

# 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

山村雅幸 (東京工業大学・大学院総合理工学研究科)

## 1. 研究の目的

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, 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 種類のテストベッドを用意し, ナチュラルコンピューティングの分子実現を試みる。さらに理論的考察との相互連携によってナチュラルコンピューティングの設計論の確立を目指してきた。

## 2. 研究項目と役割分担

本計画研究では, 大きく分けて次の 3 つの研究項目を用意した。

- (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。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た。
- (2) 伏見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した。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た。

- (3) 生物および分子の進化に対して、物理現象として統計物理からの考察、最適化手法としての考察、タンパク質の立体構造論からの考察などさまざまな角度からの理論的考察を加え、設計論の確立を目指した。

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ことを目標とした。

GA は集団中に蓄積されたビルディングブロック分布のもとで変異・交叉によりサンプル生成する適応的モンテカルロ法である。GA の分子実現による集団的分子計算の計算能力の解明、およびタンパク質というそれ自身強力な機能を持った分子機械の設計は、分子プログラミングの方法および対象として新しい展開である。構築された設計論は既存の分子計算の計算能力や精度の改善にも役立つことが期待できる。

実現可能な実験系は計算機上での恣意的な問題設定とは異なる制約を持つ。ナチュラルコンピューティングのシステム論的普遍性にとって、知見の移植性の検討は重要な課題である。本計画研究では GA の忠実な実現ではなく、適応的モンテカルロ法として、あるいは熱力学的最適化法としての設計論の適用を試みてきた。実験系の手順は狭義の GA には当てはまらないかもしれないが、計算機だけでは得られなかった設計論の補強が期待できる。関連研究に Woods ら(米)、Baeck ら(蘭)のものがあるが、汎用の GA の忠実な実現には適応度評価に無理がある。分子進化を対象とし設計論にまで発展させる試みは例がない。

メンバーの役割分担は次の通りである。最後の数字は関連する研究項目である。

- ・ 全体の取りまとめ
  - 山村雅幸(東京工業大学): 進化計算の理論設計(1,2,3)
- ・ ウェット・ドライ実験サブグループ
  - 坂本健作(東京大学 理化学研究所, H17 まで): ウェット進化計算の実現(1,3)
  - 染谷博司(統計数理研究所, H15 から): ウェット進化計算シミュレーション(1,3)
  - 伏見 譲(埼玉大学): 進化リアクターの構築(2,3)
  - 春木 満(日本大学, H16 から): タンパク質熱安定化の進化実験(2,3)
  - 小宮 健(東京工業大学, H17 から): DNA 計算からの検討(2, 3)
  - 木賀大介(東京工業大学, H18 から): 構成的生物学からの検討(2, 3)
  - 研究協力者: 若林健一(東京工業大学 理化学研究所, H17 から): ウェット進化計算および進化リアクタの実験(1,2)
- ・ 理論サブグループ
  - 樺島祥介(東京工業大学): 統計物理による理論解析(3)
  - 井上真郷(東京工業大学, H16): 統計物理による理論解析(3)

- 喜多 一 (大学評価・学位授与機構 京都大学): 最適化の側面からの理論解析(1,3)
- 太田元規 (東京工業大学, H17 まで): 立体構造論からの検討(3)
- 新田克己 (東京工業大学): 配列処理論からの検討(3)

### 3. 活動状況と班内の連携状況

#### 平成14年度

計画研究の初年度として,全体会合にあわせて実施した班会合で,それぞれの研究内容に関する情報交換を行った.

#### 平成15年度

ウェ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ェット・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.

- 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\*

#### 平成16年度

分子進化実験に参加する春木氏,前年度に重要性を確認した統計力学的理論解析を担当する井上氏の2名を新たなメンバーに加えた.前年非常に有効であった班会合は,全体会合と合同で行うこととした.分子コンピューティングの主要国際会議であるDNA10において,研究項目(1)(2)のそれぞれの成果を報告した.

#### 平成17年度

3年目の中間報告を終えて,研究課題の絞込みと新たな展開の両面を試みた.坂本氏の理研への移転に伴い,研究協力者として東京工業大学大学院総合理工学研究科特別研究員の若林健一氏にウェット進化計算の実現実験を依頼した.また小宮氏を分担者として迎えた.

#### 平成18年度

ウェット進化計算は理論・実験両面で成果をあげはじめた.木賀氏を分担者として迎え,新たな展開の方向として構成的生物学について検討した.木賀を中心としてTokyo Allianceと称するチームを率いて,第3回国際人工遺伝子回路コンテストに参加し,最優秀部品賞,最優秀共同研究賞のふたつを受賞した.

#### 4. 研究成果概要

##### 平成14年度

計画研究の初年度にあたり、主として今後の研究のための情報交換を行った。研究項目ごとに整理した研究成果は次のとおりである。

##### (1) ウェット進化計算の分子実現

タンパク質の進化を、複雑な組合せ最適化問題の「分子計算」とみなす立場から研究を行った。DNA 分子（遺伝子）はタンパク質の機能・構造を符号化した「データ・テープ」とみなされ、タンパク質の発現は「目的関数の計算」プロセスと見なされる。分子計算の並列性をもってしても「最適解」の風潰し的な探索は不可能だが、一方でタンパク質の目的関数が知られていないため、現在の高速な電子コンピュータでもタンパク質の設計には有効ではない。そこで、「進化計算」の計算スキームを取り入れ、コンピュータ・シミュレーションによるドライ実験と、分子生物学のウェット実験の両面からのアプローチを進めた。

坂本によるウェット実験の成果は次の通りである。ウェット実験の目的の1つは、タンパク質の現実の適応地形を調査することである。アミノ酸配列の配列空間の大きさを考えると、実際の実験で踏査できる領域は極めてわずかであり、領域踏査の方針は極めて重要になる。従来、野生型（WT）タンパク質の近傍の地形を調べることが行われてきた。しかし、すでに「解」（=WTタンパク質）の得られている空間の近傍を調べるのでは、解探索法としての進化計算の働きを検証することができない。ここでは野生型タンパク質とは明瞭に区別される活性を持つ変異タンパク質を得て、野生型から変異体間の地形を調査するという方針を実地形測定の方針とした。具体的には tRNA に L-チロシン（アミノ酸）を共有結合させる酵素であるチロシル tRNA 合成酵素（TyrRS）を取り上げた。アミノ酸特異性が WT TyrRS と異なるものを得るために、TyrRS の立体構造を X 線結晶解析法によって決定し、L-チロシン部位の立体構造から改変すべきアミノ酸を3残基ずつ2セット選んだ。可能なアミノ酸置換は、 $20 \times 20 \times 20 \times 2 = 1,600$  通りである（遺伝子レベルではこの3倍程度の数になる）。遺伝学的な実際の実験によって、16,000 個の変異 TyrRS をスクリーニングし、特異性の変化した変異体を2種類得ることができた。WT と変異体を含む領域の地形の詳しい調査を始めた。

- (1) Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, T., Saito, K., Shirouzu, M., Hirao, I., Yokoyama, S.  
“Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells” *Nucleic Acids Research* 30, 4692-4699 (2002).
- (2) Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., Shirouzu, M., Harada, Y., Nakayama, H., Takio, K., Hasegawa, Y., Endo, Y., Hirao, I., Yokoyama, S.  
“An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system” *Proc. Natl. Acad. Sci. USA*, 99, 9715-9720 (2002).

## (2) 進化リアクタの作成

伏見による進化リアクタの作成に係わる成果は次の3つである。

a) *in vitro* Virus 法の改良: 試験管内の遺伝子型表現型対応付け手法である *in vitro* virus 法は、リボソーム上で新生タンパクを mRNA にリンカーを介して結合させるものであるが、このリンカーを工夫することによりウイルス粒子 (mRNA - タンパク結合体) 形成効率を改善した。今回のリンカーは、3'末端にピュロマイシンを結合したスペーサーをぶら下げた OMe 核酸オリゴマーが mRNA の 3'末端に Y 字状にハイブリダイゼーションしたものである。この Y 字状の一本鎖部分の 2 つの末端を RNA リガーゼで閉じてヘアピンループにすることにより (Y ライゲーションと呼ぶ)、ピュロマイシンと mRNA が共有結合で結ばれたものが形成される。この方法によれば、mRNA とリンカーは当量でも 100% 近く進行し、形成効率は従来法に比べて遙かに高い。また、*in vitro* virus のライフサイクルを 1 回回したところ、増殖率も従来法より高いことが確認された。

*In vitro* virus の増殖段は、従来 RT-PCR 法で行われているが、これを 3SR 等の等温核酸増幅法に転換すれば、自然淘汰型進化リアクターが構築できる。3SR 法は従来 37 °C の等温過程で行われてきたが、プライマーのハイブリダイゼーション特異性を上げるために高温で行うことが望ましい。今回は耐熱性 T7RNA ポリメラーゼを用いて 50 °C の等温過程で増幅する系を構築した。進化リアクターをエミュレートする継代植継ぎ実験にも成功した。ポイントは反応液に 1.4 M トレハロースを投入することである。また、T7 プロモータの 50 °C の最適配列を RNA-Z 法進化リアクターを用いた *in vitro* selection で求めた。37 °C の野生型プロモータとハミング距離 2 だけ離れた配列であった。

b) 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ない。また、酸化によるタンパク質変性を避けるためには Met を含んでいない方がよい。あるいは、全てのアミノ酸が等確率で出現するようなライブラリーが望まれる場合もある。このように、コドンの出現頻度を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 2 種類開発した。上記と同じ Y ライゲーションを利用したブロックシャフリング法 (YLBS 法) と、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 法である。両者とも、従来法より高品質の初期ランダムペプチドライブラリーを構築できる。

c) タンパク質配列空間の中立ネットワークの探査: RNA 配列空間上でその存在が確認された中立ネットワークとネットワーク間のインターチェンジが、タンパク質配列空間上の大域的適応度地形としても存在するかどうか問題となっている。われわれは 108mer の特定のフォールド 4 種 (型, 型, + 型, / 型) の逆フォールディング問題を太田のポテンシャルを用いて解くことによりこの問題に取り組んだ。各フォールドを与える配列は配列空間をほぼパーコレートしていることがわかり、中立ネットワークは存在するようである。また、全く異なるフォールドの中立ネットワークが互いにハミング距離 5 という近距離まで接近することもわかった。また副産物として、特定のフォールドをするランダム配列球状タンパク集団のエネルギー分

布を与える統計学的公式を発見した。これにより、モンテカルロシミュレーションによってエネルギー分布を求める必要がなくなった。

- (3) Aita T., Hamamatsu N., Nomiya Y., Uchiyama, N., Shibana, Y., Husimi, Y., Surveying a Local Fitness Landscape of a Protein with Epistatic Sites for the study of Directed Evolution, *Biopolymers* 64, 95-105 (2002)
- (4) Husimi Y., Aita T., Tabuchi I., Correlated Flexible Molecular Coding and Molecular Evolvability, *J. Biol. Phys.* 28, 499-507 (2002)
- (5) Tabuchi I., Soramoto S., Suzuki M., Nemoto, N., Husimi, Y., An efficient ligation in the making of in vitro virus for in vitro protein evolution, *Biological Proc On-Line* 4, 49-54 (2002)
- (6) Kitamura, K., Kinoshita Y., Narasaki, Nemoto, N., Husimi, Y., Nishigaki, K., Construction of block shuffled libraries of DNA for evolutionary protein engineering: Y-ligation based block shuffling, *Protein Engineering* 15, 843-853 (2002)
- (7) Husimi, Y., Evolutionary molecular engineering and the evolvability of biopolymers, 7th International Symposium on Bio-Nanoelectronics: Creation of Nano-Micro Structures in Bioscience & Technology, 7, 14-28 (2002)
- (8) Soramoto, S., Ueno, S., Tabuchi, I., Husimi, Y., Design of Various High Quality Random Libraries for in vitro Protein Evolution, *Genome Informatics*, 13, 527-528 (2002)

### (3) 理論解析

山村の成果は次のとおりである。創発システムとは、何もかも人為で作りこむ従来の工学から、システムがボトムアップに生み出す秩序を積極的に利用する新しいシステムの考え方である。遺伝的アルゴリズムに代表される、生命から啓発されたシステムは、そのための強力な要素技術を提供する。遺伝的アルゴリズムにおける交叉というユニークなオペレータの効果について確率過程論を用いた理論解析を行い、進化計算の設計規範として機能分担仮説を提案して、特に関数最適化において定量的な理論設計を可能とした。

- (9) 染谷博司, 山村雅幸, 探索オペレータの機能分担を考慮した進化型計算による関数最適化, *電気学会論文誌 C*, Vol.122-C, No.3, pp.363-373 (2002).
- (10) Hiroshi Someya, Masayuki Yamamura, Robust Evolutionary Algorithms with Toroidal Search Space Conversion for Function Optimization, In *Proceedings of the Genetic and Evolutionary Computation Conference 2002 (GECCO-2002)*, 553-560, (2002).

樺島の成果は次のとおりである。情報化社会の進展に伴い様々な分野でデータが大規模化している。そのため、それらを適切に処理する統計モデルへの期待が高まっているが、多くの場合それらの実行は計算論的に困難であるため質の良い近似アルゴリズムの開発が必要となる。1980年代末、確率推論の研究で開発された信念伝播法は大規模な確率モデルの汎用的近似アルゴリズムとして期待されている計算手法の一つである。この手法は多数の要素間の単純な相互作用に基づき状態を反復することで計算を実行するアルゴリズムであり、分子計算など「自然現象

を真似た計算に示唆するところも大きい。従来、信念伝播法は確率モデルにおける要素間の依存関係が木で表現される場合の動作について詳しく調べられてきたが、依存関係に循環経路が存在する場合に関しては相互作用による反復が収束するか否かも含めてその動作が十分には明らかにされていない。本研究では、信念伝播法の動作についての知見を蓄積するために、統計力学で知られているスピングラスモデルに適用した場合のダイナミクスを調べた。その結果、巨視的な軌跡の発展規則ならびに微視的な収束性の判定条件を明らかにすることに成功した。また従来、信念伝播法は確率モデルにおける要素間の依存関係が木や疎結合グラフで表現されている場合の有効性が示されていたが、無数の循環経路が発生する密結合グラフに関してはそれほど良い性能を与えないのではないかと考えられてきた。本研究では、密に結合した確率モデルからの推定問題となる CDMA マルチユーザー復調問題に対して信念伝播法を適用し、従来法を凌ぐ高性能な復調アルゴリズムを開発することに成功した。

(11) A statistical-mechanical approach to CDMA multiuser detection: propagating beliefs in a densely connected graph Yoshiyuki Kabashima, cond-mat/0210535 (2002)

(12) Propagating beliefs in spin glass models Yoshiyuki Kabashima, cond-mat/0211500 (2002)

太田の成果は次の4つである。

a) タンパク質の立体構造予測サーバの開発：タンパク質の立体構造認識用プログラムの開発を行った。立体構造由来の 3D プロフィールと、配列由来の 1D プロフィールから混合プロフィールを作成し評価した結果、スーパーファミリーレベルの認識能で、1D プロフィールより良い精度を示した。適合度をポテンシャル関数で再評価するプログラムを実装し (PILOT)、自動構造認識コンテスト (CAFASP3) に参加した。結果はそれほど良くはなかったが、今後の課題を得ることができた。

b) タンパク質の立体構造からの活性部位推定：活性部位の推定は配列のマルチプルアラインメントからの情報抽出に大きくよっている。しかし、配列の進化パターンはかなり複雑で、配列情報のみによる予測には限界がある。そこで、立体構造の情報を機能部位推定に組み入れることを考えた。活性部位は構造安定性からの要請を免除されているので、構造を不安定化する部位によく見られる。つまり、部位置換を導入すると変異体が安定化する可能性が高い。また、タンパク質表面の窪み、例えば穴やクレフトといった、若干内部に埋もれた部位に位置することが多い。上記のような情報を利用して、マルチプルアラインメントから予測される活性部位候補 (保存部位) のうち、構造情報の観点からより確からしい部位を選ぶアルゴリズムを考案した。

c) ヒト完全長 cDNA のマルチドメインの解析：産総研、生命情報解析センターが 8 月に主催したヒト完全長 cDNA のアノテーションジャンボりに参加し、マルチドメイン構成の解析を行った。ヒトのタンパク質のドメイン構成を大域的に調べたところ、同じドメインが多く繰り返すを持つような、典型的な "動物型" のものが多いことが再確認された。その他、イントロンのフェーズとタンパク質の局在箇所との関係などを調べた。

d) 小ペプチドのフォールディングシミュレーション：20 残基からなるペプチドのシミュレーションを分子動力学法を使って実行した。今後は軌道の解析などを統計的に実行する。

- (13)Y. Isogai, M. Ota, A. Ishii, M. Ishida and K. Nishikawa, Identification of amino acids involved in protein structural uniqueness: Implication for de novo protein design, Protein Eng.15 (2002) 555-560
- (14)K. Homma, S. Fukuchi, T. Kawabata, M. Ota and K. Nishikawa, A systematic investigation identifies a significant number of probable pseudogenes in the Escherichia coli genome, Gene 294 (2002) 25-33
- (15)T. Kawabata, S. Fukuchi, K. Homma, M. Ota, J. Araki, T. Ito, N. Ichiyoshi and K. Nishikawa, GTOP: a database of protein structures predicted from genome sequences, Nucleic Acids Res.30 (2002) 294-298.
- 喜多の成果は次の通りである。評価値にノイズが加わったシステムの最適化は実験やシミュレーションを通じて行うシステム最適化には必須の技術である。遺伝的アルゴリズム(GA)はこのように用途に適した手法であるが、これまで必ずしも実応用を念頭にアルゴリズムの開発は行われてこなかった。本研究では、評価値にノイズが加わっており、なおかつシステムの評価回数が厳しく制限されている状況下で、探索履歴をメモリに蓄え、未知の評価値を探索履歴を利用しながら推定する遺伝的アルゴリズムの開発を行ったものである。
- (16)佐野泰仁,喜多 一:探索履歴を利用した遺伝的アルゴリズムによる不確実関数の最適化, 電気学会論文誌, Vol. 122-C, No.6, pp. 1001-1008 (2002).
- (17)Yasuhito Sano, Hajime Kita: Optimization of Noisy Fitness Functions by means of Genetic Algorithms using History of Search with Test of Estimation, Proc. CEC2002, pp. 360-365 (2002)

平成15年度

ウェ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ェット・ドライ実験でそれぞれ進展を見た。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。研究項目ごとに整理した研究成果は次の通りである。

(1) ウェット進化計算の分子実現

坂本の成果は次の3つである。

- a) 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。
- b) 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。
- c) 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。

(18) 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).

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。

(19) 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.

## (2) 進化リアクタの作成

伏見の進化リアクタの作成に係わる成果は次の3つである。

### a) 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。

### b) 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、長鎖化法を検討した。

### c) 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。

- (20) Aita T., Ota M., Husimi Y., An in silico exploration of neutral network in protein sequence space, *J.Theor.Biol.* 221, 599-623 (2003)
- (21) 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)
- (22) 伏見 謙, 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)
- (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.

### (3) 理論解析

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ア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。

- (24) 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).
- (25) Park, S.J. Yamamura, M.: GA-based Generic Method for Protein Structure Comparison, *Proc. 2003 Congress on Evolutionary Computation (CEC 2003)*, 1528-1535 (Dec. 2003).
- (26) 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).

伏見は, 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の理論解析においても次のような成果をあげた。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アプノフ関数 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関係に類似の関係が成立する。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関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。

- (27) 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)

(28)Aita T., Husimi Y., Thermodynamic Interpretation of evolutionary dynamics on a Mt. Fuji-type fitness landscape, Bulletin Math. Biology, 66, 1371-1403 (2004)

樺島の成果は次の5つである。

a) 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。

b) 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。

c) 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。

d) 公開鍵暗号に対する攻撃の統計力学的评价

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を評価した。

e) 低密度パリティ検査符号の統計力学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。

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

(29)Y. Kabashima and D. Saad, Statistical mechanics of low-density parity-check codes, J. Phys. A 37, R1 (2004)

(30)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)

(31)Y. Kabashima, A CDMA multiuser detection algorithm on the basis of belief propagation, J. Phys. A 36, 11111 (2003)

(32)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)

(33)Y. Kabashima, Propagating beliefs in spin glass models, J. Phys. Soc. Jpn. 72, 1645 (2003)

太田の成果は次の2つである。

- a) 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すところにある。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/> で提供した。
- b) 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。

(34) 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.

(35) 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

平成16年度

分子進化実験に参加する春木氏、前年度に重要性を確認した統計力学的理論解析を担当する井上氏の2名を新たなメンバーに加えた。前年非常に有効であった班会合は、全体会合と合同で行うこととした。分子コンピューティングの主要国際会議である DNA10 において、研究項目(1)(2)のそれぞれの成果を報告した。

(1) ウェット進化計算の分子実現

坂本の成果は次の通りである。SA 的探索を実際の分子進化実験に用いるためのプロトコールを確立した。好熱古細菌 (*M. jannaschii*) の チロシル tRNA 合成酵素 (MjYRS) の遺伝子と、この酵素によってアミノアシル化されるアンバー・サプレッサー tRNA の遺伝子を1つのプラスミド上にクローニングした。このプラスミド pSup はカナマイシンに対する薬剤耐性を有している。pSup を、もう1つのプラスミド pTest と一緒に大腸菌内に導入した。pTest はアンピシリン耐性と、アンバー・コドンによって分断された CAT 遺伝子 (CAT-Am) を有している。MjYRS が期待される活性を持っているときには、CAT-Am 遺伝子が翻訳されて、MjYRS の活性の大きさに対応して合成される CAT 遺伝子産物の量が増える。CAT 遺伝子産物は、3つめの薬剤であるクロランフェニコール (Cm) を分解する働きがあるので、CAT 遺伝子が盛んに合成されるほど大腸菌は高い Cm 耐性を示す。この性質は、様々な濃度の Cm の入った培地上で大腸菌を生やすことで用意に検定することができる。実際には、培地 1 ミリリットル中に x ミリグラムの Cm を含む培地を 7 通り (x = 0, 50, 100, 250, 500, 750, 1000) 作成して、MjYRS 変異体のスクリー

ニングに用いた。Cm0,50では1千万個の大腸菌コロニーが得られた(1つのコロニーは1匹の大腸菌が増殖した結果生じるので、コロニー中の大腸菌はみな同じ性質を持つ)。Cm250で120個、Cm500で60個、Cm750で6個という分布が得られた。Cm濃度が高くなると、生育できるコロニーの数が急速に減少することがわかる。このように多数の検体に対して活性の分布を一度に得る方法は、SAのアルゴリズムに基づいたタンパク質の人工進化実験に有用である。

染谷の成果は次の通りである。近年、タンパク質工学の重要性が高まっている。しかし、望みの機能をもつタンパク質を得るための標準的な方法は確立されていない。本稿では、タンパク質工学における望みの機能をもつタンパク質を最適解または準最適解とみなし、また、それらの発見を最適化問題とみなすことで、タンパク質工学への情報科学的接近を試みる。まず、計算機上における最適化とタンパク質工学との違いについて述べ、これらを考慮した遺伝的アルゴリズムによる一手法を示した。次に、提案手法を最適化問題に適用しその有効性を確認した。また、提案手法のウェットな実現への可能性を議論した。

これらは共同研究の成果としてDNA10で報告された。

(36)K. Sakamoto, M. Yamamura, H. Someya : Toward "Wet" Implementation of Genetic Algorithm for Protein Engineering, In Preliminary Proceedings of 10th international Workshop on DNA-Based Computers, Milan, Italy, 7-10 June 2004.

(37)坂本健作, 山村雅幸, 染谷博司 : "タンパク質工学のための遺伝的アルゴリズムの設計", 電気学会 電子・情報・システム部門大会 2004 講演論文集, (OS5), Sep 2004.

## (2) 進化リアクタの作成

山村は複数の遺伝子と調節領域を含むプラスミドを用いて、各遺伝子の調節領域を交叉を用いて進化させることにより、遺伝子群からなるシステムを進化させるウェット進化計算の発展形を開発した。大腸菌中で、複数の遺伝子を含むプラスミド2つが一箇所交叉して合体し、その後もう1箇所交叉して2つに分かれる2段階の処理によって、プロモータ領域を交換することにより、発現量を調整する実数値ベクトルの最適化を実現した。伏見は前年に引き続いて進化リアクタの改良に取り組んだ。

(38) Wakabayashi K., Yamamura M.: A design for cellular evolutionary computation by using bacteria, Preliminary Proceedings of 10<sup>th</sup> International meeting on DNA based computers (DNA10), 304-313 (2004).

## (2) 理論解析

春木および大田によるタンパク質立体構造論からの考察が加えられた。

(39)Suzuki., Y. Haruki, M., Takano, K., Morikawa, M., Kanaya, S.:Possible involvement of an FKBP family member protein from a psychrotrophic bacterium *Shewanella* sp. SIB1 in cold-adaptation. *Eur J Biochem.* 271, 1372-81.

(40)Y. Hioki, K. Ogasahara K, S. Lee, J. Ma, M. Ishida, Y. Yamagata, Y. Matsuura, M. Ota, M. Ikeguchi, S. Kuramitsu, K. Yutani. The crystal structure of the tryptophan synthase beta subunit from the hyperthermophile *Pyrococcus furiosus*, *Eur. J. Biochem.* 271

(2004) 2624-2635

(41) T. Imanishi et al. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. PLoS. Biol. 2 (2004) E162

平成 17 年度

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ェット進化計算の実現実験を依頼した。

各研究項目の研究成果は次のとおりである。

(1) ウェット進化計算の分子実現

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、ウェット 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化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。

春木の成果は次の 2 つである。

a) 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 H I 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。

b) 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、 $\alpha$ -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 に置換した変異体が得られ、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。

(42) 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).

(43) Someya, H. Yamamura, M.: "A Robust Real-Coded Evolutionary Algorithm with Toroidal Search Space Conversion", *Soft Computing*, Vol.9, No.4, pp.254-269, Apr 2005.

(44) 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 論文集, pp.466-469 (OS3-4), Sep 2005.

(45) 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)

(46) 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)

## (2) 進化リアクタの作成

伏見の成果は次のように実験・理論の 2 つである。

a) 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。

#### b) 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」 $I_{fit}$  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」 $I_{Sha}$  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」 $I_{fit}$  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  $I_{bio}$  はこの進化系の Lyapunov 関数であることが証明でき、進化は  $I_{bio}$  の最大状態に向かって進行する。そこで、我々は  $I_{bio}$  を「生物学的情報量」と呼ぶことにした。 $I_{bio} = I_{fit} + I_{Sha}$  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。

(47) 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)

(48) 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)

#### (3) 理論解析

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

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。

太田の成果は次の4つである。

- a) 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。
- b) 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置換体はそれぞれヘリックス含量をしめし、弱くヘムと結合した。
- c) 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。
- d) 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。

- (49) S. Uda and Y. Kabashima, Statistical Mechanical Development of a Sparse Bayesian Classifier, *J. Phys. Soc. Jpn.* 74, 2233-2242 (2005)
- (50) Y. Kabashima, Replicated Bethe Free Energy: A Variational Principle behind Survey Propagation, *J. Phys. Soc. Jpn.* 74, 2133-2136 (2005)
- (51) 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)
- (52) Y. Kabashima and T. Hosaka, Statistical Mechanics of Source Coding with a Fidelity Criterion, *Prog. Theor. Phys. Suppl* 157, 197-204 (2005)
- (53) D. Saad, N. Skantzos and Y. Kabashima, Security and Reliability of LDPC Based Public-Key Cryptosystems, *Prog. Theor. Phys. Suppl.* 157, 229-236 (2005)
- (54) 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
- (55) 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

(56)K. Kinoshita and M. Ota, P-cats: Prediction of catalytic residues in proteins from the tertiary structures, *Bioinformatics* 21 (2005) 3570-3571.

平成18年度

ウェット進化計算は理論・実験両面で成果をあげはじめた。木賀氏を分担者として迎え、新たな展開の方向として構成的生物学について検討した。木賀を中心として Tokyo Alliance と称するチームを率いて、第3回国際人工遺伝子回路コンテストに参加し、最優秀部品賞、最優秀共同研究賞のふたつを受賞した。

#### (1) ウェット進化計算の分子実現

染谷の成果は次のとおりである。タンパク質工学への応用を念頭においた分子計算の実装として、「探索が任意のある一点から開始される」という条件のもとでの組合せ最適化のための確率的最適化の手法を検討した。タンパク質工学における望みの機能をもつタンパク質の発見の過程を情報科学における最適化とみなし、まず、マルコフ連鎖モンテカルロ法を応用したシミュレートッド・アニーリングなどの既存の確率的最適化手法の DNA 分子による実装、すなわち分子計算による実現の可能性について主に理論的観点から検討し考察を示した。次に、遺伝的アルゴリズムに基づいた一手法を提案した。提案手法の有効性は計算機実験により確認され、DNA 分子を用いた実現の可能性を議論した。

研究協力者の若林の成果は次のとおりである。wetGA によるタンパク質の機能改変を目指した。野生型 TyrRS の遺伝子にランダムミューテーションを導入して構成した変異体集団を初期集団とし、PCR による交叉と新規活性強度を指標とした評価・選択を行った。その結果、本来の野生型 TyrRS が認識しない基質を認識可能な変異型 TyrRS が多数得られた。これらの変異体と認識能を持たない変異体の両者を用いて次世代の親集団を構成し、2 世代目以降も同様の手順に従って交叉と選択を繰り返したところ、1 世代目よりも活性の高い個体を得ることができた。

研究協力者の坂本の成果は次のとおりである。アミノアシル tRNA 合成酵素の進化実験と基質特異性の改変を進めた。この結果、新規非天然型アミノ酸を認識する変異体(1 種類)の作成に成功している。また、昨年度までに作成した、ヨードチロシン特異的な古細菌チロシル tRNA 合成酵素変異体の結晶構造を明らかにし、変異体中のアミノ酸置換がどのようにして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なり、本来の基質であるチロシンを認識しないようになったのかを明らかにした。改変されたこれらのアミノアシル tRNA 合成酵素は、大腸菌、動物細胞、無細胞翻訳系において非天然型アミノ酸をタンパク質に組み込むために使用された。このような非天然型アミノ酸含有タンパク質は、X 線結晶構造解析、酵素の触媒機構の解明、細胞内のタンパク質相互作用ネットワークの解析などに役立っている。

春木の成果は次の3つである。

#### a) ランダム変異および DNA シャッフリングによる大腸菌リボヌクレアーゼ H の耐熱化

大腸菌リボヌクレアーゼ HI (RNase HI) にランダム変異を導入した断片を制限酵素処理 (BamHI) によって切断した後に再結合し、耐熱化変異体のセレクションを行った。その結果、シャッフリングは行わずさらに変異を導入した場合に比べて、耐熱化変異体の割合が大きく増加

した。シャッフリングにより安定化を損なうような変異が除去されたため、耐熱化したものの割合が増加したと考えられる。セクションによって得られたクローンについて、導入されている変異の解析を行ったところ、His83 Arg 変異、Glu91 Arg 変異が見いだされ、これらの変異は pH 5.5 における  $T_m$  を、それぞれ 1.0、1.1 向上させた。

#### b) 大腸菌リボヌクレアーゼ HI の耐熱化機構の熱力学的解析

熱安定性を向上させるアミノ酸置換が、蛋白質の熱力学的性質に与える影響について解明することを目指し、熱安定性を向上させる 5 種類のアミノ酸置換 (Gly23 Ala, His62 Pro, Val74 Leu, Lys95 Gly, Asp134 His) を組み合わせることにより、野生型酵素に比べて変性温度 ( $T_m$ ) が 20.2 向上した大腸菌 RNase HI 変異体 (5H) について、変性の自由エネルギー変化 ( $\Delta G$ ) の温度依存性の解析を行った。その結果、5 H 変異体では  $\Delta G$  の最大値は野生型より約 4.4 kcal/mol 大きくなり、 $C_p$  は野生型と変わらず (2.7 kcal mol<sup>-1</sup> K<sup>-1</sup>)、野生型の温度依存曲線が上方へそのままシフトした形となった。これに対し、高度好熱菌由来の RNase HI では大腸菌の野生型と比べて  $C_p$  が小さくなっており (1.8 kcal mol<sup>-1</sup> K<sup>-1</sup>)、曲線が上方へシフトし、傾きも緩やかとなっている。従って、上記の 5 種類のアミノ酸置換の導入は、高度好熱菌 RNase HI の耐熱化戦略のうち、 $\Delta G$  温度依存曲線を上方へシフトさせることによる耐熱化に貢献していると考えられる。

#### c) 進化工学的手法による大腸菌エステラーゼ活性向上変異体の探索

大腸菌エステラーゼ遺伝子にランダム変異を導入し、ハロー形成能の向上を指標として活性向上変異体の探索を行った。その結果、Arg48 Ser 変異が見いだされ、この変異体酵素の酵素反応速度論的解析を行ったところ、野生型に比べて  $K_m$  値が 1.4 倍、 $k_{cat}$  値が 32 倍に増大しており、 $k_{cat}/K_m$  値は 22 倍に向上していた。大腸菌エステラーゼの立体構造モデルにおいて、Arg48 は基質のエステル結合部分に近接しているため、正電荷をもつ Arg から電荷をもたない Ser へ置換されたことにより、負電荷をもつ生成物のカルボン酸イオンが解離しやすくなり、その結果  $k_{cat}$  値が向上したと考えられる。

木賀の成果は次のとおりである。最近、工学一般における部品や作業の標準化、という観点から遺伝子工学を改めて見直すことで、遺伝子工学に工学一般の成果を適用しよう、という方針が提唱されはじめている。遺伝子プログラムに「どのような部品を組み込むか」ということを記した情報として、各種タンパク質の遺伝子が統一的な操作で組み合わせられるようなライブラリ、BioBrick が整備されつつある。このライブラリには種々の部品が整備されつつあるものの、「いつ部品を作るのか」というプログラムについてはまだ整備が不十分である。特に、複数の条件を同時に判定する命令を自由に書き込むためのフォーマットが整備されていない。本研究では、「いつ部品を作るのか」という条件を遺伝子プログラムに書き込むための、下記で詳述するフォーマットを提案し、実際に条件判定を行うためのパーツを構築した。さらに、個々のパーツを組み合わせることで AND 論理演算が可能であることを示した。

本研究の遺伝子プログラムのフォーマットは、LacI や AraC などの転写制御タンパクと特異的に結合する条件判定用 DNA 配列群を GFP コード配列の上流に導入することを基本としてい

る。このフォーマットを可能にするために、特別なプラスミドを構築した。複数の規格化オリゴを順々に組み合わせて導入して、多様な小分子の存在パターンによって制御可能な系を容易に構築することができる。こうして構築したプラスミドで大腸菌を形質転換し、GFP 発現の小分子依存性を調べた。文献をもとに大腸菌内で働く転写制御タンパク 10 種類それぞれの結合配列をデザインし、規格化された部品とした。またそれらをプラスミドに組み込んで蛍光の変化を測定したところ、lacI, AraC, LuxR, EmrR タンパク質が結合する DNA では、それぞれ対応する小分子(IPTG, アラビノース, アシルホモセリンラクトン, サリチル酸)の添加により蛍光量が上昇することが確認された。さらに LuxR 結合配列と LacI 結合配列二種類を同時に組み込んだプラスミドでは、両方の小分子を同時に添加したときのみ蛍光量が大きく上昇することが確認された。これは出力が二つの入力の共存に依存する AND 回路であるといえる。本研究で提案した手法による制御部品のライブラリが多様化すれば、より多彩な制御が可能になる。その結果より複雑で有用な人工遺伝子回路の構築も可能となると期待される。

(57)染谷博司：“有望探索領域を重点探索する交叉と形質遺伝性を考慮した世代交代による実数値 GA”，計測自動制御学会 システム・情報部門学術講演会 2006 講演論文集, pp.3-8, Nov 2006. (計測自動制御学会 システム・情報部門 奨励賞)

(58)Hiroshi Someya：“Promising Search Regions of Crossover Operators for Function Optimization”, In Proceedings of The 20th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2007), to appear, June 2007.

(59)Morikawa, M., Kagihiro, S., Haruki, M., Takano, K., Branda, S., Kolter, R., Kanaya, S., Biofilm formation by a *Bacillus subtilis* strain that produces gamma-polyglutamate, *Microbiology*, 152, 2801-2807 (2006).

(60)Mukaiyama, A., Haruki, M., Ota, M., Koga, Y., Takano, K., Kanaya, S., A Hyperthermophilic Protein Acquires Function at the Cost of Stability, *Biochemistry*, 45, 12673-12679 (2006).

## (2) 進化リアクタの作成

伏見の成果は次の2つである。

### a) 自律的に複製する in vitro ウイルスに向けて

蛋白質を進化させる場合、突然変異蛋白質とその遺伝子とを対応付けなければ蛋白質の機能の評価が、遺伝子増幅にフィードバックできず、ダーウィン淘汰の原理が働かない。この対応付け戦略には次の3種がある。(1) 両者を物理的に結合する戦略、これは f d フェージなどの単純なウイルスが採用している戦略なのでウイルス型戦略と呼んでいる。(2) 両者を一つのコンパートメントに入れる戦略、これは単細胞生物が採用している戦略なので細胞型戦略と呼んでいる。(3) 外部知性が最適蛋白質を選択分離し、そのアミノ酸配列を読み取る戦略。このうち、ウイルス的戦略は、ピューロマイシンをリンカーとして、mRNA と新生蛋白質を無細胞翻訳系で結合する in vitro ウイルス (IVV) として実現した。これは蛋白の C 末端と、mRNA の 3' 末端

を結合したものである。

今回は、mRNA の 5' 末端と、蛋白質の N 末端付近を結合することによる IVV を作製した (Ueno,S., et al, 投稿中)。リンカーは、5' 末端に Y ライゲーションでヘアピン状に末端を結合したスペーサーのもう片方の末端に結合した sup tRNA である。非天然アミノ酸を取り込む技術を応用して、スペーサーにリンクしたアミノ酸が蛋白質の N 末端付近に取り込まれ、ウイルスの対応づけが完成する。His タグをもつ IVV と FLAG タグをもつ IVV との混合物から、Ni-NTA または抗 FLAG 抗体によって選択的濃縮が起こることを確認した。従来の IVV は、C 末端が自由端であることが活性に必須であるような蛋白質や、C 末付近に活性中心がある蛋白質の進化実験には使えなかったが、この IVV は使える。また、RNA レプリカーゼの IVV を念頭に置いた場合 RNA レプリカーゼが自己の mRNA を複製するには自由端の 3' 末端を要求する。従って、従来の IVV では、実現できなかったウイルス粒子が自律的複製をする RNA レプリカーゼ IVV が、今回の IVV では実現できそうである。

#### b) 等温核酸増幅進化リアクタープロセスによるプロモーターの進化

1.4M トレハロース存在下では、50 の高温条件下で、RNA-Z 等温核酸増幅が安定に行いうる。これを継代植継ぎ実験することによって、TT7 RNA ポリメラーゼに対する 50 における最強のプロモータを取得したが、それは上流コンテキスト依存であった。上流のコンテキストをもランダム配列化すると、集団は一気に最終産物に収束した。下流の転写開始部位は通常 G であるが、これを A に変えて実験すると、やはり、A が G に置換する進化が起こった。コンテキストの変化を伴うコンテキスト依存の淘汰が逐次起こる進化過程が再び観察された。50 の最適コンテキスト中の最強のプロモータは RNA-Z 反応を安定化するので、これを用いて、ウイルス粒子が自律的複製をする RNA ポリメラーゼの IVV を構築出来そうである。

(61)Hamamatsu,N., Suzumura,A., Nomiya,Y., Sato,M., Aita,T., Nakajima,M., Husimi,Y., Shibanaka,Y., Modified substrate specificity of pyrroloquinoline quinone glucose dehydrogenase by biased mutation-assembling with optimized amino acid substitution, *Applied Microbiology and Biotechnology* 73, 607-617 (2006)

(62)Hamamatsu,N., Nomiya,Y., Aita,T., Nakajima,M., Husimi,Y., Shibanaka,Y., Accumulation of amino acid substitutions for thermostabilization of lactate oxidase with less deterioration of catalytic activity, *Protein Eng. Design and Selection* 19, 483-489 (2006)

(63)Biyani, M., Husimi, Y., Nemoto, N., Solid-phase translation and RNA-protein fusion: a novel approach for folding quality control and direct immobilization of protein using anchored mRNA, *Nucleic Acids Res. Advanced publication online* (2006)

(64)Naimuddin M, Kitamura K., Kinoshita Y, Honda-Takahashi,Y., Murakami,M., Ito,M., Yamamoto,K., Hanada,K., Husimi,Y., Nishigaki,K., Selection-by-function: efficient enrichment of cathepsin E inhibitors from a DNA library, *J.Mol.Recognition*, published online (2007)

(65)Hayashi,S., Aita,T., Toyota,H., Husimi,Y., Urabe,I., Yomo,T., Experimental Rugged Fitness Landscape in Protein Sequence Space, *PLoS ONE*, accepted (2007)

(66) Aita, T., Hayashi, S., Toyota, H., Husimi, Y., Urabe, I., Yomo, T., Extracting characteristic properties of fitness landscape from in vitro evolution: A case study on infectivity of fd phage to E. coli, J. Theor. Biol. accepted (2007)

(67) 相田 拓洋, 伏見 譲, 実験室内分子進化過程の熱力学的および情報論的解釈, 生物物理, 46, 137-143 (2006)

### (3) 理論解析

樺島の成果は次のとおりである。多数のユニットが複雑に結合した系を数理的に解析するための方法論について研究した。1つ目は、系の相互作用を結合行列の固有値分布によって特徴づける解析法である。これにより、系の振る舞いを一定の条件下結合行列の固有値分布により特徴づけることが可能になった。2つ目は、系の性質を平均場近似により解析するための効率的な数値最適化アルゴリズムの開発である。これを3体スピングラス模型に応用し、最近提唱された理論予想を支持する結果を得た。

(68) Y. Tonosaki, K. Takeda and Y. Kabashima, Numerical study of Thouless-Anderson-Palmer metastable states in three-body Ising spin glasses, To appear in Phys. Rev. B

(69) K. Takeda, S. Uda and Y. Kabashima, Analysis of CDMA systems that are characterized by eigenvalue spectrum, Europhysics Letters 76, 1193-1199 (2006)

(70) Y. Kabashima, Tutorial on Brain-Inspired Computing Part 5: Statistical Mechanics of Communication and Computation, New Generation Computing 24, 403-420 (2006)

(71) JPL. Hatchet and Y. Kabashima, Survey propagation for the cascading Sorkin code, JOURNAL OF PHYSICS A-MATHEMATICAL AND GENERAL 39 (34): 10659-10672 (2006)

## 5. 今後の課題

平成16年度にDNA10で公表した具体的課題に対する2つの分子実現は、本計画研究にとって重要な段階である。情報科学から生物学への貢献について目処がたったことは本研究課題の大きな成果と考えている。今後はこれらの分子実現を改良するとともに、理論解析についてもこれらの分子実現をサポートできるモデルが必要である。また、本研究課題の成果は広義の構成的生物学上の成果に直接発展するものであり、今後の展開が期待できる。具体的に研究項目ごとにまとめた課題は次の通りである。

### (1) ウェット進化計算の分子実現

a) 進化実験系を改良する。実験的には、開発したプロトコルを本格的に適用し、高い活性を持つ酵素を得る。理論的には、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)を行い、場合によってはすべての個体のシーケンスを調べて、集団に蓄えられている突然変異の分布が、世代交代に伴ってどのように変化してゆくか追跡する。そのデータから最適な世代交代方式をおよび各パラメータの感度を調べる。

b) 他のタンパク質の進化に応用する。プロトコルとしての汎用性を調べるとともに、特定のタンパクに特化された部分を改善する。

#### (2) 進化リアクタの作成

a) 複数遺伝子からなるシステムの進化について、具体的例題を選んで分子実現を改良する。毒性を有するタンパクなど、微妙な発現量調整を要する例をいくつか検討し、実際に進化させる。

b) 分子人工生命としての試験管ウイルスを改良する。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。

#### (3) 理論解析

生体高分子の大域的適応度地形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興味深い知見を得たので、この理論を発展させる。

#### (4) 構成的生物学

構成的生物学における機能実現の主要手段である人工遺伝子回路が、現代の機械部品や電子部品のように、利用者の目的に応じて自在に組み合わせられるためには、汎用化・標準化の方向への展開が必要である。近現代の工学の発展には、特定の用途に特化されない部品の汎用化・標準化の長い伝統が大きく貢献している。現在の分子生物学はその意味で手工業の段階にある。このことは生物学者自身も感じており、パラダイム上でのブレークスルーを待望している。ゲノム研究は汎用化・ハイスループット化の端緒ではあったが、人海戦術の単なる規模の拡大でもあった。構成的生物学は、アイデアの実現のために本質的に汎用化・標準化を志向しており、生物学に産業革命を引き起こす契機となるものと期待している。

## 6. 研究費の使用状況

### 平成 14 年度

主な設備品として、(1) ウェット進化計算の分子実現のために、蛍光検出装置等を購入、シミュレーションのために PC クラスタードを購入した。(2) 進化リアクタの作成のために、PCR 装置、分光光度計等を購入した。(3) 理論考察のために、ワークステーションを購入した。国内海外旅費・試薬等の消耗品のほか、データ整理のための学生アルバイトを雇用した。

### 平成 15 年度

主な設備品として、(1) ウェット進化計算の分子実現のために、ラボシェルフ、冷却遠心機等を購入、シミュレーションのためにワークステーションを購入した。(2) 進化リアクタの作成のために、高性能 PCR 装置、卓上遠心機等を購入した。(3) 理論考察のために、分子の立体構造モデリングソフト InsightII を購入した。国内海外旅費・試薬等の消耗品のほか、データ整理のための学生アルバイト、技術支援者として平山則子氏を雇用した。

### 平成 16 年度

主な設備品として、(2) 進化リアクタの作成のために窒素ガス発生装置を購入した。NA10 参加旅費を含む国内海外旅費・試薬等の消耗品の支出のほか、データ整理のための学生アルバイト、技術支援者として平山則子氏を引き続き雇用した。

平成17年度

主な設備品として、(1) ウェット進化計算の分子実現のために紫外線照射機コントローラを購入した。(2) 進化リアクタの作成のために蛍光スキャナーを購入した。DNA11 参加旅費を含む国内海外旅費・試薬等の消耗品の支出のほか、データ整理のための学生アルバイト、技術支援者として平山則子氏を引き続き雇用した。

平成18年度

主な設備品として、(1) ウェット進化計算の分子実現のためにグラジエント付サーマル細工らーを購入した。DNA12 参加旅費を含む国内海外旅費・試薬等の消耗品の支出のほか、技術支援者として平山則子氏を引き続き雇用した。

以上