Published in final edited form as:

J Phys Chem Lett. 2012 September 20; 3(18): 2604–2610. doi:10.1021/jz3010688.

An Efficient Method to Evaluate Intermolecular Interaction Energies in Large Systems Using Overlapping Multicenter ONIOM and the Fragment Molecular Orbital Method

Naoya Asada¹, Dmitri G. Fedorov^{2,*}, Kazuo Kitaura^{1,2}, Isao Nakanishi³, and Kenneth M. Merz Jr.⁴

¹Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

²NRI, National Institute of Advanced Industrial Science and Technology (AIST), Central 2, Umezono 1-1-1, Tsukuba, 305-8568, Japan

³Department of Pharmaceutical Sciences, Kinki University, 3-4-1, Kowakae, Higashi-Osaka, Osaka 577-8502, Japan

⁴Quantum Theory Project, The University of Florida, 2328 New Physics Building, P.O. Box 118435, Gainesville, Florida 32611-8435, USA

Abstract

We propose an approach based on the overlapping multicenter ONIOM to evaluate intermolecular interaction energies in large systems and demonstrate its accuracy on several representative systems in the complete basis set limit at the MP2 and CCSD(T) level of theory. In the application to the intermolecular interaction energy between insulin dimer and 4'-hydroxyacetanilide at the MP2/CBS level, we use the fragment molecular orbital method for the calculation of the entire complex assigned to the lowest layer in three-layer ONIOM. The developed method is shown to be efficient and accurate in the evaluation of the protein-ligand interaction energies.

OMC-ONIOM with FMO

=

ONIOM up to CCSD(T)/CBS

+

many-body polarization with FMO

*corresponding author: d.g.fedorov@aist.go.jp.

Keywords

ONIOM; FMO; multicenter; CBS; CCSD(T); protein-ligand; molecular interaction

A detailed analysis of the molecular recognition of ligands by proteins is essential to our understanding of their functions *in vivo* as well as rational drug design. There are various types of the protein-ligand interactions we must consider including strong hydrogen bonds and weak van-der-Waals interactions.¹ An accurate electronic structure theory like CCSD(T) with a large basis set can be used to accurately evaluate various interactions in the ligand binding region of a protein. Molnar *et al.*² reported that the average standard deviation of the interaction energies between small dimers at the MP2/aug-cc-pVDZ level versus CCSD(T) in the complete basis set limit (CBS) was 2.5 kcal/mol, which corresponds to more than 60 times larger binding affinity, and this inaccuracy can accumulate in a protein-ligand system with multiple interactions.³ Hence, the ability to do very accurate calculations as benchmarks for lower levels of theory is essential to improve our understanding of biomolecular recognition.

Intermolecular interactions^{4, 5} have been studied for many years. Hobza and coworkers have estimated the CCSD(T)/CBS interaction energies and geometries of the DNA and RNA base pairs, and a standard set S22.^{6,7,8,9} Gordon and coworkers have demonstrated the efficiency of the effective fragment potential (EFP) in describing molecular interactions.^{10, 11, 12, 13} Sherrill and coworkers have efficiently used a combination of basis sets at a high level of theory.^{14,15} These studies show that CCSD(T) is essential for obtaining systematically accurate and reliable results. Alternatively to the correlated wave function approach, density functional theory (DFT) has been used to accurately estimate intermolecular interactions.¹⁶

CCSD(T) is very time consuming, and it is practically impossible to apply it to large molecular systems. To overcome the difficulty, multicenter approaches in integrated quantum mechanics (QM) hybrid scheme of QM:QM type such as our own N -layer integrated molecular orbital molecular mechanics (ONIOM) methodology offer a solution. ONIOM is a hybrid scheme, in which the system is divided into layers treated at different levels of theory.¹⁷ Recently, several extensions have been proposed,^{18,19,20,21,22} where a number of regions are defined in one layer. It enables one to use high-level QM methods for some regions reducing the computational cost because the number of atoms in a region is small. Tschumper *et al.* generalized ONIOM to allow more than one region in a layer¹⁸ with a possible overlap among centers in the overlapping multicenter (OMC): OMC-ONIOM¹⁹ scheme. Regions are also called centers, with the same meaning; we call them centers throughout this paper. We employ this scheme as a QM:QM hybrid method to achieve an estimate of intermolecular interaction energy in large systems. Alternatively to ONIOM, combined QM and molecular mechanics (QM/MM) methods²³ have been used to study large molecular systems.

In this study, we use the overlapping multicenter two-level ONIOM to treat molecular interactions (MI). Our goal is to accurately estimate protein-ligand interaction energy at a high QM level. We show that with the use of the fragment molecular orbital (FMO) method^{24,25,26,27} as the lowest layer in this scheme, an accurate intermolecular interaction energy between a protein and ligand can be estimated. Alternatively to FMO, other fragment-based methods can be used in principle, whose detailed review is given elsewhere.⁴ To avoid confusion, we note that in this work we follow the general ONIOM recipe with the so called mechanical embedding (no embedding electrostatic potential).

In OMC-ONIOM, the total energy E of a system is given as

$$E = E_{\text{low}}(\text{Real}) + \sum_{i=1}^m \{E_{\text{high}}(\text{Model}_i) - E_{\text{low}}(\text{Model}_i)\} - \sum_{i=1}^m \sum_{j>i}^m \{E_{\text{high}}(\text{Model}_i \cap \text{Model}_j) - E_{\text{low}}(\text{Model}_i \cap \text{Model}_j)\} \quad (1)$$

where m is the number of centers. The third term in eq 1 is a correction for a double counting due to the overlap between centers (in this work an overlap of at most two centers is considered). Dividing separate molecules A and B into non-overlapping centers, the total energy of each of them is written as

$$E(X) = E_{\text{low}}(X:\text{Real}) + \sum_{i=1}^{n_X} \{E_{\text{high}}(X:\text{Model}_i) - E_{\text{low}}(X:\text{Model}_i)\} \quad (2)$$

where X is A or B and n_X is the number of centers in X.

For complex AB, in addition to centers in A and B, extra centers ($\text{Model}_{a,b}$) consisting of center a (Model_a) in A and another center b (Model_b) in B are introduced (see Figure 1). The additional centers overlap with the original centers a and b . With this definition of centers, the intermolecular interaction energy becomes

$$\begin{aligned} \Delta E_{A,B}^{\text{int}} &= E(AB) - E(A) - E(B) \\ &= \Delta E_{\text{low}}^{\text{int}}(\text{Real}) + \sum_{a=1}^{n_A} \sum_{b=1}^{n_B} \{ \Delta E_{\text{high}}^{\text{int}}(\text{Model}_{a,b}) - \Delta E_{\text{low}}^{\text{int}}(\text{Model}_{a,b}) \} \\ &= \Delta E_{\text{low}}^{\text{int}}(\text{Real}) + \sum_{a=1}^{n_A} \sum_{b=1}^{n_B} \Delta \Delta E^{\text{int}}(\text{Model}_{a,b}) \end{aligned} \quad (5)$$

and

$$\Delta E_{\text{low}}^{\text{int}}(\text{Real}) = E_{\text{low}}(AB:\text{Real}) - E_{\text{low}}(A:\text{Real}) - E_{\text{low}}(B:\text{Real}) \quad (6)$$

$$\Delta E_Y^{\text{int}}(\text{Model}_{a,b}) = E_Y(AB:\text{Model}_{a,b}) - E_Y(A:\text{Model}_a) - E_Y(B:\text{Model}_b) \quad (7)$$

$$\Delta \Delta E^{\text{int}}(\text{Model}_{a,b}) = \Delta E_{\text{high}}^{\text{int}}(\text{Model}_{a,b}) - \Delta E_{\text{low}}^{\text{int}}(\text{Model}_{a,b}) \quad (8)$$

where Y is low or high.

Thus, the intermolecular interaction energy is expressed as a sum of the energy change in the real system for the low layer and the contributions of the intermolecular centers (model molecules), expressed as differences between the low and high layer energies (see eq 8). This scheme avoids the computations of intramolecular interactions (between centers in X), and the intermolecular interaction energy can be evaluated by accurate and expensive methods such as CCSD(T). OMC-ONIOM for molecular interactions can be easily extended to the case of more than two layers, for instance, the intermolecular interaction energy for three layers (low, middle, and high) is written as

$$\Delta E_{A,B}^{\text{int}} = \Delta E_{\text{low}}^{\text{int}}(\text{Real}) + \sum_{a=1}^{n_A} \sum_{b=1}^{n_B} \Delta \Delta E_{\text{middle-low}}^{\text{int}}(\text{Model}_{a,b}) + \sum_{a'=1}^{n'_A} \sum_{b'=1}^{n'_B} \Delta \Delta E_{\text{high-middle}}^{\text{int}}(\text{Model}_{a',b'}) \quad (9)$$

where a different set of n'_x centers is defined for the high layer.

It is clear from eqs 5–9 that we are able to achieve linear scaling in ONIOM, for all layers above the lowest, because the number of pairs computers at those levels is linear with the system size. Furthermore, because FMO is of nearly linear scaling,²⁸ the total scaling of OMC-ONIOM for MI is also nearly linear, provided that FMO is used for the lowest layer.

Accuracy tests of OMC-ONIOM for MI were carried out on four representative systems: water tetramer (system 1, Figure 2a), *N*'-acetyl-*N*-methyl-L-alaninamide (Ace-Ala-NMe) and water (system 2, Figure 2b), truncated insulin-4'-hydroxyacetanilide complex (system 3, Figure 3) and a complex of the truncated FK506 binding protein (FKBP) with 1,3-diphenyl-1-propyl-1-(3,3-dimethyl-1,2-dioxypentyl)-2-piperidine carboxylate (system 4, Figure 4). Then, the scheme was applied to the entire insulin-ligand complex (system 5) and FKBP-ligand complex (system 6). These systems are too large (743 and 1734 atoms, respectively) to perform high-level *ab initio* calculations therefore we did not use them for an accuracy test. Instead, we estimated the intermolecular interaction energy of these real protein-ligand complexes at the MP2/CBS level.

The structure of system 1 was prepared by putting one water molecule (w4) 7.86 Å apart from the optimized water trimer consisting of w1–w3 (see Figure 2a). The MP2/6-31G* level of theory was used for geometry optimizations of systems 1 and 2. For system 5, an X-ray experimental geometry was taken from Protein Data Bank²⁹ (PDBID:1TYL).³⁰ Hydrogen atoms were added to the experimental data and their positions were optimized with the MMFF94x force field³¹ using the modeling software MOE.³² Refinement of the structure was performed with FMO at the Hartree-Fock (HF) level with empirical dispersion,³³ HF-D/6-31G. The structure of system 6 was prepared as described elsewhere.³⁴ Test systems 3 and 4 were generated by truncating the corresponding systems: the amino acid residues located more than 4 Å apart from the ligand were removed and their dangling bonds were capped with hydrogen atoms, whose positions were optimized with the MMFF94x force field (see Figure 3).

Calculations using OMC-ONIOM for MI were performed as follows: the system was divided into several fragments, and they were capped with hydrogen atoms using MOE. For systems 1 and 2, MP2 with the cc-pVDZ and aug-cc-pVDZ basis sets³⁵ was employed for the calculation of the intermolecular interaction energy of the real system in the low layer. CCSD(T)/CBS values were evaluated by the method proposed by Tsuzuki *et al.*^{36,37} Various methods have been developed to eliminate or reduce the basis set superposition error (BSSE).^{38,39,40} In all the calculations of low and high layers, BSSE was corrected by the counter-poise (CP) method of Boys and Bernardi³⁸ except for FMO calculations used for system 4. For system 1 and 2, The computational methods employed were CCSD(T)/CBS:CP-MP2/cc-pVDZ and CCSD(T)/CBS:CP-MP2/aug-cc-pVDZ in the ONIOM notation (high:low layer).

System 3 (178 atoms) and system 4 (283 atoms) are very large to apply CCSD(T) to the entire complex. Even MP2/cc-pVTZ (3884 and 6825 basis functions) calculations proved impossible to perform for the entire complex. Therefore, the accuracy test was done with CP-MP2/cc-pVDZ:CP-MP2/6-31G, and the intermolecular interaction energy (which is also used as a contribution in system 5 and system 6) was evaluated at the MP2/CBS level, which

was obtained by the extrapolation scheme using the values of MP2/cc-pVDZ and MP2/cc-pVTZ level of theory.³⁷

For system 5 and system 6, three layer ONIOM scheme was employed: MP2/CBS:CP-MP2/cc-pVDZ:FMO2-MP2/6-31G, which evaluates the intermolecular interaction energy in the real protein-ligand complex at the level of MP2/CBS (FMO2 uses a two-body expansion). In the FMO calculation, BSSE was not corrected for, because the atoms in the third layer are separated by more than 6 Å from the ligand atoms, and they can be expected to have a negligible BSSE with the ligand. A pairwise CP correction scheme^{41, 42} in FMO is in our opinion not reliable for predicting intermolecular interaction energies. We used the standard peptide fragmentation at C α (one residue per fragment) and the calculations were performed with the standard options in GAMESS.⁴³

All *ab initio* calculations except the MP2 calculation of full system 3 were performed with Gaussian03⁴⁴ on the Prime Power HPC 2500 supercomputer consisting of 4 Quad Core AMD Opteron (2.3GHz) with 32GB memory. MP2 calculation of full system 3 and system 4, and FMO calculations were carried out with GAMESS on a PC cluster.

System 1 is a model of an interaction between subsystems A and B, and the water tetramer (w1–w4) was divided into two groups taking one water (w1) as A and the other three water molecules (w2–w4) as B (see Figure 2a). In the calculations using OMC-ONIOM for MI, w1 in A and w2 and w3 in B are assigned to the high layer. In the complex AB, overlapping centers composed of w1–w2 and w1–w3 were added to the high layer. Three water molecules w1–w3 are close to one other and their interaction involves a large three-body effect.^{45,46,47,25} On the other hand, the separated water molecule w4 exerts an environment effect (mainly electrostatic in nature) on the intermolecular interaction. Thus we can verify how well the many-body and environment effects neglected in the high layer can be recovered by a low layer calculations.

In the second and third columns of Table 1, the total intermolecular interaction energies are listed along with those between model molecules in OMC-ONIOM for MI. The intermolecular interaction energy in the real system was calculated to be –11.45 kcal/mol at the CCSD(T)/CBS level, which is the target value to obtain from OMC-ONIOM for MI. The interaction energy of the real system in the low layer was calculated to be –9.56 and –10.31 kcal/mol at the CP-MP2/cc-pVDZ (case 1) and CP-MP2/aug-cc-pVDZ (case 2) levels, respectively. These theories underestimated the interaction energy by 1.89 and 1.14 kcal/mol, respectively, compared to the high layer theory.

With the replacement of the interaction energies between the model molecules in the low layer by those in the high layer, OMC-ONIOM for MI gave –12.35 (–0.90) kcal/mol and –11.51 (–0.06) kcal/mol for cases 1 and 2, respectively, where the errors (deviations from the energy of CCSD(T)/CBS) are shown in parentheses. Thus, OMC-ONIOM for MI satisfactorily estimated the interaction energy. It is also worth noting that the error becomes smaller with better levels of theory for the low layer, which is the desired behavior. We can estimate the contribution of the three-body effect in the water trimer (w1–w3) and the electrostatic effect due to w4 by subtracting the sum of the model interaction energies from that of the real system. The absolute value at the CP-MP2/aug-cc-pVDZ level was relatively large (2.75 kcal/mol), which suggests that the three-body effect of water trimer is strong, and that it is accurately estimated at a low level.

On the Prime Power HPC 2500 supercomputer, the CCSD(T)/CBS calculation took 585 min, while the calculations using OMC-ONIOM for MI for case 1 and 2 took 79 and 87 min, respectively. It means that our method reduced the computational expense by 85 percent while giving nearly the same result as the high-level *ab initio* calculation.

In the calculation of the intermolecular interaction energy in system 2 between Ace-Ala-NMe (molecule A) and a water molecule (molecule B), the former was split into two fragments for assigning non-overlapping centers in the high layer and the water molecule is also assigned as a non-overlapping center. In high layer of the complex AB, the overlapping centers consisting of two centers (one center in A and another center in B) were added (see Figure 2b) This system was designed to demonstrate that artifacts arising from covalent bond fragmentation and hydrogen caps in the high layer are almost cancelled by subtracting the low layer energies.

The calculated intermolecular interaction energies are listed in the fourth and fifth columns of Table 1. The value between molecules A and B was calculated to be -8.05 kcal/mol at the level of CCSD(T)/CBS. The interaction energies obtained at the lower level of CP-MP2/cc-pVDZ (case 1) and CP-MP2/aug-cc-pVDZ (case 2) were -5.87 (2.19) and -7.04 (1.01) kcal/mol, respectively (the errors are shown in parentheses). With the replacement of the interaction energies between the model molecules in the low layer by those in the high layer, OMC-ONIOM for MI gave -8.34 (-0.29) and -8.16 (-0.11) kcal/mol for cases 1 and 2, respectively. OMC-ONIOM for MI efficiently balanced the difference between the high and low level energies for systems in which covalent bonds are broken in the process of defining centers. As mentioned above for system 1, we find that when a higher level of theory is applied to the low layer in the calculations using OMC-ONIOM for MI, the accuracy of the predicted intermolecular interaction energy is improved.

In system 3, the truncated insulin-ligand (4'-hydroxyacetanilide) complex was employed as a test system (see Figure 3) where many amino acid residues are interacting with the ligand. Some of them are large and may be difficult to calculate at a high level of theory. Therefore, an assignment of a fixed number of minimum sized centers is desirable. To analyze the importance of specific residues, we first performed FMO2-MP2/6-31G calculations on the entire system and selected amino acid residues whose correlation components of pair interaction energy with the ligand were larger than 1 kcal/mol. The selected amino acid residues were further split into functional groups and those within $4\text{\AA}$ from the ligand atoms were assigned as non-overlapping centers. The entire ligand was also assigned a non-overlapping center (see Figure 3). The resultant system has many interaction sites and is representative of a protein-ligand binding. Such a system requires well-balanced descriptions for various types of non-bonding interactions as suggested by Faver *et al.*,³ and, hence, is an important target of OMC-ONIOM for MI.

The calculated interaction energies in the real system at the low and high level are compared in Figure 5, along with those between model molecules. The interaction energy in the real system was calculated to be -19.71 and -15.10 kcal/mol at the high (CP-MP2/cc-pVDZ) and low (CP-MP2/6-31G) levels, respectively. There are 3 hydrogen bonds between the ligand and the protein (B: model₁, B: model₂ and B: model₅), and one case of a CH-O interaction (B: model₂), for which the computed energies are attractive. On the contrary, the fragments with a van-der-Waals interaction with the ligand (B: model₃, B: model₄ and B: model₆) have repulsive contributions. At the HF level the interaction energy between the truncated insulin and its ligand is positive (0.69 kcal/mol). The reason for this unfavorable contribution is attributed to the electrostatic interaction and the poor representation of the van der Waals contacts, which indicates that this ligand is immersed in the electrostatic potential from the protein. The error between the two calculations at the MP2 level (-4.61 kcal/mol) is reduced to 0.87 kcal/mol using OMC-ONIOM for MI, which indicates it has sufficient accuracy. There were seven interaction sites between them and each fragment had about a 0.1 kcal/mol error. The main contributions are the interaction energies between the ligand and two models (B: Model₄ and B: Model₅). Both of them have a large electron

correlation contribution (−3.18 and −4.96 kcal/mol) and the 6-31G basis set could not properly describe their effects.

The calculation scheme of truncated FKBP-ligand complex (system 4) is the same as that of system 3. The calculated interaction energies in the real system at the low and high level are compared in Figure 6. The interaction energy in the real system was calculated to be −53.90 and −40.47 kcal/mol at the high and low levels, respectively. There are 2 hydrogen bonds (B: model₉ and B: model₁₁), 2 CH-O interactions (B: model₁₂) and 3 CH/ π interactions (B: model₁, B: model₁₅ and B: model₁₆) between the ligand and the protein, all of which fragments have a strong attractive interaction. The error between the two calculations at the MP2 level (−13.43 kcal/mol) is reduced to 2.18 kcal/mol using OMC-ONIOM for MI within a reasonable time.

The entire insulin-4'-hydroxyacetanilide complex (system 5) and entire FKBP-ligand complex (system 6) are used to evaluate the influence of the entire protein environment on the interaction between the protein and ligand. The interaction energy for systems 5 (Figure 7) and 6 at the MP2/CBS level using OMC-ONIOM for MI was −31.93 kcal/mol and −71.76 kcal/mol, respectively, which is over 60 percent more favorable than that at the CP-MP2/cc-pVDZ level. The interaction energies at the site which formed the hydrogen bond with the ligand (B: model₁, B: model₂ and B: model₅ in system 3, B: model₉ and B: model₁₁ in system 4) change significantly and the BSSE correction is overestimated at the CP-MP2/cc-pVDZ level. This indicates that a very high level of theory will be necessary to achieve converged energies for protein-ligand complexes.^{2,3}

The contribution of the first layer energy to the intermolecular interaction energy between the protein and the ligand is calculated as follows. The intermolecular interaction energy of system 5 and system 6 obtained from FMO2-MP2/6-31G calculations without a BSSE correction is −37.75 kcal/mol and −70.03 kcal/mol, respectively. The intermolecular interaction energy of the second layer (systems 3 and 4) obtained from FMO calculations without BSSE correction is −35.69 kcal/mol and −76.27 kcal/mol, respectively. The difference between the two energies is added to obtain the intermolecular interaction energy including the effect of the entire protein environment. In system 5, the predicted interaction energy at the level of MP2/CBS with OMC-ONIOM for MI is −33.98 kcal/mol, which is accidentally close to the above mentioned FMO energy of −37.75 kcal/mol. A similar trend is found in system 6.

The contribution from the entire protein environment in systems 5 and 6 is −2.06 kcal/mol and 6.23 kcal/mol, respectively. As the system size becomes larger, the contribution increases. Moreover, this correction may be more significant for systems that have charged residues in regions far from the ligand binding site. We also performed FMO2-HF/6-31G calculation for the first layer and evaluated the difference between the HF and MP2 interaction energy in system 5. This region is far from the ligand binding site and it is much more efficient to use HF level theory for the correction of the environmental effect. The value of the polarization and charge transfer contribution evaluated with FMO for the entire protein-ligand complex with RHF was −1.48 kcal/mol, which is only −0.57 kcal/mol different compared to the MP2 result of −2.06 kcal/mol.

We have presented the overlapping multicenter ONIOM scheme for molecular interactions and demonstrated its accuracy for several representative systems. This computational scheme is a promising tool for the accurate determination of intermolecular interaction energies in large molecular systems, for example, between a protein and ligand in drug design applications.

The use of the fragment molecular orbital (FMO) method as the lowest level theory in OMC-ONIOM for MI as well as a tool to evaluate the importance of fragments in the entire system and to assign the essential parts into the high layers is very beneficial, because FMO treats many-body polarization and charge transfer in entire protein-ligand complexes. The importance of considering these effects and the need to treat full size systems has been shown by various research groups.^{48,49,50}

Within our scheme, various types of electron correlation theories can be easily used, for example, spin component scaled MP2.⁵¹ In addition, it is straightforward to use DFT, following encouraging reports of accurate calculations.^{16,52} The solvation effect plays an important role for the protein-ligand binding mechanism.^{34,53,54} We did not consider the effect in this study, but in practical applications of the method it should be included. The ability to use CCSD(T) at the CBS limit combined with the many-body polarization treated with FMO, is expected to be useful in calculating protein-ligand binding energies very accurately in future applications.

Acknowledgments

DGF and KK thank the Next Generation Supercomputing Project, Nanoscience Program and Strategic Programs for Innovative Research (MEXT, Japan), and the Computational Materials Science Initiative (CMSI, Japan) for financial support. IN and KK acknowledge a Grant-in-Aid for Scientific Research (JSPS, Japan). KMM gratefully acknowledge financial support (GM044974 and GM066859) from the United States National Institutes of Health for this project.

References

1. Bissantz C, Kuhn B, Stahl M. *J Med Chem*. 2010; 53:5061. [PubMed: 20345171]
2. Molnar LF, He X, Wang B, Merz KM Jr. *J Chem Phys*. 2009; 131:065102. [PubMed: 19691412]
3. Faver JC, Benson ML, He X, Roberts BP, Wang B, Marshall MS, Kennedy MR, Sherrill CD, Merz KM Jr. *J Chem Theory Comput*. 2011; 7:790. [PubMed: 21666841]
4. Riley KE, Pito ák M, Jure ka P, Hobza P3. *Chem Rev*. 2010; 110:5023. [PubMed: 20486691]
5. Gordon MS, Fedorov DG, Pruitt SR, Slipchenko LV. *Chem Rev*. 2012; 112:632. [PubMed: 21866983]
6. ezac J, Riley KE, Hobza P. *J Chem Theory Comput*. 2011; 7:2427. [PubMed: 21836824]
7. Grafova L, Pito ák M, eza J, Hobza P. *J Chem Theory Comput*. 2010; 6:2365.
8. Morgado CA, Jure ka P, Svozil D, Hobza P, Sponer J. *Phys Chem Chem Phys*. 2010; 12:3522. [PubMed: 20336251]
9. Jure ka P, Sponer J, Cerny J, Hobza P. *Phys Chem Chem Phys*. 2006; 8:1985. [PubMed: 16633685]
10. Slipchenko LV, Gordon MS. *J Comput Chem*. 2007; 28:276. [PubMed: 17143863]
11. Smith QA, Gordon MS, Slipchenko LV. *J Phys Chem A*. 2011; 115:4598. [PubMed: 21504175]
12. Smith QA, Gordon MS, Slipchenko LV. *J Phys Chem A*. 2011; 115:11269. [PubMed: 21877717]
13. Slipchenko LV, Gordon MS. *J Phys Chem A*. 2009; 113:2092. [PubMed: 19072625]
14. Sinnokrot MO, Vallev EF, Sherrill CD. *J Am Chem Soc*. 2002; 124:124.
15. Marshall MS, Burns LA, Sherrill CD. *J Chem Phys*. 2011; 135:194102. [PubMed: 22112061]
16. Burns LA, Vázquez-Mayagoitia Á, Sumpter BG, Sherril CD. *J Chem Phys*. 2011; 134:084107. [PubMed: 21361527]
17. Svensson M, Humble S, Froese RDJ, Matsubara T, Sieber S, Morokuma K. *J Phys Chem*. 1996; 100:19357.
18. Hopkins BW, Tschumper GS. *J Comput Chem*. 2003; 24:1563. [PubMed: 12926000]
19. Hopkins BW, Tschumper GS. *Mol Phys*. 2005; 103:309.
20. Bates DM, Smith JR, Janowski T, Tschumper GS. *J Chem Phys*. 2011; 135:044123. [PubMed: 21806106]
21. Chu ZK, Fu G, Guo W, Xu X. *J Phys Chem C*. 2011; 115:14754.

22. Kiyota Y, Hasegawa J, Fujimoto K, Swerts B, Nakatsuji H. *J Comput Chem.* 2009; 30:1351. [PubMed: 19009605]
23. Gao J, Thompson MA. Combined quantum mechanical and molecular mechanical methods. ACS symposium series. 1998; 712
24. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayashi M. *Chem Phys Lett.* 1999; 313:701.
25. Fedorov DG, Kitaura K. *J Chem Phys.* 2004; 120:6832. [PubMed: 15267582]
26. Fedorov DG, Kitaura K. *J Phys Chem A.* 2007; 111:6904. [PubMed: 17511437]
27. Fedorov DG, Nagata T, Kitaura K. *Phys Chem Chem Phys.* 2012; 14:7562. [PubMed: 22410762]
28. Scuseria GE. *J Phys Chem A.* 1999; 103:4782.
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. *Nucleic Acids Research.* 2000; 28:235. [PubMed: 10592235]
30. Smith GD, Ciszak E. *Proc Natl Acad Sci USA.* 1994; 91:8851. [PubMed: 8090735]
31. Halgren TA, Nachbar RB. *J Comput Chem.* 1996; 17:587.
32. Molecular Operating Environment (MOE), 2011.10. Chemical Computing Group Inc; 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: 2011.
33. Grimme S. *J Comput Chem.* 2004; 25:1463. [PubMed: 15224390]
34. Nakanishi I, Fedorov DG, Kitaura K. *Proteins: Struct, Funct, Bioinf.* 2007; 68:145.
35. Dunning TH Jr. *J Chem Phys.* 1989; 90:1007.
36. Tsuzuki S, Honda K, Uchimaru T, Mikami M, Tanabe K. *J Am Chem Soc.* 2002; 124:104. [PubMed: 11772067]
37. Tsuzuki S, Honda K, Uchimaru T, Mikami M. *J Chem Phys.* 2006; 124:114304. [PubMed: 16555885]
38. Boys SF, Bernardi F. *Mol Phys.* 1970; 19:553.
39. Gianinetti E, Raimondi M, Tornaghi E. *Int J Quantum Chem.* 1996; 60:157.
40. Iwata S. *J Chem Phys.* 2011; 135:094101. [PubMed: 21913747]
41. Ishikawa T, Ishikura T, Kuwata K. *J Comput Chem.* 2009; 30:2594. [PubMed: 19408278]
42. Okiyama Y, Fukuzawa K, Yamada H, Mochizuki Y, Nakano T, Tanaka S. *Chem Phys Lett.* 2011; 509:67.
43. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, et al. *J Comput Chem.* 1993; 14:1347.
44. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA., Jr; Vreven, T.; Kudin, KN.; Burant, JC., et al. GAUSSIAN03, Revision D.01. Gaussian, Inc; Wallingford, CT: 2004.
45. Xantheas SS. *J Chem Phys.* 1994; 100:7523.
46. Kitaura K, Sawai T, Asada T, Nakano T, Uebayasi M. *Chem Phys Lett.* 1999; 312:319.
47. Xantheas SS. *Chem Phys.* 2000; 258:225.
48. Sawada T, Hashimoto T, Nakano H, Suzuki T, Suzuki Y, Kawaoka Y, Ishida H, Kiso M. *Biochem Biophys Res Comm.* 2007; 355:6. [PubMed: 17292854]
49. Söderhjelm P, Aquilante F, Ryde U. *J Phys Chem B.* 2009; 113:11085. [PubMed: 19618955]
50. Tong Y, Mei Y, Li YL, Ji CG, Zhang JZH. *J Am Chem Soc.* 2010; 132:5137. [PubMed: 20302307]
51. Grimme S. *J Chem Phys.* 2003; 118:9095.
52. Antony J, Grimme S. *J Comput Chem.* 2012; 33:1730. [PubMed: 22570225]
53. Sawada T, Fedorov DG, Kitaura K. *J Phys Chem B.* 2010; 114:15700. [PubMed: 21049970]
54. Genheden S, Mikulskis P, Hu L, Kongsted J, Söderhjelm P, Ryde U. *J Am Chem Soc.* 2011; 133:13081. [PubMed: 21728337]

Fig. 1. Schematic depiction of a) the non-overlapping centers (dotted circle) in molecule A, b) the non-overlapping centers (dotted circle) in molecule B, and c) the non-overlapping (dotted circle) and overlapping centers (composed of two non-overlapping centers connected by line) in complex AB.

Fig. 2.

Test systems: a) system 1: water tetramer in which one water w1 is assigned as molecule A and three waters (w2-w4) as molecule B and b) system 2: Ace-Ala-NMe and a water molecule. Dotted ellipses indicate the non-overlapping centers and rectangles the corresponding model molecules. The interatomic distances between heavy atoms connected by dashed lines are shown in Å.

Fig. 3. Truncated insulin-4'-hydroxyacetanilide complex (system 3). The ligand forms molecule A and all peptide fragments molecule B. Dotted curve line indicate the centers and rectangles the corresponding model molecules. The interatomic distances between heavy atoms are shown in Å. Cys6 and Cys11 are in chain C, and the other residues are in chain D in the insulin tetramer.

Fig. 4. Truncated FKBP-ligand complex (system 4). The ligand forms molecule A and all peptide fragments molecule B. Dotted curve line indicate the centers and rectangles the corresponding model molecules. The interatomic distances between heavy atoms are shown in Å.

Fig. 5. Intermolecular interaction energies in the real system (Real) and between model molecules, Model_a and Model_b in system 3. The model pairs are shown as (*a*,*b*). The level of theory is CP-MP2 /6-31G for the low layer and CP-MP2/cc-pVDZ for the high layer.

Fig. 6. Intermolecular interaction energies in the real system (Real) and between model molecules, Model_a and Model_b in system 4. The model pairs are shown as (*a*,*b*). The level of theory is CP-MP2 /6-31G for the low layer and CP-MP2/cc-pVDZ for the high layer.

Fig. 7. Protein-ligand interaction energies ΔE^{int} in system 5 at the X:CP-MP2/cc-pVDZ level and the intermolecular interaction energies between the model molecules. The level of theory X is CP-MP2/cc-pVDZ for case 1, CP-MP2/cc-pVTZ for case 2 and MP2/CBS for case 3. Model pairs are shown as (a,b) .

Table 1

Comparison of the intermolecular interaction energies (kcal/mol) from OMC-ONIOM for MI and *ab initio* CCSD(T)/CBS calculations

layer-model	system1 ^a		system2 ^b	
	case1 ^c	case2 ^d	case1 ^c	case2 ^d
high-Model _{1,1}	-4.40	-4.40	-1.48	-1.48
high-Model _{1,2}	-4.36	-4.36	-5.54	-5.54
low-Model _{1,1}	-3.00	-3.81	-0.77	-0.87
low-Model _{1,2}	-2.96	-3.75	-3.77	-5.04
low-real	-9.56	-10.31	-5.87	-7.04
OMC-ONIOM for MI ^e	-12.35	-11.51	-8.34	-8.16
<i>ab initio</i> ^f	-11.45	-11.45	-8.05	-8.05
Error ^g	-0.90	-0.06	-0.29	-0.11

^a water trimer and water

^b Ace-Ala-NMe and water

^c CCSD(T)/CBS:CP-MP2/cc-pVDZ

^d CCSD(T)/CBS:CP-MP2/arg-cc-pVDZ

^e calculated from eq 5.

^f CCSD(T)/CBS, obtained by extrapolation using counterpoise corrected values for several basis sets^{36,37}

^g difference between OMC-ONIOM for MI and *ab initio* energies