

ナチュラルコンピューティングの分子実現とその設計論

研究代表者 山村雅幸（東京工業大学）

1. 研究目的

遺伝的アルゴリズム(GA), 焼きなまし法(SA)など生命・自然現象に啓発された計算技法をナチュラルコンピューティングと総称する。これらは様々な分野で一定の成功を収めており, 方法論・設計論にも知見の蓄積が見られる。ただし, 生命科学との関わりは一方的なアイデアの借用に留まり, 恣意的な問題設定を離れた普遍的な有効性は明らかではない。

一方, 生命科学ではこれとは独立に生命・自然現象に起源を持つ概念, 例えば焼きなましのよ
うなアイデアを日常的に用いており, 変異と選択によって望みの分子を設計する分子進化学は,
確立した分野として成果をあげている。ただし, 情報科学との関わりは, ほとんど同じ現象から
アイデアを得ているにもかかわらず, 無交渉なのが現状である。

分子計算には, 情報科学者と生命科学者の間での緊密な協力関係が不可欠であるという大きな特徴がある。分子計算研究は, 類似の概念を異なる意味で用いてきた不思議さを意識する契機ともなった。また, 分子計算の研究対象として, Adleman 流の組合せ問題の超並列厳密解法に加えて, より分子固有の機能を利用する方向を模索しつつある。このように, 分子計算を通じて情報科学・生命科学がついに同一の概念を共有し, 同一の課題に取り組める時期に至ったものと考ええる。こうした相互フィードバックは双方にとって有用なはずである。

本計画研究では, ナチュラルコンピューティングの分子実現を試みる。最適化問題として見たときの適応度地形の特長によって2種類の具体例を用意した。

- (1) 特定の酵素(アミノアシル tRNA 合成酵素(aaRS))の基質特異性の進化的変化を対象とした, 熱力学的遺伝アルゴリズム(TDGA)の分子実現をテストベッドとして生物系・情報系双方向からの検討を加える。aaRS はすべての生物が持っている古いタンパクで, 多種多様の立体構造を持ち, 広域多峰性の適応度地形を持っていると考えられる。TDGA に適した進化実験系を新たに構築し, 実験で得られたデータをもとに, 進化計算, 最適化, 統計物理などナチュラルコンピューティングの側面からの理論解析・シミュレーションを行うことで, 実験条件の改良と同時に理論解析の補強を試みる。
- (2) 伏見らが開発した既存の試験管ウイルスを改良し, ワンポットの試験管中でライフサイクルを回すことのできる人工ウイルスの分子系を構築し, この等温(核酸+タンパク質)増幅系を用いて, 進化リアクターを構築する。この進化リアクター中の試験管ウイルスは, そのリアクター中に設定された環境に適応するように進化するという意味において, 分子人工生命と呼び得る。この進化リアクターを用いて, タンパク質の大域的適応度地形, 並びに, 適応

度地形中腹の性質を探查する。タンパクの多くは富士山型の地形をもつことが知られている。これら地形の性質をタンパク質物性データを基に理論的に調査し、また、その上の適応歩行を考察する。

このような相互フィードバックを通じて、最終的にナチュラルコンピューティングの分子実現のための設計論を構築する。

メンバーの役割分担は次の通りである。

- ・ 全体の取りまとめ
 - 山村雅幸（東京工業大学・大学院総合理工・教授）：進化計算の理論設計
- ・ ウエット・ドライ実験サブグループ
 - 坂本健作（東京大学・大学院理学系研究科・助手）：ウエット進化計算の実現
 - 染谷博司（文部科学省・統計数理研究所・助手）：ウエット進化計算シミュレーション
 - 伏見 譲（埼玉大学・工学部・教授）：進化リアクターの構築
- ・ 理論サブグループ
 - 樺島祥介（東京工業大学・大学院総合理工・教授）：統計物理による理論解析
 - 太田元規（東京工業大学・学術国際センター・助教授）：立体構造論からの検討
 - 喜多 一（大学評価・学位授与機構・教授）：最適化の側面からの理論解析
 - 新田克己（東京工業大学・大学院総合理工・教授）：配列処理論からの検討

2. 研究成果概要

初年度はそれぞれの研究内容に関する情報交換が計画研究の主な活動を占めていた。2年目に入り、ウエット進化計算のドライ実験を担当する染谷氏、配列処理の立場から理論検討を担当する新田氏の2名の新たなメンバーを加えた。2つの具体的課題にアプローチしているウエット・ドライ実験でそれぞれ進展を見た。この他、次の一連の班会合等を通じて進化システムの統計力学的側面について理解を深めた。

班会合等の実施内容：

- | | |
|--------|---------------------------------|
| 4月12日 | 埼玉大学工学部伏見研究室見学 |
| 4月25日 | 全体会合にて特別講演時田恵一郎（大阪大学）*、企画担当 |
| 5月16日 | 全体会合にて特定領域「統計物理」との交流*、企画担当 |
| 5月29日 | 第1回班会合、特別講演春木満（日本大学）、研究報告伏見*、染谷 |
| 7月25日 | 第2回班会合、研究報告喜多*、坂本 |
| 10月3日 | 第3回班会合、研究報告太田 |
| 12月10日 | 第4回班会合、研究報告樺島* |

上で*は進化システムと統計力学に関連する話題の報告である。その様子を特にまとめておく。時田氏は進化システムにおいては多様性の創出が本質的であることを統計物理の計算法を用いて明らかにした。伏見は分子進化を適応度地形上のランダムウォークであるとみなすと、熱力学とアナロジーが成立することを示した。喜多はやはり進化計算において個体集団の振る舞いが熱

力学的にとらえなおすことができることを利用して、多様性の維持と選択圧力の制御を行う熱力学的遺伝アルゴリズムについて紹介した。樺島はこれらに啓発されて、進化システム上での統計物理の新しい応用の可能性について提案した。

分担者それぞれの個別の研究成果は次のとおりである。

研究の取りまとめ役として進化計算の理論設計にあたる山村の成果は次のとおりである。

進化計算を応用して、局所的類似性と大域的類似性を同時に扱えるたんぱく質の立体構造のラインメントを提案し、たんぱく質ファミリーと保存部位の関係について網羅的調査を行った。細胞の代謝ネットワークの成り立ちについて有機化学的知見を取り入れたネットワーク解析を行い、原始代謝はランダムネットワーク、一次代謝はクラスタ性の強いネットワーク、二次代謝はスケールフリーネットワークとなることを示した。

ウェット・ドライ実験サブグループで課題(1)にウェット面から取り組む坂本の成果は次の3つである。

(1) ウェット GA を実施するために、遺伝子操作技術を用いて試験管内で行う一点交叉法を開発した。交叉効率を 100%-に高める工夫を加えた上で、交叉部位の解析を行った。交叉した 50 個の遺伝子クローンのそれぞれの全長配列について塩基配列決定を行った。その結果、交叉部位が遺伝子の特定の部位に偏らず、全長にわたってまんべんなく分布していることがわかった。

(2) 非天然型アミノ酸であるヨードチロシンを認識するように改変した大腸菌チロシル tRNA 合成酵素 (TyrRS) の結晶構造を決定し、この改変された酵素がこのアミノ酸を特異的に認識するメカニズムを明らかにした。

(3) それぞれ 1 万個の変異体ライブラリーをスクリーニングすることで、ヨードフェニルアラニン (もう 1 つの非天然型アミノ酸) を特異的に認識するように古細菌 TyrRS を改変することができた。変異箇所を特定し、立体構造上にその場所をマップする作業を進めている。

同じく課題 1 でシミュレーションにあたっている染谷の成果は次の通りである。

分子計算の手法を応用したタンパク質の分子進化手法の開発には、コンピュータシミュレーションによる知見の収集が必要不可欠である。このコンピュータシミュレーションの目的のひとつには、開発手法の未知パラメータのチューニングが含まれる。今年度は、パラメータチューニングに有用な最適化手法の開発を行った。具体的には、次世代の計算資源として注目される計算グリッドにおいて数百程度の並列性が可能であると考えられる遺伝的アルゴリズムによる最適化を実現し、その性能評価を行った。

課題(2)にウェット面から取り組む伏見の成果は次の4つである。

(1) 適応歩行による分子進化過程の熱力学的解釈と生命情報量の概念

Rechenberg の意味の(M,N)ES という適応歩行戦略を用いて、富士山型や NK モデル型などの数種の適応度地形を山登りする歩行者のダイナミクスを理論的に研究した。突然変異率と集団サイズで決まるゆらぎの効果を「進化温度 (T)」というパラメータで表す。いずれの地形でも、Tが高いときは、歩行者は適応度 (W) 最大を目指すのではなく、「自由適応度 (G)」の最大をめざすというリャプノフ関数 G を定義できる。G の最大値のまわりのゆらぎと登坂易動度の間に

は、アインシュタインの關係に類似の關係が成立する。また、熱力学における自由エネルギーと宇宙のエントロピーの關係を表す式に類似の式として、 $G/T = W/T + S$ を得るが、この式の各項はそれぞれ、 S は獲得したシャノンの意味の情報量(生命情報の extent)、 W/T は獲得した適応度情報量(生命情報の content)、 G/T は進化の過程で獲得した生命情報の量、という解釈ができることがわかった。

(2) 高温等温核酸増幅法を用いた自然淘汰型進化リアクター

昨年度、耐熱性 T7RNA ポリメラーゼを用いて 1.4 M トレハロース存在下で 50 の等温過程で DNA/RNA を増幅する系を構築した。今年度はこの系で、A,T,G の 3 塩基からなるランダム領域を含み 3SR 増幅機構をコードしたライブラリーを初期プールとして、適応度を比増殖速度とする自然淘汰型進化リアクターを運転した。勝ち残る配列はランダム領域が AT-rich となる傾向を示す。また、以前からしばしば観測された、より速い増幅機構である RNA-Z 増幅機構をコードした突然変異体が進化してくることはなかった。しかし、ランダム領域が欠失し、増幅機構をコードした部分だけが残存する傾向は防げなかった。生存機能に寄与しない純暗号部分は遺伝的荷重であるに違いない。

(3) タンパク質初期ライブラリーの構築法

in vitro virus のゲノムに載せるべき初期ランダムライブラリーは、終止コドンを含んでいてはならず、また、対象に応じてアミノ酸組成が自由に設計できることが望ましい。コドン組成が自由に制御出来るような DNA ランダムライブラリーの作成法を開発している。そのうち、DNA 合成機を 3 台並列に運転してコンビナトリアル化学のスプリット合成を応用する MLSDS 法は配列多様性が 10^{16} に達するが、比較的短鎖である。今年度は、このライブラリーの実際の合成物を複数種いろいろな評価関数で評価しその高品質なことを確認した。また、長鎖化法を検討した。

(4) FRET を用いてプロテアーゼ機能をスクリーニングする系の開発

緑色蛍光蛋白質 (GFP) の C 末端近傍に外来配列を介して化学蛍光発色団 (BODIPY やエオシンなど) を結合した修飾蛋白質を合成した。GFP と化学蛍光発色団の間に蛍光共鳴エネルギー移動 (FRET) が起こるようにしたものである。この外来配列として、特定のプロテアーゼの認識切断配列とスペーサーを仕込んでおく。この分子は、特定のプロテアーゼの機能を、蛍光強度によって定量的に測定するためのプローブとなり、そのプロテアーゼに対する阻害剤の進化工学的設計のためのスクリーニング用ツール等として有用である。

理論サブグループで統計物理による理論解析にあたる樺島の成果は次の 5 つである。

(1) 高性能な CDMA マルチユーザー検出アルゴリズムの提案

符号分割多重接続(Code Division Multiple Access: CDMA)方式は携帯電話や無線 LAN など最新の無線通信を支える重要な要素技術である。この CDMA 方式において受信信号から信号ビットを検出する技術は通信性能を大きく左右する。しかしながら、計算量的な制約から、現状では最良性能を示す検出法は採用されていない。本研究では、検出問題を強く相互作用している磁性体の統計力学とみなすことで、ユーザー数が十分多い場合にほぼ最良の性能を達成する近似的検

出アルゴリズムを構成することに成功した。

(2) 信念伝播に基づくスピングラス模型の解析

スピングラス模型は、複雑な相互作用下にある多体問題の典型例として磁性体のみならず、情報符号化、タンパク質折りたたみ、生態系などの研究にも有用な知見を供与する重要な数理モデルである。このモデルの性質を人工知能の研究で提案された信念伝播 (Belief Propagation) を用いて解析し、その相転移構造に関する新規な知見を得た。

(3) 統計力学的手法に基づく信頼性関数の評価

情報理論の未解決問題である誤り訂正符号の信頼性関数を、誤り訂正符号とスピングラス模型との対応関係を利用し、統計力学的手法を用いて求めた。

(4) 公開鍵暗号に対する攻撃の統計力学的評価

スピングラス模型を利用した公開鍵暗号に関して、典型的な攻撃法に対する安全性の評価をおこなった。

(5) 低密度パリティ検査符号の統計力学的解析に関する総説の執筆

低密度パリティ検査符号と呼ばれる誤り訂正符号とスピングラス模型との対応を利用した。

樺島は統計物理分野での実績を認められて、統計力学的解析の総説を J. Phys. A 誌(英国物理学会刊行)から依頼され執筆している。

理論サブグループで立体構造論からの検討にあたる太田の成果は次の2つである。

(1) タンパク質の立体構造情報を利用した活性部位予測のアルゴリズムを考案し、その評価を行った。この方法のユニークな点は、立体構造から点突然変異体の熱安定性の評価を行い、安定性を向上させる部位を探るところにある。安定性テーブルと、構造上の穴やクレフトに関するデータ、および配列保存性のデータを組み合わせて、最後には k-nearest neighbor 法で予測を実施する。この方法のサービスを <http://bioinfo.tsurumi.yokohama-cu.ac.jp/p-cats/> にて行っている。

(2) TrpCage という小タンパクの大規模なフォールディングシミュレーションを東工大のグリッドコンピュータ上で実施し、結果の解析をおこなった。計90本程度のフォールディング、アンフォールディング軌道を収集し、その比較をおこなった。こういった大規模な軌道比較を行うためのアルゴリズムを考案し、それを適用した結果、フォールディングの道筋について議論をすることができた。

3. 成果リスト

< 山村雅幸 >

- (1) 栃村 剛史, 山村 雅幸: 代謝経路に着目したタンパク群の進化系統解析, 計測自動制御学会 第30回知能システムシンポジウム資料, pp.103-108 (2003).
- (2) 広戸 裕介, 山村 雅幸: PAGE を利用した DNA 分子メモリの実装, 計測自動制御学会 第30回知能システムシンポジウム資料, pp.109-114 (2003).
- (3) 梶田 睦, 山村 雅幸, 小原 雄治: 力学モデルを用いた線虫 *C. elegans* の初期胚における細

- 胞配置シミュレーション, 計測自動制御学会 第 30 回知能システムシンポジウム資料, 127-132 (2003).
- (4) Sung-Joon Park, Masayuki Yamamura: Two-layer Protein Structure Comparison, Proc. 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2003), 435-440 (Nov. 2003).
- (5) Sung-Joon Park, Masayuki Yamamura: GA-based Generic Method for Protein Structure Comparison, Proc. 2003 Congress on Evolutionary Computation (CEC 2003), 1528-1535 (Dec. 2003).
- (6) Sung-Joon Park, Masayuki Yamamura: Asynchronous Real-coded Genetic Algorithms for Simultaneous Protein Structure-based Alignment, Proc. GECCO2003: Proceedings of the Bird of a Feather Workshops, 276-279 (2003)
- (7) Sung-Joon Park, Masayuki Yamamura: Real-Coded Genetic Algorithm to Reveal Biological Significant Sites of Remotely Homologous Proteins, Proc. GECCO2003, 1602-1603 (2003).
- (8) Atsushi Kajita, Masayuki Yamamura, and Yuji Kohara: Computer Simulation of the Cellular Arrangement in Early Cleavage of the Nematode *Caenorhabditis elegans*, *Bioinformatics*, 19(6), 704-716 (2003).
- (9) 松田 大典, 山村 雅幸: 代謝ネットワーク解析に基づく形成モデルの提案, 計測自動制御学会 第 31 回知能システムシンポジウム 資料集, (2004).
- (10) 伊藤 浩史, 山村 雅幸: 位相差情報を用いた遺伝子ネットワークの推定, 計測自動制御学会 第 31 回知能システムシンポジウム 資料集, (2004).
- (11) 皆川 恵一, 山村 雅幸: 分子動力学法の力場パラメータに関する研究, 計測自動制御学会 第 31 回知能システムシンポジウム 資料集, (2004).
- < 染谷博司 >
- (12) 染谷博司: "グリッド環境に適した遺伝的アルゴリズムの設計とその性能評価", 情報処理学会・電子情報通信学会 第 2 回 情報科学技術フォーラム (FIT2003) 講演論文集第 1 分冊, pp.75-77 (A-036), Sep 2003.
- (13) 染谷博司: "グリッド環境に適した遺伝的アルゴリズムに関する考察とその実現", 電気学会 電子・情報・システム部門大会 2003 講演論文集, pp.435-439 (OS7-3), Aug 2003.
- (14) 染谷博司: "進化型計算による適応的探索およびグリッド環境への応用", 研究集会「最適化: モデリングとアルゴリズム 17」, to appear (2003 年 3 月口頭発表), 2003.
- < 坂本健作 >
- (15) Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K, Yokoyama S. "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion" *Nature Structural Biology* 10, 425-432 (2003).

< 伏見 譲 >

- (16) Aita T., Ota M., Husimi Y., An in silico exploration of neutral network in protein sequence space, *J.Theor.Biol.* 221, 599-623 (2003)
- (17) Aita T., Husimi Y., Thermodynamical Interpretation of Adaptive Walk on a Mt.Fuji-type Fitness Landscape: Einstein's Relation-like Formula holds in a Stochastic Evolution, *J.Theor.Biol.*, 225, 215-228 (2003)
- (18) Suzuki M., Ito Y., Savage H.E., Husimi Y., Douglas K.T., Intramolecular Fluorescent Resonance Energy Transfer (FRET) by BODIPY Chemical Modification of Cysteine-engineered Mutants of Green Fluorescent Protein, *Chem.Lett.* 32, 306-307 (2003)
- (19) Husimi Y., Aita T., Free fitness as a measure of biological information, the determinant of molecular evolution, Fourth East Asian Biophysics Symposium (Taipei) Abstract 4, 11, (2003)
- (20) Aita T., Husimi Y., Thermodynamic Interpretation of evolutionary dynamics on a Mt. Fuji-type fitness landscape, *Bulletin Math. Biology* (2004) in press
- (21) 伏見 譲, 田淵 一郎, In vitro virus と進化分子工学, 蛋白質・核酸・酵素 (特集: 化学と生物学の接点がつくる NEW バイオテクノロジー), 48, (11), 1481-1487 (2003)
- (22) 伏見 譲, 編著「生命の起源: 「物質の進化」から「生命の進化へ」」(パリティ・ブックス、丸善、2004)
- (23) Tabuchi I., Soramoto S., Ueno S., Husimi Y., Multi-Line Split DNA Synthesis: a combinatorial method to make a high quality peptide library, *BMC Biotech.* submitted.

< 樺島 祥介 >

- (24) Y. Kabashima and D. Saad, Statistical mechanics of low-density parity-check codes, *J. Phys. A* 37, R1 (2004)
- (25) N. Skanzos, D. Saad and Y. Kabashima, Analysis of common attacks in public-key cryptosystems based on low-density parity-check codes, *Phys. Rev. E* 68, 056125 (2003)
- (26) Y. Kabashima, A CDMA multiuser detection algorithm on the basis of belief propagation, *J. Phys. A* 36, 11111 (2003)
- (27) N. Skanzos, J. van Mourik, D. Saad and Y. Kabashima, Average and reliability error exponents in low-density parity-check codes, *J. Phys. A* 36, 11131 (2003)
- (28) Y. Kabashima, Propagating beliefs in spin glass models, *J. Phys. Soc. Jpn.* 72, 1645 (2003)

< 太田 元規 >

- (29) N. Handa, T. Terada, Y. Kamewari, H. Hamana, J. R. H. Tame, S.-Y. Park, K. Kinoshita, M. Ota, H. Nakamura, S. Kuramitsu, M. Shirouzu and S. Yokoyama, Crystal structure of the conserved protein TT1542 from *Thermus thermophilus* HB8, *Protein Sci.* 12

(2003) 1621-1632.

- (30) M. Ota, K. Kinoshita and K. Nishikawa, Prediction of catalytic residues in enzymes based on known tertiary structure, stability profile, and sequence conservation, *J. Mol. Biol.* 327 (2003) 1053-1064
- (31) T. Aita, M. Ota and Y. Husimi, An in silico exploration of the neutral network in protein sequence space, *J. Theor. Biol.* 221 (2003) 599-613
- (32) 太田元規 「生命情報学」五條堀孝編 シュプリンガーフェアラーク東京 (2003) 第5章 タンパク質の理論構造生物学 (p109-149)
- (33) 太田元規 「バイオインフォマティクスがわかる」菅原秀明編 羊土社 (2003) 第4章 タンパク質立体構造のバイオインフォマティクス (p44-51)

4 . 代表的な論文

以下の3つのうち、(1)は課題(1)に関する進展、(2)は進化システムの統計力学的側面を議論したもの、(3)は統計力学の情報科学応用に関する成果である。その別刷を添付する。

- (1) Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K, Yokoyama S. "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion" *Nature Structural Biology* 10, 425-432 (2003).
- (2) Aita T., Husimi Y., Thermodynamical Interpretation of Adaptive Walk on a Mt.Fuji-type Fitness Landscape: Einstein's Relation-like Formula holds in a Stochastic Evolution, *J.Theor.Biol.*, 225, 215-228 (2003).
- (3) Y. Kabashima, A CDMA multiuser detection algorithm on the basis of belief propagation, *J. Phys. A* 36, 11111 (2003).

以上