地球シミュレータ産業利用シンポジウム2010

スーパーコンピュータでインフルエンザウイルス の変異の仕組みを探る

田中 成典

神戸大学大学院システム情報学研究科 計算科学専攻 計算生物学講座

インシリコ・バイオロジー(In Silico Biology)

- コンピュータの中で生命現象を「再構築」する。
- さらに、生命現象を外部からコントロールしたりデザインしたりする。
 - バイオインフォマティクス
 - バイオシミュレーション
 - システム生物学
- 我々のグループでは主に、原子・分子レベルから(ボトムアップで)生命現象を記述する「バイオシミュレーション」を行っています。

生体系の第一原理ボトムアップ・シミュレーション

物理学の基礎方程式(シュレディンガー方程式、ニュートン方程式)に基づき

- できるだけ恣意性なしに
- 分子・原子・電子のレベルから
- 生体系のシミュレーションを行う



General concept and architecture of the erythrocyte model.

Pathways for de novo GSH synthesis and export system of GSSG.

生体高分子の第一原理計算の必要性

- ポストゲノム時代の「計算構造生物学」
 - 生命現象の(インシリコ)ボトムアップ的アプローチ
 - 原子・分子レベルからの基本的理解
 - 分子に「語らせる」
 - 生化学反応を扱う上での電子状態・ダイナミクスの重要性
 - ▶ 医療・創薬・環境問題への応用
 - 疾患の分子メカニズム、テーラーメイド医療(SNPs)
 - 特に、生活習慣病(糖尿病、肥満、動脈硬化等)、感染症など
 - Structure-Based Drug Design, Virtual Screening
 - 化学物質の生体への影響の定量的モデリング
 - 酵素反応の分子設計エンジニアリング

タンパク質の振舞いは電子が決める

- タンパク質は生命を司る重要な構成要素
- タンパク質は通常数千個以上の原子からなる巨大分子
- 原子の間に働く力は「糊」の役割をしている電子が決めている(従来は電子を考えない近似的な取り扱い)
- 医薬品分子はタンパク質と相互作用(阻害、活性化)する(精確な 記述が必要)



電子の振舞いを記述する量子力学



Erwin Schrödinger (1887-1961) 1933年ノーベル物理学賞

シュレディンガー方程式(1926年) $H\Psi = E\Psi$ 原子核1個、電子1個の水素原子を紙と鉛筆で 解くことから始まった!

タンパク質は1万個以上の電子を含む

80年間の科学技術(物理学、化学、コンピュータ 科学、etc.)の進歩により、タンパク質の電子の 振舞いも解けるようになった!

1999年:フラグメント分子軌道(FMO)法

巨大な高分子を多数のフラグメント(アミノ酸など)に 分割して、それぞれのフラグメントを並列的に計算する → (超)並列コンピュータの活用

タンパク質の計算から薬の設計へ

■ 医薬品候補分子とタンパク質の結合を計算

- 高精度・高速のドッキング予測
- 生理条件下(水溶液中)のシミュレーション
- 分子間相互作用の定量的解析
- 阻害・活性化などの機能予測
- 創薬の経験則(QSARなど)との組み合わせ
- 薬物代謝(CYP)のシミュレーション



「次世代インシリコ創薬」へ

コンピュータの高性能化が後押し 「ムーアの法則」:1.5年で2倍の性能

ペタ級(10万コア超)並列計算







FMO計算によって次世代スパコンの能力をフル活用



量子力学の基礎:シュレディンガー方程式



(シュレディンガー)

 $H\Psi = E\Psi$

$$H = -\sum_{A}^{N} \frac{\nabla_{A}^{2}}{2M} - \sum_{i}^{n} \frac{\nabla_{i}^{2}}{2} - \sum_{A,i}^{N} \frac{Z_{A}}{r_{Ai}} + \sum_{i>j}^{N} \frac{1}{r_{ij}} + \sum_{A,B}^{N} \frac{Z_{A}Z_{B}}{R_{AB}}$$

 $\Psi \equiv \Psi(r_1, r_2, \cdots, r_n; R_1, R_2, \cdots, R_N)$ 電子を波動関数として記述する

ハートリー・フォック近似:軌道と平均場

$$\Psi^{HF} = \hat{A} \{ \varphi_1 \overline{\varphi_1} \cdots \varphi_M \overline{\varphi_M} \}$$

$$\varphi_i = \sum_p^N c_{pi} \chi_p \qquad f \varphi_i = \varepsilon_i \varphi_i$$

$$f = h + \sum_j (2J_j - K_j)$$

$$(N - hy -) \qquad (7\pi y f)$$

・電子の波動関数を1個の電子配置で近似 (ハートリー・フォック) ・原子の<mark>軌道</mark>から分子の<mark>軌道</mark>を線形結合で表現 ⇒ 係数を決定 ・電子間の相互作用は平均的な場として考慮される

電子相関(電子間相互排斥効果)の考慮

$$\Psi^{Post-HF} = \sum_{I} T_{I} \Phi_{I} \quad \{\Psi^{HF}, \Psi^{a}_{i}, \Psi^{ab}_{ij} \cdots\} \in \Phi_{I}$$
$$E^{Corr} = E^{Post-HF} - E^{HF}$$

・平均場近似からの差を取込む(電子相関)
 ・HF基底配置に励起配置を加えて展開する
 ・メラー・プレセットの2次摂動が便利(MP2)



(メラー)





・・・・さらに多電子の励起

MP2では2電子励起のみ考慮

電子数の壁

- 電子相関効果を入れた高精度計算では、計算時間は 電子数(N)の5乗(N⁵)程度で増加する。
- もし、電子10個の系の計算に1秒かかるとすると、電子 1000個の系の計算には、317年(!)かかる。
- ■この壁を克服するには →「困難を分割する」
- 1000個の電子を10個ずつ100組に分けて、それぞれが ほぼ独立に計算できるようにすれば、理想的には100 秒程度で計算が終了する。
- これを計算精度を落とさずに実現するには?
- ■「フラグメント分子軌道法」による(超)並列計算

Fragment molecular orbital (FMO) method



Fragment Molecular Orbital Method

Divide a molecule into fragments
N pieces of fragments (monomers)
[N(N-1)/2] pieces of fragment pairs
(dimers)
FMO Total Energy (FMO2)

$$E = \sum_{I} E_{I} + \sum_{I < J} (E_{IJ} - E_{I} - E_{J})$$

Environmental
 $H_{I} \Psi_{I} = E_{I} \Psi_{I}$
 $H_{IJ} \Psi_{IJ} = E_{IJ} \Psi_{IJ}$
 $H_{IJ} \Psi_{IJ} = E_{IJ} \Psi_{IJ}$
 $H_{IJ} = \sum_{i \in I, J} \left\{ \left(-\frac{1}{2} \nabla_{i}^{2} \right) + \sum_{A} \left(-\frac{Z_{A}}{|\mathbf{r}_{i} - \mathbf{A}|} \right) + \sum_{i \neq J, J} \frac{\rho_{A}(\mathbf{r}')}{|\mathbf{r}_{i} - \mathbf{r}'|} d\mathbf{r}' \right\} + \sum_{i \in I, J} \sum_{i > j \in I, J} \frac{1}{|\mathbf{r}_{i} - \mathbf{r}_{j}|}$
Environmental

electrostatic potential

FMO Method and Its Energy Analysis (IFIE)



Divide a molecule into fragments

N pieces of fragments
[N(N-1)/2] pieces of fragment pairs

Total Energy: Calculated from energies of fragment and fragment pair

$$\mathbf{E} = \sum_{\mathbf{I} > \mathbf{J}} \mathbf{E}_{\mathbf{I}\mathbf{J}} - (\mathbf{N} - 2)\sum_{\mathbf{I}} \mathbf{E}_{\mathbf{I}}$$

Inter-Fragment Interaction Energy (IFIE):

$$\Delta E_{IJ} = (E'_{IJ} - E'_{I} - E'_{J}) + (V_{IJ} - V_{I(J)} - V_{J(I)})$$

- E_X : energies of a fragment and a fragment pair.
- V_X : Electrostatic potential from surrounding fragments.

$$E'_X = E_X - V_X$$

FMO計算の組合せ並列化(on ABINIT-MPX)









なぜ、電子相関効果が重要か?

DNA Base Pair Stacking

塩基スタッキング:分散力の重要性



HF or DFT cannot describe the weak attraction due to van der Waals (dispersion) interaction appropriately.

As for FMO-MP2 description for vdW interaction, see, e.g., K. Fukuzawa *et al.*, J. Comput. Chem. <u>27</u> (2006) 948.

Electron Correlations

Electron correlations play important roles for the descriptions of weak molecular interactions associated with hydrogen bonding and van der Waals (dispersion) forces.

- Watson-Crick pair
- Ligand binding
- DNA base stacking



Implementation of MP2 into FMO (ABINIT-MPX)

インフルエンザ ~迫る脅威~

間断なき新型の出現とパンデミック(大流行)の脅威

1918年: スペイン風邪(H1N1亜型)の世界的流行 6億人が感染, 4000-5000万人に上る死者 1957年: アジア風邪(H2N2亜型)の出現・流行 1968年: 香港風邪(H3N2亜型)の出現・流行 それぞれ死者100万人

2005年: H5N1亜型鳥インフルエンザウイルスのトリからヒト, さらに, ヒトからヒトへの感染が確認

2008年:1月現在,新型ウイルスでの死者100名超

2009年6月: WHOが新型ウイルス(ブタ由来H1N1亜型)によるパンデミックを宣言

2010年5月9日現在、214の感染国と地域で18,036人以上の 死者を確認(8月10日、WHOがパンデミック終結を宣言)

A型インフルエンザウイルス 亜型と分布 (感染特異性)



喜田宏:細胞工学 19,27,2000

インフルエンザウイルスの構造



HA:ウイルスが吸着する過程で重要 NA:増殖したウイルスが脱出する過程で重要

- ウイルスは膜に覆われた球状の構造
- ・ 膜表面からは、三量体タンパク質へマグル チニン(HA)や四量体タンパク質ノイラミニ ダーゼ(NA)がスパイク状に突き出している
- HAはウイルスの細胞侵入時に宿主細胞表 面のシアル酸を認識し、細胞膜に結合する
- HAは抗原として中和抗体に認識される (抗原抗体反応)



インフルエンザウイルスHAとシアル酸の結合





HA-糖	[FMO-MP2/6-310			
受容体	НЗトリНА	H1EFHA	H1ブタHA	Н5トリНА
トリ	-352.9	-293.3	-363.3	-299.2
۲۲	-292.4	-335.9	-390.5	-283.9

トリH3HAと糖鎖受容体の相互作用解析

フラグメント間相互作用エネルギー(IFIE): $\Delta E_{IJ} = (E'_{IJ} - E'_{I} - E'_{J}) + Tr(\Delta \mathbf{P}^{IJ} \mathbf{V}^{IJ})$ ▶ 全てのフラグメントペアに対して自動的に算出 ▶ 複雑な生体高分子の相互作用をフラグメント単位で解析できる ▶ 自由に組み合わせることが可能(IFIE sum)



Avian : -242.6 kcal/mol Human : -158.1 kcal/mol H3 avian HAのReceptor結合特異性:

2. Avian receptor とHuman receptorでは TYR98, SER137, ALA138, HIS183, GLU190, GLN226との相互作用に違い



ヒトH1HAと糖鎖受容体の相互作用解析





 $(\angle$ IFIE = IFIEavian-IFIEhuman)

- 1. IFIE_{total}はHumanが大きい Avian : -175.3 kcal/mol Human : <u>-243.1 kcal/mol</u>
- H1 human HAのReceptor結合特異性: Human>Avian
- 2. Avian receptor とHuman receptor LYS222, ASP225との相互作用に 違い



シアル酸認識の宿主結合特異性の理論予測と重要なアミノ酸の特定が可能

インフルエンザ 抗原(HA)抗体複合体

- PDB ID:1EO8
- H3N2 A/Aichi/2/68
- インフルエンザHA(HA1およびHA2)と, BH151抗体のV領域の軽鎖・重鎖のタンパク 複合体*
- HA1上には、大きく五つの抗原領域と呼ばれる、抗体が抗原を特異的に認識して結合する部位が存在することが実験的に確認されている。
- 抗原領域上ではHAの変異が特に頻繁に起 こっている.
- *D.Fleury et al., PROTEINS: Structure, Function, and Genetics <u>40</u>, 572-578 (2000).
- 計算はFMO-MP2/6-31G, 6-31G*で行った.

1EO8とその抗原領域



インフルエンザウイルスの抗原-抗体のFMO-MP2計算

・インフルエンザ ヘマグルチニン(HA) タンパク質とFabフラグメントの複合体 ⇒ 抗原:HA1(緑)、HA2(赤) 抗体:H鎖(ピンク)、L鎖(黄) 世界最大のFMO-MP2/6-31G計算 ⇒ 921残基、14,086原子 78.390基底 •PCクラスタ(Opteron 2GHz, 16プロセッサ) でのベンチマークタイミング ⇒ 198.9時間(8.3日) ・地球シミュレータ 512ノード(4.096プロセッ) サ)でのベンチマークタイミング(NES) ⇒ 53.4分! ・MP2-IFIE(フラグメント間相互作用)によって 抗原-抗体反応の様相を詳細に解析 ⇒ ウイルスの変異予測や薬の開発



抗原領域ごとのIFIE Sumの値

1EO8のHA1上に存在す	各抗原領域ごとの IFIE Sum			
る、各抗原領域ごとに	抗原や	左記の領域に	IFIE	
IFIE sumの値を求めた。	ドメインの	該当する	Sum	
E領域とそれに挟まれた	領域名	残基番号	[kcal/mol	
間の部分の残基の			Ĵ	
IFIE sumが顕著に大	А	121–147	51.7	
きい.	В	155-198	-30.6	
	С	53-57,275-278	50.8	
	D	200-215	63.6	
E領域とその間の残	E	62-65,78-85	-345.7	
基は抗体から特異的	E+Eの間	62-85	-593.8	
認識(引力的相互作 用)を受ける!	HA1 Total	1-327	-258.9	
(以下62-85を「E近傍」と定義する)	HA2 Total	1–175	-223.6	

IFIE Sumと抗原領域の対照





 変異体のウイルスの生存には、以下の制約が伴う と考えられる、(「有利」な変異)



第二条件:抗体に認識されにくい性質をもつ

- 変異の発生についての理論的仮定と予測手法
- 第一条件, 第二条件をともに満たしたものが変異体ウ イルスとして出現しやすいのではないか.
- 変異体の出現は自然淘汰の結果で、抗体との相互作用を減弱させる(引力を弱める/反発を強める)方向に進みやすいのではないか。
- 本研究では、第一条件をランダムーアミノ酸置換実験によって、第二条件をFMO計算により検討した.

変異予測のスキーム



アミノ酸変異の許容・非許容

- ーアミノ酸変異体作成実験*
- *K. Nakajima et al., J. Virol. 77 (2003) 10088.

HA上の約200のアミノ酸部位にランダムに一アミノ酸変異を導入



変異導入部位と自然界での主要な変異部位を対照⇒26部位が一致 一致した26部位のうち24部位が許容変異.

⇒ウイルスの変異体は,許容部位のアミノ酸が変異したものである. ⇒**変異体は第一条件をほぼ満たしていることが確認された**.

抗原領域E近傍での変異部位の変遷

変異年代と置換実験結果



抗原領域E近傍では、1968年のH3型インフルエン ザウイルスの出現から、1997年までの間に主として 左図の残基が順に変異し、その部位が変異した系 統が「生存」している.

なお、左図で緑に塗りつぶしたものは,許容部位で あることが実験で示されている.

本研究では抗原領域Eおよびその間に存 在する24の残基について,抗体との引力・ 斥力の指標としてIFIE sumの値を使用し, 過去の変異の理論的説明を試みた.

Comparison to historical events



(K. Takematsu *et al.*, J. Phys. Chem. B 113 (2009) 4991.)

地球シミュレータを用いた大規模FMO-MP3計算



HA3量体-Fab抗体(1KEN;2351残基)

NA-タミフル複合体(2HU4;386残基)

HA3量体:128ノード(計1024プロセッサ)を用いてFMO-MP3/6-31Gが5.8時間 NA単量体:64ノード(計512プロセッサ)を用いてFMO-MP3/6-31Gが1.1時間

Hybrid Parallelization with MPI-OpenMP

- Inter-fragment: MPI; Intra-fragment: OpenMP
 - \Rightarrow Hybrid parallelization



- Memory saving by sharing MP3 arrays among threads
 - \Rightarrow OpenMP is promising for acceleration on multi-core chips



Computation time on ES2

SYSTE	M LEVEL	NODES	TIME(h)	(MP3/2)	NODE RAT.	TFLOPS
HA1	FMO-MP2	64	1.7			0.97
	FMO-MP3	64	2.7	(x1.6)		2.27
	FMO-MP2*	64	4.4			1.19
	FMO-MP3*	64	8.7	(x2.0)		3.02
	FMO-MP2	128	0.8		2.1	2.06
	FMO-MP3	128	1.3	(x1.6)	2.1	4.67
HA3	FMO-MP2	64	9.4			0.83
	FMO-MP3	64	11.9	(x1.3)		1.66
	FMO-MP2	128	4.3		2.2	1.83
	FMO-MP3	128	5.8	(x1.3)	2.1	3.44
NA	FMO-MP3	64	1.0			3.04
	FMO-MP3*	64	4.4			3.09

• 64 nodes = 512VPUs, 6-31G or 6-31G* basis set, Cys-Cys = 1 fragment

• HA1 (14086 atoms, 921 residues, 78390 AOs for 6-31G, 121314 AOs for 6-31G*)

• HA3 (36160 atoms, 2351 residues, 201276 AOs for 6-32G)

• NA (5792 atoms, 386 residues, 32549 AOs for 6-31G, 50447 AOs for 6-31G*)



(-30~30kcal/molで色付)

左図: 赤色のアミノ酸残基 ⇒抗体と引力的な相互作用

青色のアミノ酸残基 ⇒抗体と反発的な相互作用

赤色のアミノ酸残基が突然変異を起こせば抗体圧から逃れることができる ⇒ウイルス変異の予測、ワクチン開発へ

中•右図:

単量体間の相互作用解析 ⇒三量体としての機能の理解に役立つ

HA三量体複合系ドメイン間相互作用



黄色部分 vs. 各アミノ酸のIFIE(フラグメント間相互作用エネルギー)

IFIE sum(kcal/mol)

相互作用ドメイン	HF	MP2	MP3
Fab(I):HA(I)	-288.8	-367.0	-352.8
Fab(I):HA(II)	177.5	155.6	158.7
Fab(I):HA(III)	134.3	134.2	134.3
Fab(II):HA(I)	137.0	137.0	137.0
Fab(II):HA(II)	-292.7	-380.4	-363.7
Fab(II):HA(III)	170.8	157.0	159.5

相互作用ドメイン	HF	MP2	MP3
HA(I):HA(II)	-1022.4	-1280.3	-1237.1
HA(II):HA(III)	-981.7	-1245.7	-1200.6
HA(I):HA(III)	-1189.0	-1469.8	-1421.3
Fab(I):Fab(II)	210.8	197.8	199.5
HA全体:Fab全体	38.1	-163.6	-127.0

インフルエンザウイルス表面タンパク質のFMO計算

FMO計算による相互作用エネルギー解析は、インフルエンザHAおよびNAに 関わる特異的分子認識の解明に役立つ

(1) 宿主受容体のシアル酸認識の種特異性および認識に重要なアミノ酸の特定
 ヒト型ウイルスH1: ヒトα 2-6結合 > トリα 2-3結合 Lys222, Asp225
 トリ型ウイルスH3: ヒトα 2-6結合 < トリα 2-3結合 Glu190, Glu226, etc
 ブタ型ウイルスH1: ヒトα 2-6結合 > トリα 2-3結合

(2) 抗原変異予測への展開

FMO計算と血球吸着反応実験を組み合わせることで、ウイルス変化の歴史と 理論予測値が一致する結果が得られた

これらの成果を踏まえ、国立感染症研と共同で新型インフルエンザのワクチン開発の ための計算を実行中

(3) HA3量体-Fab抗体およびNA-タミフル複合体の大規模MP3計算

タンパク質の機能ドメインの丸ごと計算が可能となり、今後様々な生命現象のメカニズム解明に寄与すると期待できる

生命科学における量子化学計算が実用化の領域に!

ペタスケール並列シミュレーション技術による次世代創薬手法の創成



謝辞(共同研究者)

- 望月祐志(立教大学)
- 福澤薫(みずほ情報総研)
- 中野達也(国立医薬品食品衛生研究所)
- 山下勝美(NECソフト)
- 中島捷久(名古屋市立大学)
- 信澤枝里(国立感染症研究所)
- 沖山佳生(東京大学)
- 渡邉博文(OpenEye Japan)
- 岩田達則(神戸大学)
- 竹松和友(神戸大学)
- 吉岡彬生(神戸大学)