第5回 HPCI戦略プログラム合同研究交流会、平成25年1月16日（神戸）

タンパク質の大規模量子化学計算

北浦　和夫
神戸大学・システム情報学研究科

内容：

1. フラグメント分子軌道法（FMO法）の概要と創薬への応用を目指した機能の実装
2. FMO法の並列計算の仕組みと効率
3. インフルエンザタンパク質ヘマグルチニンの基質認識の解析
4. まとめ

フラグメント分子軌道（FMO）法

分子をフラグメントに分割して計算する近似的量子化学計算法

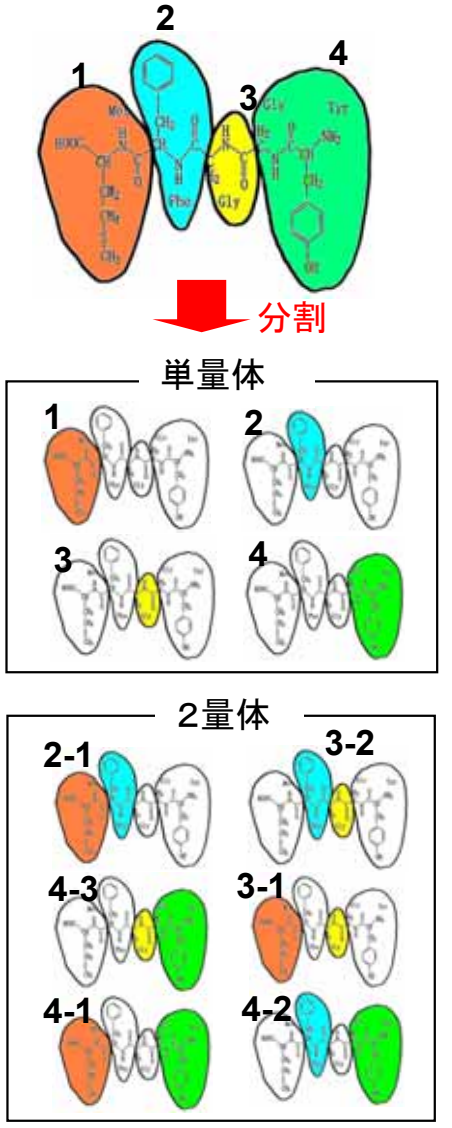
FMO法の概要:

- 分子をN個のフラグメントに分割する
- フラグメント単量体（N個）と2量体{$N(N-1)/2$個}について、他のフラグメントが及ぼす静電ポテンシャルの下で、*ab initio* MO計算を行い、それぞれの全エネルギー（E_IとE_{IJ}）を求める
- 分子の全エネルギー（E）は、E_IとE_{IJ}を用いて次式で計算する

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

FMO法の特徴:

- 数千から数万原子系の電子状態計算が可能（通常の*ab initio* MO法は数百原子から千原子）
- システムサイズのほぼ比例した計算資源ですむ
- 京コンピュータなどで大規模並列計算が容易



FMO has been applied to various systems

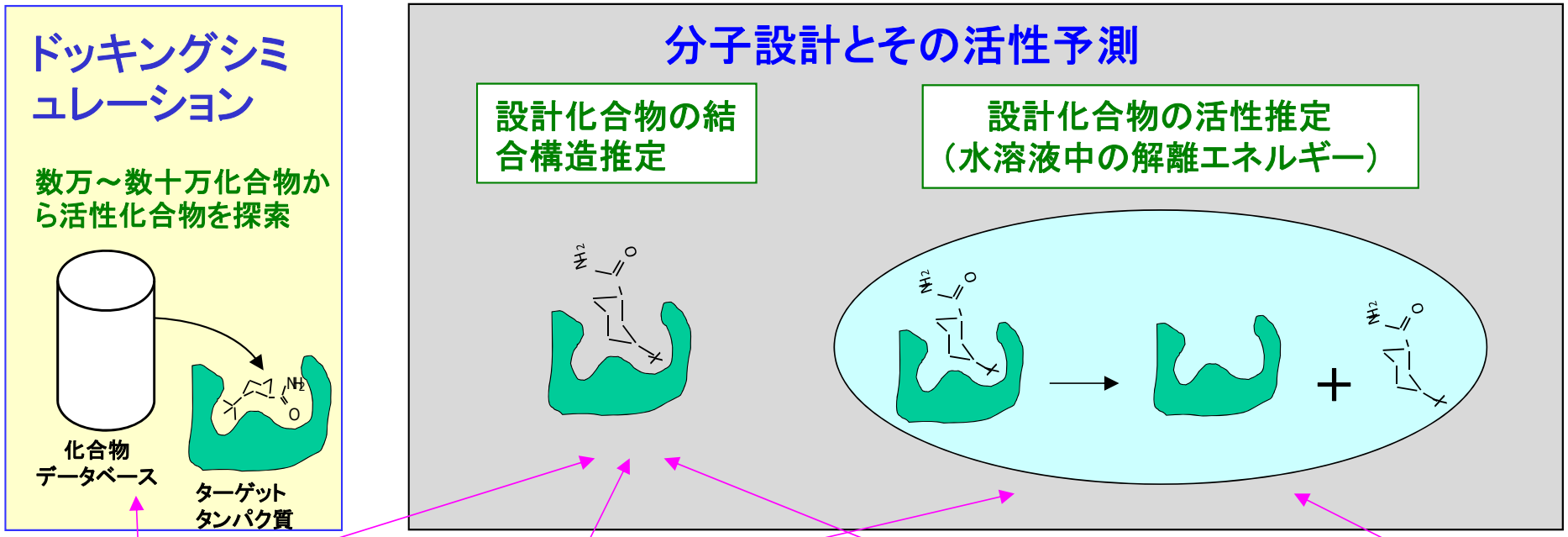
- Protein-ligand interactions (FMO-RHF, FMO-MP2)
- Geometry optimizations of small proteins (FMO-RHF)
- Relative stabilities of polypeptide conformations in solution (FMO/PCM)
- Analysis of transition state stabilization of enzymatic reactions (FMO-MP2, FMO-DFT)
- NMR chemical shift of small proteins (FMO-RHF level)
- Molecular dynamics simulations of chemical reactions in solvent (FMO-RHF)
- Excitation energy calculations of proteins (FMO-CIS, FMO-CIS(D), FMO-TDDFT)
- Adsorption of molecules on zeolite surface (FMO-RHF with adoptive frozen orbitals at fractioned bonds)

See references

1) Fedorov et al., JPCA, 111, 6904 (2007)
2) Fedorov et al, eds, "The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems", CRC press, 2009.

創薬への応用を目指した機能

―医薬品開発の各研究ステージで実用的な計算法―



ドッキングシミュレーションのための高速複合体構造・エネルギー計算

- Frozen domain FMO (FMO/FD)
  Fedorov et al.,JPCL, **2** (2011).

分子間相互作用の計算と解析

- 分子間相互作用の解析（FMO/PIEDA）
- 高精度分子間相互作用エネルギー計算 QM:QM融合法 (ONIOM+FMO)
  Asada et al., J. Phys. Chem. Lett.,3 (2012)

複合体の構造計算

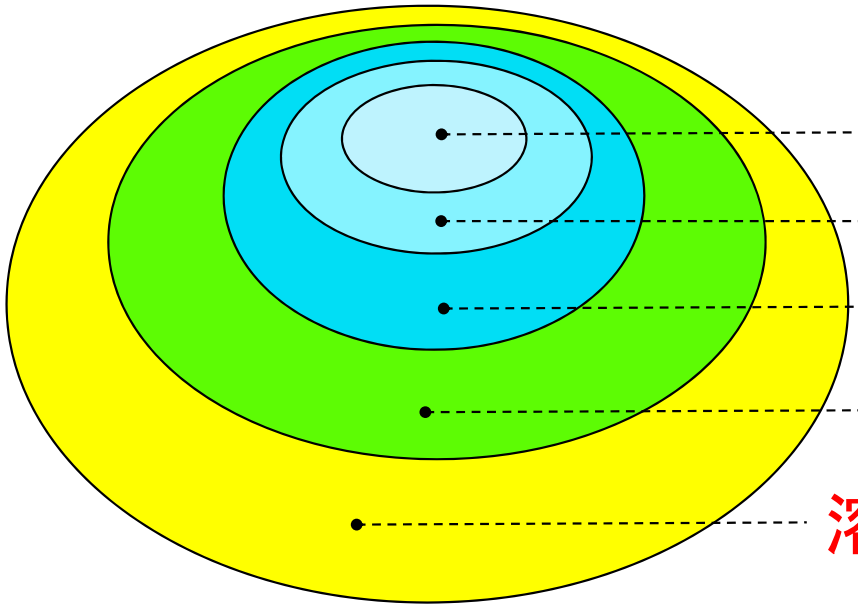
- 完全解析微分
- 量子・古典融合法 (FMO/MM)
  Asada et al., in preparation
- 水溶液中の構造計算(FMO/PCM勾配) Nagata et al., in press

溶媒和自由エネルギー

- PCMとの融合法 (FMO/PCM)
- 溶液中の結合エネルギー解析 (FMO/PCM/PIEDA) Fedorov et al., JPCA, 116 (2012).

高精度から高速計算法までを統合するスキーム

ー マルチレベル量子・マルチレベル力場融合法 ー



超高精度QM:数十原子
高精度QM：数百原子
} ONIOM
中精度QM: 数万原子
MM、EFP：数万から数百万原子
溶媒の連続体モデル：無限大

これらの融合法の解析的エネルギー勾配

必要に応じて各階層のサイズを変えたり、計算レベルの組み合わせを変えて、現実的な計算精度と計算時間を選択できる

ONIOM:M. Svensson, K. Morokuma, et. al., J. Phys. Chem. 1996, 100, 19357-19363

FMO has interfaced various wavefunctions and solvent models and implemented in GAMESS

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

FMO2,3-RHF : Kitaura et al., *CPL*,313,701(1999), Fedorov et al., *JCP*,120, 6832 (2004)

FMO2,3-DFT: Density functional theory, Fedorov et al., *CPL*,389, 129(2004).

FMO2,3-MP2 : 2nd order Møller-Plesset perturbation theory, Fedorov et al., *JCP*, 121, 2483 (2004), Mochizuki et al.,*CPL*,396, 473(2004)

FMO2-MCSCF: FMO-based MCSCF, Fedorov et al.,*JCP*,122,54108(2005).

FMO2,3-CC: Coupled cluster theory, Fedorov et al., *JCP*, 123, 134103 (2005)

FMO1-CIS and CIS(D) : Configuration interaction singles, Mochizuki et al., *CPL*, 406, 283 (2005).

FMO1,2-TDDFT : Time dependent DFT, Chiba et al., *JCP*, 127, 104108 (2007).

MFMO : FMO-based multilayer method, Fedorov et al., *JPCA*,109,2638 (2005).

FMO/PCM : Combined with polarizable continuum model (PCM), Fedorov et al., *JCC,* 27, 976 (2006).

FMO/EFP: Combined with effective fragment potential method, Nagata et al., *JCP*, **131**, 024101 (2009). Nagata et al., *JCP,* 131, 024101 (2009).

FMO/MM: Integrated FMO and MM (FMO version of IMOMM) , Asada et al., to be submitted

Parallelization Scheme for FMO in GAMESS

Generalized Distributed Data Interface (GDDI)

Fedorov et al., JCC 25, 872 (2004)

◆ DDI

◆ GDDI : Two-level hierarchical parallelization scheme

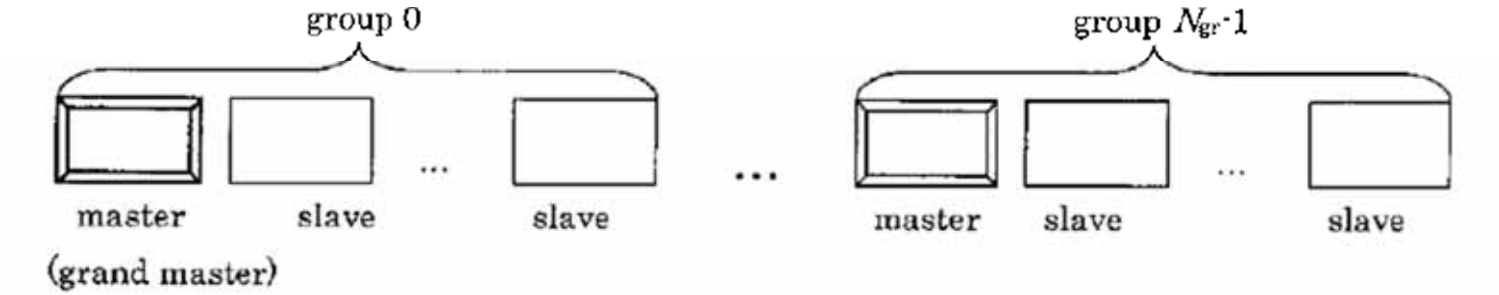
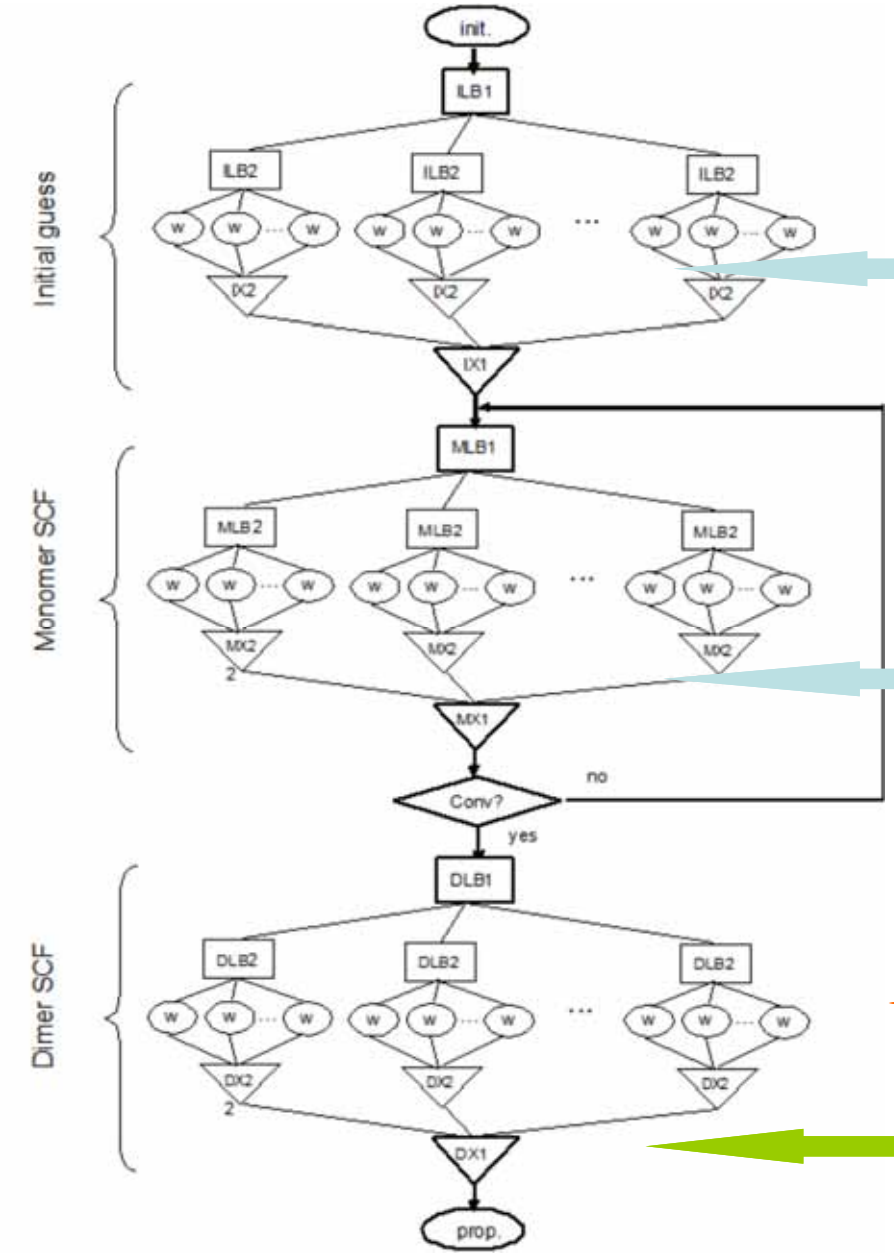


Figure 1. Comparison of notation for DDI and GDDI. Individual nodes are denoted by rectangular boxes, marked below as master and slaves.

◆ Similar two-level hierarchical parallelization scheme was developed using MPI and implemented in ABINIT-MP program.
(Sato et al., IPSJ Transaction on High Performance Computing Systems 2000, 41 No. SIG5, 104 (in Japanese)).

Parallelization of FMO in GAMESS

Flow of FMO calculations



Load balancing

Data exchange (density matrix)

Load balancing

Data exchange(density matrix)

Load balancing

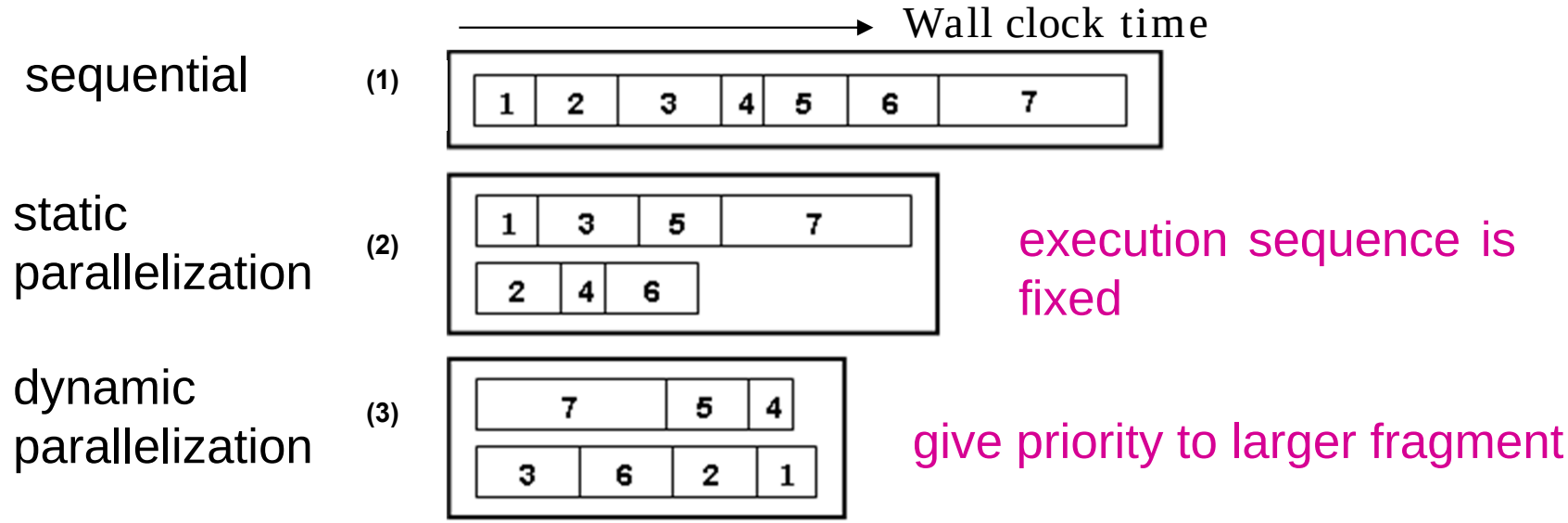
Data collection (properties)

Figure 3. The diagram for the GDDI parallelization of the FMO method. "init" stands for initialization, *Y*LB*n* is the *n*-th level load balancing of *Y* = I initial guess, *Y* = M monomer SCF, *Y* = D dimer single point runs, *n* = 1 (intergroup) and 2 (intragroup). For all *n* = 2 parts the diagram is simplified not showing loops occurring during those steps (such as, e.g., SCF iterations). "W" is some work performed [illegible] fraction of integrals). *Y*Xn is first (*n* = [illegible] data exchange. "conv." stands for the test of convergence of the monomer SCF. "prop." stand for properties, currently computed (redundantly) by all group masters. The [illegible] the total load that is divided among available nodes, whose number is arbitrary.

Dynamic Load Balancing in FMO calculations

Fedorov et al., JCC 25, 872 (2004)

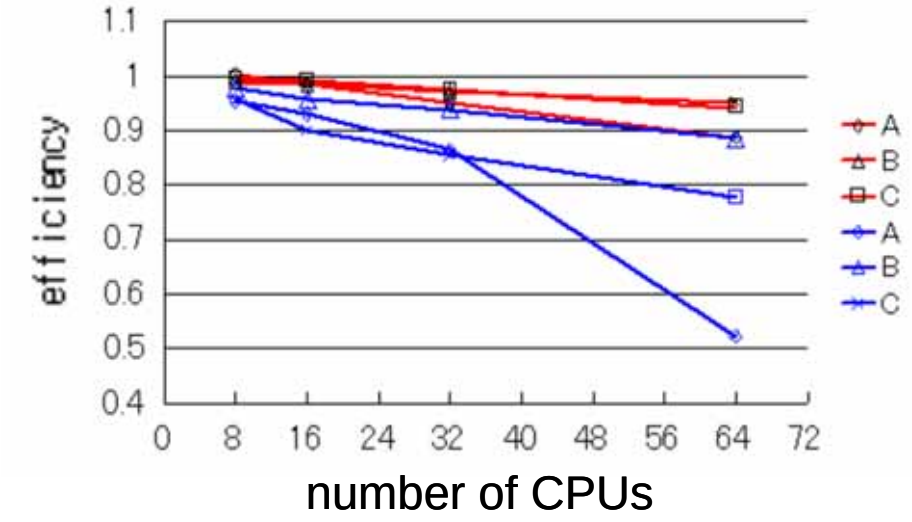
- Example: Sequential and Parallel execution on two CPU groups



execution sequence is fixed

give priority to larger fragment

- Parallel efficiency (FMO-RHF)

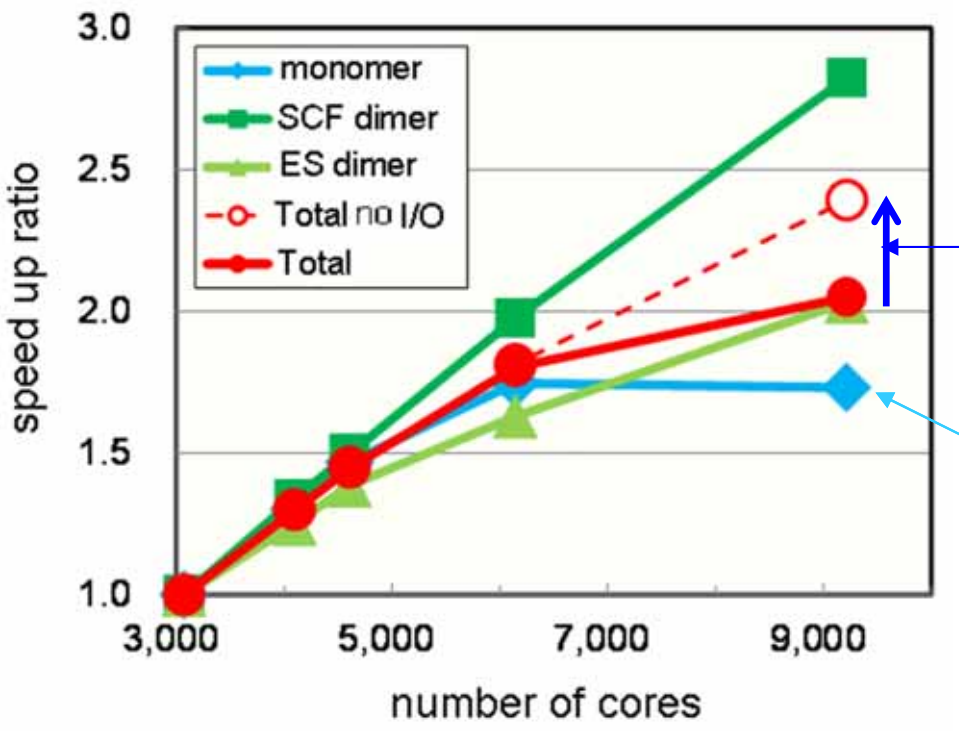


A:(H2O)256, 768atoms,HF/6-31G*,
B:(H2O)1024, 3072atoms,HF/STO-3G,
C:lysozyme,236atoms, HF/STO-3G

For all the cases, dynamic load balancing performs better.

Parallel Efficiency of GAMESS-FMO on 10,000 cores (relative to 3,000 cores)

- T2K supercomputer at Center for Computational Sciences, Tsukuba Univ.
- The cluster computer consists of 648 nodes (Opteron Barcelona 10368 CPU cores) connected fat-tree network with Infiniband (total 147 Tera Flops)
- Parallel environment: ARMCI+ mvapich1+GDDI (Umeda et al., introduced thred parallelization)
- Test system: protein trimer consisting of about 23,460 atoms (1,445 fragments)
- Computational method: FMO-RIMP2/6-31G* (RIMP2: resolution-of-identity MP2, Katohda et al., Ther.Chem.Acc.130,449(2011)



Improved by suppressing I/O

Due to load imbalance
(fragment size is different)

GAMESS-FMOの京コンピュータでの性能

筑波大・佐藤G・梅田氏を中心に、富士通・糸野氏ら、分子研・永瀬G・河東田氏ら、産総研・Fedorov氏ら、理研・南氏ら、「ナノ統合プロジェクトの中核アプリ高度化グループ・GAMESS-FMOサブグループ」のメンバーの支援の下に実施

京コンピュータでの性能（H24年6月時点）

- 768ノードを基準にした時の1,536ノードでの並列化効率が98%
- 1,536ノードを基準にした時の3,840ノードでの並列化効率92%
- 2万4千原子系のRI-MP2/6-31G*計算が12,288ノード（98,304コア）で約17分（効率4.40%）、24,576ノード（196,608コア）で約11分(3.66%)。
- ストロングスケーリングでは、これが限界？ロードインバランスの健在化。1グループ当たりが分担するフラグメント数が数個になる。また、1グループが64ノードを越えると並列効率が低下するので大きくできない。
- 石村和也氏が新規開発した高速2電子積分ルーチンを組み込んで（1グループを大きくできる）、性能改善を図る予定。

Calculation of Binding Energy Between Influenza Virus Hemagglutinin and Sialoside

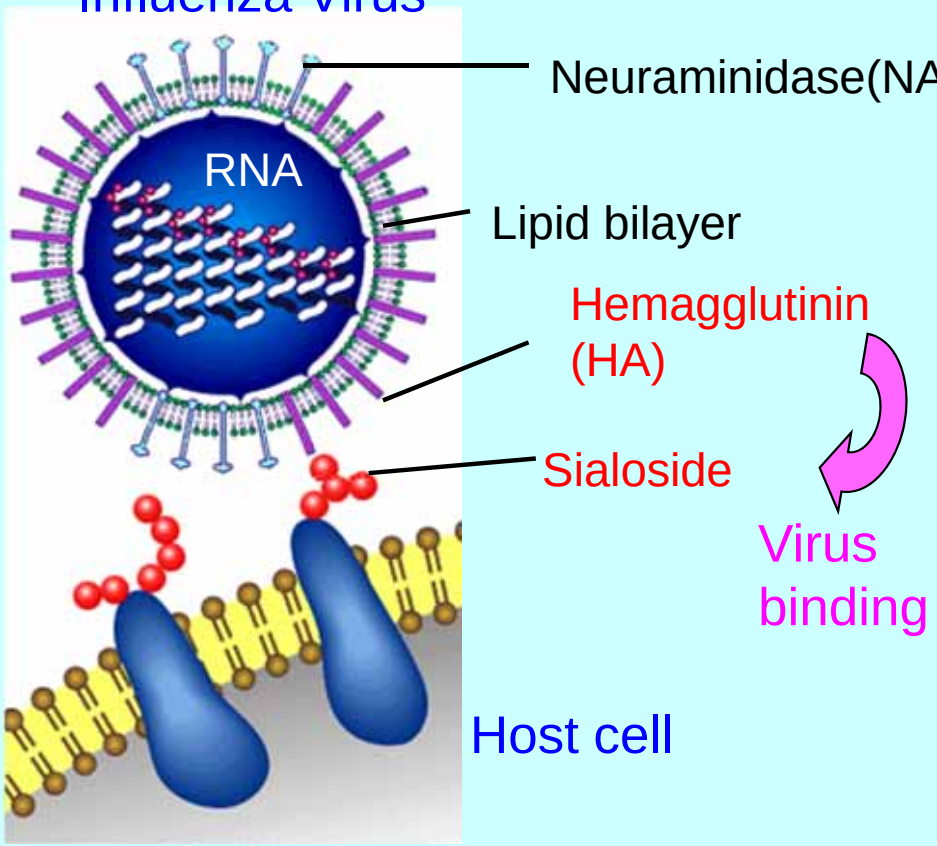
Sawada et al., JPCB, 114, 15700 (2010), JACS 132, 16862 (2010)

Motivation:

♦Influenza virus has two kinds of surface proteins, hemagglutinin (HA) and neuraminidase (NA), immersed in the lipid bilayer.

♦ For the first step of infection, HA bind to a sugar chain sialic acid on the host cell surface.

♦ Revealing the detailed binding mechanism may be helpful for the development of anti-influenza drug.



Modified from
http://www.gak.co.jp/FCCA/glycoword/GL_J.html

Calculated HA protein and model sialoside

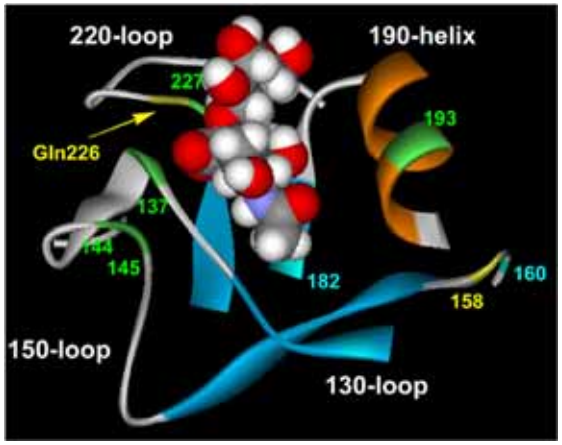
23,480 atoms

Structure of HA trimer-sialoside complex (PDB ID:1HGG)

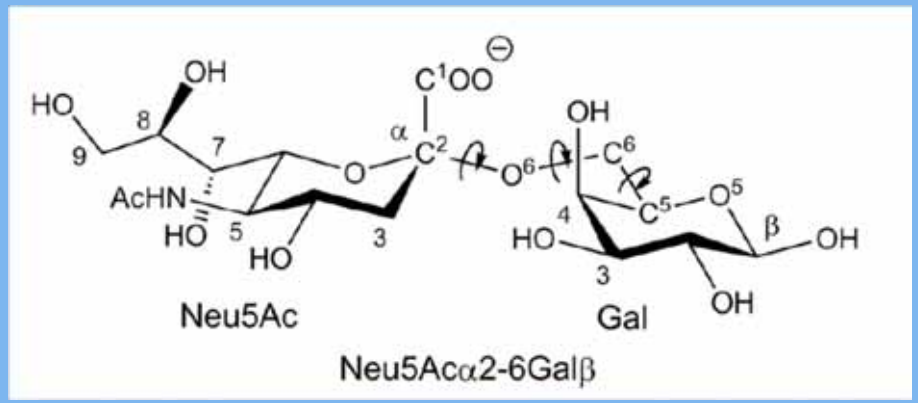
8,260 atoms

HA monomer-sialoside complex

Binding site

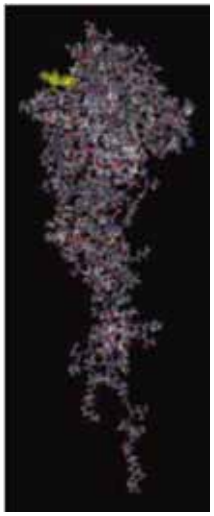


The real ligand in the complex (Neu5Acα2-3Galβ1-4Glc) was replaced with a disaccharide (Neu5Acα2-6Galβ, N-acetylneuraminic acid)



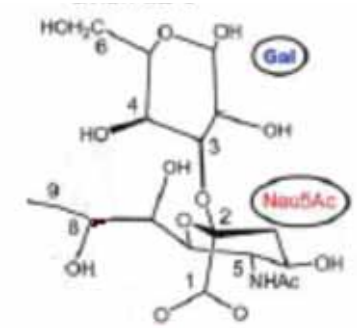
Calculated Four Model Complexes

avian HA monomer

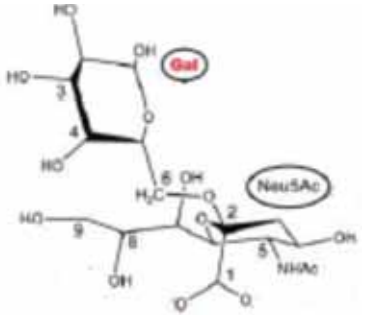


human HA monomer

avian type sialoside



human type sialoside



Structure Modeling of HA-Sialoside Complex

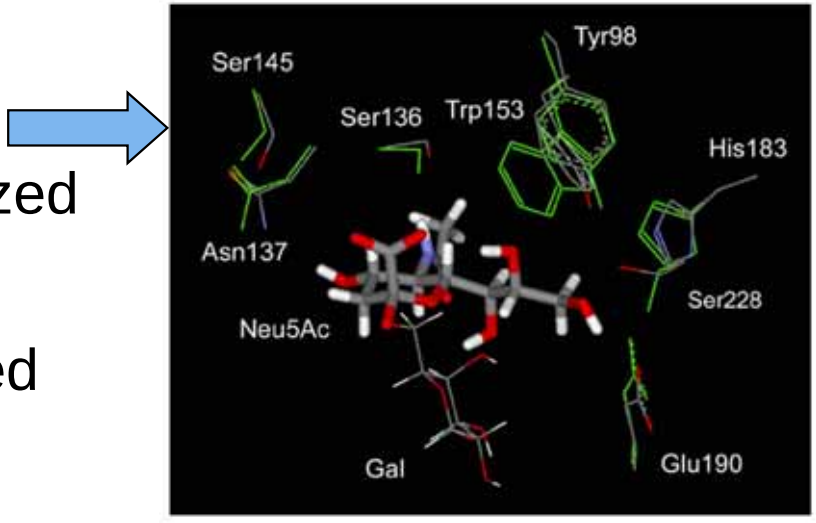
1) Starting from crystal structure (PDB:1HGG),Neu5Acα2-6Galβ was docked into the binding site of the protein.

2) The prepared complex was geometry-optimized by utilizing the class II force field and generalized Born implicit solvent model implemented in the Discovery Studio program package *(Discovery Studio 1.5.1, Accelrys*; SanDiego, CA, USA).

3) Constant NPT(296K, 1atm) MD simulation was performed for 200 ps by using NAMD2.6 with PARM99 (protein), GLYCAM6e (sialoside), and TIP3P (water molecule) parameters.

4) During the simulation, the structure deviation was small (RMSD:0.2Å), suggesting the prepared energy-minimized structure was reasonable.

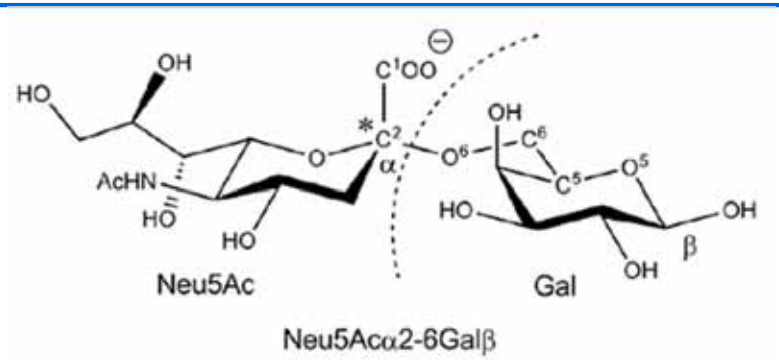
5) The modeled complex structure was used in FMO calculations.



Fragmentation of Molecules for FMO Calculations

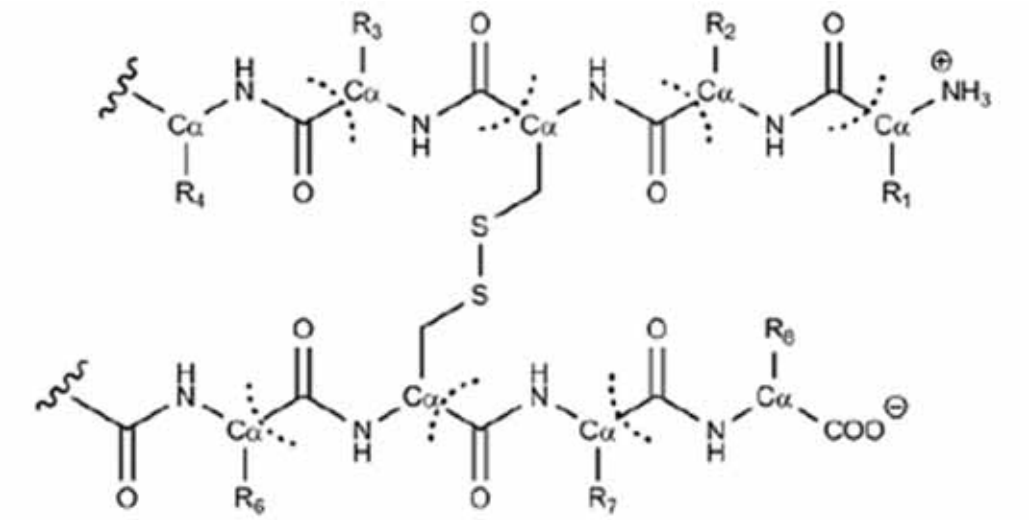
Ligand

(charge -1)



HA protein

(charge +1)



- Total number of atoms and fragments are, respectively 8,260 and 480 for HA monomer complex.
- FMO input data were generated using Facio (Suenaga, http://www1.bbiq.jp/zzzfelis/Facio_Jp.html)

溶媒効果の評価

FMO combined with Polarizable Continuum Model (FMO/PCM)

- PCM solvation free energy;

$$G_{\mathrm{PCM}} = G_{\mathrm{internal}} + G_{\mathrm{es}} + G_{\mathrm{rep}} + G_{\mathrm{disp}} + G_{\mathrm{cav}}$$

G_{internal} : internal energy change of solute

G_{es} : solute-solvent electrostatic interaction energy

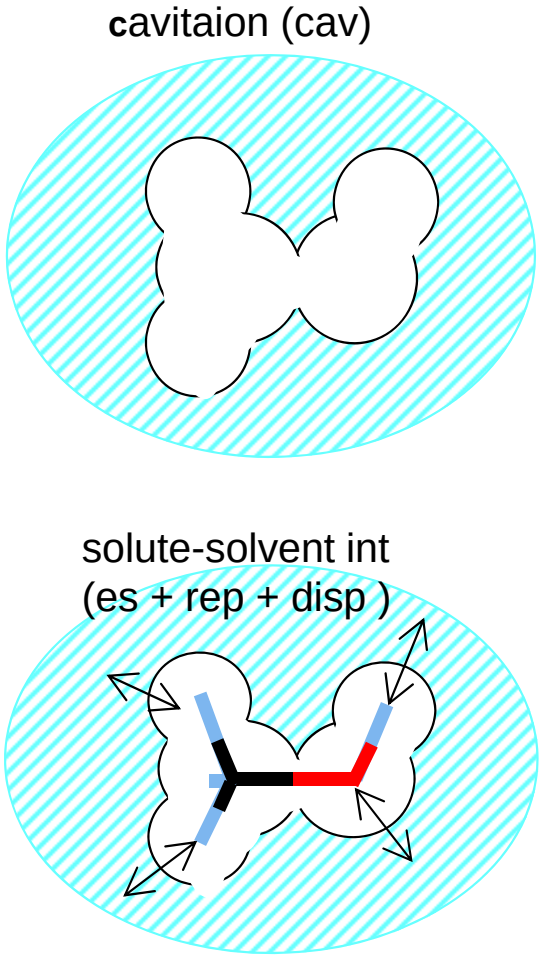
G_{rep} : solute-solvent exchange-repulsion energy

G_{disp} : solute-solvent dispersion interaction energy

G_{cav} : cavitation energy

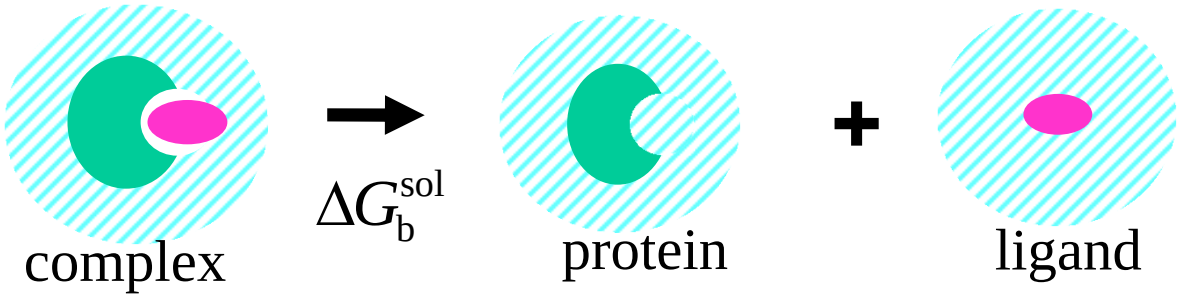
J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev*., 105 (2005) 2999.

- In FMO/PCM, internal energy change of solute G_{internal} and solute-solvent electrostatic interaction energy G_{es} is calculated using FMO many-body expansion.



Binding free energy in solution

- Binding energy of protein-ligand complex (PL) in solution



$$\Delta G_b^{sol} = G_{PL}^{sol} - G_P^{sol} - G_L^{sol}$$

$$G_X^{sol} = G_X^{internal} + G_X^{ele} + G_X^{cav} + G_X^{disp} + G_X^{rep} \quad (X = PL, P, L)$$

- 1) FMO-MP2/PCM[1(2)]/6-31G*
  2) conductor-like PCM in GAMESS
  3) cavity was created with the simplified united atomic radii, H:0.01Å, C:1.77Å, N:1.68Å, O:1.59Å, and S:2.10Å
  4) 60 tesserae per sphere.

Binging energy ΔG_{bind} and its components FMO2/PCM[1(2)]/6-31G(d) level

entry	components of ΔG_{bind}	①avian-avian avian H3/avian α2-3	②avian-human avian H3/human α2-6	③human-avian avian H3 Gln226Leu/human α2-6	④human-human human H3/human α2-6
1	ΔE_{gas}	−237.1	−226.0	−214.6	−230.3
2	ΔG_{pold}	−34.5	−28.4	−27.3	−28.2
3	$\Delta G_{internal}$	−271.6	−254.4	−241.9	−258.5
4	ΔG_{es}	210.7	199.7	191.4	200.3
5	ΔG_{cav} [a]	−5.5	−6.1	−5.1	−3.1
6	ΔG_{disp}	50.1	47.4	47.7	53.5
7	ΔG_{rep}	12.4	12.3	12.6	13.5
8	$\Delta G_{solvation}$ [b]	208.4	200.3	194.1	209.0
9	ΔG_{PCM} [c]	−28.7	−25.7	−20.5	−21.3
10	$-T\Delta S_{solute}$ [a]	28.2	31.0	20.0	17.0
11	ΔG_{bind} [d]	−0.5	5.3	−0.5	−4.3

[a] 1.0 atm, 298 K. [b] Sum of entries 2 and 4–7. [c] Sum of entries 1 and 8 (entry 8 is the sum of entries 3–7). [d] Sum of entries 9 and 10.

The order of calculated binding affinity ④＞①～③＞②
is in agreement with experiment

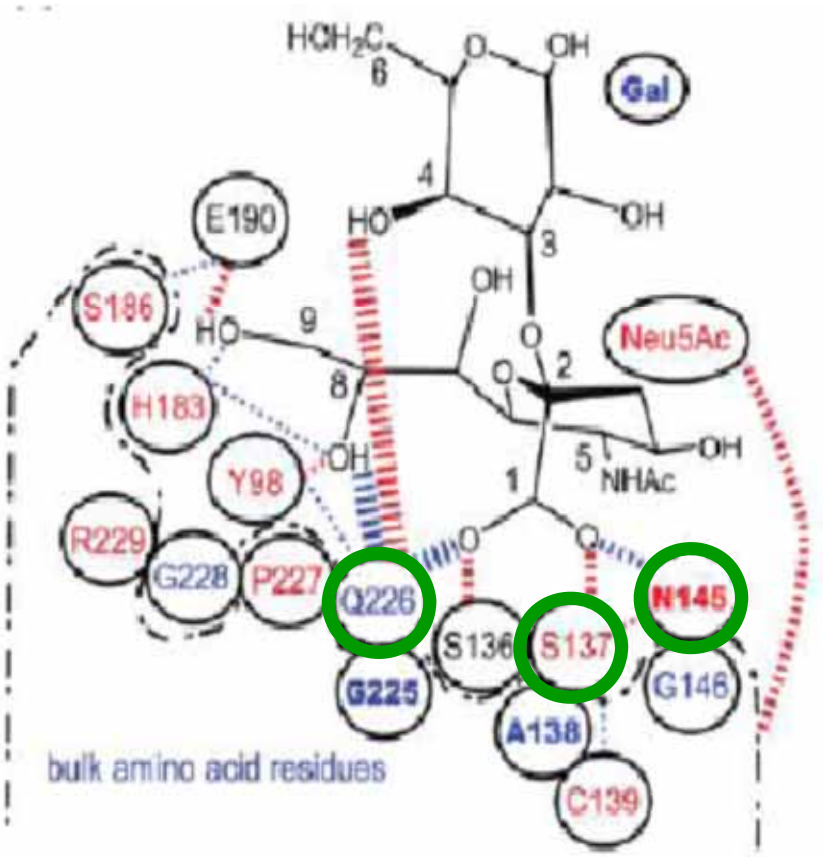
Difference between HA and sialoside interactions

favor

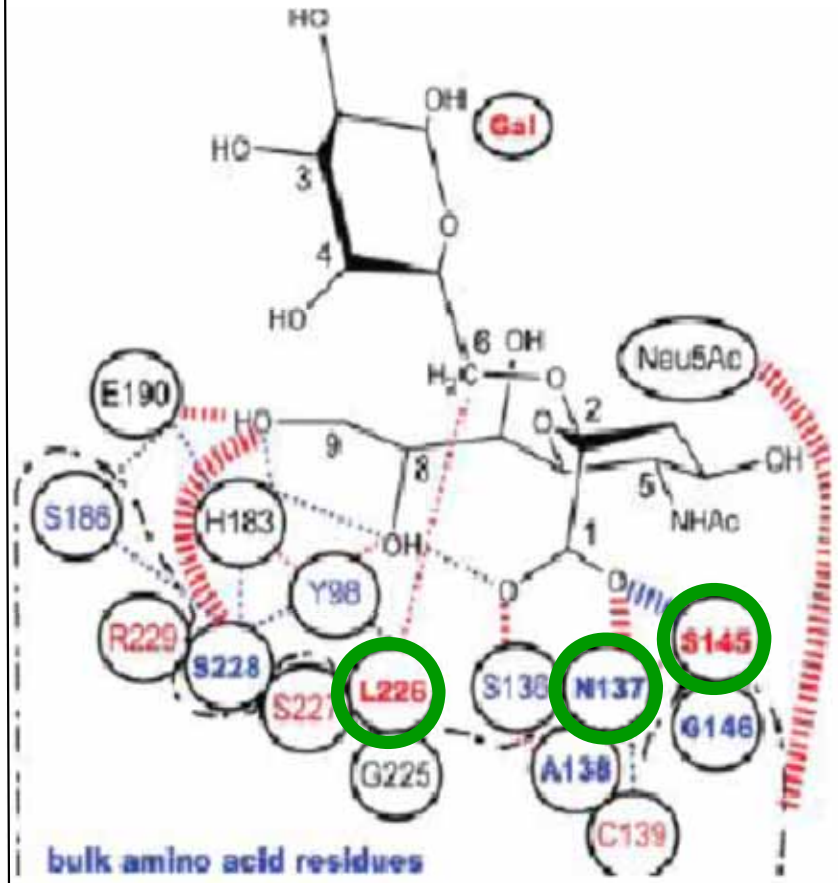
disfavor

bold: ~10kcal/mol
normal:2-10 kcal/mol
thin: < 2 kcal/mol

avian HA - avian sialoside
relative to avian HA-human sialoside



human HA - human sialoside
relative to avian HA-human sialoside



mutated residues: Q226L(Gln226Leu),S137N(Ser137Asn), N145S(Asn145Ser)

巨大・複雑なタンパク質の量子化学計算における問題点

京コンピュータにより、10万原子系の計算が数時間でできる見込み。しかし、巨大なタンパク質系の計算準備に数ヶ月を要している。このギャップを埋めないと、巨大分子計算を日常化できない

入力データ作成

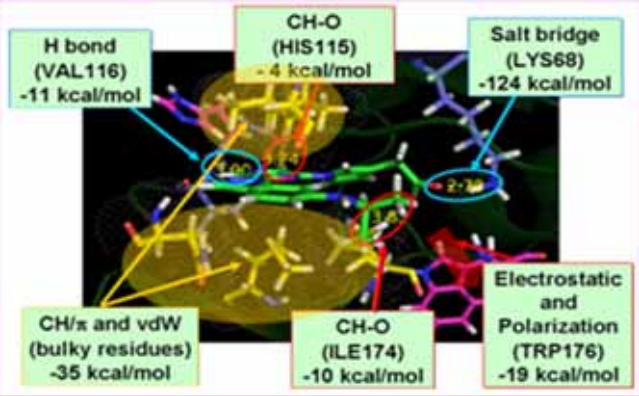
・構造モデルを作成するのが困難

計算結果の可視化

・タンパク質のリガンド結合（分子認識）を分かりやすく表現する方法が必要

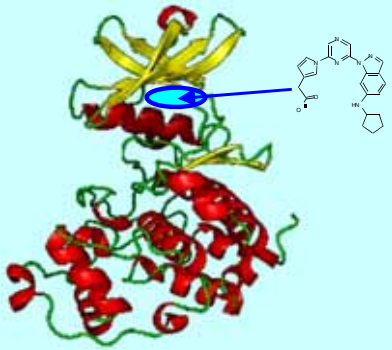
10万原子超のタンパク質の計算に対応できる、入力データ作成支援、計算結果の可視化のためのGUIプログラム（python＋OpenGL）を新たに開発中

まとめ： FMO in drug design



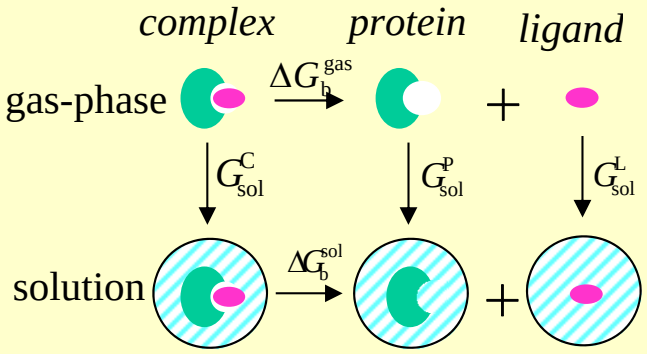
Analysis of intermolecular interactions by FMO/PIE, PIEDA

- hydrogen-bond, CH/π, CH/O, halogen/π, π/π and other weak non-bonding interactions



Geometry optimization of protein-ligand complexes by FMO/MM

- refinement of low resolution experimental structure and predicted docking structure



Solvent effects on binding energy by FMO/PCM

- Binding energy in solution and solute-solvent interaction energy

謝　辞

共同研究者

Dmitri G.Fedorov博士（産総研)

故・永田武史博士 (神戸大, 産総研)

澤田敏彦博士（産総研、現・富士フィルム

京コンピュータでのプログラムの高速化・チューニング

梅田宏明博士（筑波大）

糸野雄一郎氏（富士通）

サポート

HPCIプロジェクト、戦略分野2（CMSI）（文科省、東京大）

Three approximation levels in FMO/PCM

Fedorov et al., *JCC,* 27, 976 (2006).

♦ Solute electrostatic field V on cavity surface is calculated from many-body series expansion;

$$V_i = \sum_{I=1}^{N} V_i^I + \sum_{I>J}^{N} \left(V_i^{IJ} - V_i^I - V_i^J \right) + \text{L}$$

$$V_i^x = -Tr\left(\mathbf{D}^x \cdot \mathbf{w}^i\right) + \sum_{\alpha \in X} \frac{Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{R}_i|}$$

$$w_{\mu\nu}^i = \left\langle \mu \left| \frac{1}{|\mathbf{r} - \mathbf{R}_i|} \right| \nu \right\rangle, \quad \mu\nu \in x$$

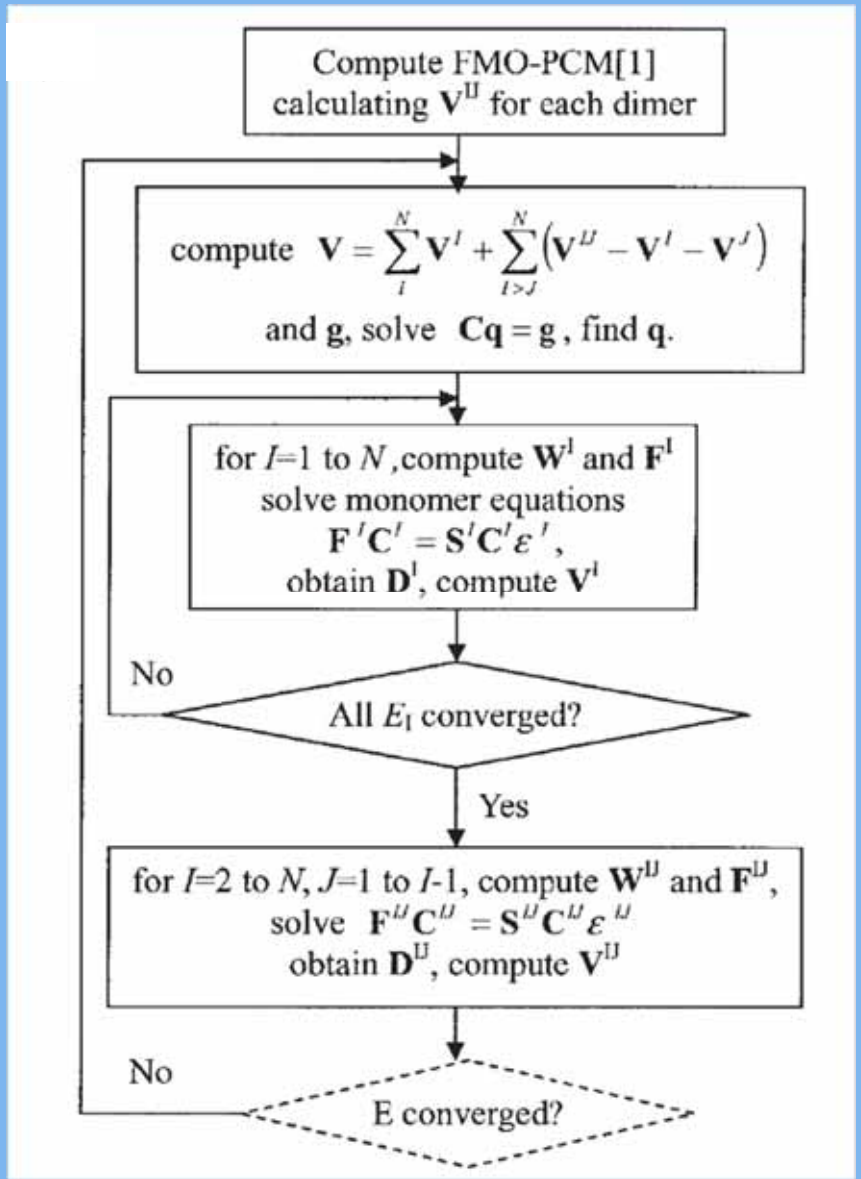
♦Several approximation levels are possible depending on V and electron density (D^x) expansion:

(1) FMO/PCM[1] uses one-body V and one-body density

(2) FMO[2] uses two-body V and two-body density

(3) FMO[1(2)] uses two-body V and two-body density (without self-consistent)

FMO/PCM[1(2)] was used in this work



flowchart of FMO/PCM[2]