Fragment MO-based Correlation Methods for Large Molecules

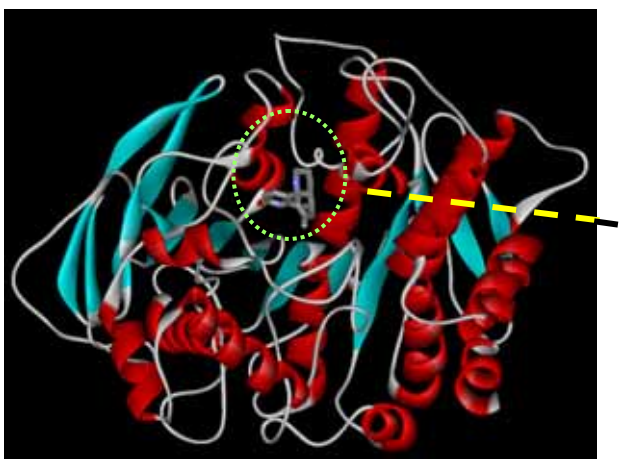
Kazuo Kitaura
Graduate School of Pharmaceutical Sciences,
Kyoto University, Japan

Outline:

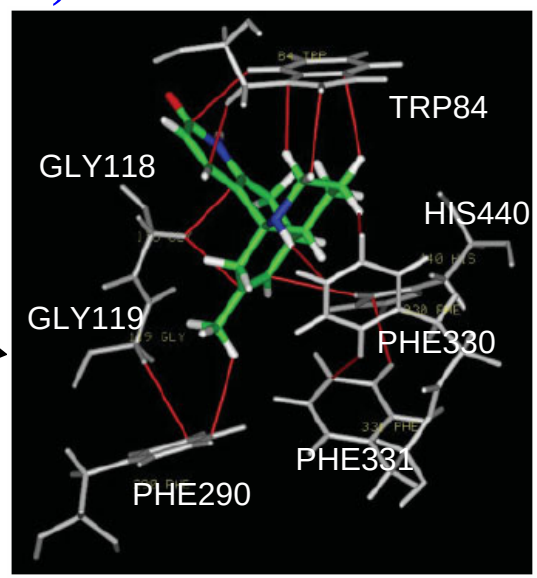
1) Importance of electron correlations in proteins
2) Fragment MO(FMO) method at HF level
3) FMO at MP2 and CC levels
4) Application to protein-ligand interactions
5) Summary

Correlation is essential for proteins and protein-ligand bindings

There are many weak interactions in proteins and protein-ligand, such as CH…X(X=O,N) and CH/π, and in these interactions dispersion interaction is essential. For instance, based-on the structural data Umezawa et al.(Biopolymer, 79, 248 (2005)) have shown the CH/ π network in the complex of Acetylcholine esterase (AchE) and huperizine B inhibitor (a drug for Alzheimer's disease).



AchE/(-)-huperizine B complex (PDB code:1GPN)



CH/π network in the binding pocket. Red sticks indicate CH/π interactions.

The HF level of theory can not describe CH/π interactions

Shibasaki et al., J. Phys. Chem. A 2006,110,10583.

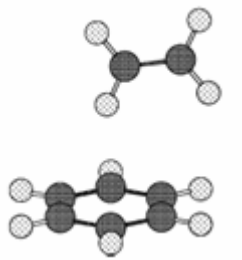
TABLE 1: Calculated MP2 and CCSD(T) Interaction Energies for the Benzene−Ethylene and Benzene−Acetylene Clusters[a]

method	$C_6H_6-C_2H_4$	$C_6H_6-C_2H_2$
HF/aug-cc-pVDZ	1.401	−0.072
HF/aug-cc-pVTZ	1.413	−0.104
HF/aug-cc-pVQZ	1.412	−0.100
MP2/aug-cc-pVDZ	−2.427	−2.900
MP2/aug-cc-pVTZ	−2.744	−3.313
MP2/aug-cc-pVQZ	−2.790	−3.421
CCSD(T)/aug-cc-pVDZ	−1.832	−2.279
$E_{MP2(limit)}$[b]	−2.823	−3.499
ΔCCSD(T)(1imit)[c]	0.658	0.747
$E_{CCSD(T)(limit)}$[d]	−2.165	−2.752
ΔZPE[e]	0.431	0.366
D_0 (calcd)[f]	1.734	2.386
D_0 (exptl)[g]	1.4 ± 0.2	2.7 ± 0.1

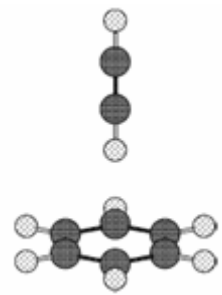
[a] Energy in kcal/mol. BSSE was corrected by the counterpoise method.

[f] Binding energy of cluster. ($D_0 = D_e - \Delta ZPE$)

C_6H_6-C_2H_4



C_6H_6-C_2H_2



Quantum Mechanical approaches to large molecules

- Truncated models have been used since early stage of quantum chemistry and they are still active.
- Quantum mechanical/molecular mechanical hybrid method (QM/MM) become widespread since the middle of 1990s.

  J. Gao, Reviews in Comp.Chem., Volume 7, 1996.
- For quantum mechanical methods for large systems have been developed since 1990s,
  - linear scaling methods (mostly based on DFT),

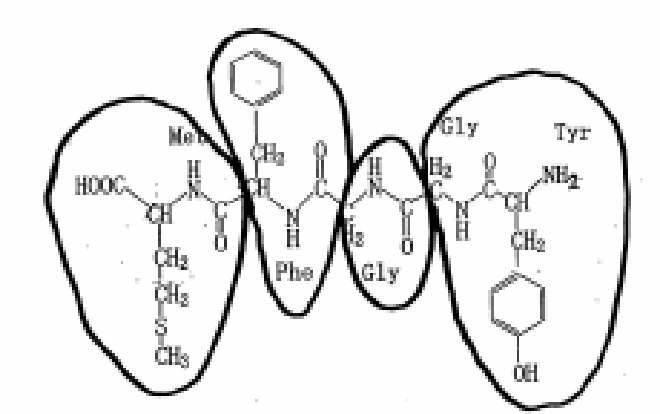
    Christian Ochsenfeld, J et al., Reviews in Comp. Chem., Volume 23, 2007.
  - and fragment-based methods.

    for a brief review, see D.G. Fedorov et al., *J. Phys. Chem. A*, *111*, 6904-6914 (2007) .
  - We are developing the fragment molecular orbital (FMO) method. The goal of FMO is to be able to treat real size proteins.

Outline of FMO

FMO is a fragment-based MO method for large molecules

- A molecule is divided into *N* fragments and *ab initio* MO calculations on the fragments (monomers) and fragment pairs (dimers) are performed under electrostatic potential from other monomers.



- The total energy of a molecule (E) is calculated using the energies of the monomer (E_I) and dimer (E_{IJ});

$$E = \sum E_I + \sum (E_{IJ} - E_I - E_J)$$

Advantage of the method:

- reproduces ab initio properties with good accuracy,
- is efficient on massively parallel computers.

FMO at RHF level (FMO-RHF)

Fock equation for fragment (monomer) and fragment pair (dimer)
(x=I for monomer and x=IJ for dimer)

$$\tilde{\mathbf{F}}^{x}\mathbf{C}^{x}=\mathbf{S}^{x}\mathbf{C}^{x}\tilde{\boldsymbol{\varepsilon}}^{x}$$

$$\tilde{\mathbf{F}}^{x}=\tilde{\mathbf{H}}^{x}+\mathbf{G}^{x}\ ,$$

$$\tilde{\mathbf{H}}^{x}_{\mu\nu}=H^{x}_{\mu\nu}+V^{x}_{\mu\nu}+\sum_{i}B\left\langle\mu\middle|\varphi_i^h\right\rangle\left\langle\varphi_i^h\middle|\nu\right\rangle\ ,$$

$$V^{x}_{\mu\nu}=\sum_{K\neq x}\left\{\sum_{A\in K}\left\langle\mu\left|-\frac{Z_A}{\left|\mathbf{r}-R_A\right|}\right|\nu\right\rangle+\sum_{\rho\sigma\in K}D^{K}_{\lambda\sigma}(\mu\nu|\rho\sigma)\right\}$$

Total energy of monomer and dimer,

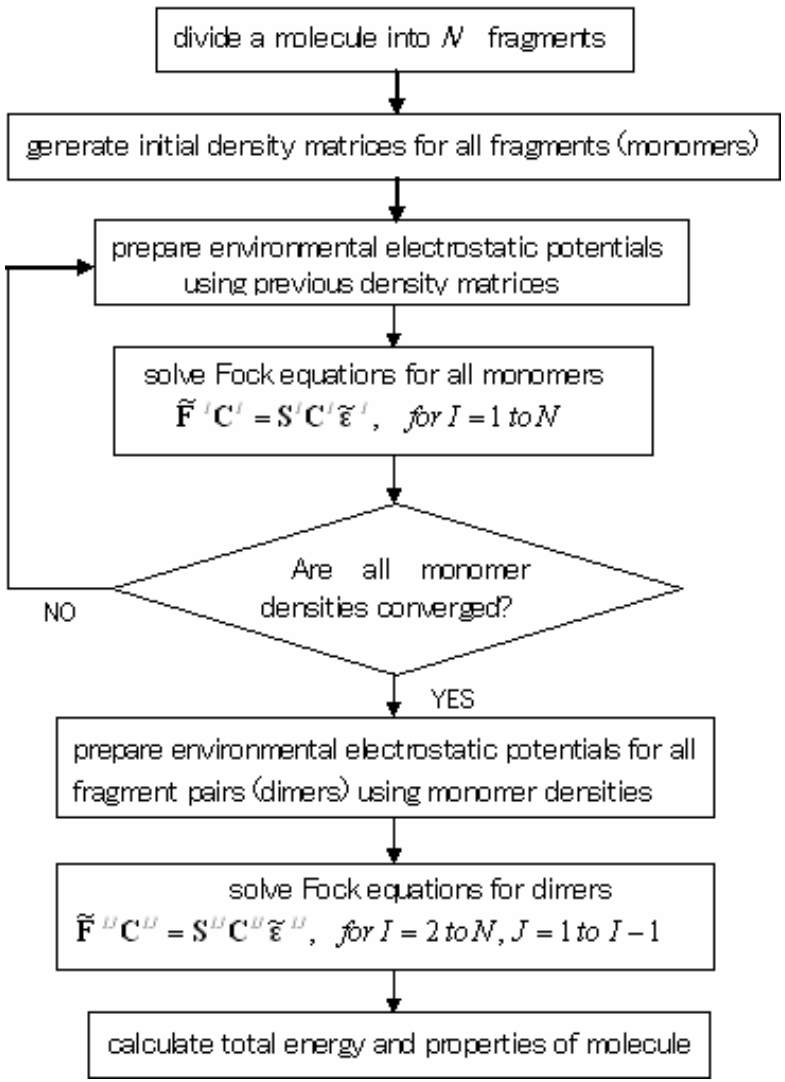
$$E_x=\frac{1}{2}Tr\left\{\mathbf{D}^{x}\left(\tilde{\mathbf{H}}^{x}+\tilde{\mathbf{F}}^{x}\right)\right\}$$

The monomers are solved self-consistently and the dimers are solved once in the monomer electrostatic environments.

Total energy of the whole molecule,

$$E=\sum_{I}E_I+\sum_{I>J}\left(E_{IJ}-E_I-E_J\right)$$

Flowchart of FMO calculations



FMO includes higher body interactions

Total energy of the whole molecule,

$$E = \sum_I E_I + \sum_{I>J} \left(E_{IJ} - E_I - E_J \right)$$

Note that the monomer and dimer energy include electrostatic interaction energy with the environment. By subtracting this energy, we obtain,

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

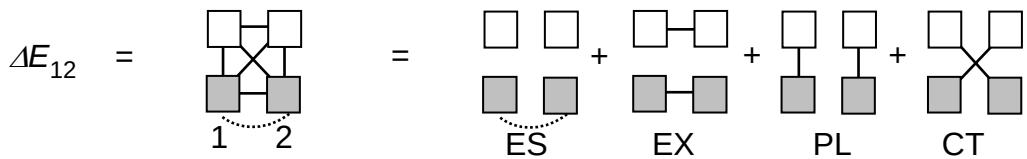
where, $E'_x = E_x - Tr\left(\mathbf{D}^{IJ} \mathbf{V}^{IJ} \right)$ is internal energy of monomer/dimer and $\Delta \mathbf{D}^{IJ} = \mathbf{D}^{IJ} - \mathbf{D}^{I} - \mathbf{D}^{J}$ and $\mathbf{V}^{IJ}$ are the difference densty and the environmental electrostatic potential, respectively.

Thus, FMO is not a simple 2-body interaction model. It includes higher body interactions!

The basic idea of FMO comes from the energy decomposition analysis (EDA) for molecular interactions

Kitaura and Morokuma, IJQC,10,325(1976)

The EDA scheme based on orbital interactions (at HF level).



where ▢ and □ are the occupied and unoccupied MOs of isolated molecule, respectively.

ES：electrostatic:

$$E_{\mathrm{ES}} = \left\langle \Phi_1^0 \cdot \Phi_2^0 \middle| H_{12} \middle| \Phi_1^0 \cdot \Phi_2^0 \right\rangle - E_1^0 - E_2^0$$

EX：exchange-repulsion:

$$E_{\mathrm{EX}} = \left\langle \mathrm{A}\left(\Phi_1^0 \cdot \Phi_2^0\right) \middle| H_{12} \middle| \mathrm{A}\left(\Phi_1^0 \cdot \Phi_2^0\right) \right\rangle - \left\langle \Phi_1^0 \cdot \Phi_2^0 \middle| H_{12} \middle| \Phi_1^0 \cdot \Phi_2^0 \right\rangle$$

PL：polarization:

$$E_{\mathrm{PL}} = \left\langle \Phi_1 \cdot \Phi_2 \middle| H_{12} \middle| \Phi_1 \cdot \Phi_2 \right\rangle - \left\langle \Phi_1^0 \cdot \Phi_2^0 \middle| H_{12} \middle| \Phi_1^0 \cdot \Phi_2^0 \right\rangle$$

CT：charge-transfer:

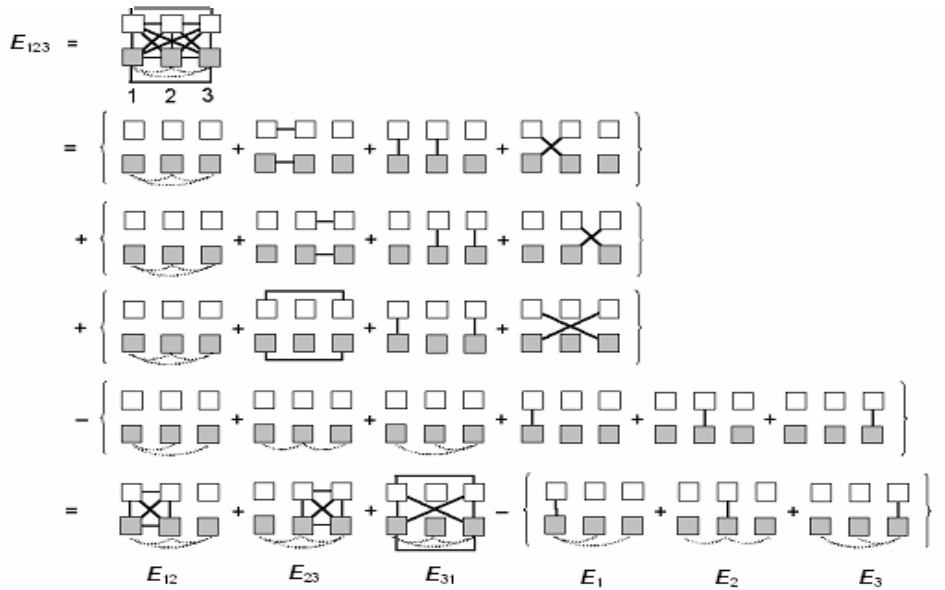
$$E_{\mathrm{CT}} = \left\langle \mathrm{A}\left(\Phi_1^{\mathrm{CT}} \cdot \Phi_2^{\mathrm{CT}}\right) \middle| H_{12} \middle| \mathrm{A}\left(\Phi_1^{\mathrm{CT}} \cdot \Phi_2^{\mathrm{CT}}\right) \right\rangle - \left\langle \Phi_1^0 \cdot \Phi_2^0 \middle| H_{12} \middle| \Phi_1^0 \cdot \Phi_2^0 \right\rangle$$

Total interaction:

$$\Delta E_{12}^{\mathrm{HF}} = E_{12}^{\mathrm{HF}} - E_1^0 - E_2^0 = E_{\mathrm{ES}} + E_{\mathrm{EX}} + E_{\mathrm{PL}} + E_{\mathrm{CT}} + E_{\mathrm{MIX}}$$

EDA scheme applied to many-body molecular interactions → FMO D.G.Fedorov et al.,JCP,120,6832(2004)

If the orbital interactions are pairwise additive, then the total energy can be decomposed into the following contributions.



$$E_{123} = E_{12} + E_{23} + E_{31} - E_1 - E_2 - E_3$$
$$= E_1 + E_2 + E_3 + (E_{12} - E_1 - E_2) + (E_{23} - E_2 - E_3) + (E_{31} - E_3 - E_1)$$

$$E = \sum_I E_I + \sum_{I>J} (E_{IJ} - E_I - E_J)$$ This is the energy expression of FMO.

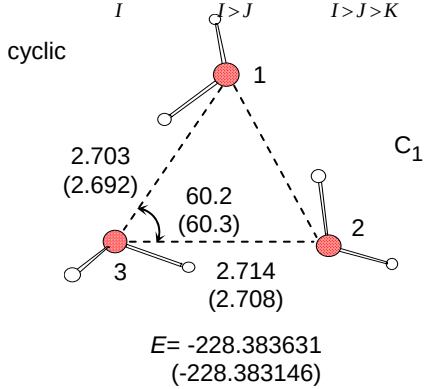
FMO includes many-body interaction energies with good accuracy

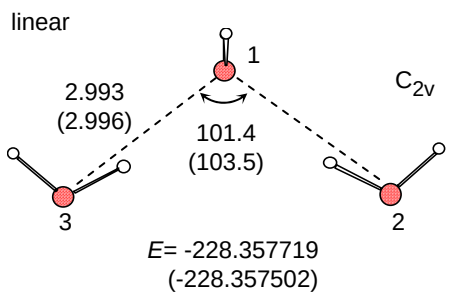
Many-body interaction energies (kcal/mol) of water clusters (RHF/6-31G)

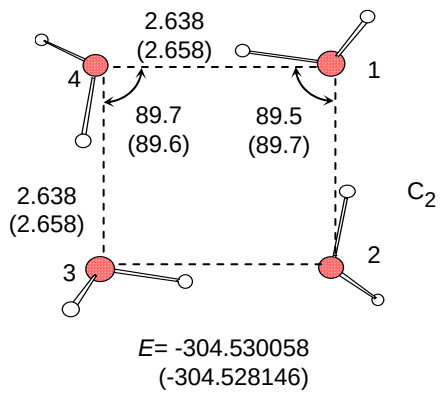
	ab initio				FMO
	2-body	3-body	4-body	total	
cyclic trimer	-21.86	-4.07		-25.93	-25.91
linear trimer	-11.03	0.85		-10.18	-10.33
tetramer	-35.26	-10.56	-0.80	-46.63	-47.45

Energy of isolated molecule E_I^0 , 2-body $e_{IJ}^0 = E_{IJ}^0 - E_I^0 - E_J^0$, and 3-body $e_{IJK}^0 = \left(E_{IJK}^0 - E_I^0 - E_J^0 - E_K^0\right) - e_{IJ}^0 - e_{JK}^0 - e_{KI}^0$ interaction energies in the series expansion

$$E_{12\cdots N} = \sum_I E_I^0 + \sum_{J>J} e_{IJ}^0 + \sum_{I>J>K} e_{IJK}^0 + \ldots + e_{12\cdots N}^0$$







Addition of explicit 3-body contribution (FMO3) improves accuracy, although 2-body expansion FMO(FMO2) is already accurate

FMO3:

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} \Delta E_{IJ} + \sum_{I>J>K}^{N} (E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK})$$

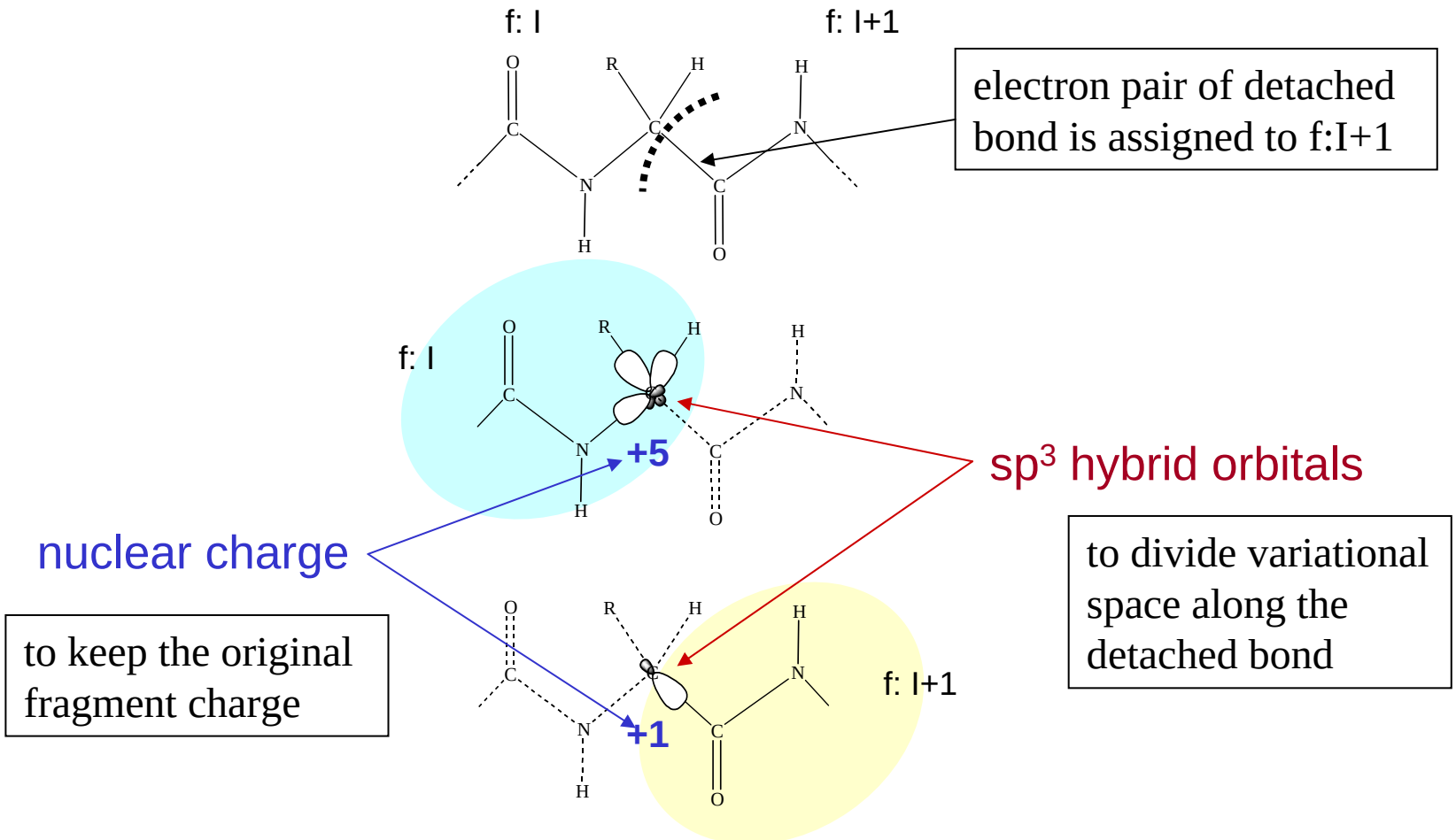
explicit 3-body term

where,

$$\Delta E_{IJ} \equiv E_{IJ} - E_I - E_J$$

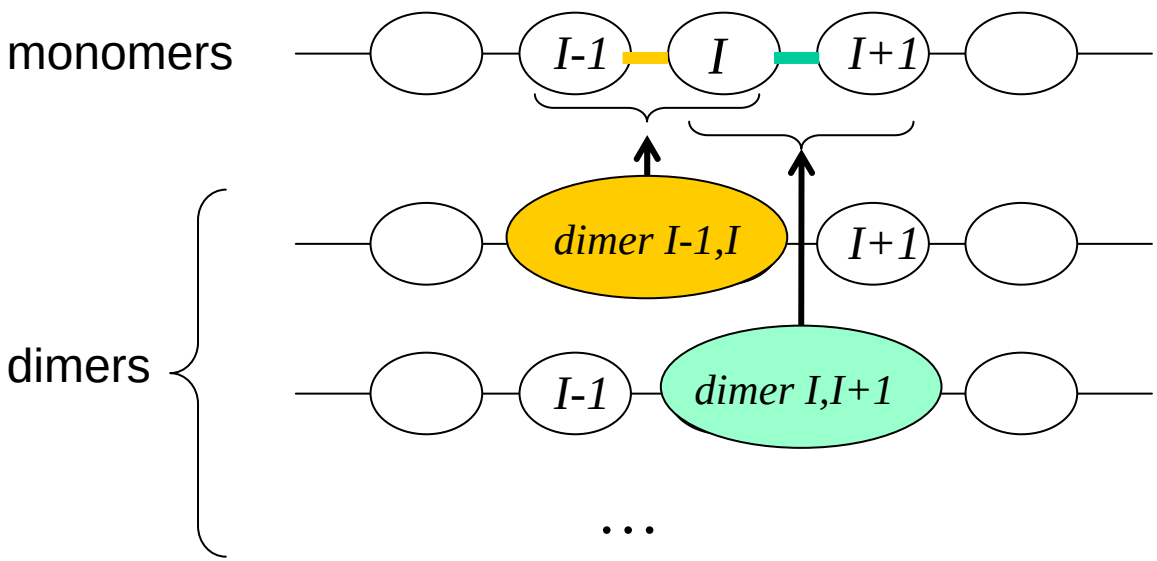
Extension to covalent-bonded fragments

Divide and assignment of basis functions and nuclear charge of boundary atom (bond-detached atom, BDA)



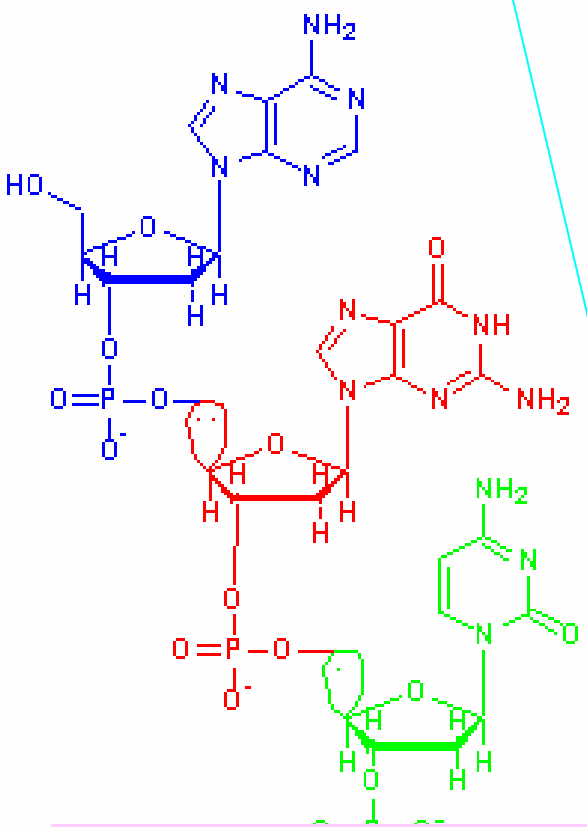
The defect of the fragmentation is almost completely patched, by replacing the fractionated covalent-bonded monomers with corresponding dimer recovering short-ranged quantum effects (exchange-repulsion and charge-transfer interactions).

$$E = \sum_I E_I + \sum_{I>J} \left(E_{IJ} - E_I - E_J \right)$$



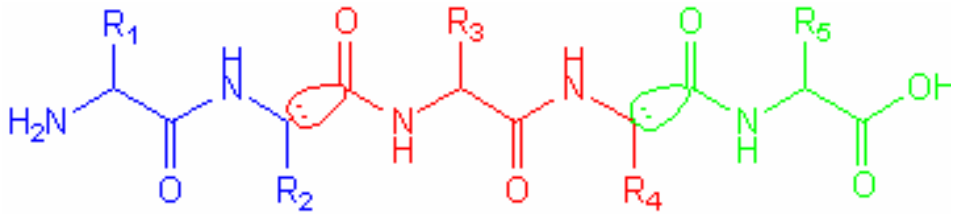
Partition of biomolecules

DNA/RNA

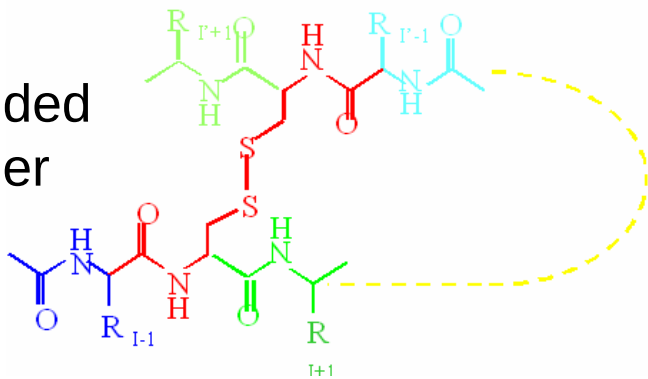


1 nucleotide/frg

polypeptide and protein



S-S bonded
Cys dimer



- Split molecule at sp^3 carbon atoms with several tens atoms per fragment.
- 2res/1frg is a reasonable choice for polypeptides.

Approximations employed in FMO-RHF

Nakano et al.,CPL,318,614 (2000)

- Approximations for electrostatic potential

$$V_{\mu\nu}^{x} = \sum_{K \neq x} \left\{ \sum_{A \in K} \left\langle \mu \left| -\frac{Z_A}{|\mathbf{r} - R_A|} \right| \nu \right\rangle + \sum_{\rho\sigma \in K} D_{\lambda\sigma}^{K} (\mu\nu|\rho\sigma) \right\}$$

1) Mulliken approximation to two-electron integrals

$$V_{\mu\nu}^{K} \cong \sum_{\lambda \in K} (\mathbf{D}^K \mathbf{S}^K)_{\lambda\lambda} (\mu\nu|\lambda\lambda) \qquad for\ R_{\min}(x,K) \geq R_{\mathrm{ESPAP}}$$

2) Point charge approximation to largely separated monomers

$$V_{\mu\nu}^{K} \cong \sum_{A \in K} \left\langle \mu \left| \left(Q_A / |\mathbf{r} - \mathbf{r}_A| \right) \right| \nu \right\rangle \quad for\ R_{\min}(x,K) \geq R_{\mathrm{ESPPC}}$$

- Electrostatic interaction approx. for largely separated dimers

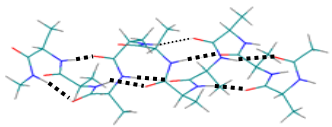
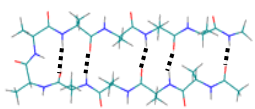
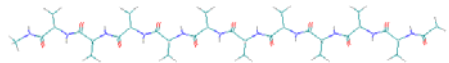
$$E_{IJ} \cong E_I^{'} + E_J^{'} + Tr\left(\mathbf{D}^I \mathbf{u}^J\right) + Tr\left(\mathbf{D}^J \mathbf{u}^I\right) + \sum_{\mu\nu \in I} \sum_{\lambda\sigma \in J} D_{\mu\nu}^{I} D_{\lambda\sigma}^{J} (\mu\nu|\lambda\sigma)$$

$$for\ R_{IJ} \geq R_{\mathrm{ESDIM}}$$

With these approximations, the computational time is reduced drastically.

Accuracy of FMO relative to regular *ab initio* MO

Geometry and total energy of polyalanine with two residues per fragment partition

α-helix  β-turn  extended 

Fedorov et al., J. Phys. Chem. A2007,111,2722

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]
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 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
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.

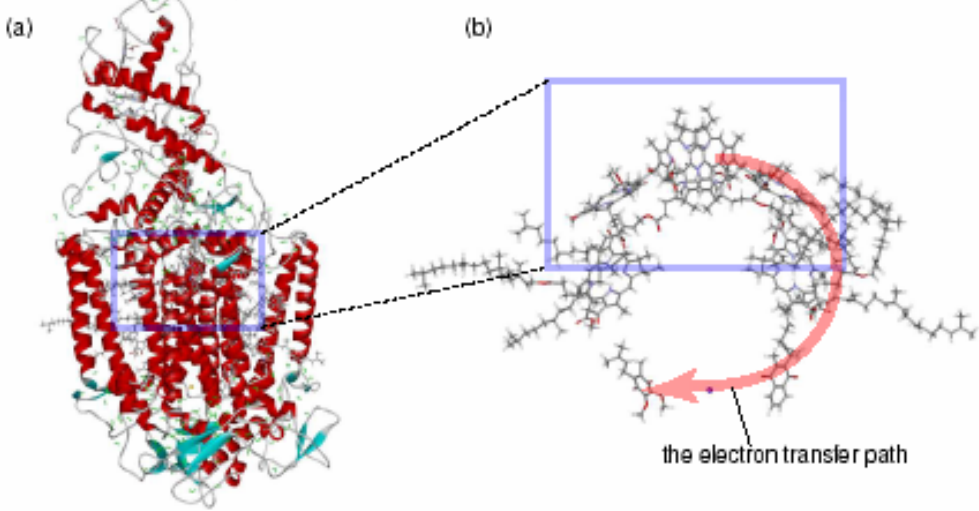
- RMSD is ~0.2Å and the error in the total energy is ~2 kcal/mol.
- The very small error of the extended conformer indicates that the fractionation itself does not cause large errors.

FMO is applicable to real size proteins

FMO-RHF/6-31G* calculation was performed on Photosynthetic reaction center protein complex of *Rhodopseudomonas viridis*

T.Ikegami et al., *Proc. Supercomputing 2005*, ACM and IEEE

- the complex has 20,581 atoms and 77,754 electrons.
- the number of basis functions is 164,442 (6-31G*).
- the computational time was 3 days on dual Opteron 300 nodes (600CPU).
- GAMESS was used.



FMO at MP2 Level (FMO-MP2)

Fedorov et al., *J.Chem.Phys*., 121, 2483-2490 (2004)

- Total Hamiltonian of monomer (*x*=I) and dimer (*x*=IJ)

$$H_x = \sum_i^{n_x} \left\{ \left(-\frac{1}{2}\nabla_i^2 - \sum_{s\in x} \frac{Z_s}{|\mathbf{r}_i - \mathbf{r}_s|} \right) + \sum_{i>j}^{n_x} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \left(\sum_{s\in J(\neq x)} \frac{Z_s}{|\mathbf{r}_i - \mathbf{r}_s|} + \sum_{J\neq I}^{N} \int d\mathbf{r}' \frac{\rho_J(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} \right) \right\}$$

- MP2 correlation energy of monomer (*x*=I) and dimer (*x*=IJ)

$$E_x^{corr} = -\frac{1}{4} \sum_{i,j}^{occ} \sum_{p,q}^{unocc} \frac{|(ij \,||\, pq)|^2}{\tilde{\varepsilon}_p^x + \tilde{\varepsilon}_q^x - \tilde{\varepsilon}_i^x - \tilde{\varepsilon}_j^x}$$

- Total correlation energy (in case of 2-body expansion)

$$E^{corr} = \sum_{I>J} E_{IJ}^{corr} - (N-2) \sum_I E_I^{corr}$$

This expansion is similar to that used in the incremental method proposed by Stoll et al.(Phys. Rev. B, 1992,46,6700). The difference is in zero-th order state; our state is not HF!

- Total energy of molecule at correlated level

$$E = E^{HF} + E^{corr}$$

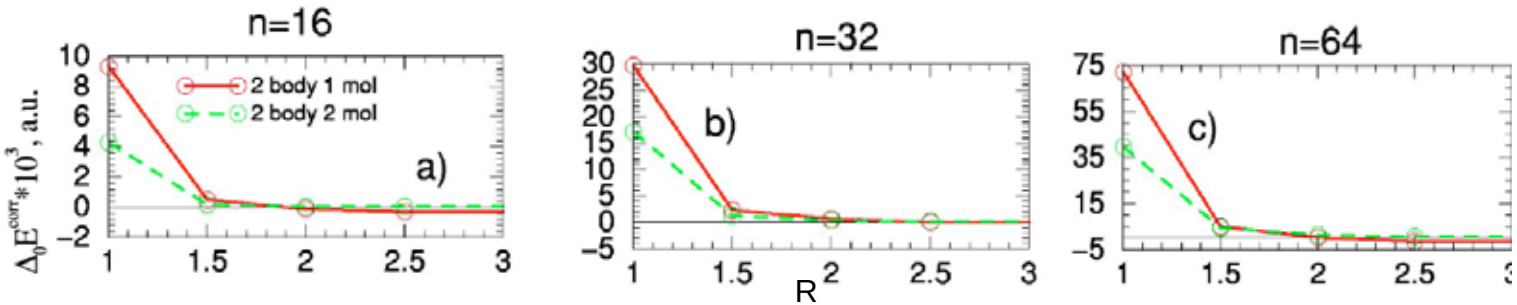
- Separated dimer approximation

contribution from far separated dimer (R_{IJ}>Rcorsd) is neglected.

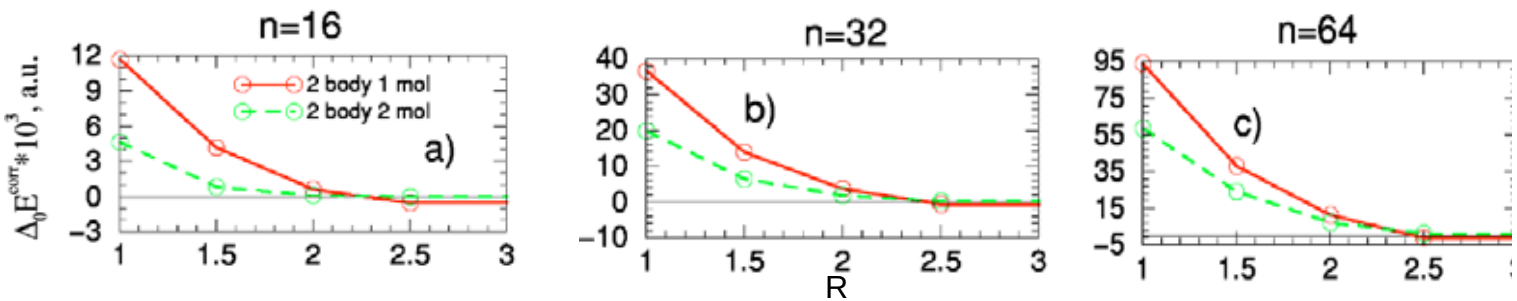
Error dependence of correlation energy on cutoff distance (Rcorsd)

Water clusters $(H_2O)_n$ (n=16,32,and 64) used for calculations

MP2/6-31G*



MP2/6-311+G*



Rcorsd=2.0 is reasonable choice.

R (distance between closest contact atoms) is measured relative to the sum of their vdw radii.

Error in MP2 correlation energy: 6-31(+)G*

Water clusters, α-helix,β-strand polyalanine and small protein(1L2Y).

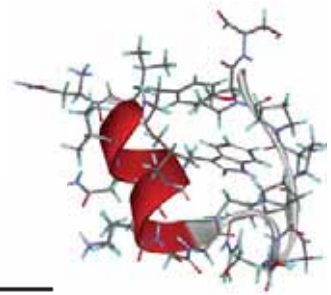


Table 3. FMO*n*/*m* Errors (in mhartree) in the MP2 Correlation Energy, Compared with *Ab Initio* MP2 (in hartree), where the *n*-body Expansion Is Used with *m* Molecules/Residues per Fragment.

System	FMO2/1[a]	FMO2/2[a]	FMO3/1[b]	FMO3/2[b]	MP2
$(H_2O)_{16}$	−1.290	−0.724	0.022	0.020	−3.00561269
$(H_2O)_{32}$	−2.520	−1.984	0.098	−0.085	−6.02565675
$(H_2O)_{64}$	−4.719	−2.001	0.605	0.171	−12.08009436
α-$(ALA)_{10}$	1.037	−0.305	−0.628	−0.026	−7.94486866
β-$(ALA)_{10}$	0.027	0.200	−0.081	0.013	−7.89106368
α-$(ALA)_{20}$	1.710	−0.824	−1.924	−0.339	−15.17568264
β-$(ALA)_{20}$	0.471	0.621	−0.157	0.051	−15.06666439
α-$(ALA)_{40}$	3.275	−1.777	−4.810	−1.175	−29.63499871
β-$(ALA)_{40}$	1.407	1.493	−0.303	0.130	−29.41790233
1L2Y	2.052	3.328	−1.981	−0.248	−22.08696021

6-31(+)G* is employed throughout.

[a]Usual two-body approximations (RESPAP = 1.5, RESPPC = 2.5, RESDIM = 2.5, and RCORSD = 2.0).

[b]Usual three-body approximations (RESPAP = 0, RESPPC = 0, RESDIM = 5.0, RITRIM = 2.5, RCORSD = 4.0, and RCORST = 2.0).

FMO2/2 and FMO3/2 errors are 3.3 and 1.2 millihartree at most, respectively.

Error in MP2 correlation energy: 6-311(+)G*

Table 4. FMO*n*/*m* Errors (in mhartree) in the MP2 Correlation Energy, Compared with *Ab Initio* MP2 (in hartree), where the *n*-body Expansion Is Used with *m* Molecules/Residues per Fragment.

System	FMO2/1[a]	FMO2/2[a]	FMO3/1[b]	FMO3/2[b]	MP2
$(H_2O)_{16}$	−5.983	−3.348	0.941	0.523	−3.30616641
$(H_2O)_{32}$	−14.190	−8.061	3.777	1.312	−6.62839056
$(H_2O)_{64}$	−33.167	−17.019	10.695	5.031	−13.28825778
α-$(ALA)_{10}$	−17.877	−4.073	−4.807	−0.066	−8.55185764
β-$(ALA)_{10}$	−8.325	0.348	−0.395	0.036	−8.49400568
α-$(ALA)_{20}$	−42.920	−11.519	−12.485	−0.378	−16.33225639
β-$(ALA)_{20}$	−18.027	1.052	−0.855	0.143	−16.21745914
α-$(ALA)_{40}$[c]	−93.743	−25.853	−27.255	0[c]	(−31.89153831)[c]
β-$(ALA)_{40}$[c]	−37.742	2.131	−2.127	0[c]	(−31.66403704)[c]
1L2Y	−55.687	−4.612	−15.076	−0.812	−23.74016031

6-311(+)G* is employed throughout.

[a]Usual two-body approximations (RESPAP = 1.5, RESPPC = 2.5, RESDIM = 2.5, and RCORSD = 2.0).

[b]Usual three-body approximations (RESPAP = 0, RESPPC = 0, RESDIM = 5.0, RITRIM = 2.5, RCORSD = 4.0, and RCORST = 2.0).

[c]The FMO3/2 energy is used as the reference instead of the unavailable MP2 energy.

The error becomes larger for larger basis set.

For reduce the error fragment size should be increased.

Plot of MP2 correlation energy error

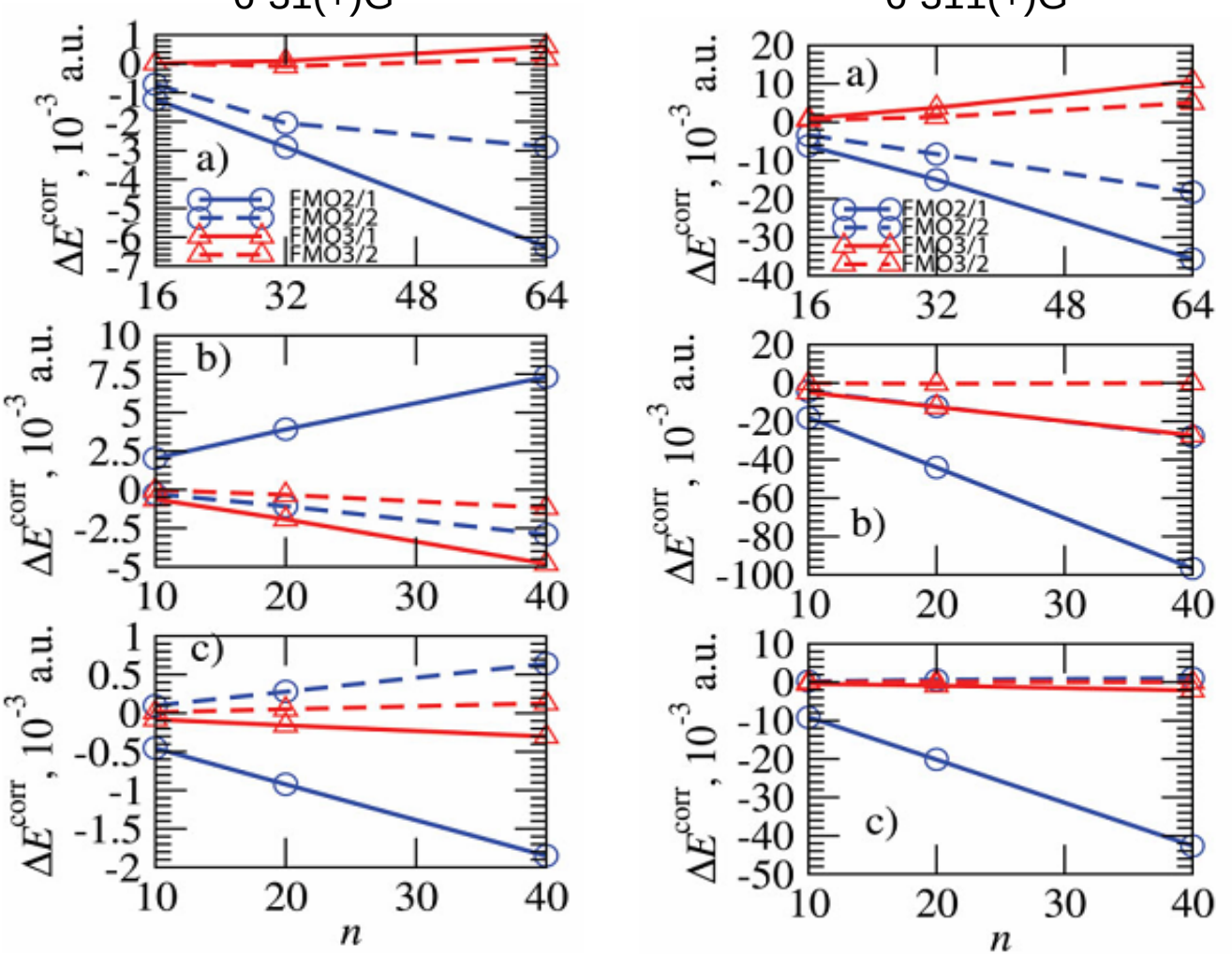
6-31(+)G*

6-311(+)G*

$(H_2O)_n$

α-$(Ala)_n$

β-$(Ala)_n$



Error depends on system size (nearly constant per fragment) and conformation.

FMO-based coupled cluster (FMO-CC)

D.G.Fedorov et al., J. Chem. Phys. 123,134103 (2005)

Calculated systems

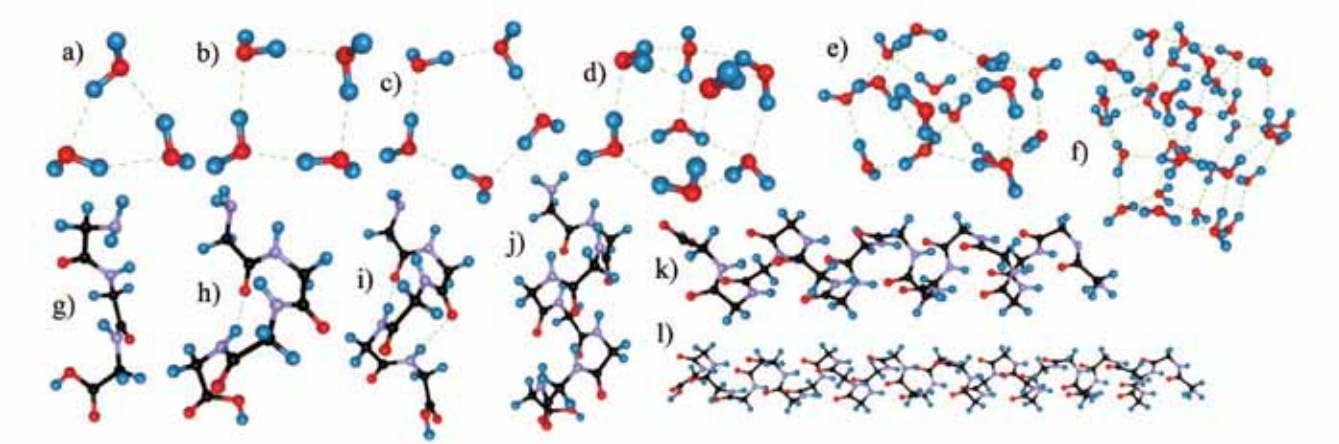


FIG. 2. (Color) Structures used in calculations. Parts (a)–(f) depict $(H_2O)_n$, $n=3,4,5,8,16,32$, correspondingly, and parts (g)–(l) depict $(GLY)_n$, $n=3,4,5,8,16,32$, correspondingly.

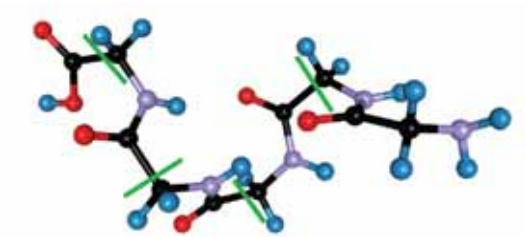


FIG. 3. (Color) Fractioning of $(GLY)_3$ into five fragments, conventionally performed at the C–C bond adjacent to a peptide bond. Inner fragments have seven atoms and terminal fragments have fewer or more.

TABLE I. The number of correlated electrons N_e and the number of spherical atomic orbitals N_{BF}.

		N_{BF}			
System	N_e	cc-VDZ	cc-PVDZ	cc-PVTZ	cc-PVQZ
$(GLY)_3$	74	139			
$(GLY)_4$	96	181			
$(GLY)_5$	118	223			
$(GLY)_8$	184	349			
$(GLY)_{16}$	360	685			
$(GLY)_{32}$	712	1357			
$(H_2O)_3$	24		72	174	
$(H_2O)_4$	32		96	232	
$(H_2O)_5$	40		120	290	
$(H_2O)_8$	64	104	192	464	920
$(H_2O)_{16}$	128		384	928	1840
$(H_2O)_{32}$	256		768	1856	3680

Error in FMO-CCSD(T) correlation energy and timings

TABLE III. The error in the correlation energy ΔE^{corr} for FMO*n*-CC, relative to *ab initio* values, where *n* denotes the *n*-body FMO expansion. *m* is the number of molecules/residues per fragment. Timings *T* in minutes are for a single node of 3.2-GHz Pentium4 with 1-Gbyte RAM, except where otherwise indicated. No approximations were used in either FMO-based or *ab initio* CC.

System	Basis set	*m*	ΔE^{corr}_{FMO2} [a]	ΔE^{corr}_{FMO3} [a]	$E^{corr}_{ab\ initio}$ [b]	T_{FMO2} [c]	T_{FMO3} [c]	$T_{ab\ initio}$ [c]
				CCSD(T)			CCSD(T)	
$(H_2O)_8$	cc-VDZ	1	−1.279 32	0.314 59	−1.099 440 38	0.7[d]	6.7[d]	90.5
$(H_2O)_8$		2	−0.419 11	0.023 97	−1.099 440 38	3.3[d]	22.7[d]	90.5
$(H_2O)_3$	cc-pVDZ	1	−0.267 91	0.000 00	−0.647 995 36	1.1	3.8	2.7
$(H_2O)_4$		1	−0.218 59	0.116 75	−0.866 102 80	2.1	13.2	16.6
$(H_2O)_5$		1	−0.027 43	0.142 05	−1.081 991 61	3.3	30.7	64.1
$(H_2O)_8$		1	−0.761 24	0.213 64	−1.744 307 45	5.1[d]	88.7[d]	14 817.3
$(H_2O)_8$		2	−0.214 19	0.010 98	−1.744 307 45	56.5[d]	502.9[d]	14 817.3
$(H_2O)_3$	cc-pVTZ	1	−0.425 47	0.000 00	−0.831 520 83	63.0	227.7	164.7
$(GLY)_3$	cc-VDZ	1	0.117 73	0.000 00	−1.421 171 57	134.9	493.9	359.0
$(GLY)_4$		1	−0.155 51	−0.036 83	−1.852 664 85	254.9	2061.7	12 185.4

[a] In 10^{-3} a.u.
[b] In a.u.
[c] In minutes.
[d] On two nodes.

FMO2 gives reasonable correlation energy with small computer time.

Relative accuracy and timings of FMO-CCSD(T)

$$\delta E^{\mathrm{corr}}_{\mathrm{FMO}n} = \frac{|\Delta E^{\mathrm{corr}}_{\mathrm{FMO}n}|}{E^{\mathrm{corr}}_{ab\ initio}} = \frac{|E^{\mathrm{corr}}_{\mathrm{FMO}n} - E^{\mathrm{corr}}_{ab\ initio}|}{E^{\mathrm{corr}}_{ab\ initio}}, \qquad \delta T_{\mathrm{FMO}n} = \frac{T_{\mathrm{FMO}n}}{T_{ab\ initio}}$$

TABLE IV. The relative accuracy $\delta E^{\mathrm{corr}}_{\mathrm{FMO}n}$ in the CCSD(T) correlation energy, and relative timings $\delta T_{\mathrm{FMO}n}$ for FMO*n*-CC (based on *m* molecules/residues per fragment) measured against *ab initio* CC.

System	Basis set	*m*	$\delta E^{\mathrm{corr}}_{\mathrm{FMO2}}$, %	$\delta E^{\mathrm{corr}}_{\mathrm{FMO3}}$, %	δT_{FMO2}, %	δT_{FMO3}, %
$(H_2O)_8$	cc-VDZ	1	99.883 64	99.971 39	1.55[a]	15.81[a]
$(H_2O)_8$		2	99.961 88	99.997 82	7.29[a]	50.16[a]
$(H_2O)_3$	cc-pVDZ	1	99.958 66	100.000 00	40.74	140.74
$(H_2O)_4$		1	99.974 76	99.986 52	12.65	79.52
$(H_2O)_5$		1	99.997 46	99.986 87	5.15	47.89
$(H_2O)_8$		1	99.956 36	99.987 75	0.07[a]	1.20[a]
$(H_2O)_8$		2	99.987 72	99.999 37	0.76[a]	6.79[a]
$(H_2O)_3$	cc-pVTZ	1	99.948 83	100.000 00	38.25	138.25
$(GLY)_3$	cc-VDZ	1	99.991 72	100.000 00	37.58	137.58
$(GLY)_4$		1	99.991 61	99.998 01	2.09	16.92

[a]FMO*n* timings collected on two nodes were multiplied by a factor of 2 to compare with *ab initio* timings obtained on one node.

FMO2 recovers correlation energy more than 99.88% and FMO3 99.97%.

FMO-CCSD(T) correlation energy errors of very large water clusters (relative to FMO3/2)

TABLE V. The error in the CCSD(T) correlation energy ΔE^{corr} (in 10^{-3} a.u.) for $(H_2O)_n$, cc-pVDZ, relative to FMO3/2. FMOm/k denotes the m-body FMO-CC method based on k molecules per fragment. Some very large calculations without approximations were not performed. Timings $T_{FMOm/n}$ are given in minutes on clusters of two, four, and eight 3.2-GHz Pentium computers for n=8, 16, and 32, respectively.

n	$\Delta E^{corr}_{FMO2/1}$ [a]	$\Delta E^{corr}_{FMO2/2}$ [a]	$\Delta E^{corr}_{FMO3/1}$ [a]	$E^{corr}_{FMO3/2}$ [b]	$T_{FMO2/1}$ [c]	$T_{FMO2/2}$ [c]	$T_{FMO3/1}$ [c]	$T_{FMO3/2}$ [c]
8[d]	−0.772 22	−0.225 17	0.202 66	−1.744 296 47	5.1	56.5	88.7	502.9
16[d]	−1.272 44	−0.476 18	0.069 13		10.4	129.5	417.4	
16[e]	−1.084 23	−0.412 47	0.033 19	−3.490 746 48	7.9	112.0	302.7	3195.9
32[d]	−3.412 70	−1.471 10			21.8	275.0		
32[e]	−2.610 40	−1.181 10	0.216 68	−7.008 545 48	13.4	203.1	756.4	11975.6

[a] In 10^{-3} a.u.
[b] In a.u.
[c] In minutes.
[d] Without approximations. Note that $(H_2O)_8$ is too small to apply useful approximations.
[e] With approximations (R_{CORSD}=2.0 for FMO2; R_{CORSD}=4.0 and R_{CORST}=2.0 for FMO3).

3.4 h | 12.6 h | 8.3 d

on 8 3.2-GHz CPUs

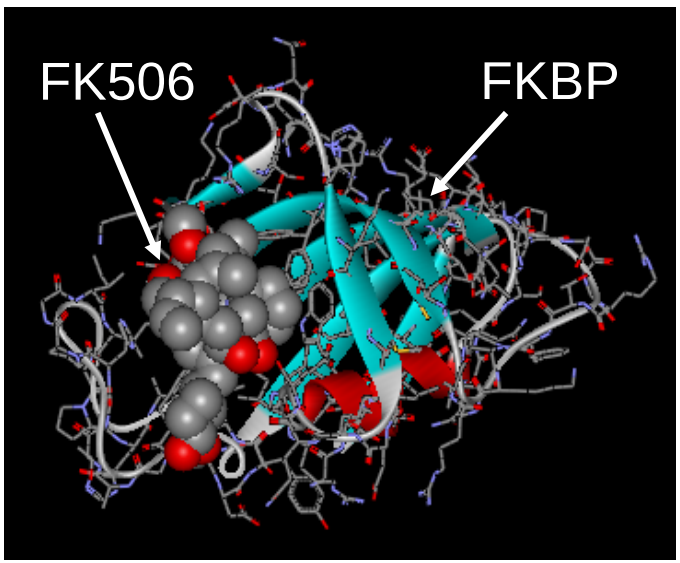
FMO3/2-CCSD(T) calculations of $(H_2O)_{32}$ with cc-pvdz (3680 basis) took 8.3 days on eight 3.2 GHz pentium PCs.

Application: protein-ligand binding

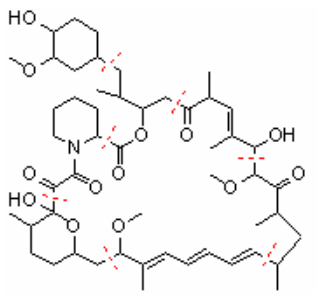
Binding energy calculation FK506 binding protein(FKBP) and its ligands

Isao Nakanishi et al., *Proteins: Struct., Funct., Bioinf.* 68, 145 (2007)

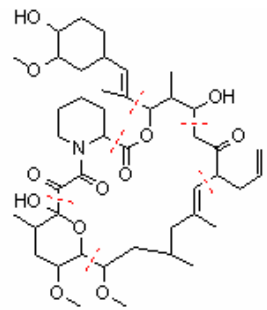
- FK506 is an immunosuppressant.
- The ligand geometry in the complex was optimized at FMO2-RHF/3-21G.
- The energies were obtained from FMO2-MP2/6-31G* calculations.



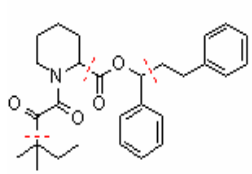
FKBP-FK506 complex (PDB:1fkf)



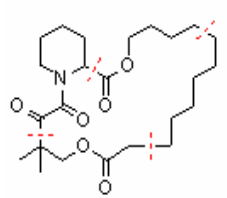
1fkb



1fkf (FK506)

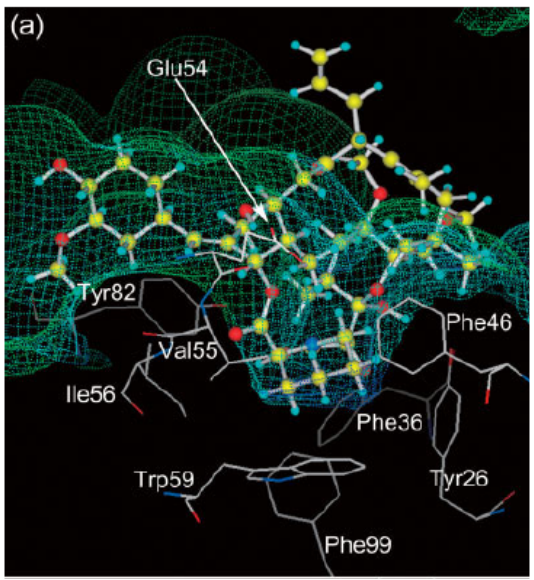


1fkg

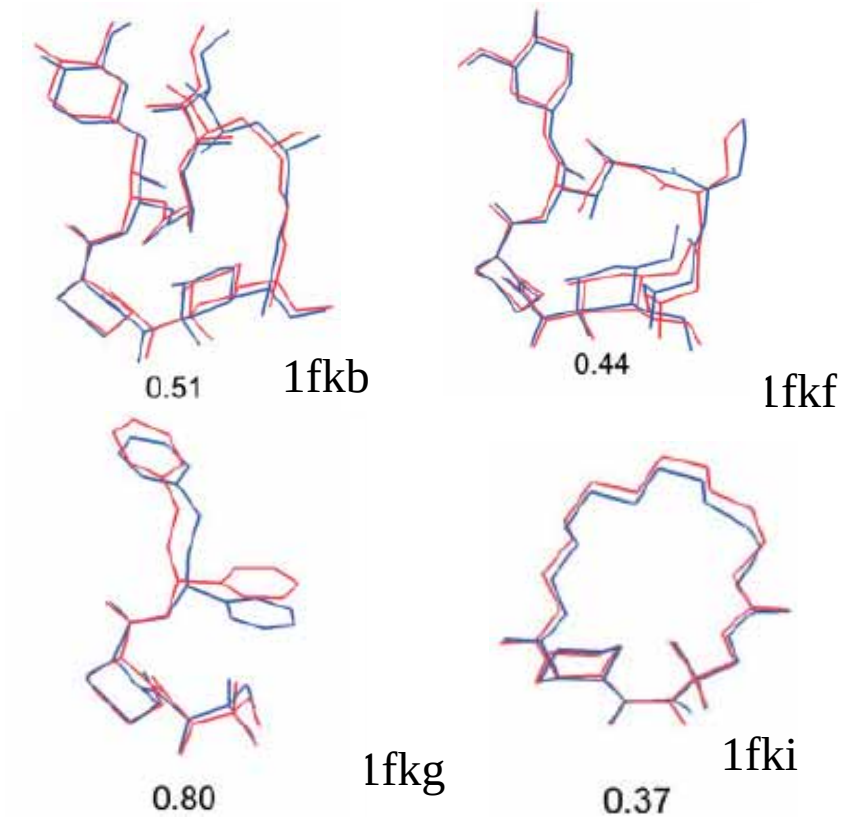


1fki

Optimized ligand geometry in the binding pocket are compared well with experimental ones, except 1fkg ligand which is distorted in crystal.



Binding pocket of FKBP



Numerical values indicate RMSD (in Å) between calculated and experimental geometries.

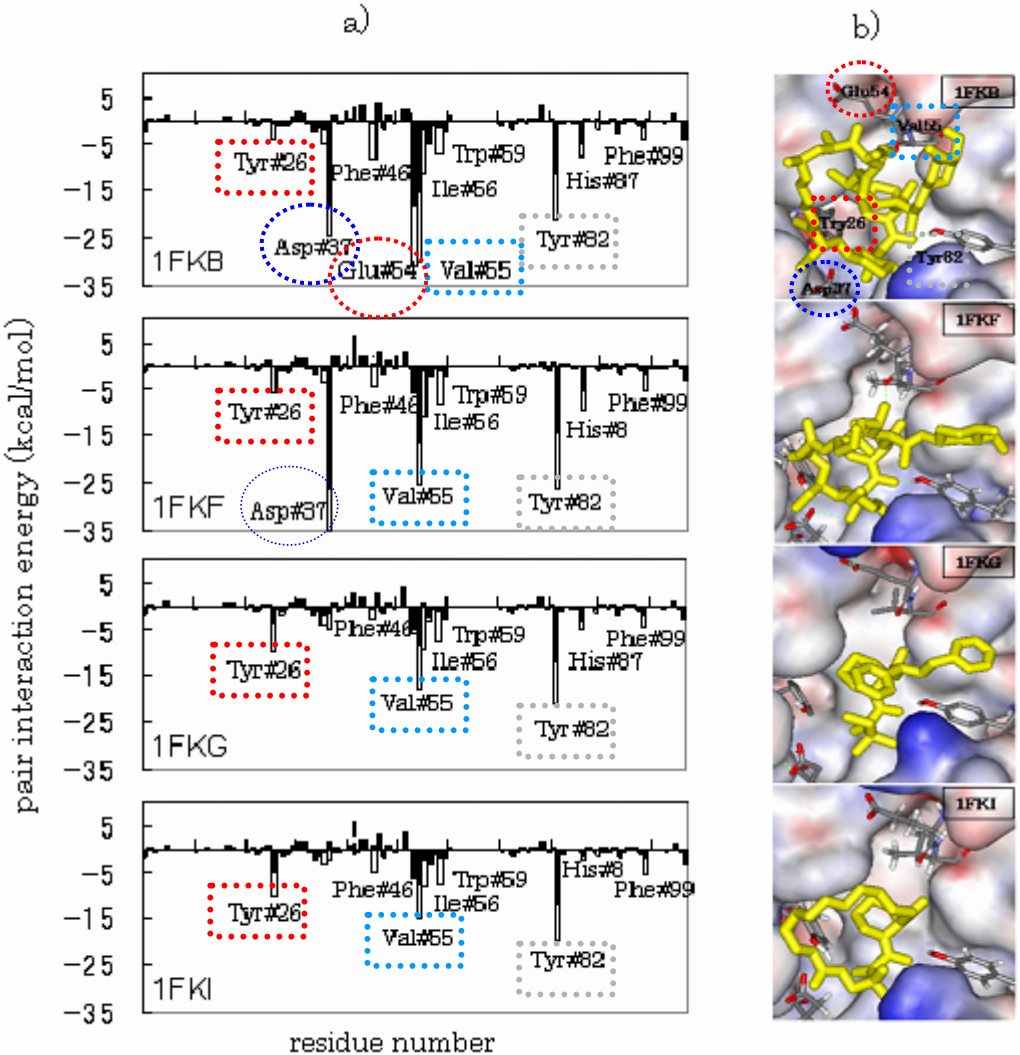
Correlation contribution to binding energy is very large: it accounts for 70-80% of binding energy

Calculated binding energy between FKBP and its ligands at FMO2-MP2/6-31G* level (kcal/mol)[a]

	ΔE_b	
System	Total	Corr[b]
1fkb	−103.9 (−101.4)	−82.0 (−80.6)
1fkf	−102.2 (−97.5)	−69.2 (−67.2)
1fkg	−70.1 (−66.4)	−57.7 (−56.7)
1fki	−71.3 (−69.5)	−55.3 (−54.1)

[a] Two residues per fragment partition is used. In the parentheses are given the energy obtained with one residue per fragment division. Note that the difference between the two is small.

Pair interactions between ligand and each residue



a) Empty bar:HF, filled bar: correlation energy contribution. b)Ligand binding modes. The proteins is shown by surface model.

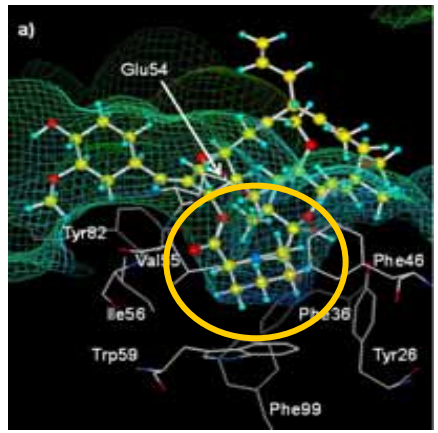
- The sum of pair interaction energies correlates well with the experimental binding affinity, 1fkb>1fkf>1fkb>1fki.
- Val55, Tyr82 and Try26 are common important residues for the all ligand bindings.
- The stronger binders have additional interactions with Asp37 (1fkb and 1fkf) and Glu54(fkb).
- The correlation contribution is very large: 70-80% of binding energy.

Pair Interactions between Important Ligand Fragment and Surrounding Residues

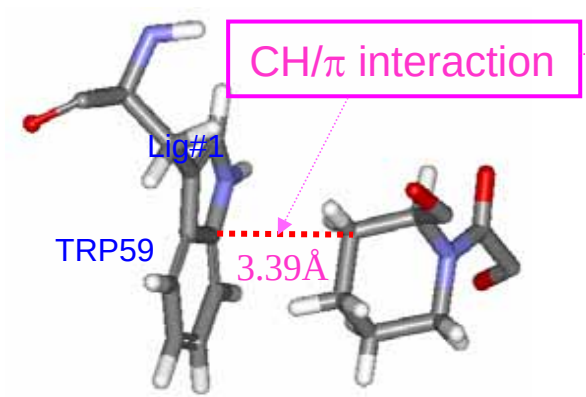
$$\Delta\tilde{E}_{IJ} = \sum_{I>J}\left\{\left(E'_{IJ} - E'_I - E'_J\right) + Tr\left(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}\right)\right\}$$

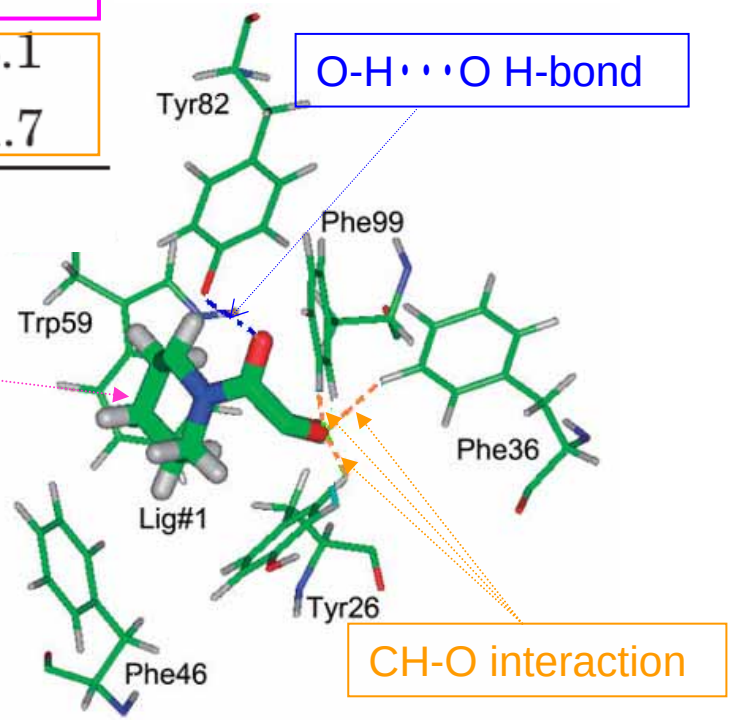
Fragment pair		$\Delta\tilde{E}^{C}_{IJ}$	$\Delta\tilde{E}^{C}_{IJ}$ (RHF)	Corr (%)[a]
Tyr#82	Lig#1	−20.1	−15.6	22.3
Trp#59	Lig#1	−7.4	−0.6	91.7
Tyr#26	Lig#1	−6.9	−3.6	48.1
Phe#99	Lig#1	−4.9	−1.4	71.7

[a]The percentage of correlation energy in $\Delta\tilde{E}^{C}_{IJ}$.



Binding pocket





List of FMO-based Methods

FMO2-RHF : original FMO (2-body expansion)

Kitaura et al., *Chem.Phys.Lett.*,313,701(1999)

FMO3-RHF : generalized to include explicit 3-body contribution

Fedorov et al., *J.Chem.Phys*., 120, 6832 (2004)

FMO2,3-DFT: FMO-based density functional theory

Fedorov et al., *Chem. Phys. Lett*., 389, 129 (2004).

FMO2,3-MP2 : FMO-based 2nd order Møller-Plesset perturbation theory

Fedorov et al., *J.Chem.Phys*., 121, 2483 (2004).

Mochizuki et al., *Chem. Phys. Lett*., 396, 473-479 (2004).

FMO2-MCSCF: FMO-based MCSCF

Fedorov et al.,*J. Chem. Phys*., 122, 54108 (2005),

FMO2,3-CC: FMO-based coupled cluster theory.

Fedorov et al., *J. Chem. Phys*., 123, 134103 (2005)

FMO-CIS : FMO-based configuration interaction singles

Mochizuki et al., *Chem. Phys. Lett*., 406, 283 (2005).

MFMO : FMO-based ONIOM like multilayer method

Fedorov et al., *J. Phys Chem. A* , 109, 2638 (2005).

FMO/PCM : Combined FMO and polarizable continuum model (PCM)

Fedorov et al., *J. Comp. Chem*, 27, 976 (2006).

FMO programs

1) FMO in **GAMESS**, coded by D.G.Fedorov

http://www.msg.ameslab.gov/GAMESS/GAMESS.html

- FMO method : FMO2, FMO3, Multilayer FMO
- Wavefunction : RHF, DFT, MP2, MCSCF, CC
- Basis function : All BFs supported in GAMESS
- Energy gradient：RHF, DFT, MP2
- Property : dipole moment, Mulliken population,
  analysis of intra- and inter-molecular interaction
- Number of atoms : no limitation
- Parallel processing : two-level parallelization (GDDI)

D.G.Fedorov *et al.*, J.Comp.Chem., **25**, 872 (2004)

2) **ABINIT-MP**, coded by T.Nakano et al.

http://www.fsis.iis.u-tokyo.ac.jp/en/result/software/

Summary

- Single point MP2/6-31* calculations of several thousand atomic systems becomes routine with FMO, while CC is still too heavy to be applied to proteins.
- For real life biochemical studies, many improvements are needed, such as use of larger basis set, better correlation theories and so on.
- The most important issue is to allow molecular dynamic simulations on solvated proteins (FMO-MD) within practical computational time. The next generation peta-flops computers might realize the dream.

Acknowledgments

Collaborators;

Dr. Dmitri G. Fedorov (RICS, AIST) … FMO-based correlation methods

Prof. Isao Nakanishi (Kyoto Univ.) … FKBP

Grants;

AIST Super Cluster Project (AIST,Japan)

Next Generation Peta-flops Supercomputer Project (MEXT, Japan)

CREST (JST, Japan)

Grant-in-Aid for Scientific Research (JSPS, Japan)

Asteras Pharmaceutical Company Co. Ltd.