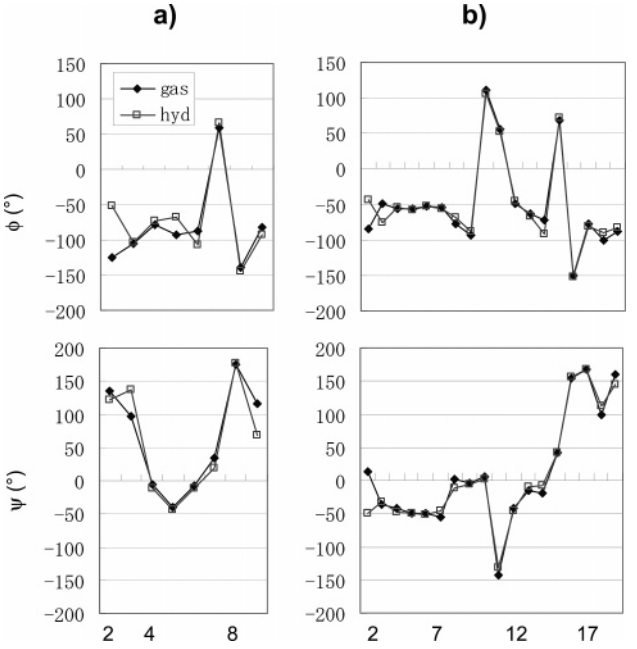

Figure 6. Dihedral angles (ϕ and ψ) of (a) 1uao and (b) 1l2y in gas phase (denoted as gas) and solvated (hyd) geometries. The horizontal axis indicates the serial number of residues.

The observed large structural difference around Pro4 comes from the fact that, in gas phase, the hydroxyl group in Tyr2 forms a hydrogen bond with C′O of Gly7, whereas in solution, the hydroxyl group of Tyr2 is stabilized by the interaction with solvent and forms no such hydrogen bond. Instead, C′O of Pro4 has a stronger hydrogen bond with NH of Thr8, resulting in a shorter separation between the two groups. This distortion extends to the main chain area around Glu5 (see the distortion of ϕ in Figure 6a).

The superimposed gas phase and solvated structures of 1l2y are shown in Figure 5d, and the dihedral angles (ϕ, ψ) are given in Figure 6b. As mentioned above, due to the structural hindrance preventing the linear arrangement of NH···O, a salt bridge between the N- and C-termini is formed in gas phase but does not lead to a proton transfer. In this solvated protein, water molecules clustered around both termini prevent the salt bridge formation. The structure distortion due to this difference is blocked by the rigid α-helix near Leu2 at the N-terminus; at the C-terminus, however, it spreads to Pro18, eliminating the hydrogen bond between the indole NH of Trp6 and the main chain C′O of Arg16. Both in the gas phase and solvated structures, a salt bridge is formed between Asp9 and Lys8, and the difference between the two in this aspect is small. The two rigid α-helices form a large part of 1l2y and, with the exception of the above-mentioned distortion due to the loss of the salt bridge, the difference between the two structures is not large.

Summarizing, the difference between the gas phase and solvated structures is in 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 extension is considerable for random coil but is largely blocked by the rigid conformations of α-helices and strong hydrogen bonds. One can expect a fairly small gas phase and solvated structure difference for those proteins that have a rigid structure, and this difference is limited mostly to the hydrogen bond formation with the polar surface residues. It should be noted that using a single configuration of explicit water molecules is a reasonable approach to obtain the structure but is not suited for calculations of solvation energies, in which

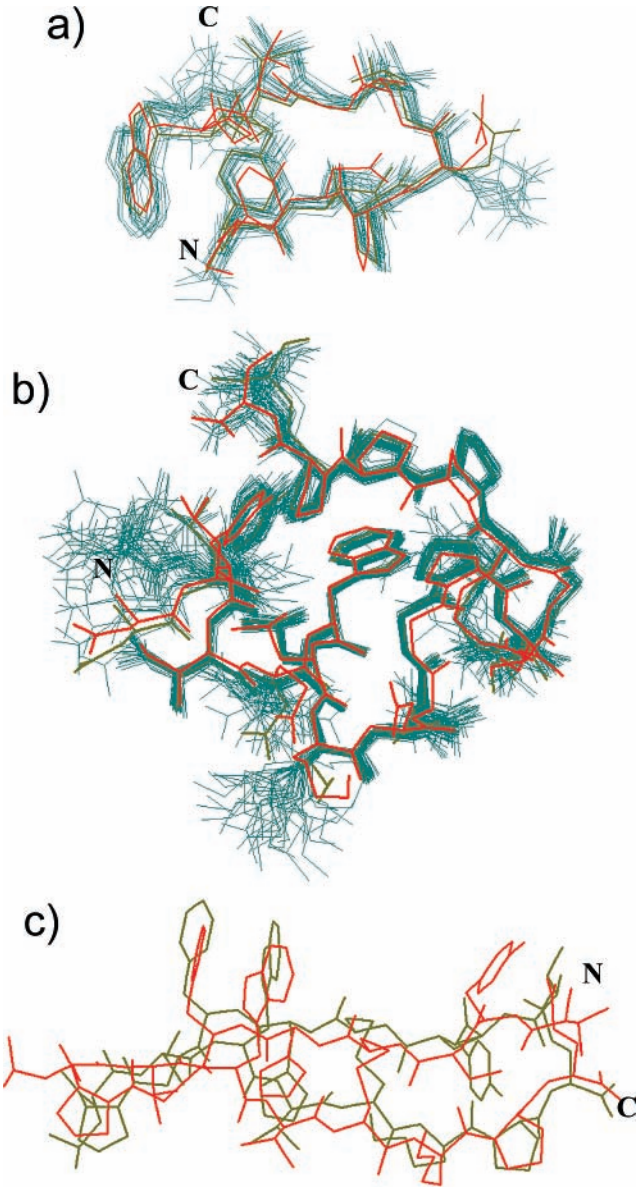

Figure 7. Superposition of optimized and experimental geometries: (a) 1uao, (b) 1l2y, and (c) emp1. For (a) and (b), all NMR geometries (blue and ochre) given in the PDB data and the solvated FMO optimized geometries (red) are overlaid. Ochre color shows the first geometry (model 1), which is used as the initial geometry for optimization calculations. For (c), the X-ray experimental geometry of the polypeptide in the protein complex (ochre) and the FMO optimized geometry in gas phase (red) are superimposed. "N" and "C" indicate the N- and C-terminus, respectively. Only heavy atoms are shown.

case, either continuum models[33] or configuration sampling should be done.

In Figure 7a,b, the computed structures of 1uao and 1l2y are superimposed with the experimental ones, which come from NMR experiment and are typically constructed with force field optimizations fitting the actually measured data. Thus, only qualitative comparison is meaningful and one can observe from the presented data that a good agreement is achieved. FMO can be used instead of force fields in such experimental structure construction.

The computed and experimental structures of emp1 are shown in Figure 7c. The latter comes from a complex (PDB: 1ebp), whereas the former is from the free gas-phase optimization. Accordingly, the purpose of the comparison here is not to discuss the computed and experimental structure difference, but rather to point out the distortion due to the complex formation. The latter effect is mostly expressed in the bent ladderlike structure of the β-turn.

4. Conclusion

To permit real protein optimizations, an important component of the FMO energy gradient was developed (the energy derivative of the electrostatic fragment pairs, forming the vast fraction of all pair calculations). With the developed code, we performed a number of geometry optimizations of the polypeptide and small protein structures. The accuracy in reproducing ab initio structures was evaluated for several conformers of polyalanine and other polypeptides. In addition, the basis set dependence of the structure accuracy was investigated. The effect of solvation upon minimum structures was elucidated in detail.

In all cases, we conclude that the achieved accuracy is sufficiently high to perform protein structure optimizations. To obtain reliable predictions, the use of one residue per fragment partition for FMO2 geometry optimizations can be complimented by single-point FMO3 calculations refining the energetics.

In general, some solvation model is needed to provide reliable structure predictions, although in case of compact rigid proteins, the gas-phase structures may be expected to be fairly reasonable. The present scheme of explicit water described by TIP3P may be difficult to apply to large proteins due to the large number of degrees of freedom and, consequently, of the very significant number of steps needed to find a minimum structure. In this respect, continuum models such as PCM may be practically favorable for geometry optimizations of proteins.

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

Supporting Information Available: Geometric parameters for polyalanines are tabulated. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Sato, F.; Yoshihiro, T.; Era, M.; Kashiwagi, H. *Chem. Phys. Lett.* **2001**, *341*, 645.
(2) Inaba, T.; Tahara, S.; Nisikawa, N.; Kashiwagi, H.; Sato, F. *J. Comput. Chem.* **2005**, *26*, 987.
(3) Alsenoy, C. V.; Yu, C. H.; Peeters, A.; Martin, J. M. L.; Schäfer, L. *J. Phys. Chem. A* **1998**, *102*, 2246.
(4) Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. *Int. J. Quantum Chem.* **2004**, *97*, 747.
(5) Wada, M.; Sakurai, M. A. *J. Comput. Chem.* **2005**, *26*, 160.
(6) Yu, N.; Yennawar, H. P.; Merz, K. M., Jr. *Acta Crystallogr., Sect. D* **2005**, *61*, 322.
(7) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* 2003, *24*, 760.
(8) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. *J. Chem. Phys.* **2006**, *124*, 024301.
(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) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A.* **2001**, *105*, 293.
(12) Stoll, H. *Phys. Rev. B* **1992**, *46*, 6700.
(13) Paulus, B. *Phys. Rep.* **2006**, *428*, 1.
(14) For a short review, see: 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.
(15) Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.

(16) Deshmukh, M. M.; Gadre, S. R.; Bartolotti, L. J. *J. Phys. Chem. A* **2006**, *110*, 12519.
(17) Zhang, D. W., Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
(18) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7251.
(19) Imamura, A.; Aoki, Y.; Maekawa, K. *J. Chem. Phys.* **1991**, *95*, 5419.
(20) Makowski, M.; Korchowiec, J.; Gu, F. L.; Aoki, Y. *J. Comput. Chem.* **2006**, *27*, 1603.
(21) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255.
(22) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
(23) Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Comput. Chem.* **2004**, *25*, 1431.
(24) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.
(25) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
(26) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
(27) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129.
(28) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.
(29) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *122*, 054108.
(30) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
(31) Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638.
(32) Kitaura, K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.* 2001, *336*, 163.
(33) Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. *J. Comput. Chem.* **2006**, *27*, 976.
(34) 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.
(35) Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.* **2007**, *28*, 222.
(36) Nemoto, T.; Fedorov, D. G.; Uebayasi, M.; Kanazawa, K.; Komeiji, Y.; Kitaura, K. *Comp. Biol. Chem.* **2005**, *29*, 434.
(37) Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. B* **2006**, *110*, 1457.
(38) Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. *J. Phys. Chem. B* **2006**, *110*, 16102.
(39) Sawada, T.; Hashimoto, T.; Nakano, H.; Suzuki, T.; Ishida, H.; Kiso, M. *Biochem. Biophys. Res. Commun.* **2006**, *351*, 40.
(40) Nakanishi, I.; Fedorov, D. G.; Kitaura, K. *Proteins: Struct., Funct., Bioinf.* **2007**, in press.
(41) Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.* **2007**, DOI: 10.1002/jcc.20645.
(42) Fedorov, D. G.; Kitaura, K., in preparation.
(43) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.
(44) 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, http://www.msg.ameslab.gov/.
(45) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
(46) Ponder, J. W. *TINKER, Software Tools for Molecular Design*, version 3.9; http://dasher.wustl.edu/tinker.
(47) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(49) *FMOutil*, version 2.0; http://staff.aist.go.jp/d.g.fedorov/fmo/fmoutil.html.
(50) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.* **2004**, *25*, 872.
(51) Yasuda, K.; Yamaki, D. *J. Chem. Phys.* **2006**, *125*, 154101.
(52) Finkelstein, A. V.; Ptitsyn, O. B. *Protein Physics: A Course of Lectures*, Academic Press: Amsterdam, 2002.