第一原理からのタンパク質の立体構造 予測シミュレーション法の開発

岡本祐幸(Yuko OKAMOTO) 分子研・理論研究系、総合研究大学院大学 (Department of Theoretical Studies Institute for Molecular Science) (The Graduate University for Advanced Studies) e-mail: okamotoy@ims.ac.jp URL: http://konf2.ims.ac.jp/

プロジェクト・リーダー (Project Leader)
 岡本祐幸 (OKAMOTO, Yuko) [分子研・総研大]

・ プロジェクトのコアメンバー

(Core Members of the Project)

平田文男 (HIRATA, Fumio) [分子研・総研大] 青柳睦 (AOYAGI, Mutsumi) [分子研・総研大]

• A 計算手法班 (Simulation Algorithm Group)

岡本祐幸 (OKAMOTO, Yuko) [分子研・総研大、班長] 中沢隆 (NAKAZAWA, Takashi) [奈良女大・理] 升屋正人 (MASUYA, Masato) [鹿児島大・工] 杉田有治 (SUGITA, Yuji) [分子研] 西川武志 (NISHIKAWA, Takeshi) [分子研] 光武亜代理 (MITSUTAKE, Ayori) [分子研]

• B 溶媒班 (Solvation Theory Group)

平田文男 (HIRATA, Fumio) [分子研・総研大、班長] 木下正弘 (KINOSHITA, Masahiro) [京大・エネルギー理工研] 入佐正幸 (IRISA, Masayuki) [九州工大] 高橋卓也 (TAKAHASHI, Takuya) [生理研] 佐藤啓文 (SATO, Hirofumi) [分子研] KOVALENKO, Andriy [分子研]

• C 電子状態班 (Electronic Structure Group)

青柳睦 (AOYAGI, Mutsumi) [分子研・総研大、班長] 南部伸孝 (NANBU, Shinkoh) [分子研] 伊藤正勝 (ITO, Masakatsu) [分子研]

• 秘書 (Secretary)

粂美和子(KUME, Miwako)

技術的サポート (Technical Support)

日立製作所 (HITACHI, Co. Ltd.)

その他の共同研究者 (Other Collaborators)

Ulrich H. E. HANSMANN [Michigan Technological Univ.] Jose ONUCHIC [U.C. San Diego] 三上益弘 (MIKAMI, Masuhiro) [物質研] 川合光 (KAWAI, Hikaru) [京大・理] 菊地武司 (KIKUCHI, Takeshi) [倉敷芸術科学大・産業科学技術] 福来正孝 (FUKUGITA, Masataka) [東大・宇宙線研]

研究目的

• 第一原理からの蛋白質の立体構造予測

(アミノ酸配列の情報のみを使い、ランダムな 初期構造からシミュレーションを始めて、自然 の立体構造を予測する。)

基礎研究の興味

- 蛋白質の折り置み原理 機構の解明
 (Protein Folding Problem)
- 蛋白質の機能発現原理の解明

応用への発展性

- ・ 新薬品の設計
- 特定の機能を持った人工蛋白質の設計
- ・ 蛋白質の誤った折り置みに起因する病理の 研究 (例えば、狂牛病やアルツハイマー病)



IBMが1億ドル規模の世界最速スパコン開発計画を発表

グランド・チャレンジに挑むペタ・フロップス級の「Blue Gene」

1999年12月6日、IBMは世界最速のスーパーコンピューターを開発する1億ドル規模の研究プロ ジェクトを発表しました。現在の最高速スーパーコンピューター^(*1)の500倍の演算処理能力を目 指します。



開発するスーパーコンピューターは「Blue Gene(ブルー・ジーン)」と呼ばれ(Geneは遺 伝子の意味)、1PFLOPS(ペタ・フロップス) 以上の演算を実行することが可能になりま す。1PFLOPSとは毎秒1,000兆回の浮動小 数点演算を実行できる性能で、1TFLOPS(テ ラ・フロップス)の1,000倍にあたります。

これは、1997年にチェスの世界チャンピオンを下した<u>「Deep Blue」</u>の1,000倍の演算処理能 カ、もしくは、デスクトップPCの約200万倍の処理能力です。

超並列コンピューター「Blue Gene」は、たんぱく質の構造解析などで使用されることになりま す。IBMでは、この生物学上の基礎的研究を、「Deep Blue」の開発以来の「グランド・チャレン ジ」として位置づけています。たんぱく質の折りたたみ構造の解析は、医療分野における病気やそ の治療法の解明につながると期待されています。

AMINO ACID



R = SIDE CHAIN







PROTEIN













 $\alpha\text{-HELIX}$ and $\beta\text{-SHEET}$

TPI



CELL



タンパク質は細胞内の リボゾームで合成される。

Anfinsen's Experiment (1960's)



Anfinsen's Experiment (1960's)



変性剤を添加

Anfinsen's Experiment (1960's)



変性剤を添加

変性剤を除去

ANFINSEN'S DOGMA

・蛋白質分子の立体構造はそのアミノ酸配列の情報で決まっている(自由エネルギーの最小状態)。

よって、 系の正しいエネルギー関数が与えられれば、 後は、計算機シミュレーションで蛋白質分子 の立体構造が予測できるはず。

やるべきことその1

厳密なエネルギー関数を用いる 蛋白質自身の構造エネルギー (静電相互作用項 + Lennard-Jones 項 +

御電伯互下用項 + Lennard-Jones 項 + 水素結合項 + ねじれエネルギー項, etc.)

+

溶媒との相互作用エネルギー

静電相互作用項等に関しては、将来的には 量子化学計算を導入する必要

やるべきことその1

特に溶媒効果が難しい

- レベル1 距離に依存した誘電率 ε(r) (大雑把だが、計算時間は気相の場合とほぼ同じ)
- レベル2 溶媒接触表面積に比例する項の和 (精度は大分まし。計算時間は気相の場合よ) 長いが、数倍程度)
- レベル3 溶媒分子を陽に取り入れる 厳密だが、計算時間が非常に長い)

 *液体の統計力学に基づく理論等

 Reference Interaction Site Model (RISM)

 Scaled-Particle Theory + Poisson-Boltzmann
 *水分子をシミュレーションに直接取り入れる

やるべきことその2

 シミュレーションがエネルギー極小状態に留まらない 強力な手法を用いる

次の2つの手法を適用することを提唱してきた

- * 徐冷法 (Simulated Annealing) 結晶を作るプロセスを利用
- * 拡張アンサンブル法 (Generalized-Ensemble Algorithms)

エネルギー空間上のランダムウォークを実現

[例えば、マルチカノニカル法、焼き戻し法 (Simulated Tempering)、

レプリカ交換法 (Replica-Exchange Method, also referred to as

Parallel Tempering)、1/k Sampling、Tsallis 統計, etc.]

利点

唯一回のシミュレーションの結果から、最小エネルギー状態ばかりでなく、

幅広い温度領域における熱力学量を計算できる





Met-Enkephalin: Global Minimum Structure in Gas Phase





MC Sweeps

— Canonical 600K



MC Sweeps
— Canonical 300K



— Canonical 50K



徐冷法 (Simulated Annealing)

S. Kirkpatrick, C. Gelatt, Jr. & M. Vecchi, Science 220, 671 (1983).

Reproduce a Crystal-Making Process on a Computer



徐冷法の導入

• H. Kawai, T. Kikuchi & Y.O., *Protein Eng.* **3**, 85 (1989).

See also:

- S. Wilson, W. Cui, J. Moskovitz & K. Schmidt, *Tetrahedron Lett.* 29, 4373 (1988).
- C. Wilson & S. Doniach, *Proteins* **6**, 193 (1989).
- A. Brunger, *J. Mol. Biol.* **203**, 803 (1988).
- M. Nilges, G. Clore & A. Gronenborn, *FEBS Lett.* **229**, 317 (1988).

For a review see:

• Y.O., Recent Res. Devel. In Pure & Applied Chem. 2, 1 (1998).





Multicanonical Algorithm



MULTICANONICAL ALGORITHM

B. Berg & T. Neuhaus, *Phys. Lett.* B267, 249 (1991).B. Berg & T. Neuhaus, *Phys. Rev. Lett.* 68, 9 (1992).

Step 1: Iterations of Short Preliminary Runs to Determine the Multicanonical Weight Factor Wmu (E)

Step 2: One Long Production Run

Step 3: Analyze the Data to Obtain:

* Global-Minimum Energy Conformation

* Thermodynamic Quantities for Desired Temperatures (by Ferrenberg-Swendsen Histogram Reweighting Techniques)

マルチカノニカル法の導入

MC Version:

• U. Hansmann & Y.O., *J. Comp. Chem.* **14**, 1333 (1993).

See also:

M. Hao & H. Scheraga, *J. Phys. Chem.* 98, 4940 (1994). (ENTROPIC SAMPLING)

MD Version:

- U. Hansmann, Y.O. & F. Eisenmenger, *Chem. Phys. Lett.* 259, 321 (1996).
 See also:
- N. Nakajima, H. Nakamura & A. Kidera, *J. Phys. Chem.* **101**, 817 (1997).
- C. Bartels & M. Karplus, *J. Phys. Chem. B* **102**, 865 (1998). (ADAPTIVE UMBRELLA SAMPLING)

For reviews see:

- Y.O., Recent Res. Devel. in Pure & Applied Chem. 2, 1 (1998).
- U. Hansmann & Y.O., *Curr. Opin. Struct. Biol.* **9**, 177 (1999).
- U. Hansmann & Y.O., Ann. Rev. Comp. Phys. VI, D. Stauffer (ed) (1999) 129.





(a)



Т





I

C

他の拡張アンサンブル法 (Generalized-Ensemble Algorithm)の導入

A 1/k SAMPLING: B. Hesselbo & R. Stinchcombe, *Phys. Rev. Lett.* **74**, 2151 (1995) Uniform distribution in entropy

- U. Hansmann & Y.O., *J. Comp. Chem.* **18**, 920 (1997).
- B EXPANDED-ENSEMBLE METHOD: A. Lyubartsev, A. Martinovski, S. Shevkunov & P. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).

Also referred to as:

SIMULATED TEMPERING: E. Marinari & G. Parisi, *Europhys. Lett.* **19**, 451 (1992).

Uniform distribution in temperature

• U. Hansmann & Y.O., *J. Comp. Chem.* **18**, 920 (1997).

See also:

• A. Irback & F. Potthast, J. Chem. Phys. 103, 10298 (1995).

他の拡張アンサンブル法 (Generalized-Ensemble Algorithm)の導入

C TSALLIS STATISTICS: U. Hansmann & Y.O., *Phys. Rev. E* 56, 2228 (1997). See also:

- I. Andricioaei & J. Straub, *Phys. Rev. E* **53**, R3055 (1996).
- U. Hansmann, *Physica A* **242**, 250 (1997).

D REPLICA-EXCHANGE METHOD: K. Hukushima & K. Nemoto, J. Phys. Soc. Jpn. 65, 1604 (1996); K. Hukushima, H. Takayama & K. Nemoto, Int. J. Mod. Phys. C 7, 337 (1996).

See also:

- R. Swendsen & J. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986). (REPLICA MC)
- C. Geyer, in *Comupting Science and Statistics: Proc. 23rd Symp.* (1991) 156.
- M. Tesi, E. van Rensburg, E. Orlandini & S. Whittington, *J. Stat. Phys.* 82, 155 (1996). (MULTIPLE MARKOV CHAIN METHOD)
- E. Marinari, G. Parisi & J. Ruiz-Lorenzo, in *Spin Glasses and Random Fields* A. Young (ed) (1998) 59. (PARALLEL TEMPERING)

他の拡張アンサンブル法 (Generalized-Ensemble Algorithm)の導入

D REPLICA-EXCHANGE METHOD (REPLICA MC, MULTIPLE MARKOV CHAIN METHOD, PARALLEL TEMPERING, etc.)

MC Version:

- U. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- A. Irback & E. Sandelin, *J. Chem. Phys.* **110**, 12256 (1999).
- M. Falcioni & M. Deem, *J. Chem. Phys.* **111**, 1754 (1999).
- Q. Yan & J. de Pablo, *J. Chem. Phys.* **111**, 9509 (1999).

MD Version:

- U. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- Y. Sugita & Y.O., *Chem. Phys. Lett.* **314**, 141 (1999).

レプリカ交換法(手法)

- 1. Generalized Ensemble
 - 互いに相互作用しない異なる温度を持つオリジナルシス テムのコピー (レプリカ)をM個用意する。
- 2. Replica-Exchange
 - 各レプリカの温度を詳細釣り合いの原理を満たす遷移確率に基づき交換。
 - ・各温度のカノニカル分布を保証。
- 3. Canonical Ensemble
 - 各温度のデータのみを集める。





並列計算機では、各レプリカの計算は各プロセッサで行う また、通信量も少ないため高い並列化率が実現可能。

レプリカ交換分子動力学法

Y. Sugita & Y.O., Chem. Phys. Lett. 314, 141 (1999).

1. Probability Distribution of the Generalized Ensemble

$$W_{REM}({X}) = \exp\left\{-\sum_{m=1}^{M} \beta_m H(q^{i(m)}, p^{i(m)})\right\}, \text{ where } H(q, p) = K(p) + E(q)$$

2. Exchange between replica $i(\beta_m)$ and $j(\beta_n)$

$$\begin{cases} X:1,2,L,i(\beta_m),j(\beta_n),L,M\\ X':1,2,L,i(\beta_n),j(\beta_m),L,M \end{cases}$$

3. Detailed Balance Condition

$$W_{REM}({X})w(X \to X') = W_{REM}({X'})w(X' \to X)$$

4. Velocity Scaling

$$p^{i(n)} \equiv \sqrt{\frac{\beta_m}{\beta_n}} p^{i(m)}, p^{j(m)} \equiv \sqrt{\frac{\beta_n}{\beta_m}} p^{j(n)}$$

5. Transition Probability

$$w(X \to X') = \begin{cases} 1 & \text{for } \Delta < 0 \\ \exp(-\Delta) & \text{for } \Delta > 0 \end{cases} \text{ where } \Delta = (\beta_n - \beta_m) \left(E(q^{(i)}) - E(q^{(j)}) \right)$$

Met-Enkephalin in Gas Phase

Y. Sugita & Y.O., Chem. Phys. Lett. 314, 141 (1999).

Time Series of Replica-Exchange and Potential Energy



Fixed Temperature



Met-Enkephalin in Gas Phase

Y. Sugita & Y.O., Chem. Phys. Lett. 314, 141 (1999).

Acceptance Ratio

 $200K \Leftrightarrow 239K$ 0.160 $239K \Leftrightarrow 286K$ 0.149 $286K \Leftrightarrow 342K$ 0.143 $342K \Leftrightarrow 409K$ 0.139 $409K \Leftrightarrow 489K$ 0.142 $489K \Leftrightarrow 585K$ 0.146 $585K \Leftrightarrow 700K$ 0.146



Met-Enkephalin in Gas Phase

Y. Sugita & Y.O., Chem. Phys. Lett. 314, 141 (1999).

Distributions of (ϕ , ϕ) of Gly-2 at T = 200 K



Canonical Simulation

Replica-Exchange Simulation

REMD-7

Alanine-Dimer in Water (Level 3)

228 Water Molecules

Y. Sugita & Y.O., Prog. Theor. Phys. Suppl. 138 (2000), in press.



Time Series of Potential Energy

C-peptide-1



C-Peptide of Ribonuclease A

Amino-Acid Sequence:

Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-

GIn-His-Met

C-peptide-2



H. Kawai, Y.O., M. Fukugita, T. Nakazawa & T. Kikuchi, *Chem. Lett.* **1991**, 213. Y.O., M. Fukugita, T. Nakazawa & H. Kawai, *Protein Eng.* **4**, 639 (1991).

RMSD = 1.4 A



U. Hansmann & Y.O., *J. Phys. Chem. B* 102, 653 (1998).U. Hansmann & Y.O., *J. Phys. Chem. B* 103, 1595 (1999).

C-peptide-4



M. Masuya, T. Nakazawa & Y.O., in preparation.






















BPTI(16-36) (N = 21): X-Ray



Fragment of Bovine Pancreatic Trypsin Inhibitor

BPTI(16-36)

Amino-Acid Sequence:

Ala-Arg-Ile-Ile-Arg-Tyr-Phe-Tyr-Asn-Ala-

Lys-Ala-Gly-Leu-Cys-Gln-Thr-Phe-Val-Tyr-

Gly



T. Nakazawa, H. Kawai, Y.O. & M. Fukugita, *Protein Eng.* 5, 495 (1992).



T. Nakazawa & Y.O., *J. Peptide Res.* **54**, 230 (1999).



Y.O., M. Masuya, M. Nabeshima & T. Nakazawa, Chem. Phys. Lett. 299, 17 (1999).

17-Residue Helical Peptide (120000-300000 MC Sweeps)



Multicanonical MC



Canonical MC: T = 200 K





17-Residue Helical Peptide (120000-300000 MC Sweeps)



Multicanonical MC



Canonical MC: T = 200 K





Hpoly-1

Helix Propensities of Homopolymers: (Ala)10, (Val)10 & (Gly)10

Average Helicity (Level 2)



Level 0: Y.O. & U. Hansmann, *J. Phys. Chem.* **99**, 11276 (1995). Level 2: A. Mitsutake & Y.O., *Chem. Phys. Lett.* **309**, 95 (1999).

Average Energy of (Ala)10

Level 0 (Gas Phase)





A. Mitsutake & Y.O., Chem. Phys. Lett. 309, 95 (1999).

Specific Heat of (Ala)10



Helix-Coil Transition Temperatures:

T = 420 K (Level 0)T = 340 K (Level 2)

A. Mitsutake & Y.O., Chem. Phys. Lett. 309, 95 (1999).

B1 Domain of Streptococcal Protein G

• 56 amino-acid IgG-binding domain of Protein G (N = 56) has an α -helix and a four-stranded β -sheet

Amino Acid Sequence:

4 5 6 7 8 9 2 3 10 1 MET THR TYR LYS LEU ILE LEU ASN GLY LYS β-strand --->| |<---THR LEU LYS GLY GLU THR THR THR GLU ALA **β-strand** |<--------> VAL ASP ALA ALA THR ALA GLU LYS VAL PHE α-helix |<----LYS GLN TYR ALA ASN ASP ASN GLY VAL ASP |<------>| GLY GLU TRP THR TYR ASP ASP ALA THR LYS **β-strand** --->| THR PHE THR VAL THR GLU |<--**β**-strand --->|

Protein G

Multicanonical MC (Level 1 Solvation)







Protein G

Multicanonical MC (Level 1 Solvation)







まとめ

第一原理からの蛋白質の立体構造予測に成功するための我々の戦略 及びこれまでの計算結果について述べた。

- 計算手法については、徐冷法及び拡張アンサンブル法の有効性を示した。
 ほぼ完成の域に達しつつあると思われる。
- 溶媒効果については、種々の可能性を試しているが、最終的にはレベル3の厳密な寄与を取り入れる必要がある。
- 2年目の本年度は、特に、本プロジェクトで最終目標としている Protein G (N=56)においてレベル 1の溶媒効果を取り入れた拡張アンサンブルシミュ レーションに初めて成功した。
- 来年度は Protein G のレベル 2及び 3の溶媒効果を取り入れた拡張アンサンブルシミュレーションに挑戦したい。
- もし、レベル3の溶媒効果を取り入れても、自然の構造が得られないとき、4年目と5年目で、電子状態計算を含めて、より厳密な系のエネルギー
 関数を独自に開発し、改めて拡張アンサンブルシミュレーションを実行する
 予定である。