

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す。

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索する。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する。

- (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、設計論の確立をめざす。

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。

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される。

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る。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・大学院総合理工・教授): 進化計算の理論設計(1,2,3)
- ・ 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): ウェット進化計算の実現(1,3)
 - 染谷博司 (統計数理研究所・助手): 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・工学部・教授): 進化リアクターの構築(2,3)
 - 春木 満 (日本大学・工学部・助教授): 分子進化論からの検討(2,3)
- ・ 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(3)
 - 井上真郷 (東京工業大学・大学院総合理工・助手): 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・教授): 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・学術国際センター・助教授): 立体構造論からの検討(3)
 - 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(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。

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細な調査を始めた。

- (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).
- (3) 坂本健作「DNA コンピュータ」蛋白核酸酵素 (共立出版) Vol. 47, No. 15, pp. 2037-2044, 2002 年

(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'末端にピュー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)、ピューロマイシンと 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 3 SR 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1.4M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の問題となっている。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。

(4) 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)

(5) Husimi Y., Aita T., Tabuchi I., Correlated Flexible Molecular Coding and Molecular Evolvability, *J. Biol. Phys.* 28, 499-507 (2002)

(6) 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)

(7) 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)

- (8) Aita T., Ota M., Husimi Y., An in silico exploration of neutral network in protein sequence space, *J.Theor.Biol.* 221, in press (2003)
- (9) Aita T., Husimi Y., Statistical Formulae for Energy Distribution among a Globular Protein Structure Ensemble, *J.Theor.Biol.* 220, 107-121 (2003)
- (10)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)
- (11)Aita,T.,Husimi,Y., Acceleration of Encoded Protein Evolution: in vitro virus, Analysis of local fitness landscapes and Adaptive jump, Santa Fe Institute Workshop on Robustness and Evolvability of Molecules and Microbes, 10 (2002)
- (12)Aita,T.,Husimi,Y., Formulation of the energy distribution among a globular protein structure ensemble for the high-throughput screening of sequences, *Genome Informatics* 13, 346-347 (2002)
- (13)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)
- (14)伏見 謙編著 , シリーズ・ニューバイオフィジクス 第 8 巻「生命の起源と分子進化の物理学」(共立出版, 2002)
- (15)Tabuchi I., Soramoto S., Ueno S., Husimi Y., Multi-Line Split DNA Synthesis: a combinatorial method to make a high quality peptide library, *Nucleic Acids Res.* submitted.

(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。

- (16)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).
- (17)Daisuke Matsuda, Masayuki Yamamura, Cascading Whiplash PCR with a Nicking Enzyme, *Proceedings of 8th International Meeting on DNA Based Computers (DNA8)*, to appear (2002).
- (18)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).
- (19)山村 雅幸, アクエラスコンピューティング - 生体分子による並列メモリの実現 -, *電気学*

会誌, Vol.122, No.3, 156-159 (2002).

(20)Sung-Joon Park, Masayuki Yamamura, FROG (Fitted Rotation and Orientation of protein structure by means of real-coded Genetic algorithm) : asynchronous parallelizing for protein structure-based comparison on the basis of geometrical similarity, Genome Informatics, 13:344-345, 2002

(21)山村雅幸, 進化計算の機能分担仮説, 電気学会ナチュラコンピュテーション共同研究委員会・産業システム情報化研究会資料, 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。

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

(23)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。

(24)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.* in press

(25)T. Aita, M. Ota and Y. Husimi, An in silico exploration of the neutral network in protein sequence space, *J. Theor. Biol.* in press

(26)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

(27)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

(28)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.

(29)太田元規 「あなたにも役立つバイオインフォマティクス」菅原秀明編 共立出版 (2002) 第 9 章 アミノ酸配列から 2 次構造を予測する (p69-76) 第 10 章 アミノ酸配列から立体構造を予測する (p77-84)

(30)若山守, 森口充暲, 太田元規, 西川建 タンパク質の立体構造予測 結晶化しない酵素への応用 *化学と生物* 40 (2002) 452-459

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)はこのような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。

(31)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).

(32)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。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) ウエ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。

(33)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。

(34)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.

(35) 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.

(36) 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.

(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 . 伏見の理論解析に関わる成果については後述する .

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 5 0 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 1 0 1 6 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 .

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 .

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

(38) 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)

(39)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)

(40)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)

(41)伏見 謙, 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)

(42)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。

(43)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).

(44)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).

(45)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).

(46)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).

(47)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).

(48)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)

(49)Sung-Joon Park, Masayuki Yamamura: Real-Coded Genetic Algorithm to Reveal Biological Significant Sites of Remotely Homologous Proteins, Proc. GECCO2003,

1602-1603 (2003).

(50) 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).

(51) 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).

(52) 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).

(53) 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).

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ャブノフ関数 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.

(54) 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)

(55) Aita T., Husimi Y., Thermodynamic Interpretation of evolutionary dynamics on a Mt. Fuji-type fitness landscape, *Bulletin Math. Biology* (2004) in press

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モデルの解析

スピングラスモデルは, 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.

e) 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.

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.

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

(57)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)

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

(59)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)

(60)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/> にて行っている.

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.

(61)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.
- (62)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
- (63)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
- (64)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)
- (65)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)

平成16年度 (平成16年9月現在)

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 の入った培地上に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。

- (66) Sakamoto, K., Ishimaru, S., Kobayashi, T., Walker, J. R., Yokoyama : The Escherichia coli argU10(Ts) Phenotype Is Caused by a Reduction in the Cellular Level of the argU tRNA for the Rare Codons AGA and AGG. S. Journal of Bacteriology (in press)
- (67) 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.
- (68) 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, pp.xxx-xxx (OS5-x), 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でいる。

- (69) Kenichi Wakabayashi and Masayuki Yamamura : A design for cellular evolutionary computation by using bacteria, Proceedings of 10th International Workshop on DNA-Based Computers, Milan, Italy, June 2004.
- (70) Aita T., Husimi Y., Thermodynamic Interpretation of Evolutionary Dynamics on a Fitness Landscape in an Evolution Reactor I, Bulletin Math. Biol. 66, 1371-1403 (2004)
- (71) Tabuchi, I., Soramoto, S., Ueno, S., Husimi, Y.: Multi-Line Split DNA Synthesis: a combinatorial method to make a high quality peptide library, BMC Biotechnology, in press, (2004)
- (72) Suzuki, M., Ito, Y., Savage, H.E., Husimi, Y., Douglas, K.T.: Protease-sensitive signalling by chemically engineered intramolecular fluorescent resonance energy transfer (FRET) mutants of Green Fluorescent Protein, Biochem. Biophys. Acta, in press, (2004)
- (73) 伏見 謙, 生命の起源と進化への生物物理学的アプローチ, 学術の動向, 9, 70-73 (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。

- (74) 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.
- (75) Y. Hioki, K. Ogasahara, 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
- (76) T. Imanishi et al. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS. Biol.* 2 (2004) E162
- (77) 太田元規, タンパク質の安定性と機能部位, 生物物理学会誌, in press

5. 今後の課題

今年度 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。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。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。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。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ことを目標とする。

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が指摘されたので、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、実験系の改善の可能性を見積もる。

(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質など、微妙な発現量調整を要する例をいくつか検討し、実際に進化させてみる。

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興味深い知見を得たので、この理論を発展させる。

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 クラスタノードを購入した。(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

設備品の購入はまだない。DNA10 参加旅費を含む国内海外旅費・試薬等の消耗品の支出がある。

以上