

特定領域研究「分子プログラミング」計画研究

平成17年度成果報告書

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いい, 様々な分野で一定の成功を収めているが, 生命科学との関わりは一方的なアイデアの借用に留まる. 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を日常的に用いており, 特に分子進化学は確立した分野として成果をあげている. 本課題の目的は, 分子計算によって培われた情報科学・生命科学の協力体制によって, 分子進化という同一の課題に取り組むことにある. 3年間の研究成果の上に立って, 引き続き次の2つの具体的対象に取り組んだ.

- (1) アミノアシル tRNA 合成酵素 (aaRS) の基質特異性の進化的改変を対象とした, 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. 前年度までに構築した分子系のプロトコルを運用し, 進化過程のシミュレーションから予想される性質の検証を行う.
- (2) 試験管ウイルスをさらに改良する. 前年度までに構築した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により, 具体的な分子進化を行う. 統計物理的手法に基づく適応度地形の歩行の理論解析から, 実際のタンパクの分子進化を裏付ける. 本年度は次の実施計画に基づいて, 上記2つの具体的対象のそれぞれについて, 実験・理論両面から接近した.

(1) aaRS の基質特異性の進化的改変について

- (a) 進化実験系を運用してデータを蓄積する. 遺伝的アルゴリズム(GA)との整合性を高めるため, 繰返しスクリーニング(選択)を行うことで, 望みの性質を持つ酵素を進化させる. 高い活性を持つ酵素を得ると同時に, シミュレーションの検証に役立つデータを収集する.
- (b) 構成される進化実験系は物理的・生物学的な制約から, 計算機上の GA とは相当に異なる問題設定となっている. 提案したモデルでいくつかの問題点について打開点が見出されたので, 進化実験系の試験データをもとにモデルのパラメータチューニングを行い, 実験系によって検証する.

(2) ワンポット分子進化系の構築について

- (a) 構築した分子人工生命としての試験管ウイルスをさらに改良する. あらかじめリンカーを5'末端に化学結合した DNA オリゴマーを mRNA の3'末端にハイブリダイゼーシ

ョンさせることにより、新生タンパク質と結合させ、遺伝子型表現型対応付けを行う試験管 DNA ウイルスを開発する。

- (b) 生体高分子の大域的適応度地形と地形中腹の性質の理論的調査を行う。生体高分子の配列空間上の適応度地形の多くは富士山型に近い。統計物理的手法によって、適応度地形の歩行を解析し、具体的課題(1)と興味深い関連性が発見されたので、この理論を発展させる。

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
 - 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
 - 春木満（日本大学・工学部・助教授）：タンパク質熱安定化の進化実験
 - 若林健一（研究協力者、東京工業大学・大学院総合理工学研究科・特別研究員）：ウエット進化計算の実現
- ・ 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（京都大学・学術情報メディアセンター・教授）：最適化からの理論解析
 - 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

3年目の中間報告を過ぎて、研究課題の絞込みと新たな展開の両面を試みた。昨年より分担者として春木満氏の参加を得て実験面を強化した。本年は坂本健作氏の理研への移転に伴い、研究協力者として若林健一氏にウエット進化計算の実現実験を依頼した。

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理論解析にあたる山村の成果は次のとおりである。

理論面において、組換えタンパク質の発現量およびフォールディングの有無について生物学実験データからのデータマイニングを行って、組換えタンパク質の生産性向上に役立つ知見を得た。

ウエット・ドライ実験サブグループで課題(1)にウエット面から取り組む、研究協力者の若林の成果は次のとおりである。

中立変異戦略に基づいた初期集団構築を行い、ウエット GA を実行するための下準備を整えた。タンパク質の分子進化において、より大域的な探索を目指すには、野生型の遺伝子への突然変異導入を繰り返し行い集団分布を広げることが要求されるが、変異蓄積によってタンパク質の構造

が崩れるなど根本的な形質の破壊が起こる危険性が懸念される。導入されたランダムな変異がタンパク質機能にとって有利にも不利にも働かない中立的な変異だけで構成されていれば、タンパク質は本来の機能を維持し、同時に構造も維持されていると考えることが出来る。そこで、変異体が機能を維持しているかどうかの評価を行い、変異導入後も機能を維持している個体だけを残すこととし、これらを GA の初期集団として用いる。

TyrRS タンパク質の遺伝子に 4.9, 6.7, 9.5, 15.1bp/gene の変異頻度で変異を導入したライブラリを用意し、それぞれの活性の有無を調べたところ、導入変異量が増えるに従って活性を保持する変異体の集団全体に対する割合が減少していくことを確認できた。15 残基を越えると、変異の蓄積が進まず、むしろ多様性が減少していくことが認められたため、15.1bp/gene のライブラリを親集団として選定した。

PyIRS 遺伝子についても同様の操作を行い、ライブラリを構築した。活性の有無を調べる際の指標は TyrRS の場合と異なり、アンバーサプレッション活性をもとにしたが、PyIRS では、4.5bp/gene 程度の変異でほぼ 9 割の変異クローンが活性を失うことが明らかとなった。

同じく課題 1 で理論解析と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。

シミュレーテッドアニーリングなどのマルコフ連鎖モンテカルロ法に基づく最適手法の分子による実現可能性について検討した。その結果、これらの実現は困難であるとの結論を得た。これにより、間接的にはあるが、我々が提案しているウェット GA によるタンパク質の進化手法の意義を確認した。

課題(2)に取り組む伏見の成果は次のように実験・理論の 2 つである。

(1) 等温核酸増幅に基づく自然淘汰型進化リアクターにおけるプロモータの進化過程

Joyce らは RNA-Z 反応を用いて 37 の T7RNA ポリメラーゼに対するプロモータの最適配列を決定したが、われわれは進化リアクターにおけるミスハイブリを減少させる目的で 50 の ThermoT7RNA ポリメラーゼに対するプロモータの最適配列を決定するための *in vitro* selection 実験を行った。ところが、我々が設定した *in vitro* selection 環境から逸脱した *in vitro* evolution が起こった。すなわち、コンテキスト依存の淘汰による中立経路を経た進化という、中立進化の適応進化に対する役割を明示する現象を複数回観察した。また、37 の最適配列と 50 の最適配列はハミング距離 2 だけ異なるものであった。

(2) 実験室内分子進化過程の熱力学的及び情報論的解釈

単純化された進化分子工学実験における分子進化過程と熱力学過程とのアナロジーをとることで、分子進化過程が外部環境からの情報獲得過程であるという概念を明確にした。進化分子工学実験における進化する対象(生体高分子)は外部環境から「その環境に適応して生存し続けるための情報」を獲得する。例えば「ある受容体タンパク質に結合するペプチドの進化的創出」に関する場合は、実験で創出された結合性ペプチドは、外部環境から「受容体に結合するための情報」を獲得したことになる。ここまではよく聞く話ではあるが、この情報の価値を「適応度情報

量” Ifit と呼び、「適応度変化である結合自由エネルギー変化 W を進化温度 TE で割ること、言い換えると、その変化の不確定性でスケールリングすること」で定量化する。ここで、進化温度 TE とは、有限サイズの集団が有限突然変異率によって行う配列空間上の拡散効果を表すパラメータである。“Shannon の情報量” $ISha$ は情報の多様性に関する側面を表現するのに対し、“適応度情報量” $Ifit$ は情報の価値的な側面を表現していると言える。その両者の和 $Ibio$ はこの進化系の Lyapunov 関数であることが証明でき、進化は $Ibio$ の最大状態に向かって進行する。そこで、我々は $Ibio$ を“生物学的情報量”と呼ぶことにした。 $Ibio = Ifit + ISha$ という表式は熱力学における自由エネルギーに対する表式： $-G/T = -H/T + S$ のアナロジーとなる。

実験サブグループでタンパク質熱安定化の進化実験に取り組む春木の成果は次の2つである。

(1) 蛋白質耐熱化についての進化シミュレーション実験

熱安定性の低下した大腸菌 RNase HI の C 末端欠失変異体と、大腸菌 RNase H 依存性温度感受性株を利用し、淘汰圧の有無により、熱安定性の向上の頻度がどのように変化するか検討を行った。ランダム変異を導入した遺伝子に、淘汰を行わずにさらにランダム変異を導入した場合と、ある程度以上熱安定性が低下した変異を淘汰した後にランダム変異を導入した場合について、熱安定性の向上の頻度を比較した。その結果、淘汰を行わなかったものは、淘汰を行ったものより、高い頻度で熱安定性の向上が見られた。さらにランダム変異を導入し同様な操作をおこなったところ、逆に淘汰を行ったものは、行わなかったものより、高い頻度で熱安定性の向上が見られた。したがって、初期段階では淘汰圧をかけない方が、淘汰圧をかけるよりも多様性が保たれる結果、進化した変異体の比率が高くなったと考えられる。他方、進化が進んだ段階では淘汰圧をかけない方が安定性を損なうような変異が導入される確率が高くなるため、進化した変異体の比率が低くなったと考えられる。

(2) タンパク質変異体安定性プロフィールの予測に基づく部位特異的およびランダム変異導入による大腸菌リボヌクレアーゼ HI の熱安定化の向上

大腸菌リボヌクレアーゼ HI (RNase HI) において、タンパク質変異体安定性プロフィール (SPMP) により予測された変異体の安定性は、実験値と相関することが示されている。この結果から、 α -ヘリックス IV の N 末端側のターン上に連続する Lys99, Asn100, Val101 への変異導入は、大きく安定性を向上させると予測され、この領域は安定化の余地が大きい可能性が考えられる。そこで、Lys99, Asn100 については、SPMP により最も安定性を向上させると予測される置換 (K99P, N100G) を導入した。その結果、K99P 変異体は野生型より、pH 5.5 における T_m が約 1.8 上昇した。Val101 については、安定性を向上させると予測される置換が多く存在するため、ランダム変異を導入し、熱安定性の低下した大腸菌 RNase HI の C 末端欠失変異体と、大腸菌 RNase H 依存性温度感受性株を利用して、安定化変異のセレクションを行った。セレクションの結果、Ala, Arg, Gln に置換した変異体を得られ、V101A 変異体は野生型より、pH 5.5 における T_m が約 2 上昇した。

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通りである。
大規模なパーセプトロン型確率モデルに対して、ベイズ学習を行なうための近似的学習アルゴリズムを開発した。パーセプトロンは機械学習における代表的な分類器であるが、ノイズを含む実データの解析では入力に応じて正負のラベルを確率的に出力する確率モデルとして取り扱うことが自然である。近年の統計的学習理論では、ベイズの定理に基づいた学習（ベイズ学習）が確率モデルに対して優れた予測能力を与えることが示されているが、モデルの大規模化に伴いベイズ学習の実行が困難になるという問題が残されている。本研究では、統計力学の知見に基づいてこの困難を実際的に解決するため近似的学習アルゴリズムを開発し、その性能を理論的に検討した。加えて、2000次元からなるマイクロアレイデータに基づいて大腸がんを分類する実問題に対しその有効性を確認した。

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4つである。

(1) フォールディング関連

軌道プロフィールを導入し、Trp-Cageのフォールディング初期過程について調べた。フォールドする軌道は二次構造があまりない初期フェーズから数ステップでフォールドするのに対し、フォールドしない軌道や擬天然構造にフォールドする軌道はステップ数が多いことがわかった。

(2) デザイン関連

ミオグロビン構造をもとにヘムに結合するフィコシアニン配列を設計した。実験的に合成を行い構造特性を調べたが、設計配列は溶解しなかった。合成時に作成された1置換体はそこそこのヘリックス含量をしめし、弱くヘムと結合した。

(3) 相互作用関連

タンパク質の会合状態を決めるアミノ酸を調べるために、同じファミリーであるが会合状態が異なるタンパク質のマルチプルアラインメントから置換テーブルを作成した。この置換テーブルをもとに相互作用面/表面の判別を試みたが、アミノ酸組成による結果とあまり変わらなかった。組成関数と置換テーブルの融合を試みる予定である。

(4) ネットワーク関連

微生物の代表的ゲノム30種を用い、オーソログを調べることでミニマム代謝マップを作成した。このマップは22種の一次必須マップとそれらの入力物質を生成する13種の二次および三次必須マップからなる。実験で調べられた必須遺伝子と比較することで、KEGGなどの問題点について考察を行った。

3. 成果リスト

<山村雅幸>

- (1) 新野 賢央, 山村 雅幸: 分子強化学習の実現に関する研究, 計測自動制御学会 第32回知能システムシンポジウム 資料集, 333-338 (2005).
- (2) 山村 雅幸, 清尾 康志, 藤本 健造: アクエラスコンピューティングのスケールアップ, 計測

自動制御学会 第 32 回知能システムシンポジウム 資料集, 339-344 (2005).

- (3) Satoshi Kira, Masayuki Yamamura and Seiki Kuramitsu: Predicting the Expression Levels of Recombinant Protein by using Reinforcement Learning, Proc. of the 8th IASTED International Conference on Intelligence Systems and Control (ISC2005), pp.471-476 (2005).
- (4) 山村雅幸, 萩谷昌己: 分子コンピューティングの現状と新展開, 日本建築学会第 28 回情報・システム・利用・技術シンポジウム論文集, pp.286-293 (2005).
- (5) 吉良聡, 磯合敦, 山村雅幸: データマイニング技術を用いた組換えタンパク質の発現量解析, 人工知能学会誌, Vol.21, No.1, pp.481-491 (2006).

< 染谷博司 >

- (6) Hiroshi Someya, Masayuki Yamamura : "A Robust Real-Coded Evolutionary Algorithm with Toroidal Search Space Conversion", *Soft Computing*, Vol.9, No.4, pp.254-269, Apr 2005.
- (7) 染谷博司, 山村雅幸, 坂本健作 : "マルコフ連鎖モンテカルロ法の分子計算による実現の一検討", 電気学会 電子・情報・システム部門大会 2005 講演論文集, pp.466-469 (OS3-4), Sep 2005.

< 伏見譲 >

- (8) Aita T., Morinaga S., Husimi Y., Thermodynamical Interpretation of Evolutionary Dynamics on a Fitness Landscape in an Evolution Reactor, II, *Bulletin of Mathematical Biology*, 67(3), 613-661 (2005)
- (9) Hamamatsu N., Aita T., Nomiya Y., Uchiyama H., Nakajima M., Husimi Y., Shibana Y., Biased mutation-assembling: an efficient method for rapid directed evolution through simultaneous mutation accumulation, *Protein Eng. Design & Selection*, 18(6), 265-271 (2005)
- (10) 伏見 譲, 試験管内での分子の自己増殖と進化, *パリティ*, 20(10), 28-29, (2005)
- (11) 伏見 譲, 分子進化の偶然と必然, *数理科学*, 44(1), 36-40, (2006)
- (12) 伏見 譲, 岩波講座物理の世界「DNA と遺伝情報の物理」(岩波書店, 2005)
- (13) 相田拓洋, 伏見 譲, 実験室内分子進化過程の熱力学的及び情報論的解釈, *生物物理*, 印刷中 (2006)

< 春木満 >

- (14) Hirano, S., Haruki, M., Takano, K., Imanaka, T., Morikawa, M., Kanaya, S. Gene cloning and in vivo characterization of a dibenzothiophene dioxygenase from *Xanthobacter polyaromaticivorans*. *Appl Microbiol Biotechnol*; 69, 672-81 (2006)
- (15) Haruki, M., Saito, Y., Ota, M., Nishikawa, K., Kanaya, S. Stabilization of *E. coli* Ribonuclease HI by the 'stability profile of mutant protein' (SPMP)-inspired random and non-random mutagenesis *J. Biotechnol.*, in press (2006)

- (16)春木 満: 表面プラズモン共鳴とその利用, 「生物工学ハンドブック」 日本生物工学会編
コロナ社, p.220-222 (2005)
<樺島祥介>
- (17)S. Uda and Y. Kabashima, Statistical Mechanical Development of a Sparse Bayesian Classifier, J. Phys. Soc. Jpn. 74, 2233-2242 (2005)
- (18)Y. Kabashima, Replicated Bethe Free Energy: A Variational Principle behind Survey Propagation, J. Phys. Soc. Jpn. 74, 2133-2136 (2005)
- (19)K. Ogure and Y. Kabashima, An Exact Analytic Continuation to Complex Replica Number in the Discrete Random Energy Model of Finite System Size, Prog. Theor. Phys., Suppl 157, 103-106 (2005)
- (20)Y. Kabashima and T. Hosaka, Statistical Mechanics of Source Coding with a Fidelity Criterion, Prog.Theor. Phys. Suppl 157, 197-204 (2005)
- (21)D. Saad, N. Skantzos and Y. Kabashima, Security and Reliability of LDPC Based Public-Key Cryptosystems, Prog.Theor. Phys. Suppl. 157, 229-236 (2005)
- (22)樺島祥介: 確率を手なずける秘伝の計算技術 - 古くて新しい確率・統計モデルのパラダイム - 小特集編集にあたって, 電子情報通信学会誌 88[9] 695 (2005)
- (23)田中和之, 田中利幸, 樺島祥介: 序文 - 確率・統計モデルが切り拓く推論と学習の新しいパラダイム -, 電子情報通信学会誌 88[9] 696-697 (2005)
<太田元規>
- (24)M. Haruki, Y. Saito, M. Ota, K. Nishikawa, S. Kanaya. Stabilization of E. coli Ribonuclease HI by the 'stability profile of mutant protein' (SPMP)-inspired random and non-random mutagenesis. J. Biotech. in press
- (25)Y. Isogai, Y. Ito, T. Ikeya, Y. Shiro, M. Ota, Design of lambda Cro Fold: Solution Structure of a Monomeric Variant of the De Novo Protein. J. Mol. Biol. 354 (2005) 801-814
- (26)Y. Asada, M. Sawano, K. Ogasahara, J. Nakamura, M. Ota, C. Kuroishi, M. Sugahara, K. Yutani, N. Kunishima, Stabilization mechanism of the tryptophan synthase alpha-subunit from Thermus thermophilus HB8: X-ray crystallographic analysis and calorimetry. J Biochem (Tokyo) 138 (2005) 343-353
- (27)K. Kinoshita and M. Ota, P-cats: Prediction of catalytic residues in proteins from the tertiary structures, Bioinformatics 21 (2005) 3570-3571.
- (28)太田元規, 池口満徳 タンパク質フォールディング博物学の曙 生物物理 (印刷中)
- (29)太田元規 タンパク質フォールディングのバイオインフォマティクス パリティ 21 (2006) 63-65
- (30)太田元規 タンパク質のフォールディングシミュレーション シミュレーション 24 (2005) 29-34

(31)磯貝泰弘,太田元規 「タンパク質科学」後藤祐児,桑島邦博,谷澤克行編 化学同人(2005)
第8章4 人工タンパク質設計 (pp363-371)

(32)太田元規,磯貝泰弘 「生物工学ハンドブック」日本生物工学会編 コロナ社(2005)I
編 3.2.3 [3] タンパク質の分子設計 (pp. 188-195)

4 . 代表的な論文

以下の3つのうち,(1)は課題(1)の理論的側面に関する進展,(2)は分子進化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) 染谷博司,山村雅幸,坂本健作:"マルコフ連鎖モンテカルロ法の分子計算による実現の一検討",電気学会 電子・情報・システム部門大会 2005 講演論文集, pp.466-469 (OS3-4), Sep 2005.

(2) Aita T., Morinaga S., Husimi Y., Thermodynamical Interpretation of Evolutionary Dynamics on a Fitness Landscape in an Evolution Reactor, II, Bulletin of Mathematical Biology, 67(3), 613-661 (2005)

(3) S. Uda and Y. Kabashima, Statistical Mechanical Development of a Sparse Bayesian Classifier, J. Phys. Soc. Jpn. 74, 2233-2242 (2005)

以上