2.2.1 当該年度(平成27年度)における研究成果

○ 研究開発課題

Ⅰ　戦略課題１：細胞内分子ダイナミクスのシミュレーション[統括：杉田　有治（理化学研究所）]

細胞質中の分子混雑、生体膜環境、膜を介した物質及び信号伝達など細胞環境を強く意識した分子および細胞スケールシミュレーションの実現を目指し、細胞内信号伝達経路の１分子粒度計算、膜タンパク質による物質輸送の解明、核内DNAタンパク質複合体の構造予測と機能解明を行う。

Ⅰ－１　実施計画

生体分子が働く細胞環境の大きな特徴の１つは、周囲の生体分子の存在による分子混雑環境である。QM/MMモデル、全原子、粗視化、1分子粒度を繋ぐマルチスケール分子モデルを用いた分子・細胞スケールシミュレーションを「京」上で行う。「細胞内環境を考慮した信号伝達経路のモデリング」と「核内混み合い環境でのヌクレオソーム、クロマチンの機能発現機構」にターゲットを絞り、細胞内分子ダイナミクスから細胞機能の理解と予測を目指す。

(1-1) 信号伝達経路上のリン酸化酵素への細胞質分子混雑の影響（杉田有治・理化学研究所）

平成27年度は、細胞質分子混雑環境に関する大規模分子動力学計算に関する解析を継続するとともに、タンパク質間相互作用や流体力学効果の影響などを調べる。また、MEK-ERK複合体の構造が分子混雑環境と溶液環境でどのように異なるのかを調べる。

(1-2) 細胞内環境下での信号伝達経路のモデリング基盤整備およびコード開発
（高橋恒一・理化学研究所）

平成27年度は、１）反応拡散シミュレーションソフトウェアpSpatiocyteの開発に関しては、細胞形状や内部の構造物等を再現するためのパイプライン構築およびプログラムのチューニングを実施し、２）EGF信号伝達経路シミュレーションに関しては、これまでに構築した数理モデルに細胞質／細胞核間のタンパク質輸送システムを追加する。そして、細胞膜上のEGFレセプターからリン酸化ERKの核内移行までの一連の経路の一分子粒度シミュレーションを京コンピュータ上で実施する。

(1-3) 信号伝達経路上のリン酸化酵素の反応性解析（林重彦・京都大学大学院理学研究科）

平成 27 年度は、信号伝達のスイッチングを行う Ras-GAP 複合体における GTP 加水分解酵素活性と、MAP リン酸化酵素である ERK と、MAPK リン酸化酵素である MEK の複合体におけるリン酸化反応に対して、混雑環境におけるシミュレーションを実施し、細胞環境が信号伝達過程に対して与える影響の分子機構を明らかにする。Ras-GAP に関しては、昨年度までに行ってきた BSA タンパク質をクラウダーとした混雑環境における QM/MM 自由エネルギー計算を続行し、反応機構解析を行う。また、MEK-ERK 間のリン酸化反応に関しては、MEK/ERK 複合体のモデリングを完了し、BSA タンパク質をクラウダーとした混雑環境における QM/MM 自由エネルギー計算を用いた反応機構解析を行う。

(1-4) 粗視化分子モデルを用いた信号伝達経路上のリン酸化酵素複合体と
クロマチンの動的モデリング（高田彰二・京都大学大学院理学系研究科）

平成27年度は、粗視化シミュレータCafeMolを用いて、信号伝達分子リン酸化酵素複合体MEK-ERKの複合体について実験検証可能な複合体構造および相互作用予測を行なうとと

もに、核内環境におけるモデルクロマチンの折れたたみ構造に対するエピジェネティックな要因の影響とモデルクロマチン構造中における転写因子等の拡散運動を定量化する。

(1-5) 全原子モデルにもとづくヌクレオソームポジション変化の自由エネルギープロファイル計算（河野秀俊・日本原子力研究開発機構）

平成 27 年度は、ヒストンタンパク質に巻き付いている DNA が完全にほどける過程の自由エネルギープロファイル計算の精密化を行い、研究をまとめる。計算は、カノニカルヒストンとヒストンヴァリアントを含むヌクレオソームの２つについて行う。また、H3 ヒストンテールの翻訳後修飾のひとつ、アセチル化の数とテールの構造分布の関係を調べ、アセチル化の影響を定量的に明らかにする。これにより、ヌクレオソーム凝集など高次クロマチン構造形成におけるヒストンテールの役割を探る。また、チーム内の MD/SAXS 計算(池口)や粗視化シミュレーション(高田)との連携や実験科学者との連携を図り、マルチスケールの観点からクロマチン動態を明らかにしていく。

(1-6) 核内 DNA 結合タンパク質の機能ダイナミクス研究

（池口満徳・横浜市立大学大学院生命医科学研究科）

平成 26 年度に行った粗視化分子動力学シミュレーションと X 線小角散乱(SAXS)実験の連携研究（CG-MD-SAXS）法の開発とヌクレオソーム構造への適用結果を踏まえ、平成２７年度には、CG-MD-SAXS 法の開発をさらに進めるとともに、ヌクレオソーム等核内 DNA 結合タンパク質への応用において、サンプリングなど研究上の不足点を補い、研究結果を成果として最終的にまとめる。

Ｉ－２　実施内容（成果）

本研究課題では主に基礎科学として科学的な成果を追求してきた。しかし、「京」の利活用のために（スクラッチから）開発したシミュレーションソフトウェアが複数あるため、これらの利用は実用的な観点からするとアカデミアのみならず、将来的には創薬など産業界へも大きな貢献と言える。実際、ポスト京の利活用に向けた重点課題では理研と戦略分野で開発した分子動力学ソフトウェア GENESIS の創薬応用計算に向けた利用がコデザインを通して進められている。このソフトウェア開発と「京」の計算資源がなければ不可能である大規模計算を通して、HPC が生命科学においても重要な役割を果たすことを示した。また、この課題では、課題内の理論研究者が異なるモデルを持ち寄り、２つのサブ課題に集中して密接に議論を交わしながら研究を行った。さらにそれぞれの課題で実験研究者と共同研究を行い、大規模シミュレーション結果を検証した。

(1-1) 信号伝達経路上のリン酸化酵素への細胞質分子混雑の影響（杉田有治・理化学研究所）

【実施内容】

平成 27 年度は、細胞質分子混雑環境に関する大規模分子動力学計算に関する解析を継続し、タンパク質間相互作用や流体力学効果の影響などを調査した。また、MEK-ERK 複合体の構造が分子混雑環境と溶液環境でどのように異なるのかについて調査を行った。

【成果】

細胞内環境は 70%が水で占められており、残りの 30%をタンパク質や RNA などの生体高分子や各種の代謝物などの低分子化合物が占めている。この分子混み合い環境におけるタンパク質などの動態はこれまで十分に理解されていなかったが、近年、In-cell NMR などの実験によって新しい知見が得られつつ

ある。我々は、1170 万原子および 1 億原子を含む巨大な分子混雑系（図１ａと図１ｂ）を構築し、「京」を用いた大規模分子動力学計算を行うことによりこの問題に取り組んだ。この分子混雑系は、プロテオーム解析等でタンパク質の濃度等が明らかになっているバクテリア細胞（*Mycoplasma Genitalium*）の細胞質環境をモデル化したものであり、個々のタンパク質等は共同研究者の Michael Feig（Michigan State University）が Homology Modeling で構築した。水や代謝物も含む全原子分子動力学計算は、理研で開発中の分子・細胞シミュレータ GENESIS を用いて行った。GENESIS は並列性に優れたソフトウェアとして論文発表を行い、また計算機構からフリーソフトウェアとして公開している。その際、理研からプレスリリースを行った。図１ｃと図１ｄに京での GENESIS の並列性能を示した。この結果は海外の研究者にも大きなインパクトを与え、理研で開発した並列化アルゴリズムが世界的なソフトウェア(CHARMM など)に逆輸入される可能性が高い(Private communication)。

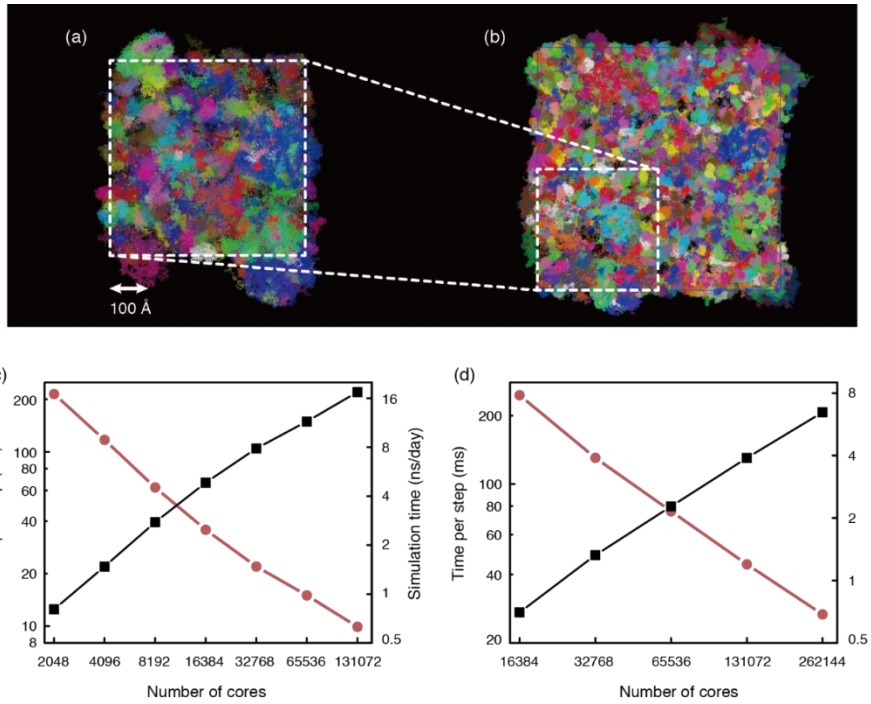


図１　細胞質内の分子混雑環境のモデル（1170 万原子、1 億原子を含む）と「京」での並列化パフォーマンス

2 年間の「京」を用いた計算によって、1000 万原子系については 130ns のプロダクションを実施し、1 億原子系については 20ns のプロダクションが終了した。

得られたトラジェクトリから、タンパク質や RNA の立体構造安定性、並進や回転拡散に関するパラメタを記述、タンパク質と代謝物の相互作用、タンパク質間相互作用の解析を行った。最も大きな知見は、従来用いられてきた球状の分子混雑物では得られない様々な相互作用に起因する影響が全原子分子動力学シミュレーションによってはっきりと見えてきたことである。例えば、PGK というタンパク質の酵素活性は分子混雑中の方が希薄溶液中よりも高くなることが実験的に知られている。今回の分子混雑系に含まれる PGK の構造を調べたところ、希薄溶液中の構造と比較して２つある活性部位間の距離が短くコンパクトな構造を取っていることがわかった。興味深いのはこの変化が濃厚な塩溶液中でも見られたことである。そこで、タンパク質構造とイオンの分布、分子混雑系における代謝物の分布を調べたところ、活性部位近傍にいずれもマイナスの電荷を持つイオンや代謝物の一部が凝集していたことが明らかになった。このようにエントロピー変化に起因する排除体積効果のみならず、複数の分子のかかわる詳細な分子間相互作用が細胞内環境でのタンパク質の構造・機能・ダイナミクスの解析に必要である可能性が示唆された。これらは今後の実験と計算のさらなる連携研究を行う高いモーチベーションとなり、

細胞まるごとモデリングに向けたはじめの一歩を踏み出したと言える。

〈論文〉

Jaewoon Jung, Takaharu Mori, Chigusa Kobayashi, Yasuhiro Matsunaga, Takao Yoda, Michael Feig, and Yuji Sugita, “GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations”, *WIREs computational molecular science*, 5, 310-323 (2015). (IF=11.885)

（1-2）細胞内環境下での信号伝達経路のモデリング基盤整備およびコード開発

（高橋恒一・理化学研究所）

【実施内容】

平成 27 年度は、１）反応拡散シミュレーションソフトウェア pSpatiocyte の開発に関して、細胞形状や内部の構造物等を再現するためのパイプライン構築およびプログラムのチューニングを実施した。２）EGF 信号伝達経路シミュレーションに関しては、これまでに構築した数理モデルに細胞質／細胞核間のタンパク質輸送システムを追加した。そして、細胞膜上の EGF レセプターからリン酸化 ERK の核内移行までの一連の経路の一分子粒度シミュレーションを京コンピュータ上で実施した。

【成果】

細胞内信号伝達経路全体のモデリングを実現するために、最も粗視化したモデル（一分子粒度）を用いた計算を行った。一分子粒度とはいえ細胞丸ごとのスケールで、細胞表面の受容体が信号分子を感知してから核内で転写因子が遺伝子の発現のスイッチを入れるまでの一連の反応の計算を連成させた例は世界的に例がなく、画期的な結果である。この課題では pSpatiocyte という反応拡散シミュレーションソフトウェアを開発し、MPI と OpenMP を利用したハイブリッド並列を実現することで、最終的には 82944 ノード（663,552 コア）までの高並列化を達成した（図 2 図 A）。また、pSpatiocyte においては細胞内構造物等を再現するコンパートメント機能を含み、図 2 に示したように共焦点レーザー顕微鏡で得られた画像からの情報を取り込むことができる。

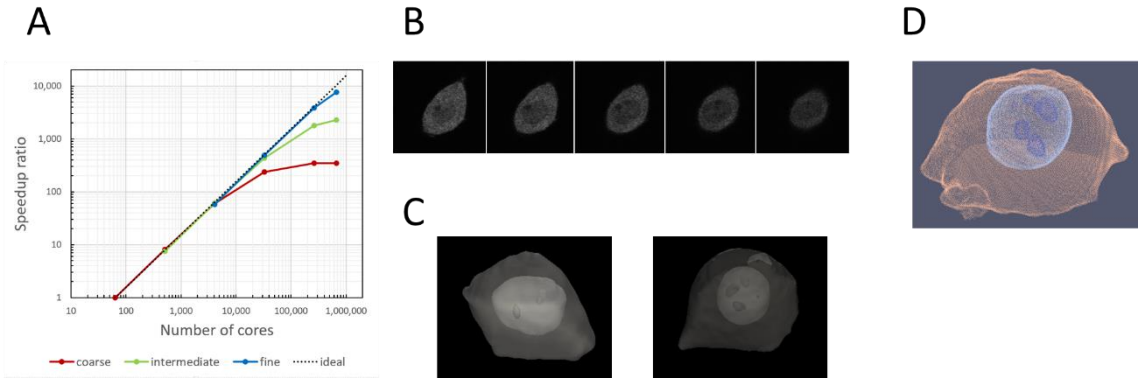


図 2　「京」での並列化パフォーマンスと細胞形状の再構成

A) 京でのストロングスケーリングの結果

B) 共焦点レーザー顕微鏡により観察した PC12 細胞の 3 次元スライス画像

C) 3 次元スライス画像から再構成した細胞形状および細胞核の構造

D) pSpatiocyte 上で再現された細胞形状および細胞核

EGF 信号伝達経路をサブ課題における統一的なターゲットとして設定し、高橋らは細胞内環境（特に

分子混雑）・細胞間のタンパク質量の違い等がその経路の応答におよぼす影響を検証した。シミュレーション結果は、分子混雑が詳細な反応過程にも大きな影響をおよぼしていること示唆した。また、異なる EGF 濃度における核内 ERK 応答のばらつきを実験とシミュレーションで比較した。EGF 高濃度条件の方が、低濃度条件よりも分布の幅が狭くなっており、ばらつきが抑制されていることが分かった。同条件でのシミュレーションは、実験と非常によく似た分布を示しており、核内 ERK 応答のばらつきがよく再現できた。そして、各 EGF 濃度における核内 ERK タンパク質の分布が、パスウェイのどのタンパク質のばらつきに起因するかを調査した。その結果、EGF 低濃度条件においては、複数のタンパク質のばらつき（EGFR, Ras, Raf, MEK）により、核内 ERK 応答のばらつきが生み出されていることが分かった。一方、EGF 高濃度条件では、ERK タンパク質のみのばらつきにより、核内 ERK タンパク質のばらつきが生み出されていることが分かった。

〈論文〉

Shindo Y*, Iwamoto K, Mouri K, Hibino K, Tomita M, Kosako H, Sako Y, and Takahashi, K*, “Conversion of graded phosphorylation into switch-like nuclear translocation via autoregulatory mechanisms in ERK signaling”, *Nature Communications*, 2016, doi: 10.1038/ncomms10485

（1-3）信号伝達経路上のリン酸化酵素の反応性解析（林重彦・京都大学大学院理学研究科）

【実施内容】

平成 27 年度は、信号伝達のスイッチングを行う Ras-GAP 複合体における GTP 加水分解酵素活性と、MAPK リン酸化酵素である MEK のリン酸化活性に対して、分子シミュレーションを実施し、その分子機構を明らかにした。Ras-GAP に関しては、昨年度までに行ってきた BSA タンパク質をクラウダーとした混雑環境における QM/MM 自由エネルギー計算を続行し、反応機構解析を行った。また、発がん性変異体と考えられている Leu902Ile 及び Leu902Phe 変異体に対し、酵素活性の解析を行った。また、MEK に関しては、リン酸化 MEK と共に反応活性部位の構造最適化を行い、イオン及び周辺側鎖の配位構造を明らかにした。

【成果】

EGF 信号伝達経路では複数の酵素がリン酸化反応を触媒することにより下流の酵素を次々に活性化していく。したがってこの信号伝達機構を最も詳細に解析するためには、酵素反応をシミュレーションする理論基盤が必要である。そのために、活性部位の電子状態を露わに量子力学・量子化学で取り扱い、それ以外の多数の自由度を古典力学に基づく分子力場を用いて近似する QM/MM 法を用いて解析した。林らは量子化学と古典分子力場の計算を分離することで、巨大な分子システムにも適用可能な QM/MM 自由エネルギー法を開発しており、本研究課題ではこの手法を適用した。QM/MM 自由エネルギー法の QM/MM 部分は、GAMESS を使用しているが、これまでの「京」における使用においては、GAMESS のデータサーバが 1 ノード内のコアの半分を使用しているため、演算の効率が低かった。そこで、通常の使用で用いる密度汎関数法による計算の場合に、データサーバを使用しないようにすることにより、すべてのコアを演算に使用することを可能にし、計算効率の 2 倍の向上を達成した。1,000 コアを超えても非常に良い並列化効率が得られている。

科学的な成果としては、EGF 反応経路上のリン酸化酵素である MEK1 に対して、分子動力学（MD）シミュレーションを用いて、ATPase 活性に対するリン酸化の影響を解析した。リン酸化を受けた MEK1-PP では、Arg227 のグアニジニウム基が、すぐ前の S218 及び S222 に付加したリン酸基に配位することに

より、ATP の γリン酸基周辺を大きく開け、ATP 加水分解が可能な構造となることを明らかにした。さらに、MEK-ERK 複合体のモデルとして、信号伝達経路の上流にある GAP-Ras 複合体の GTP 加水分解反応に関する研究を行った。酵素反応の反応遷移状態生成において、細胞内における混雑環境効果が酵素活性に重要となる可能性がある。そこで、BSA 混雑環境下での反応始状態の自由エネルギー構造最適化を、QM/MM 自由エネルギー法を用いて行った(図 3)。また、GTP 加水分解活性の減少によりがん化を引き起こすと考えられている Leu902Ile 及び Leu902Phe 変異体に対して、QM/MM 自由エネルギー法を用いた反応始状態及び遷移状態の自由エネルギー構造最適化を行い、反応活性化自由エネルギーを求めた。その結果、それらの変異体における反応活性化自由エネルギーは、天然のそれに比べて 3 kcal/mol 程度上昇し、反応活性が大きく低下することを見出した。詳細な解析により、Leu902 の側鎖は、反応活性部位から若干離れているものの、変異導入による側鎖の変化により、水分子を含む複合体間の分子認識が変化し、活性部位における反応活性に重要となる相互作用の変調を与えていることを明らかにした。

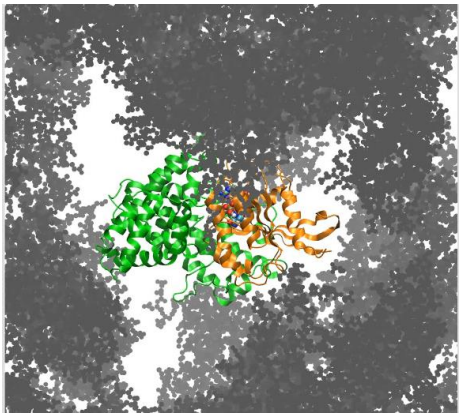


図 3　BSA 分子混雑下での Ras-GAP 複合体の QM/MM 自由エネルギー計算

(1-4) 粗視化分子モデルを用いた信号伝達経路上のリン酸化酵素複合体とクロマチンの動的モデリング（高田彰二・京都大学大学院理学系研究科）

【実施内容】

平成 27 年度は、粗視化シミュレータ CafeMol を用いて、信号伝達分子リン酸化酵素複合体 MEK-ERK の複合体について実験検証可能な複合体構造および相互作用予測を行った。また、核内環境におけるモデルクロマチンの折れたたみ構造に対するエピジェネティックな要因の影響とモデルクロマチン構造中における転写因子等の拡散運動の定量化を行った。

【成果】

(a) 粗視化分子モデルを用いた信号伝達経路上のリン酸化酵素複合体

EGF シグナル伝達経路における MAPK カスケードでの分子機構に迫る為に、(実験的手法では未だに明らかにされていない)MEK1(MAPKK)と ERK2(MAPK)の複合体の構造予測を行った。複合体の立体構造予測を全原子分子動力学で実行することは困難であるため、高田らが開発してきた粗視化シミュレーションプログラム：CafeMol を適用し、MEK-ERK ドッキングの粗視化シミュレーションを行った。

さらに、タンパク質ドッキングソフトウエア ZDOCK によりサンプリングされた MEK-ERK 複合体構造(3000 サンプル)から、KYG-PROTEIN 法を用いたスコアリングにより有力構造を抽出した。

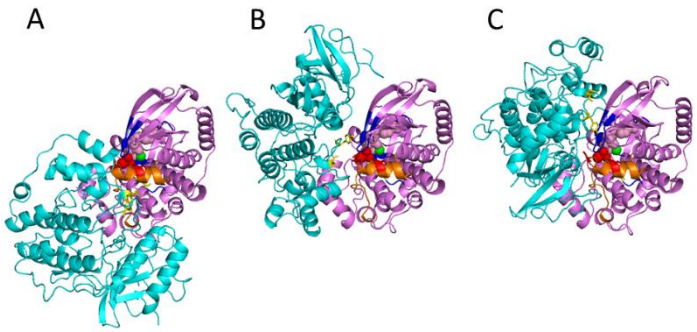


図 4　得られた MEK-ERK 複合体のモデル構造(３つの有力候補)

ここで得られた３つの有力構造を図 4 に示す。このうちの２つの構造は杉田グループ Wang 博士による全く独立な方法(ドッキングソフトウェアによるモデル構築と MMPB/SA 法による自由エネルギー評価)によって得られたモデルとよく一致していた。しかし、これらのモデルでは MEK と ERK の不活性状態を用いて複合体の構造予測を行っていることが問題であり、全原子分子動力学シミュレーションによる ERK 活性構造モデルの予測を行い、予測構造の改良が可能となった。

理研 QBiC の木川らは実験的に MEK と ERK の複合体の立体構造を溶液 NMR を用いて解析し、高田や杉田らの予測構造との比較を試みた。しかし、戦略分野の活動期間にはタンパク質発現量の問題で複合体の構造情報を実験的に得ることはできなかった。この課題は引き続き推進し、複合体予測構造モデルを実験的に評価するとともに、林らによる QM/MM 自由エネルギー計算による酵素反応機構解析を行う計画である。

(b) クロマチンの動的モデリング

クロマチン構造のモデルとしてまず、CafeMol を用いた粗視化分子シミュレーションによって、ヌクレオソーム 1，2，3 量体の構造を明らかにした（図 5)。特に、合成されたヌクレオソーム 3 量体について、X 線溶液散乱データと粗視化分子シミュレーションの結果を定量的に比較し、極めて良い一致を得た。この手法は横浜市大の池口らによって開発された MD/SAXS 法（後述）を CafeMol など粗視化シミュレーションと組み合わせるために共同で開発された。

セントロメアは、動原体が結合するサイトを作るクロマチンの特別な領域であり、そのことによって染色体の分離を可能にしている。これまでにその構造はほとんど明らかになっていなかった。我々は、セントロメア特異的なヌクレオソーム 3 量体 H3-CA-H3 の構造を求めた。これから、たくさんの H3 ヌクレオソーム多量体のなかに CA ヌクレオソームを 1 分子だけ含めた構造のモデルを作成すると、CA のある場所でクロマチンファイバーが折れた。これは、セントロメアが動原体マシナリーの結合サイトを与える足場になる可能性のある構造であり、CA ヌクレオソームの機能が示唆された。

ヒストンアセチル化は、代表的なエピジェネティック制御である。本研究では、ヌクレオソーム1,2,3量体からなる系を例にとって、様々なヒストンテールアセチル化が、その構造をどのように変えるのかを CafeMol を用いて研究した。H4テールアセチル化の影響がもっとも大きく、トリヌクレオソームをより広がった構造に変化させた。他のヒストンのアセチル化は、それよりは小さな影響であった。また、クロマチン構造と転写因子動態の連関を理解するために、核内に移行した ERK2蛋白質の核内混み合い環境における拡散運動を CafeMol を用いて調査した。核内混み合い環境を実現する為に、20ヌクレオソームからなるクロマチン構造を系に含めて計算を行った小型の転写因子 HMGB1の場合、0.5 mM ヌクレオソーム環境下でも、遅いながらも拡散可能であり、拡散係数は溶液中の約4割程度になった。また、HMGB1と同程度の非転写因子である ERK の場合、0.1 mM, 0.5 mM と徐々に拡散係数は減少するものの、0.5 mM ヌクレオソームでも溶液中の6割以上の拡散係数であった。

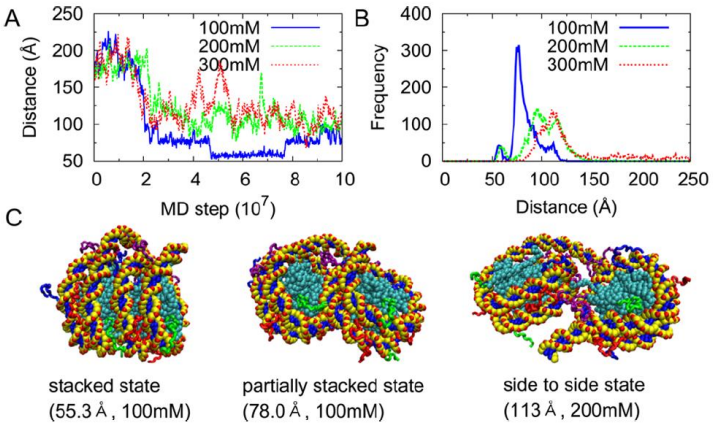


図５　ヌクレオソーム２量体の分子シミュレーション

塩濃度依存的に、２つのヌクレオソームが様々にドッキングできる(Takada et al Acc. Chem. Res 2015)

〈論文〉

Shoji Takada, Ryo Kanada, Cheng Tan, Tsuyoshi Terakawa, Wenfei Li, and Hiroo Kenzaki, **Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations**, *Accounts of Chemical Research*, **48**: 3026-3035 (2015).(IF=22.3)

（1-5）　全原子モデルにもとづくヌクレオソームポジション変化の自由エネルギープロファイル計算（河野秀俊・日本原子力研究開発機構）

【実施内容】

平成 27 年度は、ヒストンタンパク質に巻き付いている DNA が完全にほどける過程の自由エネルギープロファイル計算の精密化を行い、研究をまとめた。自由エネルギー計算では、カノニカルヒストンとヒストンヴァリアントを含むヌクレオソームの２つについて行った。どちらの場合も、Inner turn の DNA 解離に必要なエネルギーは、outer turn に比べて大きいことが示された。また、H3 ヒストンテールの翻訳後修飾のひとつ、アセチル化の数とテールの構造分布の関係を調べ、アセチル化の影響を定量的に明らかにした。アセチル化の数に応じてテールのヘリックス構造を取る率が高くなること、また、テール自体がコンパクトな構造を取る傾向が強まることがわかった。この結果として DNA がヒストンから解離しやすくなることが示唆された。これらは、ヌクレオソーム凝集など高次クロマチン構造形成が、ヒストンテール３１のアセチル化の量によって精巧に制御されていることを示している。また、チーム内の MD/SAXS 計算(池口)や粗視化シミュレーション(高田)との連携や実験科学者との連携を図り、マルチスケールの観点からクロマチン動態の一端を明らかにした。

【成果】

近年、ヌクレオソームを構成するカノニカルヒストンとその変異体の違い、ヒストンの修飾及びＤＮＡの修飾が遺伝子発現制御や細胞分化などにおいて重要な役割を果たしていることがわかってきた（エピジェネティクス)。そこで、これらの修飾や変異体とヌクレオソームポジション変化の関係を明らかにすることにより、遺伝子発現制御や細胞分化の分子メカニズムの一端を明らかにする。特に、ヒストン変異体を含むヌクレオソームについて、巻き付いているＤＮＡがほどける過程の自由エネルギープロ

ファイル計算を行った。そのために、Adaptively biased 分子動力学法と Umbrella Sampling を組み合せた効率の良い構造探索手法を原子力機構で開発している SCUBA に組み込み、「京」に最適化を行った。実際の計算ではターゲットとなるヌクレオソームの系に関する１００個のコピー（レプリカ）を発生させて同時に、分子動力学計算を同時実行する。これにより４０００ノード以上の並列化が必要となる大規模計算となったため、「京」以外の計算リソースでは実行することが不可能であった。
ヒストン組成の違いがヌクレオソーム DNA の解離にどのような違いをもたらすか調べるために、カノニカル H3 ヒストンとそのヴァリアント CENP-A（染色体の均一な分配に重要でセントロメアに局在）をそれぞれもつヌクレオソームについて、DNA 解離の自由エネルギープロファイル計算を行っている。結果、H3 と CENP-A では、乖離の自由エネルギーに大きな差が見られ、また、その要因もアミノ酸残基レベル（原子レベル）で特定された。さらに、エピジェネティクスの観点から、DNA メチル化がヌクレオソームの立体構造に与える影響や、ヒストンテールに含まれるリジン残基のアセチル化の影響を調べた。ヒストンテールの重要性は従来から知られていたが、テールは天然変性状態と呼ばれる特定の構造を取らない状態で存在するため、従来の計算方法や実験的に構造情報を得ることが困難であった。そこで、adaptive lambda square dynamics 法という構造サンプリングを効率的に行える計算方法を考案し、256 レプリカの並列計算を実施し、定量的にアセチル化の影響を調べた。結果、アセチル化することでヒストンテールはヌクレオソームに巻きついている DNA が剥がれやすくなり、テールはヘリックス構造を取る傾向がより高まることが明らかになった。本研究で得られた知見はエピジェネティクスの理解と予測に役立つものであり、この課題は新学術領域（動的クロマチン構造と機能：胡桃坂仁志（早稲田大学））と連携し、X 線及び中性子溶液散乱解析を行っている実験グループ（京都大学、杉山正明）と密接な情報交換を行ってきた。

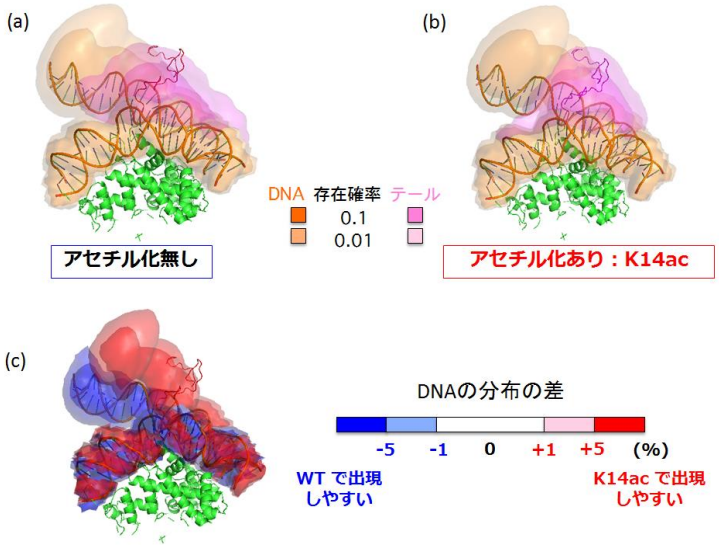


図 6　テールと DNA の空間分布。

(a) アセチル化無し (b) アセチル化あり　(c) アセチル化の有無による DNA の空間分布の差

〈論文〉

Ikebe, J., S. Sakuraba, and H. Kono. H3 histone tail conformation within the nucleosome and the impact of K14 acetylation studied using enhanced samp PLoS Comp. Biol. 12(3):e1004788 (2016)

Ikebe, J., S. Sakuraba, and H. Kono. Adaptive lambda square dynamics simulation: an efficient conformational sampling method for biomolecules. Journal of Computational Chemistry 35:39-50 (2014)

(1-6) 核内 DNA 結合タンパク質の機能ダイナミクス研究

（池口満徳・横浜市立大学大学院生命医科学研究科）

【実施内容】

平成 26 年度に行った粗視化分子動力学シミュレーションと X 線小角散乱(SAXS)実験の連携研究（CG-MD-SAXS）法の開発とヌクレオソーム構造への適用結果を踏まえ、平成 27 年度には、CG-MD-SAXS 法の開発をさらに進めるとともに、ヌクレオソーム等核内 DNA 結合タンパク質への応用において、「京」を用いた大規模分子動力学計算によってサンプリングなど研究上の不足点を補い、実験結果との比較により、計算結果の妥当性を検証した。

【成果】

ヌクレオソームに代表される核内ＤＮＡ結合タンパク質複合体の構造やダイナミクスが遺伝子の発現や維持に深く関わっていることが明らかになってきた。本研究では、そのような核内ＤＮＡ結合タンパク質の機能ダイナミクスのあり方を、実験系と連携しつつ、分子シミュレーションを用いて調査した。分子シミュレーションと実験を密に連携させるために、粗視化分子動力学シミュレーション(MD)法と X 線小角散乱(SAXS)実験法を組み合わせる粗視化 MD-SAXS 法の開発を行った。この粗視化 MD-SAXS 法のソフトウェアは、「京」において高並列の MPI 並列計算が実行できるようにした。この計算では、DNA とタンパク質の相互作用と粗視化モデルにおけるイオン濃度の２つがパラメタになり、その組み合わせは膨大になる。実際、９０個のパラメタの組み合わせについてそれぞれを１００回繰り返し、３つのヒストンバリアントに関する計算を行ったため、合計で２７０００個のシミュレーションが必要となった。これは Capacity 計算ではあるが、「京」なしでは実現が困難であった課題である。

この手法をもちいて、様々なヒストンバリアントを含んだヌクレオソームの粗視化 MD-SAXS 研究を行った。核内 DNA 結合タンパク質のうち、ヒストンバリアントの種類の異なる様々なモノヌクレオソームに着目して粗視化 MD-SAXS 解析を行った。SAXS （X 線小角散乱）については、横浜市大の佐藤衛教授のグループと連携し、粗視化 MD は高田らが開発した CafeMol を用いた。ヌクレオソームに含まれるヒストンバリアントの違いにより、遺伝子発現制御に変化が生じることが報告されており、ヒストンバリアントによるヌクレオソームの構造変化は、エピジェネティクスの観点から重要である。粗視化 MD-SAXS の結果、カノニカルヌクレオソームは、メジャーな状態として、溶液中で両方の DNA 末端がコアドメインに巻き付いた状態をとり、また、マイナーな状態として、片方の DNA 末端がコアドメインから離れた状態をとっていることが分かった（図 7）。

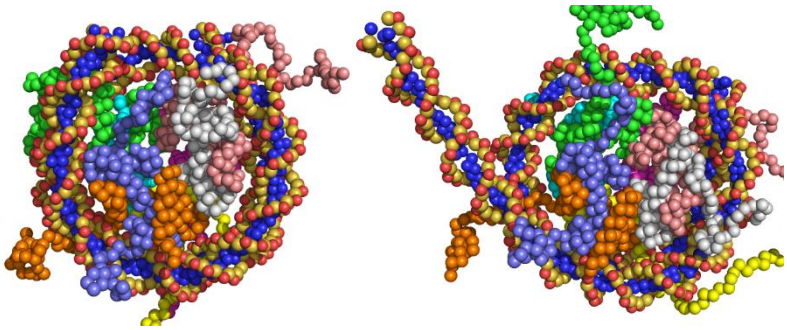


図 7　粗視化 MD-SAXS 法によるカノニカルヌクレオソームの構造

左がメジャーな状態、右がマイナーな状態

Ⅱ　戦略課題２：創薬応用シミュレーション　［統括： 藤谷 秀章（東京大学）］

分子動力学を用いた生体高分子解析のために、特定高速電子計算機施設を中核とするHPCIの計算能力を活用するとともに、最新の計算アルゴリズムによる創薬プロセスの革新を目指し、革新的な薬の活性予測シミュレーションを行う。

Ⅱ－１　実施計画

(2-1)超並列結合自由エネルギー計算法を用いた創薬応用シミュレーション

（藤谷秀章・東京大学先端科学技術研究センター）

平成27年度は、これまでに得られたヒット化合物の実験データと計算結果との比較検証を行い、新薬の動的設計に有効な情報をどのようにして獲得するかを明らかにする。医薬標的として重要なGPCRを中心とした水に溶けない膜タンパク質の高精度計算を実現する為に、今までに開発を進めて来た脂質力場とファンデルワールス力を遠距離まで取り入れる新しい計算方法を用いて、膜タンパク質の特性を高精度に記述する計算方法の開発を進め、膜タンパク質と複数の薬剤の自由エネルギー計算を高精度に行う。

Ⅱ－２　実施内容（成果）

(2-1)超並列結合自由エネルギー計算法を用いた創薬応用シミュレーション

（藤谷秀章・東京大学先端科学技術研究センター）

【実施内容】

平成27年度は化合物が入っていないアポ状態のタンパク質構造から、化合物がタンパク質に取り込まれて結合構造が形成されるまでのシミュレーションを大量に行い、Markov State Model (MSM)でこの結合過程を解析した。その結果、結合構造に至る過程で６個の準安定構造が現れる事が明らかになった。この準安定構造間の遷移確率を求め、化合物がタンパク質に結合するまでのプロセスモデルを構築した。医薬標的として重要なGPCRの高精度結合自由エネルギー計算を実現する為に、従来の脂質力場によるGPCRの振る舞いの違い調べた。その結果、GPCRのヘリクス３と６の間のionic lockが受容体の活性・不活性のスイッチになっていると信じられているが、脂質力場の違いでこのionic lockの開閉が大きく左右される事が明らかになった。これまでに開発を進めて来た脂質力場とファンデルワールス力を遠距離まで取り入れる新しい計算方法を用いる事で、計算条件の違いによる膜タンパク質特性のバラツキを抑えて、信頼性の高い計算が可能になる事が明らかになった。

【成果】

（１）超並列結合自由エネルギー計算(MP-CAFEE)法の高効率化

超並列結合自由エネルギー計算法MP-CAFEEでは、欧州の研究者が中心となって開発されている分子動力学プログラムGROMACSを動力学コアとして使用している。GROMACSはインテルやAMDのCPUチップ、IBMのPowerPCチップに対するSIMD intrinsic accelerated kernelを持っており、各CPUの演算能力を高効率で使用して高速な分子動力学計算を実現している。この機能を開発したスエーデン王立科学技術研究所のLindahlと共同して、SPARC64-HPC-ACEアーキテクチャを持ったCPUに対するSIMD intrinsic accelerated kernelを開発した。「京」と富士通のFX10がこのアーキテクチャを持っており、どちらのコンピュータでも変更なくSIMD intrinsic accelerated kernelが動作する事を確認した。またC言語で書かれたkernelコードをコンパイ

ラだけで最適化した場合と比較して、どちらのコンピュータでも 2 倍の高速化が実現出来た事を確認した。このコードを平成 25 年 5 月に GROMACS の正式リリース版に入れて一般に公開すると同時にプレスリリースを行った。

（2）結合自由エネルギー計算法の精度

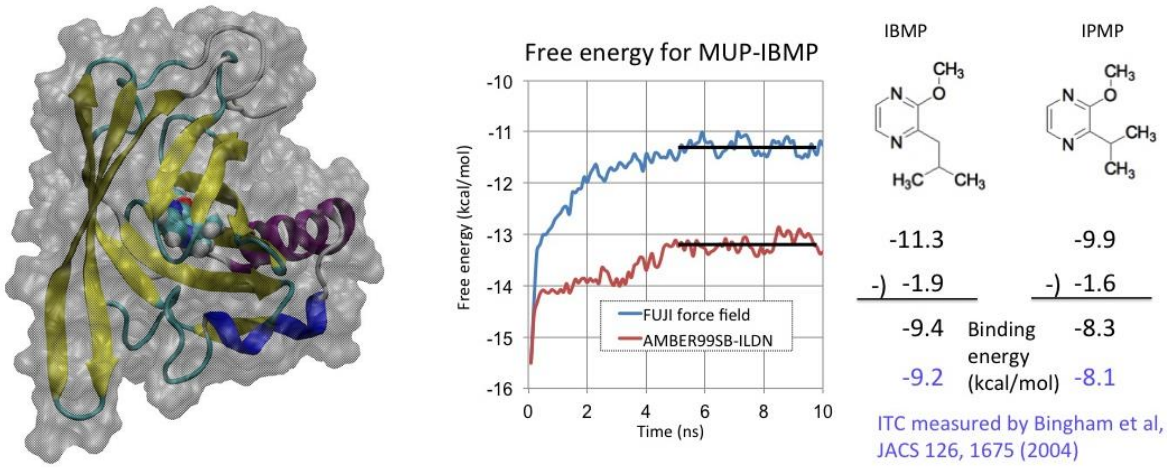


図 8　MUP の結合構造と二種類の化合物に対する結合自由エネルギー計算結果

MP-CAFEE 法で求めた結合自由エネルギーの計算精度を確かめるためにタンパク質の内側に化合物の結合部位を持つ Major Urinary Protein (MUP)に対して結合自由エネルギーを MP-CAFEE 法で計算した（図 8）。二種類の化合物(IBMP と IPMP)について、複合体の自由エネルギーから水和エネルギーを引いた結合自由エネルギーは、IBMP が-9.4 kcal/mol で IPMP が-8.3 kcal/mol となり Bingham らによって ITC で測られた結合自由エネルギーと良い一致を示している。この計算ではタンパク質の力場として我々が独自に開発した FUJI 力場を使用しているが、タンパク質の力場を D.E.Shaw グループが開発した AMBER99SB-ILDN で計算を行うと、およそ 2 kcal/mol ほど実験値より強い結合自由エネルギーが得られる。これは FUJI 力場が高精度なタンパク質主鎖の力場パラメータをもっているのに対して、従来の AMBER 力場はタンパク質が実際より硬くなってためである。タンパク質主鎖の角度分布に関する最近の分光実験と種々の力場による計算結果を比較して、FUJI 力場が最も測定分布に合致する事が明らかになっている。

（3）水溶性タンパク質に対する薬設計(ガン治療標的キナーゼ)

製薬企業と富士通と連携して実際の疾患治療標的タンパク質に対する薬設計を進めた。三者の役割分担は富士通が独自ソフトを使ったフラグメント法での新規化合物設計、東大が「京」を使った MP-CAFEE 法による結合自由エネルギー計算、製薬企業が標的タンパク質の選定と新規化合物合成と化学・細胞アッセイである。

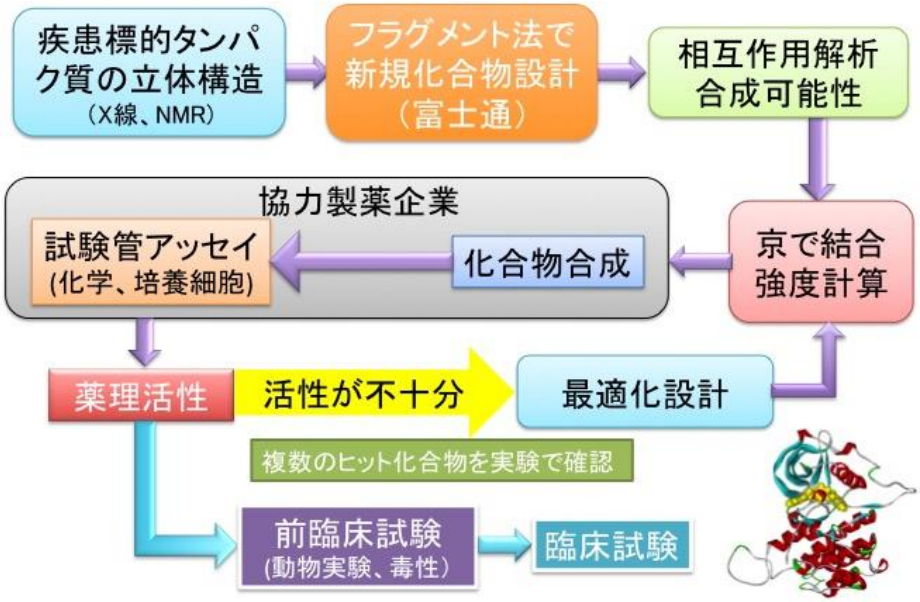


図 9　低分子薬の開発フロー

標的タンパク質を選定するにあたり、「京」で数百の低分子化合物に対する MP-CAFEE 計算が実行可能なタンパク質サイズである 300 から 500 個のアミノ酸残基を持つタンパク質から標的を選定する事にした。まだ薬が発見されていない標的で、将来的には開発費が回収できて利益がでる市場規模が望まれるガン治療標的の水溶性タンパ質であるキナーゼから選定する事になった。X 線結晶構造が沢山解かれている有名なキナーゼに対しては、既に実用化されている薬が存在しており、今回の開発対象からは除外された。少なくとも一個は X 線結晶構造が解かれていてガン治療の標的になる可能性があるキナーゼの中から標的を選定した（図 10)。

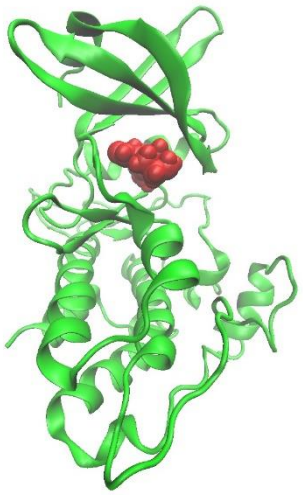


図 10　ガン治療標的タンパク質(キナーゼ)

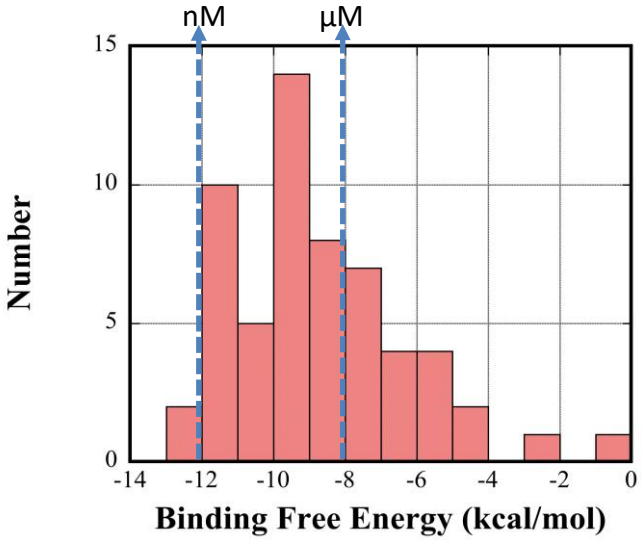


図 11　「京」で最初の MP-CAFEE 計算結果

先の開発フローに基づいて疾患標的タンパク質のX線結晶構造をベースに最適化フラグメント法で化合物を設計し、「京」で結合自由エネルギーを MP-CAFEE 法を使って最初に求めた 58 化合物の結果である（図 11)。MP-CAFEE 計算の前にタンパク質と低分子化合物を結合した系を、水の中で平衡化する為に 50 ナノ秒の分子動力学(MD)計算を実行した。この計算結果から製薬企業の研究者が 4 個の化合物を選んで化合物を合成し、細胞アッセイでその薬理活性を調べた処、凡そ二ヶ月後にウェット実験の結果が出て、全く活性が見られなかった事が報告された。この false positive の原因はX線結晶構造を初期構造とした 50 ナノ秒の平衡化 MD 計算では十分に水和したタンパク質構造が得られない事に起因する。10 倍の 500 ナノ秒の平衡化 MD を行うと右図のように結合した化合物の奥にも水分子が入り込み結合自由エネルギーも実験値に良く合うようになった。

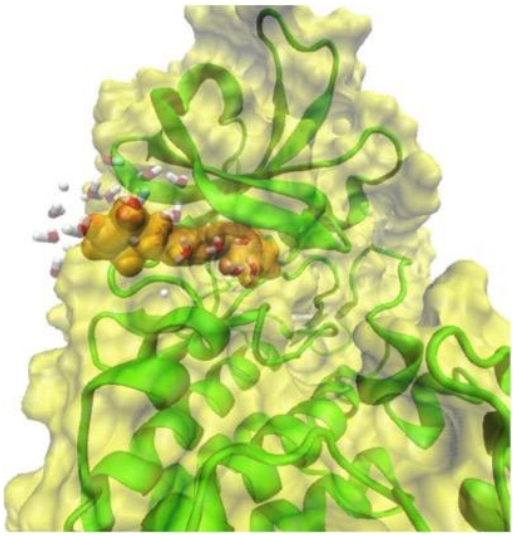


図 12　キナーゼと低分子化合物

この様な false positive に注意を払いながら、このキナーゼに対して 269 個の新規設計化合物の結合自由エネルギーを「京」で求めて、それとキナーゼと低分子化合物の相互作用エネルギー(E_{int})を比較した。右図が示す様に両者の間では全く相関は見られない。ドッキング・シミュレーションや分子軌道計算を用いた従来の計算創薬では、結合自由エネルギーを高精度に求める事は全く行われておらず、相互作用エネルギー(E_{int})で設計化合物の良し悪しを議論する方法が取られている。右図は E_{int} だけを用いた従来法では薬活性が全く予測できない事を如実に示している。高精度結合自由エネルギー計算法を考案してからの五、六年間で MP-CAFEE 法で計算したタンパク質と化合物のペア数は 50 個にもなっていなかったのに、「京」を使う事により二、三ヶ月で数百ペアの計算が可能になり、ウェット実験と連携して先の開発フローの最適化設計ループを実用的な時間で回す事が可能になった。

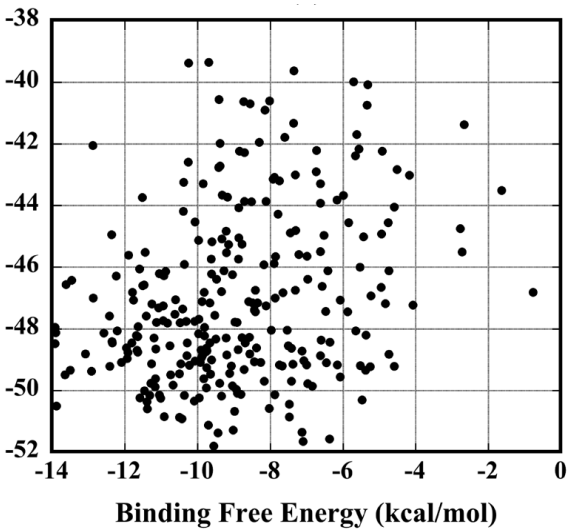


図 13　結合自由エネルギーとの相関図

本課題で連携して頂いた製薬企業と協議して創薬標的タンパク質を選定したが、複数企業との協議から二種類のガン治療標的キナーゼが標的に選定された。担当する製薬企業が異なるので Intellectual Property(IP)が混ざらない様に、阻害薬をフラグメント法で設計する人と MP-CAFEE 計算を行う人を標的タンパク質ごとに割り当てて、独立プロジェクトとして厳格に進める体制を整えた。一つのガン治療標的キナーゼではフラグメント法による設計、MP-CAFEE 計算、ウェット実験からのフィードバックなどが順調に進んだが、動物実験から標的キナーゼを阻害すると重篤な副作用が現れるなど、標的キナーゼの機能が明らかになるに従い、このキナーゼは創薬標的としては不適格である事が明確になった。この為、平成 26 年 4 月にこの標的に関するプロジェクの中止を決定した。もう一つのガン治療標的キナーゼでは、タンパク質と化合物の共結晶構造を得

るのが極めて難しく、プロジェクトで進めたX線構造解析でも、既知の一化合物との共結晶構造しか得られなかった。この為に、新規設計化合物との結合構造を推定するのに多くの計算時間を要した。このガン治療標的キナーゼに対しては前臨床試験に進む基準をクリアする新規化合物が複数個得られて、平成26年8月に富士通、製薬企業、東大の三者でプレスリリースを行った。複数個の候補化合物に優先順位をつけて製薬企業内で動物実験を進めている。

（4）エピゲノム創薬と膜タンパク質

epigenetics と呼ばれる遺伝子機能の選択的な活性化・不活性化の機構とガンなど疾患との関わりが明らかになり新たな創薬アプローチとして期待されている。東大先端研では白血病などで増えているヒストン4の20番目のリジンをメチル化するPR-SET7酵素に対する阻害薬の開発を進めて来た。PR-SET7は極めて柔らかいタンパク質で、結合する化合物により立体構造が変化するので、分子動力学で構造変化を調べながら薬を設計してリード化合物を取得した。

GPCRは7回膜貫通型のタンパク質で約3割の薬がGPCRを標的としている。GPCR創薬に結合自由エネルギー計算による定量的な評価法を導入する為に従来の力場の精度を調べた。脂質の力場の違いにより脂質二重膜に埋め込まれている GPCR の構造がどの様な影響を受けるかを精密に調べた。GPCRのヘリックス3と6の間のionic lockが受容体の活性・不活性のスイッチになっていると信じられているが、脂質力場の違いでこのionic lockの開閉が大きく左右される事が明らかになった。信頼性の高い脂質モデルを構築する事が GPCR の計算創薬にとって極めて重要である。

（5）バイオ医薬品の開発

epiregulinは大腸ガン、乳ガン、卵巣ガン、前立腺ガンの四類のガン細胞表面に特異的に発現する一回膜貫通型タンパク質でガン治療の標的タンパク質である。我々はマウスで得られたepiregulin抗体(9E5)をヒト型化し、抗体先端部のFvを繋ぎ合わせsingle chain化したscFvを作成して平成26年末から動物実験へと進んでいる。この様にバイオ医薬品としての開発は順調に進んでいるが、epiregulinと9E5が結合する過程を一つの抗原抗体反応として捉えた時に多くの基本的な疑問が生じている。結合定数をSPR(Surface Plasmon Resonance)法とITC (Isothermal Titration Calorimetry)法の二つの方法で測定すると、通常の抗原抗体反応では二つの値はほぼ一致するのに、epiregulinと9E5の反応では一桁ほど異なる値を示す。右図は9E5のX線構造で、上がアポ構造で、下がepiregulinと結合している時の構造である。驚くべき事に抗原認識で最も重要な重鎖のCDR3にある103番のproline (P103)のペプチド結合がアポ状態ではcisであり、結合状態でtransになっている。prolineのcis-trans構造変化は生体反応の速度を制御するタイマーとして使われており、その反応時間は数時間から数日に及ぶ事がある。

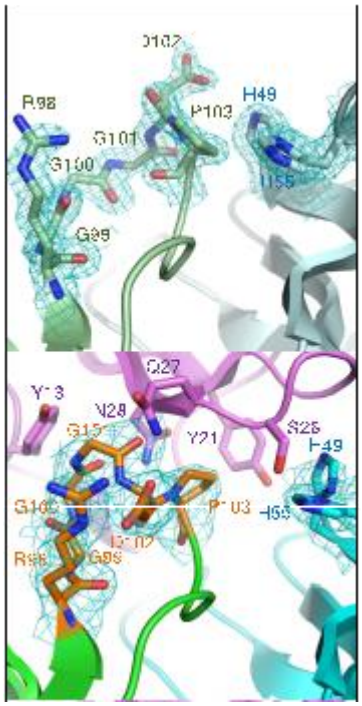


図14　9E5のCDR3近辺の構造

この様な cis-trans の変化が抗原抗体反応においてどの様な役割を担っているかを明らかにする為に、「京」でアポ構造の9E5とepiregulinの結合過程を調べる大規模シミュレーションを実行した。その膨大な計算データを解析してシミュレーション結果の意味する処を明らかにすると、抗原

抗体反応を鍵と鍵穴で説明していた従来モデルとは大きく異なった描像が浮かび上がって来るとともに、ITC と SPR の測定結果の不一致の理由が判明した。現在、この結果をアカデミック論文としてまとめている。

Ⅲ　戦略課題３：予測医療に向けた階層統合シミュレーション
［統括：高木 周（東京大学)］

これまで個別に開発が進められてきた各種生体シミュレータ（血栓症、心臓、筋骨格、脳神経系等）に対して、心筋梗塞やパーキンソン病等、様々な疾患に対してより複雑なプロセスを再現していく。そのために、基盤ツールを整備するとともに、特定高速電子計算機施設を中核とするHPCI を活用することで病態予測や治療支援を目指す。

Ⅲ－１　実施計画

特に平成 27 年度は、各サブテーマの研究内容を集約し、神経細胞レベル・筋繊維レベルからの脳神経系・全身筋骨格系を階層統合したシミュレータを完成させる。そして、パーキンソン病の病態である振戦・固縮・姿勢制御の破綻について、実際に体に現れる病態として再現することを目指す。具体的には、東大・高木の研究チームで開発を進めてきた筋繊維レベルからの骨格筋マルチスケールモデルを、東大・中村仁彦の研究チームで開発してきた体積効果を考慮した全身筋骨格系モデルへ導入し、さらに沖縄科技大・銅谷賢治の研究チームで開発を進めてきたパーキンソン病の脳モデルと結合させることにより、全身筋骨格・脳神経系の階層統合シミュレーションを実施し、パーキンソン病の病態である振戦・固縮の違いの再現、さらに、阪大の野村泰伸の研究チームによりパーキンソン病患者のデータを基に解析されたパーキンソン病患者の静止立位における重心制御の異常を再現することを目指す。

(Ⅲ-①)心筋梗塞シミュレーションに向けた血栓シミュレータと心臓シミュレータの連成及び筋繊維の集合体としての骨格筋のシミュレーション（高木周・東京大学大学院工学系研究科）

(a)「筋繊維の集合体としての骨格筋のシミュレーション」の部分については、下記の中村仁彦、銅谷賢治、野村泰伸の研究チームと協力し、パーキンソン病の病態である振戦と固縮の違いの再現を目指し、平成 27 年度は、全身筋骨格系・脳神経系の全身統合シミュレーションに向けた研究統括をする。具体的には，高木の研究チームで開発を進めてきた筋繊維レベルからの骨格筋マルチスケールモデルを、東大・中村の研究チームで開発してきた体積効果を考慮した全身筋骨格系モデルへ導入し，「京」での計算速度を向上させるためのチューニングを行う。

(b)心筋梗塞シミュレーションと関連した部分では，引き続き UT-Heart の冠循　環末梢血管系のモデリングを進める。平成 27 年度は、毛細血管内における赤血球の動きを再現したシミュレーションを実施し，赤血球の非定常な変形流動が末梢血管の抵抗に与える影響を定量的に評価し，UT-Heart に導入可能な信頼性の高い末梢血管モデルの構築を行う。血栓症のモデリングに関しては，　GPIIｂ/IIIa の阻害薬と P2Y12 の阻害薬を併用することによる血小板粘着に対する影響に関して、東海大・後藤らが平成２６年度に得た新しい知見に対して，シミュレーションを介して，その詳細なメカニズムの解明を試みる。

(Ⅲ-②)心疾患の合理的治療のためのマルチスケール・マルチフィジックス心臓シミュレーション（久田 俊明・株式会社 UT-Heart 研究所）

平成 27 年度は、昨年度に引き続きマルチスケール・マルチフィジックス心臓シミュレータ UT-Heart のサルコメア動力学モデルの一層の高度化とその応用を図る。特にミオシンヘッド首振りの確率的運動を、1 次元エネルギ地形が与えられた場合に実用的な時間ステッ

プで計算するための手法の検討に焦点を当てる。また実験系グループとの連携も継続しサルコメアモデルの実証にも注力する。

また、冠循環マルチスケールシミュレーションを通じて新たな医学的・生理学的知見を得ることを目標とし、昨年度に引き続き、冠動脈カテーテル治療（PCI）において、適応の判定のための指標として広く用いられている一方でその生理学的妥当性に疑問が呈されている FFR（Fractional Flow Reserve）の真の意味を明らかにし、より合理的な指標について検討する。更に、高木チームで開発が進められている血栓症シミュレータにおいて血小板粘着から血栓形成のプロセスを再現するため、UT-Heart から冠血管径変化、流量、圧などのデータを提供する。

（Ⅲ-③）抗血小板薬薬効予測を可能とする階層統合シミュレータの開発に向けたモデリング

（後藤信哉・東海大学）

平成 27 年度は、引き続き以下の研究開発を実施する。

1)高木チームの「心筋梗塞・脳梗塞のマルチスケールシミュレーション」モデルに対して、フローチェンバーを用いた血小板細胞接着動態の実証実験に裏打ちされた、かつ血小板細胞の GPIbα と von Willebrand 因子の分子結合から血小板細胞接着を連成する血小板細胞接着モデルを作成し精緻化する。特に、分子動力学計算の手法の応用により GPIbα と von Willebrand 因子の結合による結合ポテンシャル予測を行い、分子スケールと細胞スケールの力学的連成を行う。

2)実証実験に裏打ちされた血小板細胞代謝、活性化モデルの精緻化を継続する。

3)実証実験に裏打ちされた血小板細胞活性化と活性化血小板上の凝固系活性化を連成するモデルを作成し精緻化する。

（Ⅲ-④）全身筋骨格モデルにおける筋肉の体積効果及び人の行動モデルに関する大規模並列計算

（中村仁彦・東京大学大学院情報理工学系研究科）

平成 27 年度は、引き続き以下の研究開発を実施する。

1)脊髄の筋支配モデルの開発

上肢および下肢の脊髄反射モデルを開発する。また、前庭・足底感覚ニューロンのフィードバックによる姿勢制御系モデルの開発を行なう。

2)筋繊維モデル、脳神経モデルとの統合

銅谷Ｇとの統合計算を行い、上肢の振戦・固縮のシミュレーションを行う。また、下肢の筋を対象に、高木チームとの統合計算を実施し、筋と運動のマルチスケールシミュレーションを実現する。さらに、野村チームのＰＤ患者の姿勢制御データを用いて姿勢制御系モデルのパラメータ同定を行う。

（Ⅲ-⑤）行動制御の脳神経系の多階層モデルの構築（銅谷賢治・沖縄科学技術大学院大学学園）

平成 27 年度は、NEST により実装した大脳基底核モデルと視床-大脳皮質モデルを、東大高木チーム、中村チームが作成した脊髄神経回路モデルおよび筋骨格系モデルと統合した大規模シミュレーションを実行する。まず、単関節運動における振戦と固縮症状の発生に関与する回路、ニューロン、シナプス機構の探索を行い、病態の改善につながる介入ポイントの推定を行う。また全身モデルでの立位姿勢制御シミュレーションにおいて、健常者

とパーキンソン病患者の重心動揺パターンの違いを再現することで、健常時のしなやかな制御とパーキンソン病におけるぎこちない動きの原因を明らかにすることをめざす。

(Ⅲ-⑥)コンプライアントな生体筋・腱系に駆動されるヒト骨格系の運動制御に対して大脳基底核が果たす役割の数理モデル化（野村泰伸・大阪大学）

平成 27 年度は、（１）高木チームの「筋繊維レベルからの骨格筋モデル」、銅谷チームの「脳神経系の多階層モデル」、中村チームの「脊髄と四肢筋骨格モデル」を統合したインターフェイスの開発と利用を引き続き実施する。（２）運動に関わる関節の受動的粘弾性が低くかつ能動的神経フィードバック制御のゲインも小さい（関節がコンプライントである）という生理学的に妥当な条件の下で全身機械力学モデルの直立姿勢と歩行を柔軟かつ安定化に実現する制御機構の数理モデル化を平成 26 年度に引き続き実施し、得られた知見を（１）における大規模詳細モデリングに対してフィードバックする。特に、直立姿勢および歩行周期に現れるフラクタル的揺らぎを生成するマクロな制御数理モデルを構築する。

Ⅲ－２　実施内容（成果）

(3-1)　心筋梗塞・脳梗塞のマルチスケールシミュレーション

【実施内容】

（Ⅲ-①）(b)心筋梗塞シミュレーションに向けた血栓シミュレータと心臓シミュレータの連成

（高木周・東京大学大学院工学系研究科）

心筋梗塞シミュレーションと関連した部分では，引き続き UT-Heart の冠循環末梢血管系のモデリングを進めた。平成 27 年度は、毛細血管内における赤血球の動きを再現したシミュレーションを実施し，赤血球の非定常な変形流動が末梢血管の抵抗に与える影響を定量的に評価し，UT-Heart に導入可能な信頼性の高い末梢血管モデルの構築を進めた。血栓症のモデリングに関しては， GPII b/IIIa の阻害薬と P2Y12 の阻害薬を併用することによる血小板粘着に対する影響に関して、東海大・後藤らが平成２６年度に得た新しい知見に対して，シミュレーションを介して，その詳細なメカニズムの解明を試みた。

（Ⅲ-③）抗血小板薬薬効予測を可能とする階層統合シミュレータの開発に向けたモデリング

（後藤信哉・東海大学）

平成 27 年度は、引き続き以下の研究開発を実施した。

1)高木チームの「心筋梗塞・脳梗塞のマルチスケールシミュレーション」モデルに対して、フローチェンバーを用いた血小板細胞接着動態の実証実験に裏打ちされた、かつ血小板細胞の GPIbα と von Willebrand 因子の分子結合から血小板細胞接着を連成する血小板細胞接着モデルを作成し精緻化を行った。特に、分子動力学計算の手法の応用により GPIbα の N 末端部分と von Willebrand 因子の A1 ドメインの結合について水分子存在下での安定構造を予測した。また、重心間距離を移動させて potentials of mean force を計算した。Potentials of mean force と重心間距離のプロットから、GPIbα と von Willebrand 因子の接着力を 70 pN 程度と予測した。この予測は別途行った原子間力顕微鏡を用いた GPIbα と von Willebrand 因子の接着力とスケールにおいて相同であった。分子スケールにて計算科学的に単一分子対の接着力を予測し、その単一分子対数対により血流の両体力に抗した血小板細胞の接着が可能であることを示した。血小板細胞の接着力の視点から分子スケールと細胞スケールの力学的連成を行った。

2)実証実験に裏打ちされた血小板細胞代謝、活性化モデルの精緻化を継続した。特に、$P2Y_{12}$ ADP 受容体刺激と血小板細胞接着のモデル化において、従来、P2Y12 ADP 受容体阻害による血小板細胞接着阻害効果は GPIIb/IIIa の寄与が必須と想定されていたが、GPIbα と von Willebrand 因子のみよる接着においても $P2Y_{12}$受容体が寄与することを実証的に示した。GPIbα と von Willebrand 因子の接着には、動脈血流条件において血小板細胞を血管壁損傷部位に停止させる役割とともに、von Willebrand 因子と結合した一部の GPIbα では細胞膜のラフトなどを介して血小板活性化シグナルを惹起させる機構があることを示した。

3)実証実験に裏打ちされた血小板細胞活性化と活性化血小板上の凝固系活性化を連成するモデルを作成し、線溶系を組み込んで精緻化した。

【成果】

本年度は、心筋梗塞のシミュレーションと関連して、「冠循環末梢血管系のモデリング」と「抗血小板薬のモデリング」の２つの研究を行った。以下、これらを説明する。

(A)「冠循環末梢血管系のモデリング」

本研究では、UT-Heart の冠循環末梢血管系として利用可能な末梢血管モデルの構築に向けて、Kaneko et al.(2011)[1] の論文に掲載されている毛細血管データを著者より譲り受け、分岐を伴う毛細血管網の抽出を行ない、数値シミュレーションが可能となるようにデータの補正を行なった。図 15 に文献[1]に掲載されている毛細血管網の画像データを、図 2 に血流の数値シミュレーション用に構築された血管網を示す。図 16 に示された血管網を基に、赤血球などの血球細胞の影響を考慮せず、血液を単相の流体で構成されたニュートン流体として、この血管網を流れる血流の３次元計算を行った結果を図 17 に示す。分岐に従い、流量分配があり、また左側の入り口側から右側の出口に向けて圧力損失に伴う圧力の低下が見て取れる。この計算結果を直管の管摩擦損失のみを考慮に入れた一次元モデルと比較した結果、良好な一致を示すことが確認された。すなわち、単相での計算結果では、管摩擦損失による圧力損失に比べ、分岐部における損失は相対的に影響が少ないことが示された。

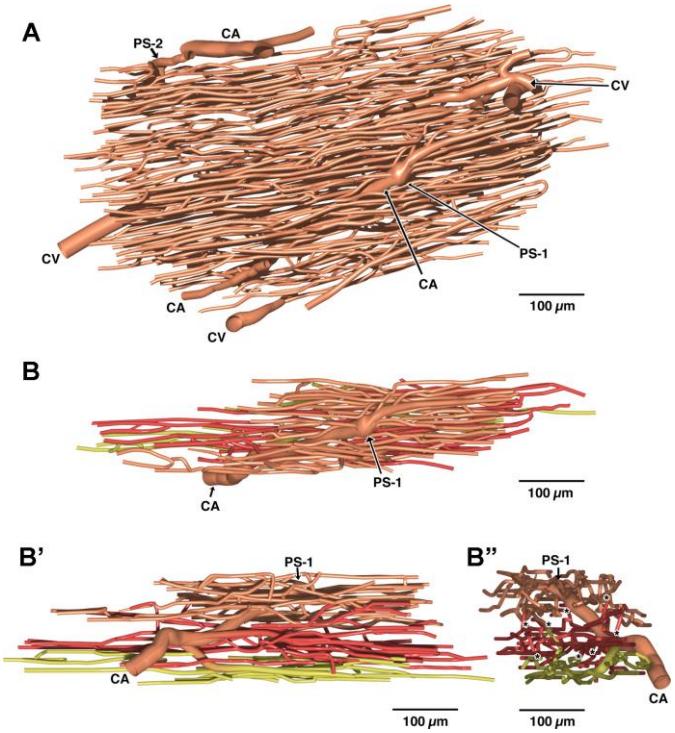


図 15　文献[1]に掲載されている冠循環微小血管系の画像

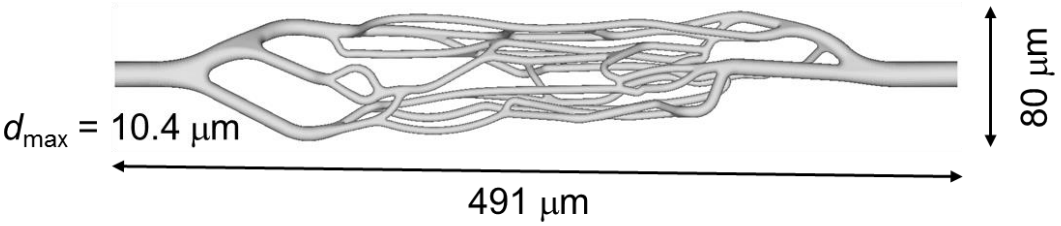


図 16　文献[1]のデータから構築された数値計算用血管網モデル

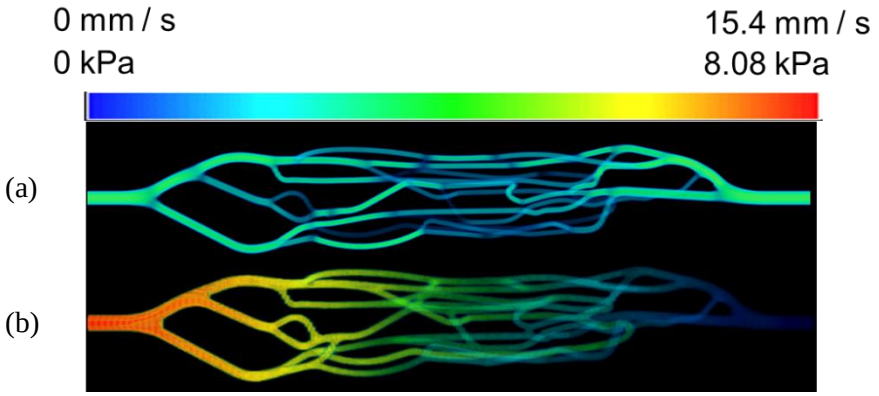


図 17　構築された血管網に対する単相の流体計算（(a) 流速分布、　(b) 圧力分布）

次に、このモデルに多数の赤血球を流した数値計算を実施するための準備を進めた。はじめに、単純な直管の系に対し、多数の赤血球が変形しながら流れていくシミュレーションを行い、赤血球の配置と末梢抵抗の関係を調べた。数値計算手法は、これまで著者らが開発を進めてきた Immersed Boundary 法を適用した。図 18 に計算系を示す。ここでは、入り口と出口の速度境界に周期境界条件を課し、両端に一定の圧力勾配を課すことにより、100 万タイムステップを超える長い時間発展の後、赤血球を含む系における流動構造がどのように変化するか計算を行った。

計算結果を図 19 に示す。図 19 の(a)～(e)は、異なる時間ステップにおける赤血球の流動状態のスナップショットである。図 19 (f)は、流量の時間変化である。ここでは、管の両端に一定の圧力勾配をかけたシミュレーションを行っているため、末梢抵抗の高い状態では流量が減る計算となっている。図より、初期にほぼ等間隔に配置されていた両凹型の形状をした赤血球が、流れとともに変形しながらスリッパ型の形状をとりながら交互の配置へと移っていく様子がわかる。t=170ms ぐらいまでは、赤血球が交互に配置したまま安定した流動構造をとり、ほぼ定常に近い状態へ達した後、一度配置がくずれ、(d) t=180ms の図に示されているような乱れた配置になる。その後、また交互に配列した安定な配置へと戻る。t=180ms のあたりで、乱れた配置になっているときの方が、流量が増えているすわなち、末梢抵抗が減っているのは興味深い。本来ならば、流動構造の非定常性が増した場合の方が、非定常性による散逸が増えるので、末梢抵抗の増加があると考えられるが、ここでは、一度赤血球位置の再配置が行われるプロセスで、赤血球が流路中央寄りを流れる傾向があるため、末梢抵抗の一時的な低下があったと考えられる。

次に、図 16 に示した分岐のある系に対して、多数の赤血球を含めた計算を実施し、流量の変化を調べるためのソフトウェア開発を行った。分岐のある場合には、分岐部での赤血球の振る舞いが管摩擦損失に影響を与える。ここでは、分岐部への赤血球の流入条件を決定するために、上記した直管系でのシミュレーションを同時に実施し、周期境界条件を課した直管系での計算の出口から流出する速度分布および赤血球分布を分岐計算側の流入条件として用いる連成計算を行った。以下では、この流入条件を決める前部分の直管の計算を助走区間計算と呼ぶ。この助走区間計算を導入することにより、分岐部での赤血球の流入が人工的な不自然なものでなく、より実際の系に近い流入条件となる。

計算結果の一例を図 20 に示す。図 20 (a)～(c)は分岐部への赤血球流入が始まってからの流動構造の変化である。図 20 (d)には、流量の変化を示す。ここで示した計算例では、t=45ms までは、助走区間より赤血球の流入がなく、主に助走区間側の流量変動により、分岐計算側の流量変動が見られるが、t=45ms をこえると、赤血球が分岐部へも流入し始め、流量の時間変化に変動の激しい高周波成分が現れるようになる。このソフトウェアを用いることにより、今後さまざまな血管網形状に対してシミュレーションを実施し、赤血球の影響を考慮した冠循環微小循環系の 1 次元モデルへとつなげることができる。

参考文献

［1］Kaneko N, et al., *Am J Physiol Heart Circ Physiol*　300: H754–H761, 2011

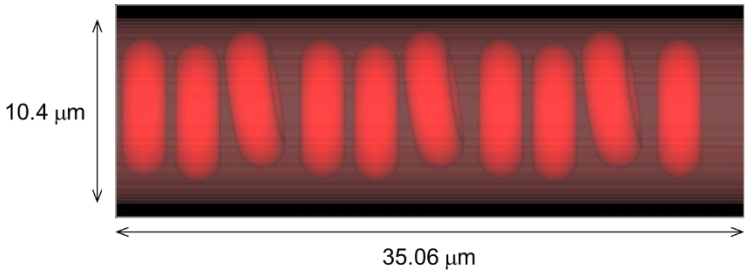
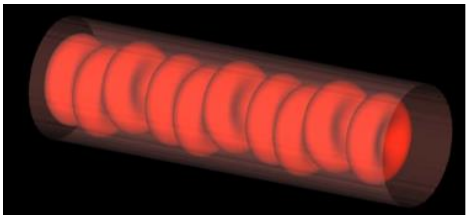
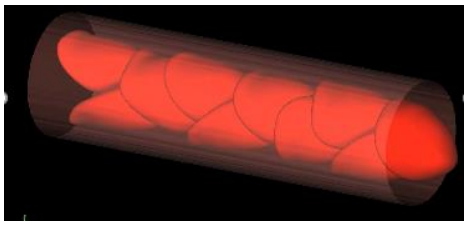


図 18　直管を用いたテスト用計算系

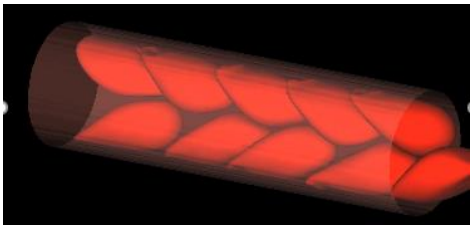
(a) t=0 (ms)



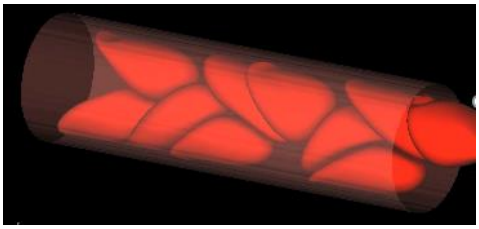
(b) t=25 (ms)



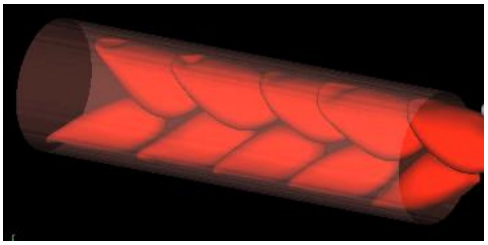
(c) t=120(ms)

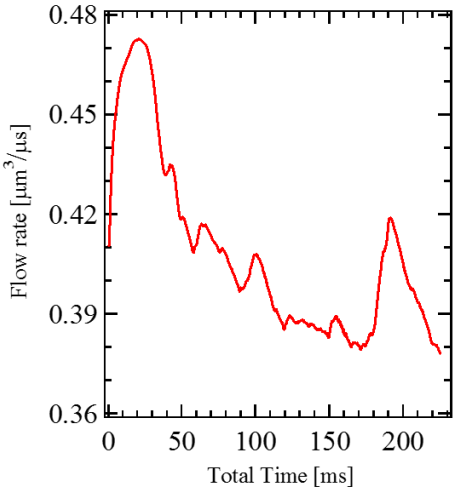


(d) t=180(ms)



(e) t=200(ms)





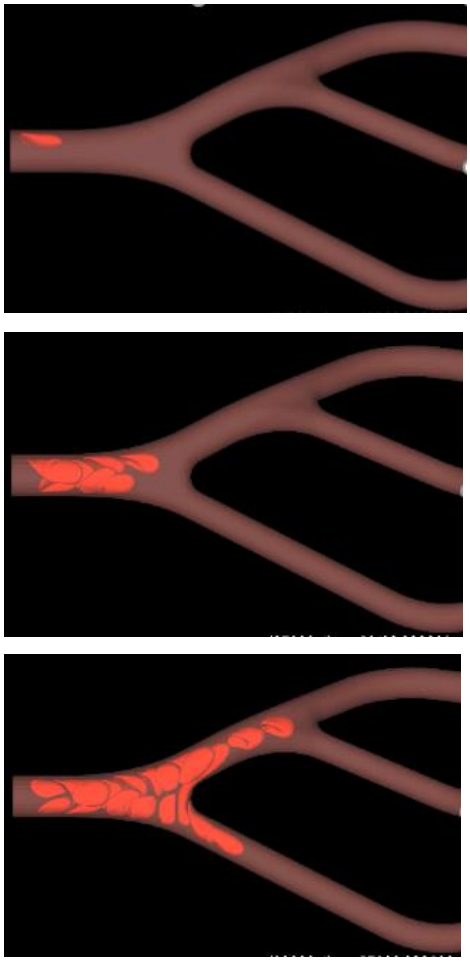
(f)　流量の時間変化

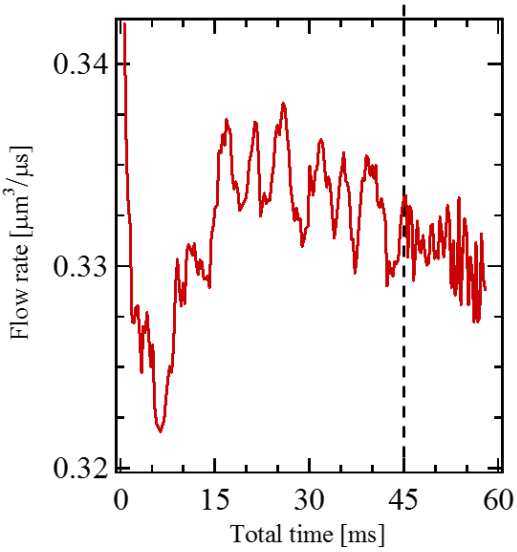
図 19　赤血球の配置の時間発展((a)～(e))と流量の時間変化(f)

(a)　t=45 (ms)

(b) t=50 (ms)

(c) t=58(ms)





(d)　流量の時間変化

図 20　流動構造の変化((a)～(e))と流量の時間変化(f)

(B)「抗血小板薬のモデリング」

（１）緒言

本研究では，心筋梗塞・脳梗塞の最終段階である血栓症を再現するため，多数の赤血球や血小板の変形流動構造に関する血流解析を実施しながら，血小板の動脈硬化層への粘着から始まる血栓の生成・成長プロセスを再現し，抗血小板薬のモデリングを行うことを目指し研究を進めてきた。

（２）血小板粘着のマルチスケールモデリング

血栓症は，心筋梗塞・脳梗塞を引き起こす重要な循環器系疾患である。動脈硬化巣において血管内皮細胞の損傷を受けた部位に血小板が粘着、集積するところから始まり，血小板の凝集さらには血栓の成長，血液の凝固へと進展し，血管閉塞に至るのが血栓症であり，その結果が心筋梗塞や脳梗塞などになって現れる

血管内皮細胞が機能的、器質的に損傷を受け von Willebrand 因子が血流に曝露されると、糖タンパク glycoprotein Ibα（GPIbα）を発現している血小板細胞が流体条件にて受動的に接着する。接着した血小板に作動する力と、スケールの異なる各種の力の解析計算手法を図 21 に示す。分子スケールのミクロな現象と流体力学レベルのマクロな現象の相互作用の結果，細胞スケールの血小板の粘着・脱離が決定される。血栓の形成過程は，このように様々な時空間スケールの現象が複雑に影響

しあいながら進行する典型的なマルチスケール問題である。さらに、疾病の発症には多数の血小板細胞が関与するため、スケール間を橋渡しする大規模な連成解析と大規模並列計算が必要となる。

本研究では，血管壁上の VWF 分子と，血小板表面の GPIbα 分子の間のタンパク質分子間の相互作用をモンテカルロ法で計算しながら，有限差分法に基づくオイラー型流体構造連成計算手法と連成させるマルチスケール血小板粘着シミュレーションを実施した。すなわち，血流中を流れる多数の赤血球や血小板などの血球細胞については，流れ場と相互作用して変形しながら流れていく状態を流体構造連成問題として詳細に解きながら，タンパク質分子間相互作用を取り入れたシミュレーションを実施した。さらに，膜表面の GPIbα 分子と血管壁の VWF 分子の結合部位について，分子間相互作用力を分子動力学シミュレーションにより詳細に解析した。また，「京」クラスのコンピュータを用いたシミュレーションにより知見を得るための新しいフローチャンバー実験法を提案し，抗血小板薬として使用されている ADP 受容体 $P2Y_{12}$ の阻害薬，日本以外の欧米諸国にて使用されている GPIIb-IIIa の阻害薬に関して，新しい知見を得ることに成功した。その結果について以下に説明する。

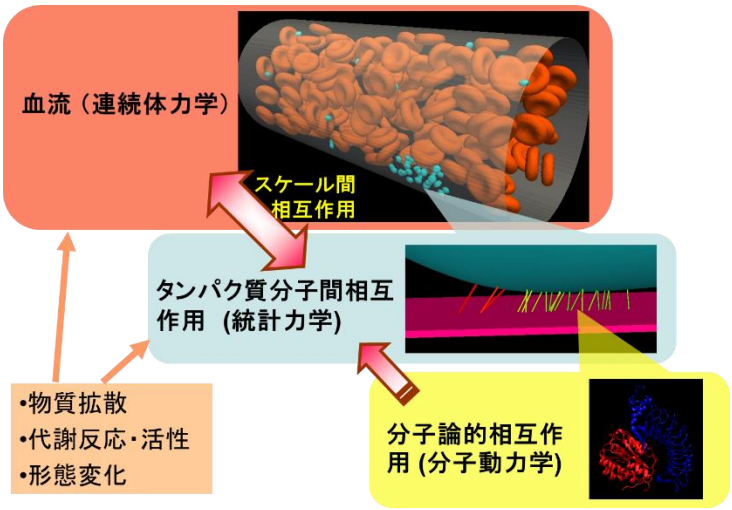


図 21　血小板粘着プロセスのマルチスケール性とモデリング

(3) 血小板粘着の大規模数値シミュレーション

本研究では，実際のフローチャンバーと同寸法の系で大規模数値シミュレーションを実施し，VWF の吸着している壁面への血小板粘着の様子を再現した。

図 22 に計算結果の一例を示す。この計算例は，赤血球の体積率を表すヘマトクリット値が 20%の条件で，実際のフローチャンバー実験に対応する 100 ミクロンの高さを持つチャネルの中に約 100 万個の赤血球が含まれている計算になっている。図手前側の壁面に VWF 分子が埋め込まれた系になっており，手前側壁面の青色の部分が GP1b-αと VWF のタンパク質分子が結合しているサイトになっている。

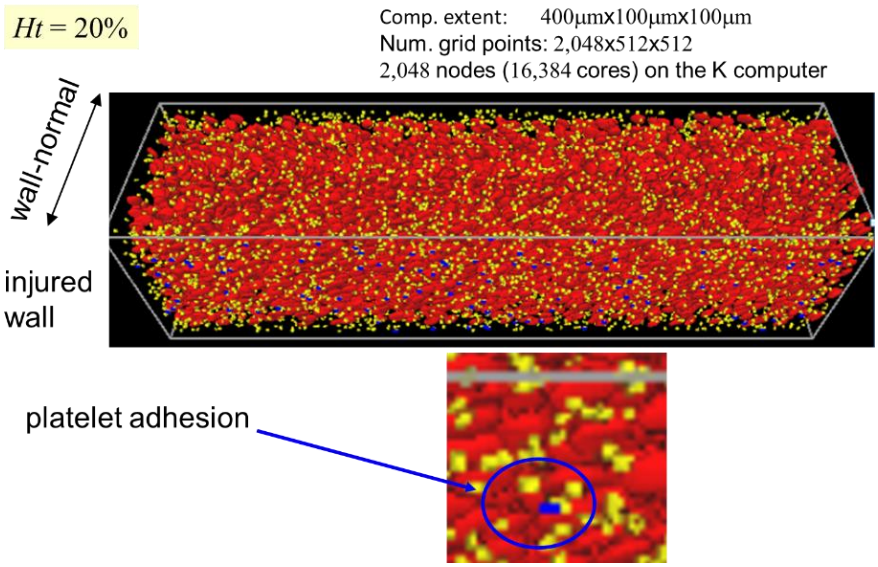
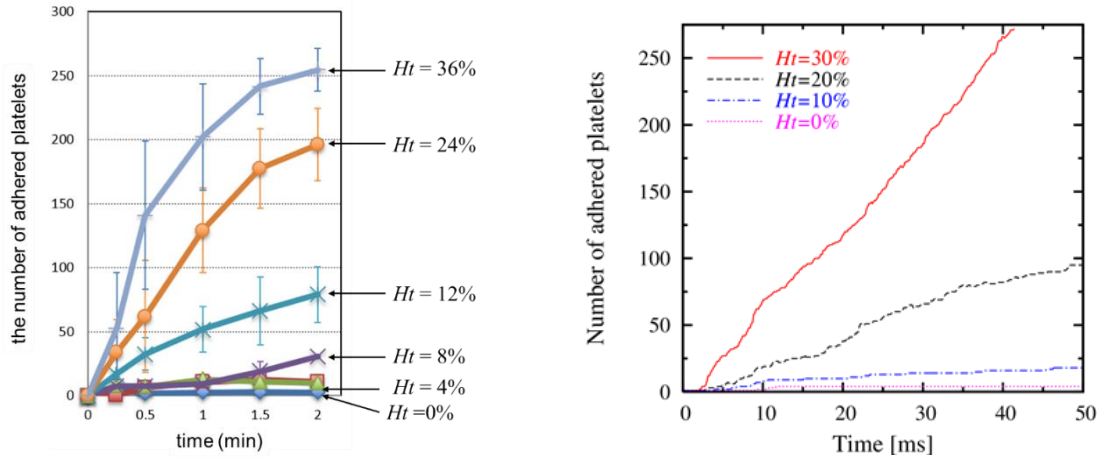


図 22　実際のフローチャンバー実験と同寸法での大規模シミュレーション

図 23 にフローチャンバー実験とシミュレーション結果の比較を示す。ヘマトクリット値の上昇すなわち，赤血球数の増加とともに血小板粘着数が大幅に増加している実証実験結果をシミュレーションによって捉えることに成功した。シミュレーション結果より，赤血球のもたらす壁面垂直方向の速度変動が極めて重要な役割をしていることが示された。



(a) 実験結果　　(b) シミュレーション結果

図 23　フローチャンバー実験における粘着血小板数のヘマトクリット値依存性
（実験結果とシミュレーション結果の比較）

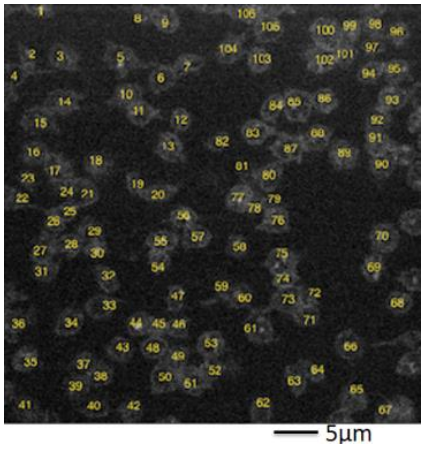
(4)抗血小板薬の薬効メカニズム解明に向けたシミュレーション

抗血小板薬として実際に利用されているクロピドグレルは，血小板表面にある ADP 受容体 $P2Y_{12}$ を阻害することにより薬効を示すことが知られている。本研究では，この効果を詳細に評価するために，シミュレーションモデル構築用の新しい実験系の提案を行ない，従来とは異なる方法で評価を行った。実験では， GPIIb/IIIa 阻害の下で VWF コーティングされた壁面に粘着する多数の血小板を用意した。(図 24 (a))様々な条件で血流を灌流し，血小板の剥離数を計測し，灌流液中の薬物が血小板粘着力に与える影響を評価した(図 24 (b))。実験結果より，GPIIb/IIIa の受容体機能が阻害された状態の血液に対して，ADP 添加のもとで血小板粘着が増加すること，強まった粘着力は $P2Y_{12}$ の阻害薬によって再び弱まることが示された。本実験データより $P2Y_{12}$ 阻害薬は，GP1bα と VWF の結合に基づく血小板

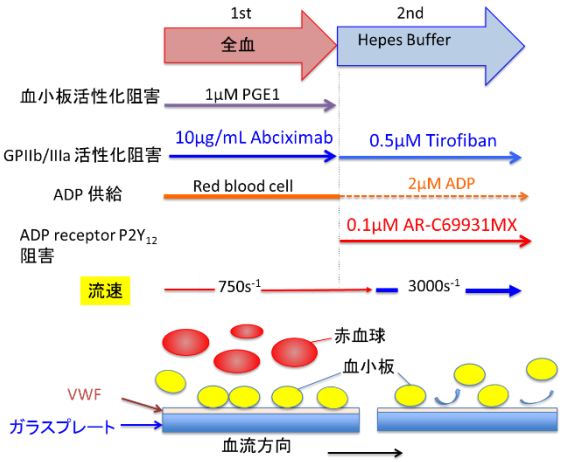
の接着に影響を与えている可能性が示唆され，従来のモデルでは考慮されていない GP1b-αと VWF 結合に対する抗血小板薬の影響をモデルに組み込むことの重要性が示された。

(5)GP1bα N末端と VWF A1 ドメインの相互作用の分子動力学シミュレーション

血小板表面の GPIbαと von Willebrand 因子(VWF)の結合部位である GPIbαの N 末端と VWF の A1 ドメインの相互作用に関する分子動力学解析を実施した。もともと VWF と GPIb α の結合は不安定なので安定構造予測が従来の方法ではできなかった。安定結合する変異体を用いた他グループの結晶構造から wild type の高次構造を予測した。水分子を配置して水溶構造を予測した。構成論的に予測した wild type の構造と、mutant の血漿構造は近似した (RMSD<2Å)。VWF と GPI α の結合による力を見積もるため，アンブレラサンプリング法により potential of mean force(PMF)を図 25 のように算出した。計算結果より力の概算をすると，両分子の結合により惹起される力が数十 pN のオーダーであることを予測した。同時に施行した原子間力顕微鏡による予測と大きな差異はなかった。本法の応用により活性化血小板における活性化構造を呈した GPIIb/IIIa と VWF、フィブリノーゲンの結合力を見積もるための基盤モデルの構築へと繋げることができる。



(a) 血小板剥離実験の初期条件



(b)シミュレーション側から提案された実験方法

図 24　フローチャンバーを用いた血小板剥離実験のデータ

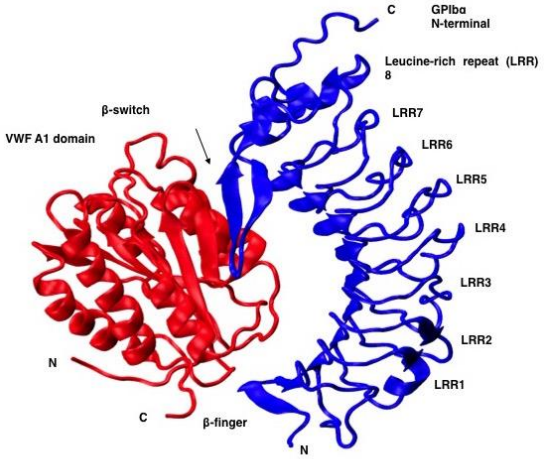
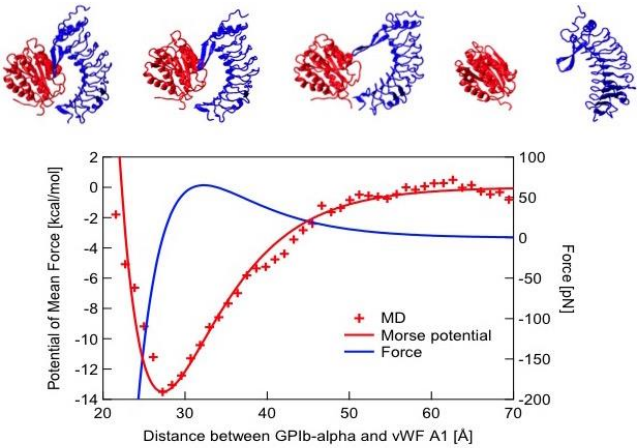


図1. von Willebrand因子とGPIbαの結合構造



(a) 結合サイトの分子構造　　(b) Potential of Mean Force のモデリング

図 25　GP1bα N末端と VWF A1 ドメインの分子動力学解析

(3-2)　心疾患の治療法・薬効評価のためのマルチスケール・マルチフィジックス心臓シミュレーション

(Ⅲ-②)心疾患の合理的治療のためのマルチスケール・マルチフィジックス心臓シミュレーション

（久田 俊明・株式会社 UT-Heart 研究所）

【実施内容】

平成 27 年度は、昨年度に引き続きマルチスケール・マルチフィジックス心臓シミュレータ UT-Heart のサルコメア動力学モデルの一層の高度化とその応用を図った。特にミオシンヘッド首振りの確率的運動を、1 次元エネルギー地形が与えられた場合に実用的な時間ステップで計算するための手法の検討を実施した。また実験系グループとの連携も継続しサルコメアモデルの実証を行った。先ず、計算手法改良においては、1 次元エネルギー地形の安定構造の谷の幅やミオシンアーム歪み変化に伴う首振り幅の変動を Kramers Smoluchovsky 近似を利用して状態遷移モデルに取り込むことにより、より少ないミオシンとアクチンの結合率でこれまでのモデルと同等の拍動性能を発揮し、エネルギー効率も向上する心筋モデルができることを確認した。また、実験系グループとのモデル実証においては、カルシウムイオン濃度中間状態におけるサルコメアの自励現象が爆発的な逆方向の首振り運動により説明できることを見出し、実験系グループが確認しているサルコメア振動の温度依存性も再現することに成功した。

【成果】

（１）モンテカルロ法に基づくサルコメアモデルの開発

心筋を単に能動的に収縮するゴムのような連続体として現象論的な方程式で記述した場合、一見医療画像と似通った心臓シミュレーションが出来たとしても、そこから得られるものは少ない。心筋の収縮は、細胞を構成する基本単位であるサルコメアと呼ばれる微小な収縮機構が根底にあり、そのメカニズムを反映したモデル化を行うことが基礎医学、臨床医学の両者にとって必要不可欠である。本研究では、サルコメアを一つの代表的方程式で表すことを放棄し、サルコメア内の一つひとつのミオシン分子先端(ミオシンヘッド)が確率的首振り運動をすることによってアーム部のばねを伸展し収縮力を発生するモンテカルロ法に基づく力学モデルを独自に開発した。これにより以下の諸点を特徴とする実装が可能となった。①ミオシンヘッドとアクチンフィラメントの間の結合・解離の遷移率が近傍の分子の状態の影響を受けて変化する協調性（Cooperativity）の導入、②ミオシンヘッドの首振り確率を、ヘッドに蓄えられる内部エネルギーとアームの歪みエネルギーから定められる Boltzmann 因子を基に決定（ATP 加水分解エネルギーがアームの歪みエネルギーに変換されるエネルギー変換過程を統計力学的に矛盾のない形で表現)、③マクロスケールからの影響を合理的にミオシンヘッドの首振り運動に反映。以上は計算機パワーを必要とするモデル化であるが、「京」により順調に開発が進められた。世界で唯一のサルコメア数理モデルとしてポスト「京」において更に発展を目指している。

（２）3 レベルマルチスケールシミュレーション手法の開発（SIAM MMS, 2013）

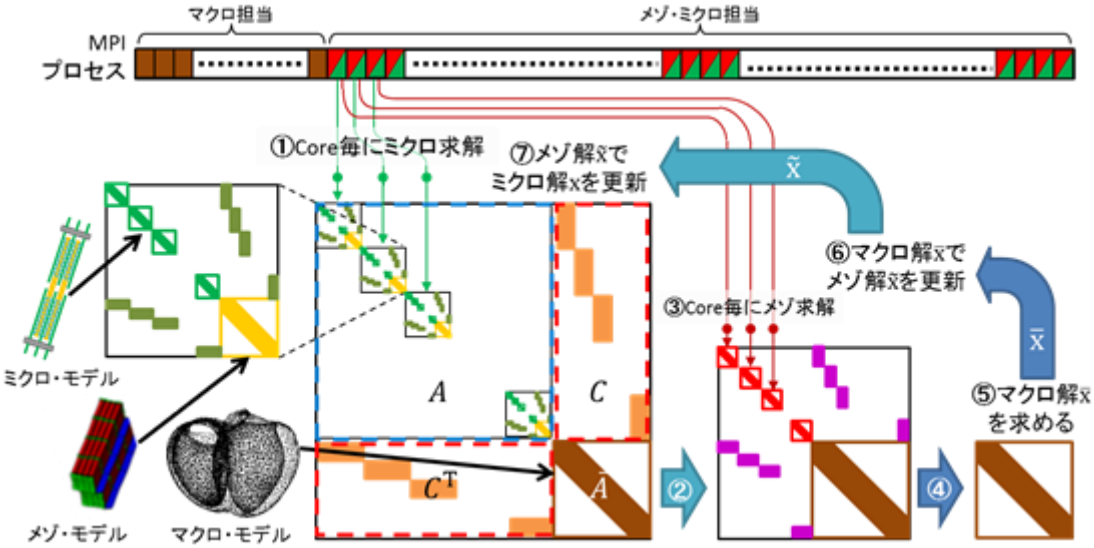


図 26　3 レベルマルチスケール解析手法

ミクロレベルの物理量の空間平均がマクロレベルの変数になるという概念を活用すると、ミクロ変数同士は直接連成せずマクロ変数を通じてのみ間接的に連成をすることになる。これから図 26 に示されるような特徴的なパターンを有する係数行列の連成方程式が得られ、ミクロモデルに関わる計算の主要部を超並列計算機で分散処理することが可能となる。また時間尺度の違いは、ミクロの仕事の時間積分をマクロの仕事と一致させることで克服した。更にメゾレベルも導入して 3 レベルに亘る現象をシームレスに接続するマルチスケールシミュレーション手法を世界で初めて開発した。以上により、サルコメアモデル、細胞モデル、臓器モデルの 3 階層連成計算をフルノードに近い構成の「京」で行い、約 17 時間連続稼働させ 1 心拍半の心臓シミュレーションを実現し、生理学的にも妥当な結果を得た。

（３）小児先天性心疾患の外科手術予後予測シミュレーション

小児先天性心疾患は多様性に富み、術後の循環動態や心臓への負荷を予測することは成人の心臓手術に較べて難しい。前項1で開発したサルコメアモデルに基づくテーラーメード心臓シミュレーションにより、先ず術前の状態を再現した後、想定される複数の術式を計算機上で再現し最適なものを選択することが出来るようになった。図 27 に術前の心臓に関する血流及び ATP 消費のシミュレーション結果を示す。ATP 消費は心臓への負荷を計る有効な指標となる。岡山大学病院心臓血管外科・佐野俊二教授の 3 症例につき検証し良い一致を見た。また長期予後予測に欠かせないリモデリングを導入するための準備として心筋線維分布最適化の理論を完成した(IJNBE, 2016)。現在、薬事承認を目指して研究継続中である。

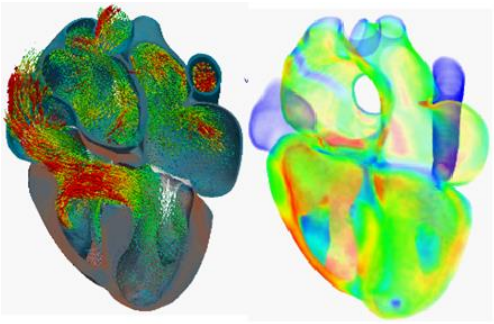


図 27　小児心臓の血流と ATP 消費

（4）心毒性解析のための心電図データベース構築の有効性に関する検討

創薬における成功率を高めるため、当研究チームではパッチクランプ実験と UT-Heart を組み合わせたハイブリッド評価系を開発し、不整脈発生のリスクが従来のどの方法より良好に予測できることを 12 種類の薬剤を用いて実証した（Science Advances, 2015）。しかし「京」のパワーを使えば、予め各種イオンチャネルの阻害率の網羅的組合せに対してシミュレーションを行い、例えば図 28 に示すようなデータベースを準備することが可能となる。製薬企業は候補化合物の各チャネル阻害率を実験等から推定するだけで、本データベースから不整脈発生領域への距離(安全率)と現在のガイドラインで求められている QT 延長の程度を直ちに知ることが出来る。また複数のイオンチャネルの相乗効果や心電図指標との関係も明らかになるなど学術的価値も高い。

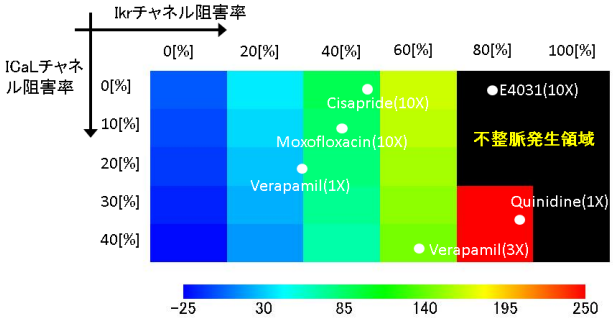


図 28　心電図データベースの一例

チャネル阻害率に対する QT 延長[色表示]と不整脈発生の関係。プロットは幾つかの具体的薬剤の位置を示す。

(3-3)　**神経疾患による運動機能障害解明のための全身筋骨格-神経系統合シミュレーション**

【実施内容】

(Ⅲ-①)(a)筋繊維の集合体としての骨格筋のシミュレーション

（高木周・東京大学大学院工学系研究科）

(a)「筋繊維の集合体としての骨格筋のシミュレーション」の部分については、下記の中村仁彦、銅谷賢治、野村泰伸の研究チームと協力し、パーキンソン病の病態である振戦と固縮の違いの再現を目指し、平成 27 年度は、全身筋骨格系・脳神経系の全身統合シミュレーションに向けた研究統括を行った。具体的には，高木の研究チームで開発を進めてきた筋繊維レベルからの骨格筋マルチスケールモデルを、東大・中村の研究チームで開発してきた体積効果を考慮した全身筋骨格系モデルへ導入し，「京」での計算速度を向上させるためのチューニングを行った。

(Ⅲ-④)全身筋骨格モデルにおける筋肉の体積効果及び人の行動モデルに関する大規模並列計算

（中村仁彦・東京大学大学院情報理工学系研究科）

平成 27 年度は，引き続き以下の研究開発を実施した。

１） 脊髄の筋支配モデルの開発

上肢および下肢の脊髄反射モデルを開発した。36,000 個からなる前庭感覚ニューロンおよび 104 個の足底感覚ニューロンのフィードバックによる姿勢制御系モデルの開発をおこなった。また，拮抗筋構造を対象としたニューラルネットワークによる姿勢制御系の学習モデルの開発をおこなった。銅谷Ｇの皮質のＰＤ病態モデルとの統合により，錐体路系を模擬したニューロン 15547 個の運動発火信号による，全身 374 本の筋を用いたシミュレーションを実現した。

２） 筋繊維モデル，脳モデルとの統合

銅谷Ｇ，高木チームとの統合計算により，上肢のＰＤ病態における振戦シミュレーションをおこなった。野村チームのＰＤ患者の振戦時筋電・加速度データから明らかになった腕関節の振戦活動，およびその筋活動度のデータに基づいてパラメータ同定を行い，それを用いて前腕の神経筋骨格モデルと脳モデルとの統合シミュレーションを実施した。

(Ⅲ-⑤)行動制御の脳神経系の多階層モデルの構築（銅谷賢治・沖縄科学技術大学院大学学園）

平成 27 年度は、NEST により実装した大脳基底核モデルおよび視床-大脳皮質モデルを、東大高木チーム、中村チームが作成した脊髄神経回路モデルおよび筋骨格系モデルと MUSIC により統合した大規模シミュレーションを実行した。大脳基底核モデルでドーパミン枯渇に相当する結合パラメタ変化により β バンドの振動が発生し、その入力を受ける視床回路モデルでは約 5Hz のバースト発火が起こること、さらにその入力を受ける大脳皮質モデルでは拮抗する筋肉を支配する錐体路細胞の間で交互の発火が起こることを確認した。その出力を脊髄神経回路および筋骨格系モデルに与えることで、腕の振動が起こることを確認した。さらに振戦と固縮症状の発生を統一的に説明する神経回路機構の探索を行った。

(Ⅲ-⑥)コンプライアントな生体筋・腱系に駆動されるヒト骨格系の運動制御に対して大脳基底核が果たす役割の数理モデル化（野村泰伸・大阪大学）

平成 27 年度は、（1）高木チームの「筋繊維レベルからの骨格筋モデル」、銅谷チームの「脳神経系の多階層モデル」、中村チームの「脊髄と四肢筋骨格モデル」を統合したインターフェイスの開発と利用を引き続き実施した。（2）運動に関わる関節の受動的粘弾性が低くかつ能動的神経フィードバック制御のゲインも小さい（関節がコンプライントである）という生理学的に妥当な条件の下で全身機械力学モデルの直立姿勢と歩行を柔軟かつ安定化に実現する制御機構の数理モデル化を平成 26 年度に引き続き実施し、得られた知見を（1）における大規模詳細モデリングに対してフィードバックした。特に、直立姿勢および歩行周期に現れるフラクタル的揺らぎを生成するマクロな制御数理モデルを構築した。

【成果】

（1）緒言

本研究テーマでは、具体的な疾患としてパーキンソン病を取り上げ、疾患を再現するモデル構築を行った。パーキンソン病は手足のふるえ（振戦），筋固縮，動作緩慢，歩行障害などの運動障害を示す神経変性疾患の一つであり、大脳基底核におけるドーパミンの不足によりその症状が現れることは知られているが、様々な異なる症状がどのようなメカニズムで生じているかについては明らかになっていない。本研究では、図 29 に示すように世界最大級の細胞数の脳神経系シミュレーションに成功した NEST と、筋線維の集合体として筋肉全体の振る舞いを再現するマルチスケール骨格筋シミュレータ Hi-MUSCLE、さらには全身筋骨格シミュレータ K-Body を統合し、パーキンソン病における振戦・固縮の違い、さらには姿勢保持障害を再現することを目指して研究を進めてきた。現在までの成果として、サルを用いた動物実験でも観測されているドーパミンの不足から生じる大脳基底核での約 15Hz の β バンドの振動を再現することに成功し、そのシグナルが視床で周波数を半分に変え、大脳皮質、脊髄から筋線維へと伝わり、パーキンソン病特有の手の震えに繋がることをシミュレーションにより再現した。

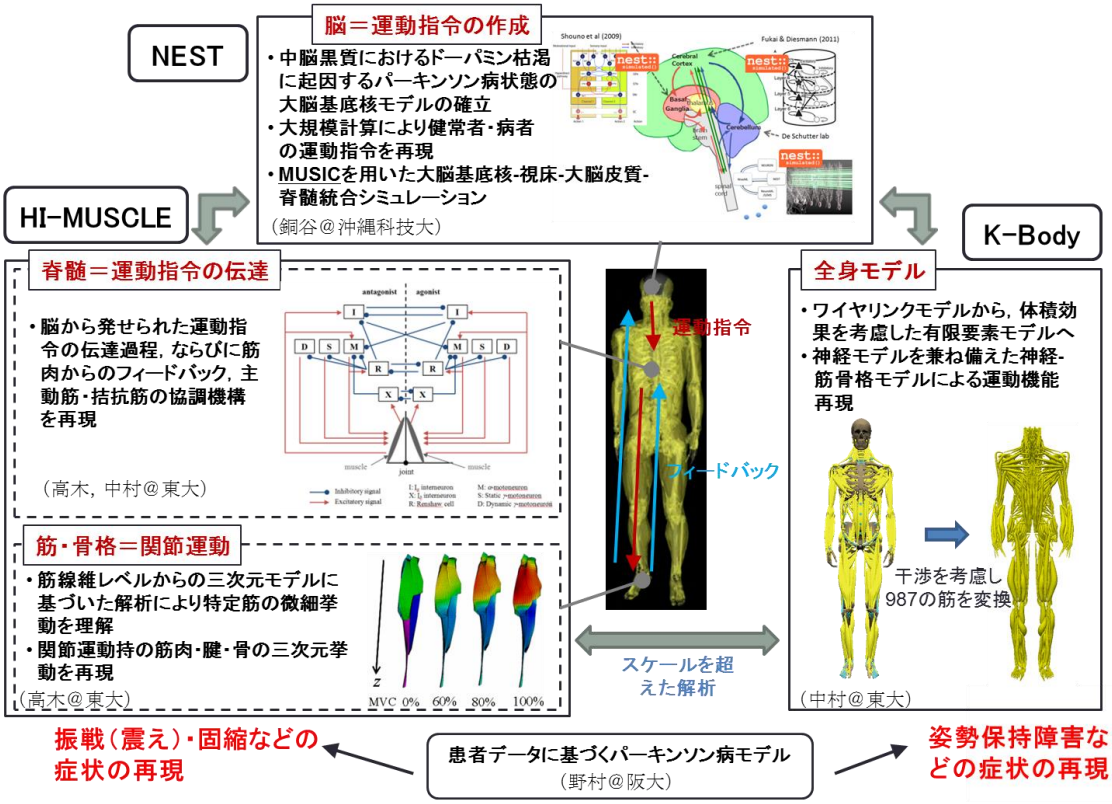


図 29　パーキンソン病の解析に向けた階層統合シミュレーション

（2）筋骨格−神経系統合シミュレータ

ここでは開発された統合シミュレータについて説明する。

ヒトの随意運動では，脳で生成された運動指令は大脳皮質から錐体路を通り脊髄へと送られる。脊髄において，運動指令は筋紡錘などの感覚受容器からのフィードバック情報と統合・調整され，脊髄前角の運動神経へと伝達される。骨格筋筋線維は運動神経に支配されており，運動神経の活動により筋線維は収縮し，最終的に，関節運動が生じることになる。本研究で開発してきた筋骨格-神経系統合シミュレータは，これら一連のダイナミクスの再現を目的とし，脳神経系シミュレータ，脊髄神経系シミュレータおよび筋骨格系運動シミュレータを統合し，ヒト全身の筋骨格系へ拡張したものである。

脳神経系シミュレータは、銅谷らの研究グループが中心になり開発を進めてきた。このシミュレータは、大脳基底核モデルおよび視床-皮質ネットワークモデルから構成される。大脳基底核モデルでは，300 万以上の神経細胞からなる線条体，視床下核，淡蒼球外節および淡蒼球内節の神経ネットワークモデルを構築する。神経細胞モデルはコンダクタンスベースモデル［1］を用いている。視床‐皮質ネットワークモデルは，4 種の神経細胞からなる視床と 18 種の神経細胞が 6 層に分布した一次運動野から構成されている。神経細胞の数，空間配置および神経細胞同士接続は実験データ［2, 3］に基づいて設定し，神経細胞モデルには積分発火型ニューロンモデル［4］を用いている。一次運動野の 1.6mm 四方の領域に相当するモデルを 18 万の神経細胞により構築している。脳神経シミュレータで作成された運動指令は最終的に，一次運動野第 5 層の錐体路細胞の発火情報として，脊髄神経系シミュレータに送られる。

脊髄神経系シミュレータとして，一次運動野の錐体細胞および骨格筋の筋紡錘につながる感覚神経（Ia 神経線維）の発火情報を受け，それらを統合・調整し，骨格筋筋線維につながる運動神経の活動（発火）を計算する神経シミュレータを開発した。運動神経のモデルには積分発火型ニューロンモデル［3］を用いた。運動神経は運動の強さによって動員される神経細胞の数は変化する。ここでは，運動神経モデルの静電容量を変化させることで，運動神経の活動の閾値を再現した。また，脊髄神経ネットワークの反射回路を考慮するため，感覚神経からのフィードバック情報を，主動筋への興奮性の入力および拮抗筋への抑制性の入力として与えるネットワークモデル導入した。感覚神経の情報は脳神経シミュレータへもフィードバックされ，脳神経内でも同様の反射回路を構成する。これら脳神経系および脊髄神経系シミュレータの神経ネットワークの発火シミュレーションはオープンソース・ソフトウェア NEST (http://www.nest-initiative.org)を用いて行った。

筋骨格系シミュレータとして，中村らの研究グループと協力して剛体リンク系で表される骨格モデルと 3 次元有限要素メッシュで表される弾性体の骨格筋モデルを連成した，筋骨格系運動シミュレータを開発した。骨格筋モデルにおいて，運動神経つながる筋線維は，運動神経の発火にしたがい収縮して力を発揮する。ここでは，筋収縮力のモデルに Hill-type モデル[5]を用いた。本モデルでは，運動の強さは筋活動度として，運動神経の発火情報と関連したモデル[6]で与えられる。また，骨格筋には筋自身が引き伸ばされたことを感知するセンサーとして筋紡錘がある。筋紡錘は筋線維の長さ変化を感知し，その情報は感覚神経を通して脊髄に発火シグナルとして伝達される。ここでは，感覚神経の発火頻度モデルに，筋紡錘のひずみ速度で記述されるモデル[7]を用いた。

これらシミュレータの統合では，シミュレータ間で神経細胞の発火情報を伝達する必要がある。そのために，MUSIC ライブラリ(MUlti SImulation Coordinator; http://software.incf.org/software/music)を用いた。図 30 に，肘関節モデルにおける脳神経系‐脊髄神経系‐筋骨格系統合シミュレーションの接続図を示す。

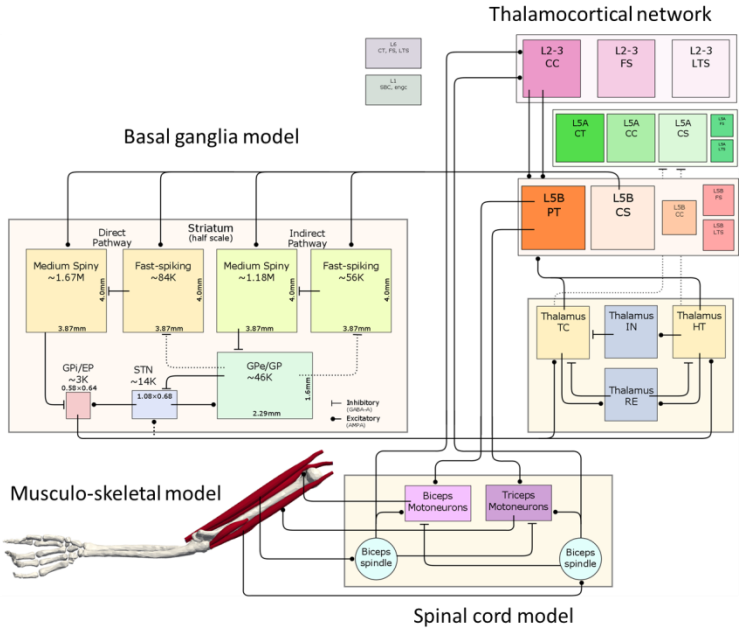


図 30　肘関節モデルにおける脳神経系 - 脊髄神経系 - 筋骨格系の統合シミュレーション

（３）パーキンソン病振戦のシミュレーション

パーキンソン病振戦は，安静時に，関節を挟んで拮抗する筋が 5Hz 程度で交互に活動することにより，4-6Hz の規則的な関節運動が生じるふるえである。

統合シミュレータによるパーキンソン病振戦のシミュレーション例を図 31 に示す。統合シミュレーションはスーパーコンピュータ「京」の 512 ノード（4096 コア）を使って行った。計算には上腕二頭筋と上腕三頭筋からなる肘関節の有限要素モデルを用いた（図 31 (a)）。パーキンソン病の脳神経の活動状態は，大脳基底核モデルおよび視床-皮質ネットワークモデルにおいて，中脳黒質におけるドーパミン枯渇による影響を反映することで再現した。

図 31 (b)に脳神経系シミュレータによる，大脳基底核，視床および大脳皮質の発火パターンを示す。大脳基底核において，サルを用いた実験[8]でも観測されているパーキンソン病特有のβバンドの振動（8~16 Hz）が再現されている。実際にパーキンソン病の症状として表れる手の震え(振戦)は、5Hz 程度と大脳基底核で観測されるβバンド振動の半分程度の周波数となるが、本シミュレーションでは、大脳基底核に現れるβバンド振動の入力が、視床の神経細胞では 5 Hz 程度のバースト発火を引き起こすことを確認した。視床からの入力を受ける大脳皮質の錐体路細胞 L5BPT では，脊髄運動神経を経て上腕二頭筋と上腕三頭筋に投射する集団の間で交互の発火がみられ、これが 4-6Hz 手の振戦を引き起こしていると考えられる。

図 31 (c)は脊髄神経系シミュレータによる，上腕二頭筋および上腕三頭筋の運動神経の発火パターンである。錐体路細胞の発火の違いにより，動員される運動神経の数は異なっている。錐体路細胞の発火と同様に，部分的に拮抗する筋の交互の発火がみられる。

図 31 (d)は筋骨格系シミュレータによる，筋紡錘とつながる感覚神経の発火パターンである。筋紡錘は引き伸ばされると，発火頻度が高くなる。筋収縮が生じると筋紡錘は短縮するため，筋収縮直後の筋紡錘の発火頻度は低下している。図 31 (e)は肘関節の関節角度の変化である。筋収縮による力が骨に伝達され，ふるえのような往復運動が生じている。

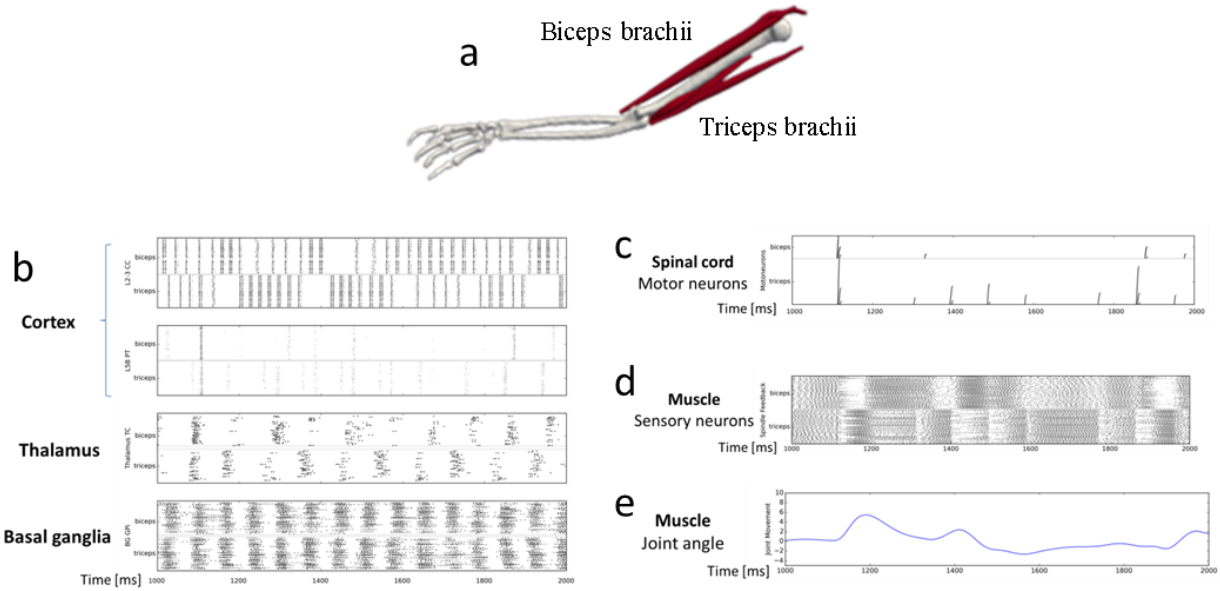


図 31　Simulation of Parkinsonian Tremor.

(b) Spikes of neurons in brain output from brain nerve simulator.

(c) Spikes of motor neurons output from spinal cord nerve simulator.

(d) Spikes of sensory neurons output from muscle spindles of musculo-skeletal simulator.

(e) Time course of elbow joint angle variability during tremor movement simulated by musculo-skeletal simulator.

（4）　結言

本研究では脳神経疾患の運動機能障害の病態予測と治療支援を目指して，神経細胞レベル、筋線維レベルからの神経-筋骨格系の階層統合シミュレータの開発を行い、ヒトの運動に関連した，脳神経系，脊髄神経系および筋骨格系の一連の統合シミュレーションに成功した。具体的な疾患例として、パーキンソン病を対象とし、振戦・固縮・姿勢制御保持障害の再現を目指した。現時点においては、パーキンソン病振戦のシミュレーションに関して、大脳基底核における β バンド振動の再現と視床における 5Hz 程度の振動、その結果として上腕に震えが生じることを再現することに成功した。今後、振戦に関しては、継続的に周期的発火が発生することの再現や、患者データをより正確に再現するモデルの開発、固縮や姿勢制御保持に関しては、固縮と振戦の違い生み出す因子の抽出や、姿勢制御の大規模シミュレーションを実施していくことが重要となる。

参考文献

[1] Shouno et al., *The Basal Ganglia* IX. 58 (2009), pp. 191-200.

[2] Weiler et al., *Nature Neuroscience*. 11 (2008), pp. 360-366.,

[3] Kätzel et al., *Nature Neuroscience*. 14 (2011), pp. 100-107.

[4] Gerstner, et al., *Neuronal Dynamics* , Cambridge University Press, (2014), Chapter II.

[5] Johansson et al., J. Theore. Biolo., 206 (2000), pp. 131–149.

[6] Cisi et al., J. Comput. Neurosci., 25 (2008), pp. 520–542.

[7] Prochazka et al., J. Physiology, 507 (1998), pp. 277–291.

[8] Tachibana et al., *European Journal of Neuroscience* 34.9 (2011): 1470-1484.

Ⅳ　戦略課題４：大規模生命データ解析
（統括：宮野悟・東京大学医科学研究所）

特定高速電子計算機施設を中核とするHPCIに最適化した最先端・大規模シークエンスデータ解析基盤を整備した上で、生命プログラムの複雑性・多様性や進化をゲノムによって理解する研究と同時に、ゲノムを基軸とした生体分子ネットワーク解析研究を行う。それにより、薬効・副作用予測、毒性の原因の推定、オーダーメイド投薬、予後予測などへの応用に貢献することを目指す。

Ⅳ－１　実施計画

(4-1) 大規模データ解析によるがんのシステム異常の網羅的解析とその応用

（宮野悟・東京大学医科学研究所）

平成27年度は、平成26年度までの研究開発を継続し、配分される「京」のリソースと「京」の運用の制約内で「数万検体規模の大規模がんサンプルデータの利用を目指し、ゲノムシークエンス、RNAシークエンス、機能性RNA、網羅的なメチル化データなどトランスオミックスデータのシステム的統合解析を実施し、薬剤感受性や予後の良・不良などがんの多様な個性をシステムの違いとして解明する。更に、平成26年度までに実施した腫瘍内ヘテロ性を再現する大規模シミュレーションにより、薬剤耐性獲得による再発を防止する新規投薬レジメンに関しての知見を得る。また、課題全体の統括を行う。

(4-2)大規模生体分子ネットワーク解析による脂肪細胞組織の刺激応答の網羅的解析とその応用

（松田秀雄・大阪大学）

平成27年度は平成26年度までに得られた成果をさらに深化させ、micro RNAを含めた、脂肪細胞の褐色化と熱産生の制御機構の解明を目指す。具体的には、網羅的なネットワーク解析により、褐色化で特異的に働くmicro RNAとその制御機構を解明する。

(4-3)次世代シークエンサデータ解析のための情報処理システムの開発（秋山泰・東京工業大学）

平成26年度までに、次世代シーケンサから産出される大量のゲノム配列情報の超高速解析ソフトウェアとして、配列相同性解析のための大規模並列ソフトウェア（GHOST-MP）の開発、エクソーム解析パイプラインを「京」上に構築してきた。平成27年度は、ヒトロ腔内菌叢のメタゲノム解析の結果をまとめるとともに、東大医科学研究所と連携してGHOST-MPを活用して腸内細菌叢のメタゲノム解析を実施する。

Ⅳ－２　実施内容（成果）

(4-1) 大規模データ解析によるがんのシステム異常の網羅的解析とその応用

（宮野悟・東京大学医科学研究所）

【実施内容】

平成 27 年度は、平成 26 年度までの研究開発を継続し、配分される「京」のリソースと「京」の運用の制約内で「数万検体規模の大規模がんサンプルデータの利用を目指し、ゲノムシークエンス、RNA シークエンス、機能性 RNA、網羅的なメチル化データなどトランスオミックスデータのシステム的統合解析を実施した。その結果、100 以上の薬剤に対して、遺伝子発現データから個々人の薬剤耐性・感受性を予測する世界最高精度の方法の構築、ATL における遺伝子異常の全体像の解明、がん遺伝子 MYC を制御するノンコーディング RNA を世界で初めて発見、未分化甲状腺癌の全ゲノムシークエンス解析による遺伝子異常の網羅的解明などを達成した。更に、腫瘍内不均性解明のための大腸がん統合解析及び進化シミュレーションにより、分子標的医薬治療及び再発の再現により薬剤耐性獲得による再発を防止する新規投薬レジメンに関しての知見を得た。これらのベースとして、GFK（Genomon-Fusion for K）による大規模シークエンス解析とその高速化の実現、Genomon2 などの開発をした。また、戦略課題４全体の統括を行った。

【成果】

５０歳を過ぎたころから、がんはだれにも巣くい始めるという「不都合な事実」が明かになってきた(Jaiswal S *et al*. N Engl J Med. 2014 Dec 25;371(26):2488-98; Xie M *et al*. Nat Med. 2014 Dec;20(12):1472-8)。がんは時空間で進化する多様な不死細胞集団で、一旦効いた抗がん剤も、がんはやがて薬剤耐性を獲得し、その薬は効かなくなる。ウイルスによる感染が原因でがんになることもある。治療により寛解したといわれたがんも、数年後に再発することがある。あるクローンがひっそりと長い年月をかけて進化し、再びがんとして出現する。その「がんのシステム異常の全貌」を「京」で解析し、「私に、この抗がん剤が効くか」を予測できたら、よい治療法のなかったがんの創薬標的が見つけられたら、がんの進化の状況とその進化原理を捉えることができたら再発を防げるのでは、と考えることは自然である。また、ヒトの遺伝子は 2 万と少しと言われてきた。しかし、ゲノムの 70%は RNA に転写されている。機能を有するノンコーディング RNA の未踏の大陸も出現してきた。がんの複雑さが次第にわかるにつれ、これまでの「美しいがん生物学」が終焉を迎えるかもしれない。この研究項目では「京」により、世界最大規模のデータ解析を行い、がんのシステム異常の多様性の実態を網羅的に解明した。

（１）「京」によるかつてない規模の網羅的がんの薬剤感受性・耐性遺伝子ネットワーク解析で、１００以上の薬剤に対して、遺伝子発現データから個々人の薬剤耐性・感受性を予測する世界最高精度の方法を構築(Park H et al. PLoS One. 2014 Oct 17;9(10):e108990; Park H et al. J Comput Biol. 2015 Feb;22(2):73-84; Park H et al. PLoS One. 2015 Nov 6;10(11):e0141869)

グランドチャレンジで開発した SiGN-L1 というソフトウェアに新たな方式的改良を行い高性能化することで、約 600 がん細胞株データ、約 100 の化合物、それぞれが１万３千以上の遺伝子からなる 60,000 以上の遺伝子ネットワークを「京」で推定し、抗がん剤が効く細胞と効かない細胞のネットワークの違いを明らかにすることができた（図 32)。データは英国サンガー研究所の“Sanger Genomics of Drug Sensitivity in Cancer”を利用した（13,435（genes）× 728（がん細胞株）遺伝子発現データ)。薬剤感受性･耐性を表現するモジュレータとしては IC50 をもちいた。IC50 は、抗がん剤などの阻害剤が、その標的の半数（50%）の働きを阻害する濃度であり、阻害剤がターゲットタンパク質の半数を阻害する際にどれだけの濃度が必要かを示し、この数値が小さけ

れば小さいほど、阻害剤はターゲットタンパク質に対して阻害活性が高いことを示している。「京」を使った最初の解析で、サンガー研究所のデータには外れ値が多く、また一様ではないため、これが高精度予測の障害になっていることが判明した。そこで、主成分空間で計算された Mahalanobis 距離を利用し、高次元データに混入している外れ値をコントロールし、外れ値に対してロバストな新しい Kernel-based Lasso タイプの正則化推定化法（Robust kernel-based L1-type regularized regression: RKLRR）開発し、この問題を解決した。こにより、各薬剤に対して細胞の感受性・耐性を決めている転写因子群を基にした約 100 の薬剤ランドスケープを描くことに成功した。これらの成果に基づき、遺伝子発現プロファイルデータから抗がん剤の感受性・耐性を予測する方法を開発した。この方法は個々人に対する抗がん剤の感受性・耐性のバイオマーカも提示するものである。そして、これまで最高精度とされていた“Garnett *et al*. Nature. 2012 Mar 28; 483（7391）:570-5” の方法(Elastic Net と呼ばれる方法を利用）を超える精度を叩き出した。投薬の個別化及びバイオマーカ―探索の方法論が「京」で作れたといえる。

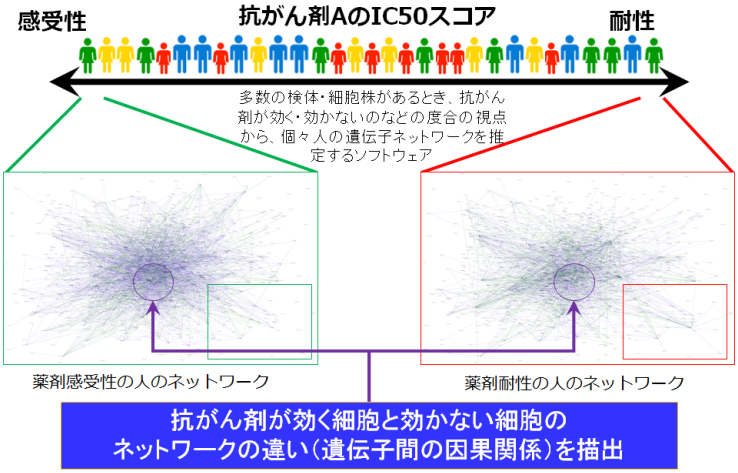


図 32　SiGN-L1 による遺伝子ネットワーク解析の概念図

（２）大規模ながんゲノム解析を中心としたがんのオミクスデータ解析によるがんのシステム異常の網羅的解明

（２）－１：乳児期に HTLV-1 ウイルスに感染し、数十年後に発症し、発症後半年で死にいたる成人 T 細胞白血病リンパ腫(ATL)における遺伝子異常の全体像の解明(Kataoka K et al. Nature Genetics. 47(11):1304-1315, 2015)

ATL に対する現行の化学療法による長期生存率は 15%であり、治療成績は ATL が疾患として認知された約 40 年前からほとんど向上していない。それは、遺伝子異常の全体像が不明のままであり、その病理理解の不足に起因していると考えられていた。京都大学大学院医学研究科の小川誠司教授と新学術領域「システムがん」で開発してきたがんゲノム解析パイプライン Genomon の計算負荷の大きい部分を「京」で担い、それ以外を東大医科学研究所ヒトゲノム解析センターの Shirokane2 で細やかに解析し、48 例の全ゲノムシークエンスデータ解析を含む、史上最大規模である 426 例の ATL 検体の解析を行った。その結果、ATL において計 50 個の遺伝子に有意に変異が認められることを明らかになった。これらの遺伝子の多くは ATL において初めて変異が報告される遺伝子であり、13 個の遺伝子は 10%以上の症例に変異を認めた。次に、426 例を対象としてマイクロアレイを用いたコピー数解析を行い、計 26 個のゲノム欠失領域および計 50 個のゲノム増幅領域を同定し、その多くで原因となる遺伝子を同定した。これらの遺伝子変異とコピー数異常を合わせると、ATL の 98%

の症例に少なくとも一つの遺伝子異常を認められ、その遺伝子異常の全貌が明らかになった（図33)。要約すると以下のようになるが、その発見は、ATL の病気の仕組みの解明に大きな進展をもたらすのみならず、本疾患を克服するための診断や治療に関して大変重要な知見を与えるものである。小川誠司教授らは、本研究成果に基づいて、京都大学、宮崎大学、武田薬品工業による産学連携医療イノベーション創出プログラム「成人 T 細胞白血病リンパ腫に対する新規テーラーメード治療」が AMED によって採択され、ATL の克服にむけた、新規治療薬の開発を実施する予定である。(a) 全体の 90%以上を超える症例に T 細胞受容体シグナリング/NF-κB 経路の少なくとも一つの遺伝子異常を認め、この経路の異常が ATL の病態において中心的な役割を果たすことが示唆された。(b) Genomon-fusion による RNA シークエンスデータの解析により CTLA4-CD28 または ICOS-CD28 融合遺伝子が認められ、また Genomon-ITD による解析により全ゲノム解析でも同じ部位にタンデム重複(ゲノムの一部の領域が重複して縦列に並ぶ構造異常)が認められ、融合遺伝子の原因と考えられるに至った。そしてこの融合タンパク質が T 細胞受容体シグナリングを活性化させるために重要な B7/CD28 共刺激経路を増強させる異常を起こしていることが示唆された。(c) 現在 ATL に対して唯一使用可能な分子標的薬(ポテリジオ®)の標的は CCR4 分子であるが、同様の機能を持つ CCR7 (11%)にも高頻度に分子が途中で切断されるタイプの変異が認められ（ケモカイン受容体の活性型変異)、有望な創薬ターゲットの可能性が見えてきた。(d) Genomon-SV による全ゲノム解析によりスプライシング異常を伴う微小欠失が同定され、これまで発見が困難であったゲノムの構造異常が発見された。特に、IKZF2 (35%)、TP73 (10%)、CARD11 (8%)遺伝子の特定の部位に集中して起こる欠失(遺伝子内欠失)は特筆すべき発見であり、これらの遺伝子内欠失はスプライシング異常を引き起こし、ATL の進展に寄与していると考えられる。このような新しいタイプの遺伝子異常の発見はスーパーコンピュータを用いた解析による画期的な成果であり、今後のがん研究全体の進展に寄与することが強く期待される。

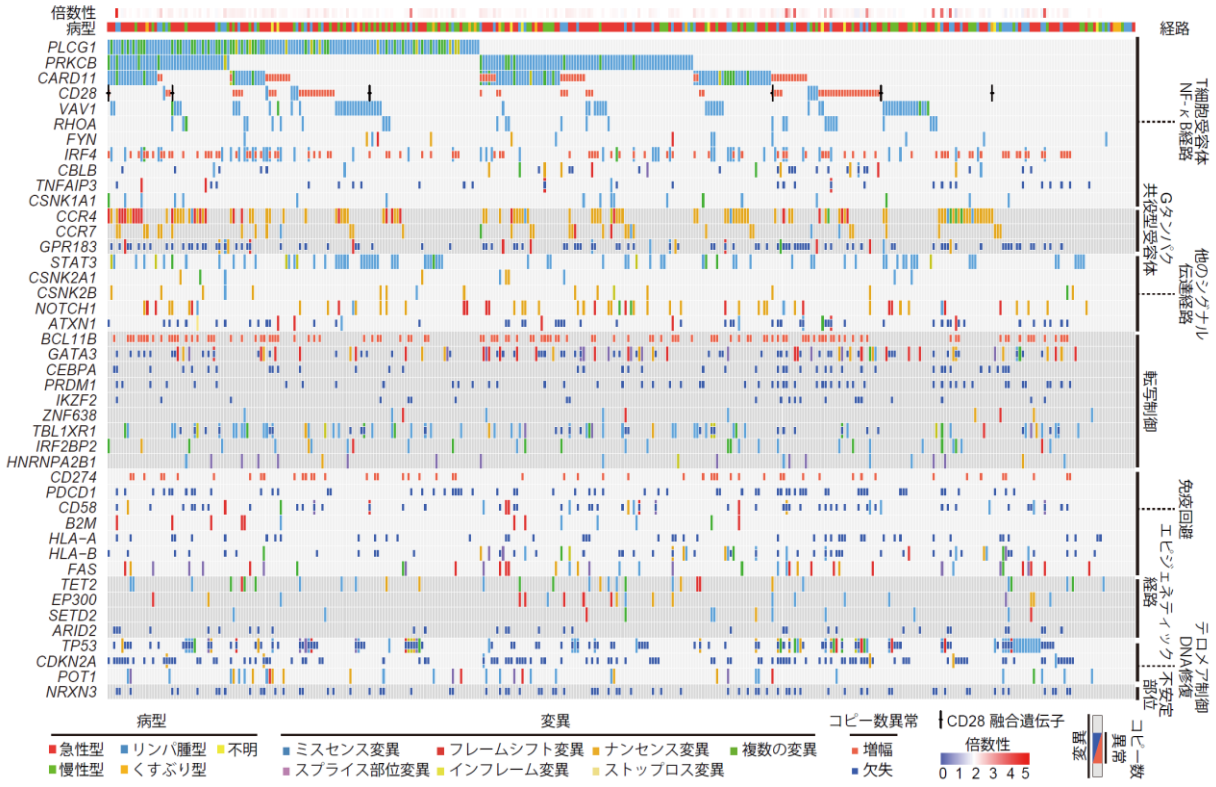


図 33　ATL における遺伝子異常の全体像（計 370 例の解析結果）

（２）－２：未分化甲状腺癌の全ゲノムシークエンス解析による遺伝子異常の網羅的解明（論文準備中）

甲状腺がんのほとんどはヒトのがんのなかで最も予後のよい分化型甲状腺がんであるが、まれにヒトのがんのなかで最も予後の悪い未分化型甲状腺がんに転化する。分化型がんにさらなる遺伝子異常が蓄積された結果、未分化転化が引き起こされていると考えられているが、その詳細は明らかになっていない。連携研究先である野口病院（大分県別府市））より甲状腺未分化癌組織 12 例、正常甲状腺組織 7 例提供を受け、全ゲノムシークエンスを施行した。「京」を用いた変異解析及びゲノム構造異常解析を行ったところ、分化型甲状腺がんに近い変異率やコピー数異常を示していた一方で、分化型甲状腺がんでは報告されていない新規の遺伝子異常や構造異常が明ら

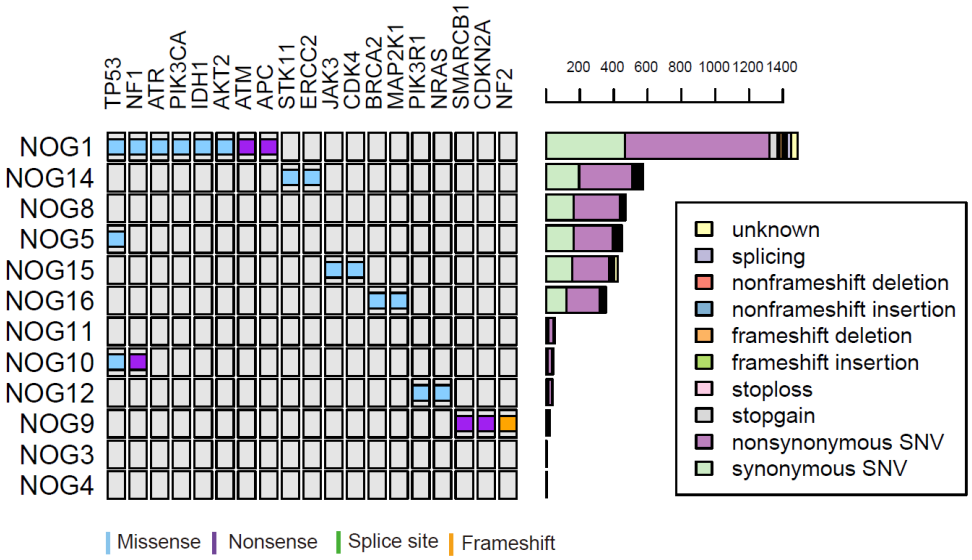


図 34　変異解析によって明らかとなった治療標的として期待される遺伝子異常

かとなった。これらの新規異常について九大別府病院において分子生物学的実験を行い、未分化甲状腺癌の治療標的となるかの検討を行う。さらに今回解析を行った症例 12 例中 8 例において他癌腫において既知の薬剤標的遺伝子異常が個別に存在することが明らかとなった（図 34）。未分化甲状腺癌は効果的な治療法の確立がされていないため、今後発展すると考えられる個別化医療によって、本研究で明らかとなったような各標的遺伝子への治療効果が期待される。

（２）－３：食道扁平上皮がんのシステム異常の網羅的解明(Sawada G, Niida A et al. Gastroenterology. pii: S0016-5085(16)00134-00137, 2016)

食道扁平上皮がんは日本を含む東アジアで多い難治性のがんである。これまでの疫学及び遺伝学的研究により喫煙、飲酒行動及びアルコール代謝遺伝子上の SNP が日本人の食道扁平上皮がんのリスクファクターとして知られている。しかしながら、体細胞変異の全体像はこれまでに明らかになっていなかった。本研究で我々は連携研究先である九大別府病院とともに日本人食道扁平上皮がん 144 症例の全エキソーム解析及びコピー数プロファイリングを行うことにより体細胞変異の全体像を明らかにした。一塩基体細胞変異の置換パターンには TpCpN サイトの C>T 置換(CpG シグネチャー)及び TpCpN の C>G/T 置換（APOBEC シグネチャー）が高頻度に認められた。さらに体細胞変異の置換パターンに基づいて 114 症例は 3 つのクラスターに分けられ、そのクラスターには喫煙、飲酒行動、ニコチン及びアルコール代謝遺伝子上の SNP が相関を示していた。統計解析によりドライバー遺伝子として細胞周期制御遺伝子(TP53, CCND1, CDKN2A, FBXW7)、エピジェネティクス制御遺伝子(MLL2, EP300, CREBBP, TET2)、NOTCH 経路遺伝子(NOTCH1, NOTCH3)、WNT

経路遺伝子（FAT1, YAP1, AJUBA）、 RTK-PI3K 経路遺伝子（PIK3CA, EGFR, ERBB2)等が同定された（図 35)。同定したドライバー遺伝子のうち、EP300 及び TET2 変異は患者の生存時間と負に相関し、ZNF750 の変異は APOBEC シグネチャー変異の数と正に相関を示した。また、細胞株実験により TET2 の変異によりがん細胞の浸潤能を増加させることを明らかにした。更に、「京」を用いた EEM 解析により食道扁平上皮がんトランスクリプトーム中の NFE2L2 発現モジュールの同定に成功し、ゲノムデータと統合解析を行うことで NFE2L2 の変異は食道がんにおいて一連の抗酸化経路遺伝子の発現を誘導していることを明らかにした（図 36)。

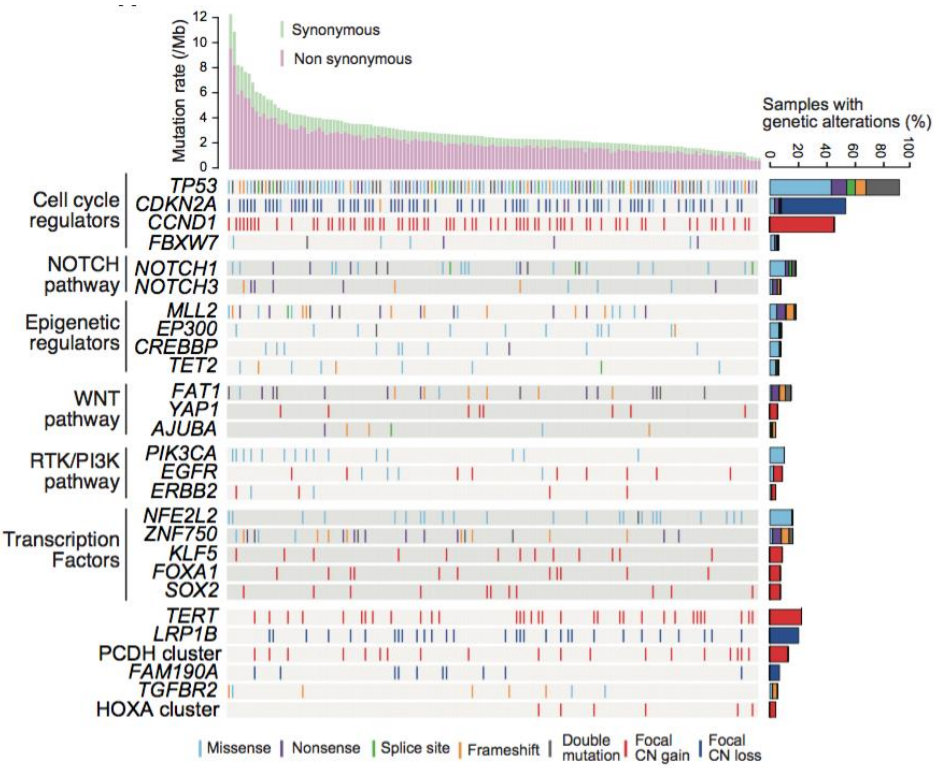


図 35　日本人食道扁平上皮がん 144 症例のドライバー遺伝子の変異プロファイル

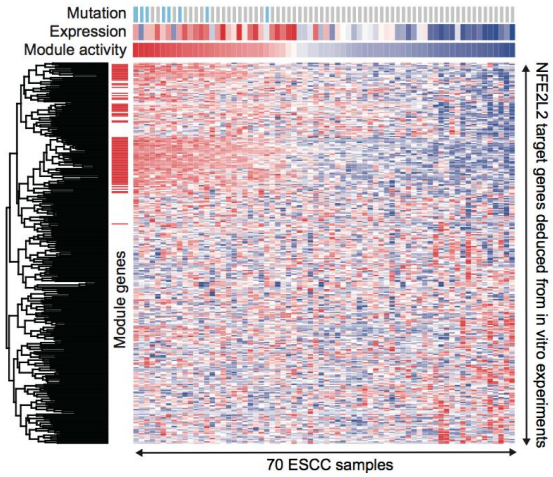


図 36　EEM 解析により同定された食道扁平上皮がんトランスクリプトーム中の NFE2L2 発現モジュール

（２）－４：胃がん腹膜播種制御モジュールの同定(Kurashige J, Hasegawa T, Niida A et al. Sci Report.6:22371, 2016)

胃がんはしばしば腹膜に転移（腹膜播種）することで最終的に人を死に至らしめるが、その分子機構の詳細はいまだわかっていない。前年度までに連携研究先である九大別府病院で取得した高腹膜播種能を獲得した胃がん細胞株の発現プロファイル及びシンガポール大学で取得した胃がん発現データセットを「京」上でEEMを用いて統合解析を行うことにより腹膜播種制御モジュールの同定に成功している。腹膜播種制御モジュールの活性が高い患者群は有意に腹膜播種が多く、生存期間も短かった。また腹膜播種制御モジュールの活性は発生、細胞接着及び細胞外マトリクス遺伝子群の発現と相関していた。今年度は腹膜播種制御モジュールに含まれる type-I collagen receptor tyrosine kinase である discoidin domain receptor 2 (DDR2)遺伝子に注目して腹膜播種能への寄与を分子生物学的に検討した。DDR2 はプロモーターの脱メチル化により発現上昇し、ゼノグラフトモデルで DDR2 の発現をノックダウンすると腹膜播種が抑制された（図37)。また DDR2 経路の阻害剤である Dasatinib の投与よっても同様に腹膜播種が抑制された。以上の結果により DDR2 が腹膜播種能の制御因子であり、新たな治療標的となりうる可能性が示された。

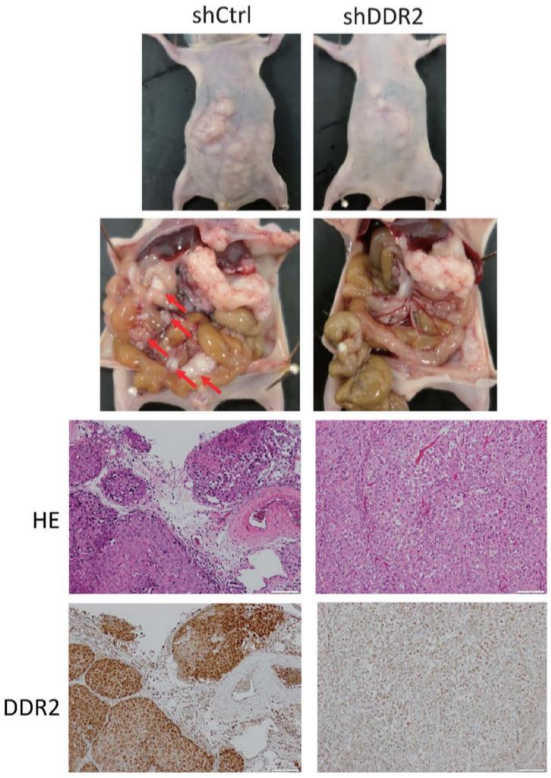


図 37　ゼノグラフトモデルにおいて DDR2 の発現をノックダウンすると腹膜播種が抑制される。

（３）がん遺伝子 MYC を制御するノンコーディング RNA を世界で初めて発見：「京」を使ったシステムがん研究でなければ、あり得ない発見（成果の詳細は今後発表予定）

1982 年に発見された重要ながん遺伝子の１つであるがん遺伝子 MYC の研究は３０年以上ジャングルを這い回るように研究されてきた。その結果、30 年以上にわたって１万８千報以上の論文が発表されているが、MYC の決定的制御メカニズムは不明のままであった。MYC は数千個のターゲット遺伝子のプロモーターに結合し、転写を制御していることが示唆されていた。MYC と協調して働くモジュレータ因子の多くは未知であり、モジュレータ因子の網羅的探索はがんの分子メカニズムの解明、さらには治療標的となる遺伝子を探索する上で極めて重要と考えた。そこで、名大医学研究科の高橋隆教授らのグループと連携して、大規模サンプルデータを使って探索するチャレンジを実行した。そのために新学術領域研究「システムがん」及びグランドチャレンジプログラムで開発した SiGN-L1 の高性能化ソフトウェア(GIMLET)を開発した。GIMLET (Genome-wide Identification of Modulators with Local Energy statistical Test)は、転写因子、ターゲット遺伝子、モジュレータ因子の３項関係を数理モデル化し、観測された遺伝子発現データからモジュレータの制御調節の変化を推定する（図 38)。そして、モジュレータの候補遺伝子の発現が変化したときに、どれくらい転写因子とそのターゲット遺伝子の制御調節を変化させるかを発現データから統計的に検定する。しかし、極めて膨大な計算を必要とし、「京」を使ってしか計算が不可能であり、「京」以前では発想すらできないアイディアであった。「京」で 7988 検体の超大規模遺伝子発現データのネットワーク解析（7988 のネットワーク）をして（世界最大規模)、１万以上もある膨大なノンコーディング RNA から MYC を操っている MYMLR（ミムラー）と高橋隆教授が名付けたロングノンコーディング RNA (lncRNA)をついに突き止めた。システムがんの方法論と「京」がなければ達成できなかった偉業と考えている。

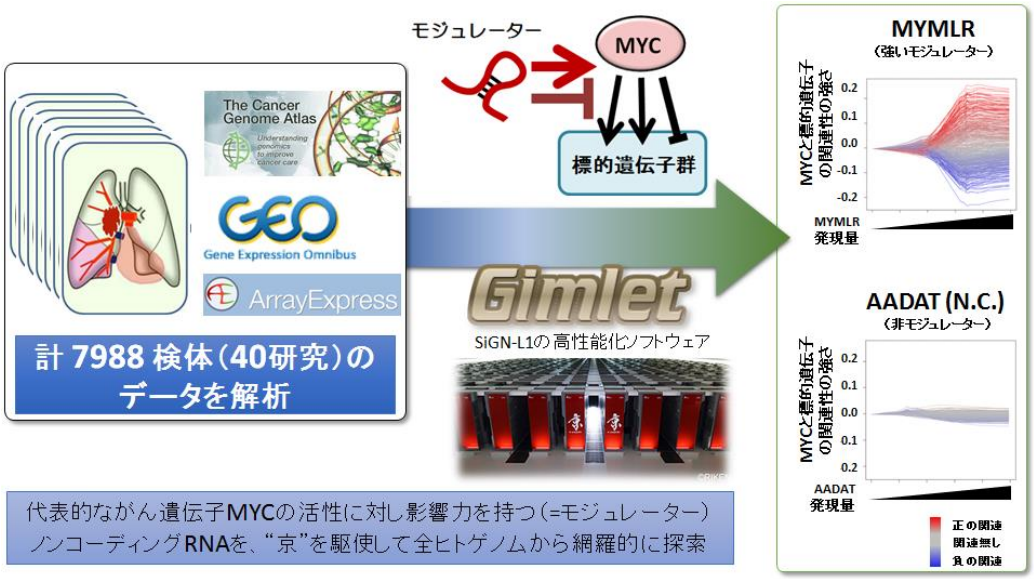


図 38　GIMLET による解析概念図

（４）腫瘍内不均性解明のための大腸がん統合解析及び進化シミュレーション(Uchi R, Takahashi T, Niida A et al. PLoS Genetics. 12:e1005778, 2016)

腫瘍内不均一性は治療抵抗性の一因であると考えられ、その理解は臨床的にも重要な問題である。前年度までに連携研究先である九大別府病院と大腸がん 9 症例各腫瘍の複数部位から取得したサンプルよりエクソーム・シークエンスデータ、DNA コピー数、DNA メチ化、mRNA 発現データを取得し、統合的に解析することによりゲノム、エピゲノム、トランスクリプトームが協働して広汎な腫瘍内不均一性を構成していることを見出して居る。また、BEP モデルを用いたがんの進化のシミュレーションを「京」上で行うことによりゲノムの腫瘍内不均一性は中立進化により生み出されている可能性を見出している。今年度はさらに、実データとシミュレーションモデルがあたえられたときに実データを生じるパラメーター値の確率分布を計算できる approximate Bayesian computation（ABC）法を用いて、BEP モデルによるシミュレーションにおいて，実データと同様の変異プロファイルを生じるパラメーター値の探索を行った。その結果、変異率の高いパラメーター値領域で中立進化により、実データと同様の変異プロファイルが再現できることが確認できた（図 39)。また BEP モデルによるシミュレーションにおいて、治療抵抗遺伝子、一定用量の分子標的医薬の投与を仮定することにより、中立進化により生み出された無数のクローン中から治療開始による環境変化に適応する治療抵抗性クローンが現れ再発が起こり得ることを見出した（

図 40)。今後この系を用いて治療抵抗性クローンの出現を阻止または遅延させる分子標的医薬の投与レジメンの探索を行うことが可能となった。

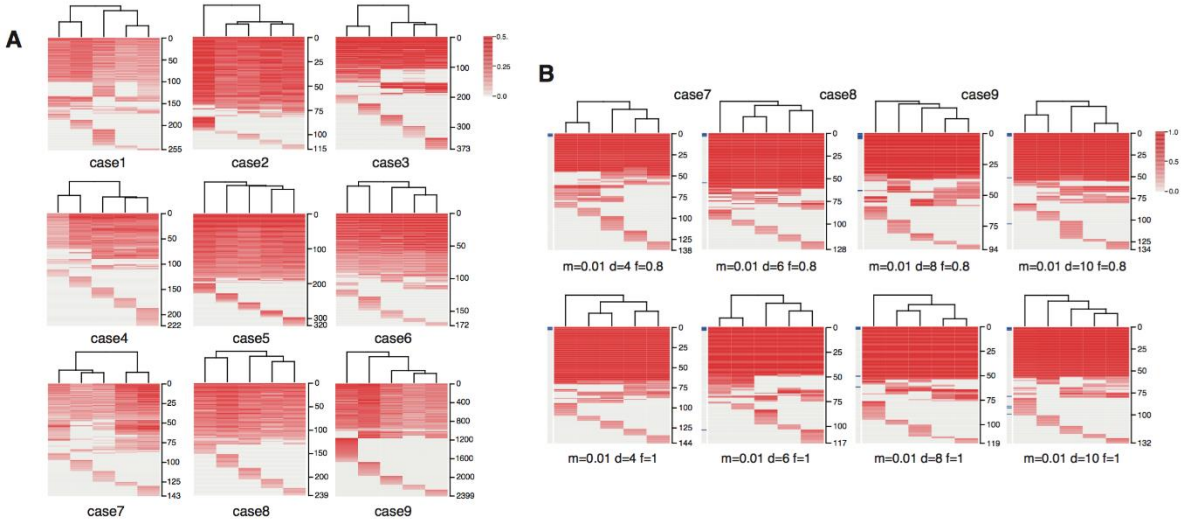


図 39　大腸がん 9 症例各腫瘍の複数部位から得た変異プロファイル(A)及び ABC 法により探索されたパラメーター値で再現した変異プロファイル(B)

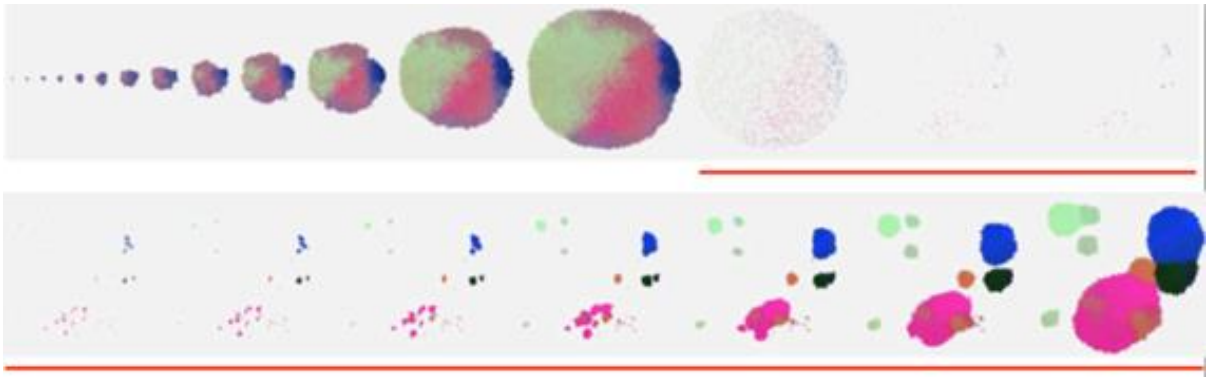
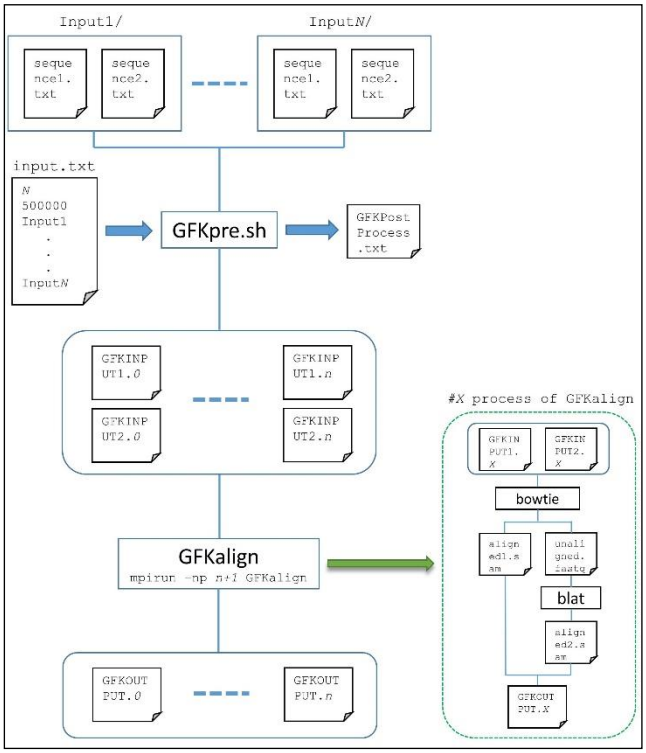


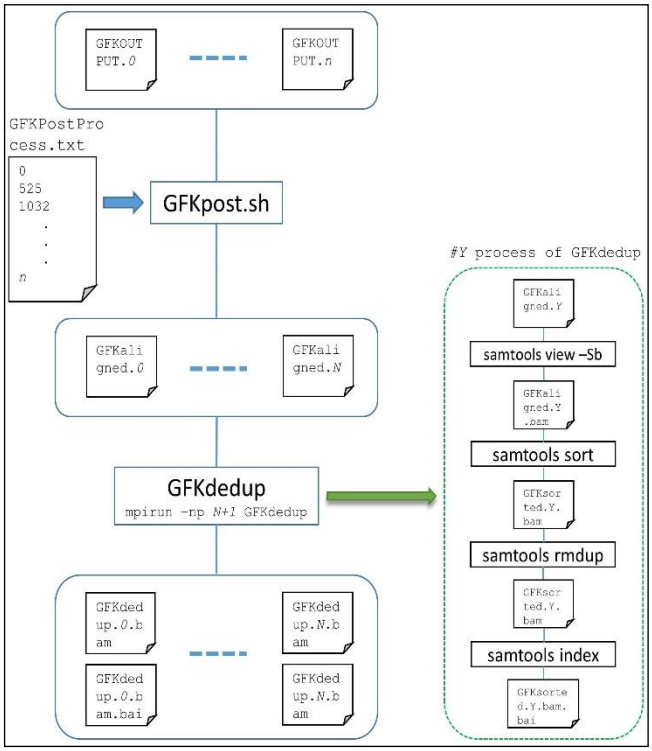
図 40　BEP モデルを用いたがんの進化シミュレーションによる分子標的医薬治療（下線で示す）及び再発の再現

（５）ソフトウェアの「京」への移植作業、及びソフトウェアの改良、新たな大規模生命データ解析の方式の研究

（５）－１：GFK (Genomon-Fusion for K)による大規模解析とその高速化(Ito S, Shiraishi Y et al. BIBM. 1441-1447, 2015; 鈴木惣一朗, 伊東聰, 他. HPCS2016 採択)

平成 26 年度に開発した Genomon-fusion の京移植版である GFK を用いた大規模解析を実施した。データは Cancer Cell LineEncyclopedia (CCLE: http://www.broadinstitute.org/ccle/home) の RNA-seq、サンプル数は 669 である。GFK のフローチャートを図 41 に示す。





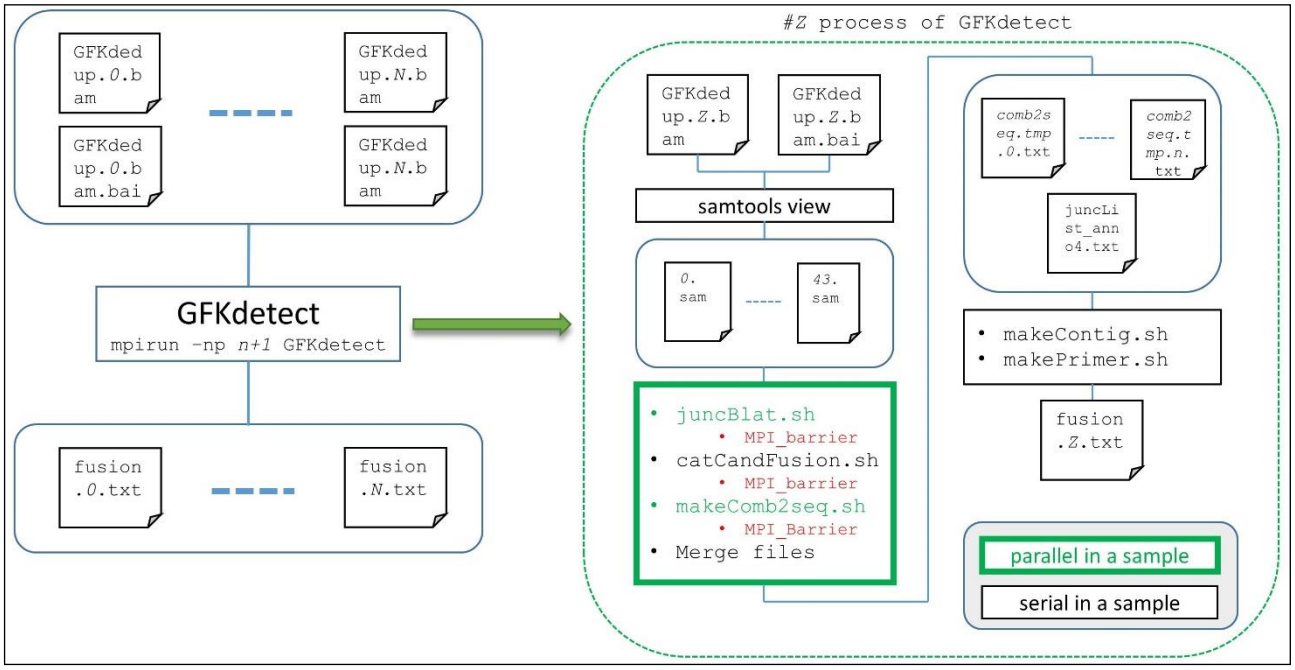


図 41　Flowchart of GFKalign, GFKdedup, GFKdetect

GFK は 3 つのステップで構成されており、GFKalign、GFKdedup、GFKdetect の順に計算を実行する。約 20TB のインプットデータを 50 検体ずつの 14 のジョブグループに分割し、解析を実施した。これは、京全系を利用しても全サンプルを 1 ジョブでは実行できないという物理的な理由と、2 万プロセス前後の並列数のジョブが実行しやすい（待ち時間が少ない）という二つの理由からである。

表 1　計算時間のサマリー

	Elapse time (node-hours)	CPU core
GFKalign	875,122.2	1,574,540
GFKdedup	4,512.8	669
GFKdetect	53,143.0	29,436
Total	932,778.0	1,604,645

表 1 に各ステップごとの解析に利用した京のリソース総量を示す。全解析におよそ 160 万 CPU コア、約 93 万ノード時間を使用した。この値を元にパイプラインの処理性能を見積もると、京全系を占有利用した場合、11,422 サンプルの解析能力となる[1]。本解析において問題となったのは GFKdedup 部分である。京コンピュータでは 1 ジョブ当たり最大 24 時間の割り当てが可能であるが、数サンプルにおいてこの制限時間を超えるものが存在した。上記解析では、失敗したサンプルについては個別にプレポストノードを用いて GFKdedup の処理を実施して問題を回避している。

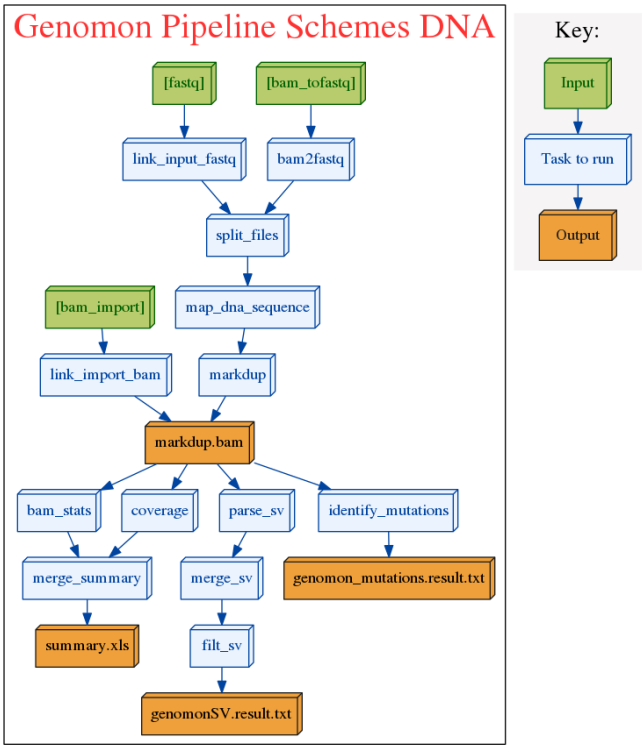
GFKdedup はソートおよび重複除去というデータ処理を主に実施するが、その性質から検体ごとに逐次処理で実装されていた。このボトルネックを解消するため、我々は染色体に着目した。ソートおよび重複除去は注目する染色体の中でのみ依存関係が発生し、異なる染色体間ではお互いに干渉しない。そのため、ここではデータを染色体ごとに分割する方式を開発した。具体的には、

GFKalign でマッピングされたリードデータを染色体ごとに個別のファイルに出力する。それぞれのファイルに対しソートおよび dedup 処理を実施し、最後に染色体の順番に結果をマージする。こうすることで逐次処理を染色体数分の並列処理（ヒトの場合は 26 並列）出来るようになり、およそ 20 倍という劇的な性能向上（パイプライン全体では 6.5 倍）を達成した[2]。本パイプラインはインターネット上ですでに公開済みである(https://github.com/HumanGenomeCenter/GFK)。

（５）－２：Genomon2 の開発(Shiraishi Y et al.)

これまで白石等が開発してきたがんゲノムシークエンス解析プログラム Genomon-Exome (http://genomon.hgc.jp/exome/en/) を包括的に改善し，効率的に全ゲノム解析を可能とする新規プラットフォーム，Genomon2 (http://genomon-project.github.io/GenomonPages/)の開発を進めた。Genomon2 は、周辺のソフトクリッピング情報や再アラインメントにより偽陽性変異を効果的にフィルターするプログラム GenomonMutationFilter (https://github.com/Genomon-Project/GenomonMutationFilter)、 多数のコントロール検体におけるエラー率の統計学的な評価により効率的に利用して偽陽性を除くプログラム EBFilter (https://github.com/Genomon-Project/EBFilter)を備え、特に複雑な配列パターンを有する non-coding 領域においても高度に正確な変異検出が可能となった。また、 GenomonSV (https://github.com/Genomon-Project/GenomonSV) により、染色体間の転座を始めとする大規模な構造変異だけではなく、数十塩基から数百塩基の中程度のサイズの構造変異の検出が可能となった。さらに、ruffuse, drmma など最新の job management ツールを効率的に利用した GenomonPipeline (https://github.com/Genomon-Project/GenomonPipeline)により、多数検体の効率的な解析を容易にしている。現在 Genomon2 はヒトゲノム解析センターのスーパーコンピュータシステムにおいて性能評価が進んでおり、また現在多くの共同研究者に利用されている。

Genomon2 のイラスト（辻田幸宏氏による）



Genomon2 の DNA 解析のスキーマ図：配列のアラインメント，点変異・欠失・挿入，構造変異の検出を行う。

（５）－３：大規模遺伝子ネットワークアプリの高性能化とその応用

（５）－３－１：大規模・網羅的がん遺伝子ネットワークデータベースの拡張

平成 26 年度までに、がん関連公開データ 256 データセット 30,261 サンプルに対してベイジアンネットワークを用いた２種類の遺伝子ネットワーク推定手法SiGN-BN HC+BootstrapおよびSiGN-BN NNSR を適用し 512 個の遺伝子ネットワークに加え，EGF 関連 1520 遺伝子リストを用いた 256 データセットを新たに構築し、SiGN-BN HC+Bootstrap を用いて 256 データセット分の計算を「京」を用いて完了し，合計 768 個の遺伝子ネットワークの推定を完了している。昨年度までに追加で計算を行った EGF 関連 256 ネットワークについては、ネットワークの推定が完了しただけの状態で、公開しているウェブデータベースサイト tcng.hgc.jp への反映は未だであった。平成 27 年度中に、これらのネットワークのデータベースへの反映を行った。新規に追加したネットワークは前述の通り 256 個で、それぞれ EGF 関連 1520 遺伝子のうち公開データ 256 データセット中で有効な発現データを持つ遺伝子を抽出しているため、それぞれのネットワークの遺伝子数は異なる。具体的には 547 遺伝子のネットワークから 1422 遺伝子のネットワークまで様々である。これらのネットワークには総計 4,934,791 本の遺伝子間の二項関係、すなわちネットワーク中の枝があり、データベースへ登録された二項関係は総計 16,545,373 個となった。これらのネットワークは TCNG (The Cancer Network Galaxy)のウェブサイト http://tcng.hgc.jp/で全て検索可能となっている。

（５）－３－２：京トーラス構造インターコネクトを利用した最適 DAG 探索大規模並列アルゴリズムの開発

SiGN-BN はベイジアンネットワークの DAG 構造探索により遺伝子ネットワークを推定している。この構造探索問題は NP 困難なため大規模な遺伝子ネットワーク推定には発見的な構造推定アルゴリズムを用いている。一方で、最適な構造探索もスーパーコンピュータと専用の並列アルゴリズムを駆使することにより徐々に大きな遺伝子ネットワークサイズへの適用が可能になりつつある。これまで最適構造探索は他グループによる 33 変数が最大であったが、京で採用されているトーラスインターコネクトネットワークに最適化された新しい並列アルゴリズムを考案し、離散モデルを用いて 36 変数の構造探索の実証に成功した（論文準備中）。京互換機である東京大学情報基盤センター所有の FX10 全 4800 ノードを用いて計算を行い、36 変数 200 サンプルのデータに対し 11 時間 38 分 21 秒で全探索を完了した。これは主に科研費若手(B)「大規模ベイジアンネットワーク構造探索並列アルゴリズムの研究」の成果であるが、実装したプログラムを京で動作するように調整を行い、30 変数 200 サンプルのデータに対して京の 324 ノード 2592 CPU コアを用いて、1741 秒で計算が終了することを確認した。FX10 324 ノード 5184 CPU コアを利用した場合は、1365 秒であった。これらの結果は前述の通り離散モデルによる計算であるため、今後 SiGN-BN で用いられている BNRC スコア関数を実装し、遺伝子ネットワーク推定に適用できるように今後改良予定である。

（５）－３－３：遺伝子ネットワーク解析・処理プログラムの京ログインノードへの移植

SiGN によって推定した遺伝子ネットワークの解析・処理プログラムである SiGN-Proc はこれまでヒトゲノム解析センターのスーパーコンピュータでのみ動作するものを提供していたが、ユーザからのリクエストを受け、京及びその互換機のログインノード上で動作するよう移植作業を行った。SiGN-Proc は遺伝子ネットワークファイルのフォーマット変換や、遺伝子間の最短経路の探索、ブートストラップ法によって推定された遺伝子ネットワークの後処理などを行うツールである。このツールの京環境への移植により、京上で様々な処理が推定結果のファイルを移動することなく行えるようになった。

（５）－３－４：TCNG を用いた腎臓における TOPK 阻害剤、MELK 阻害剤遺伝子ネットワーク解析 (Kato et al. Oncogenetics. 2016. 印刷中)

TCNG は SiGN-BN によって公開データから推定した癌遺伝子ネットワークのデータベースである。シカゴ大学の中村祐輔教授の研究室からの依頼により、開発中の TOPK 阻害剤及び MELK 阻害剤に関連する遺伝子ネットワーク解析を行った。これまでの実験により、TOPK ノックダウンによって FOXM1 の発現抑制が起こり、その逆も起こる。また FOMX1 ノックダウンにより TOPK, MELK の発現が下がるという複雑なフィードバックループの可能性が示唆されていた。そこで TOPK, MELK, FOXM1 の３遺伝子の関連をがん遺伝子ネットワークデータベース TCNG で検索を行った。TCNG には５１２個の様々な癌種のデータから推定された遺伝子ネットワークが登録されており、これらの中から３遺伝子が距離２以内で繋がる遺伝子ネットワークを探索した。既存の検索システムでは対応できなかったため新しく探索用プログラムを作成し、３つの遺伝子のうち、いずれか２つが距離２以内で繋がるネットワークを 512 個中 106 個抽出した。図 42 は、GSE12667 杯腺癌関連データから推定した遺伝子ネットワークから抽出したサブネットワークで、探索結果のうち最大のものである。

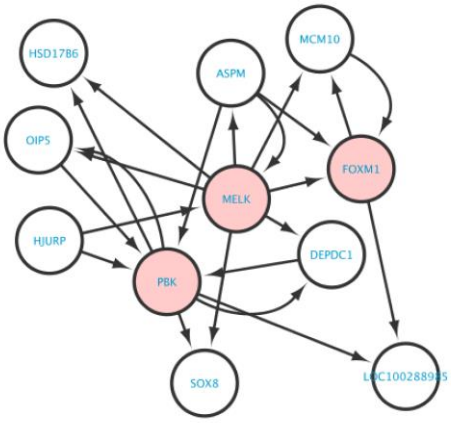


図 42　GSE12667 部分遺伝子ネットワーク

（５）－４：深層学習に基づく RNA-seq データバイアス補正技術の開発（論文投稿中）

個別化ゲノム医療への応用を目指し、先進的な機械学習技術、特に深層学習（ディープラーニング）に基づく、シークエンスデータ解析技術の開発を行った。近年、従来の機械学習技術では解けなかった様々な分野における問題に対して、深層学習が顕著な有効性を示している。その大きな特徴の一つとして、データから自動的かつ階層的抽象化により抽出される feature 情報が挙げられる、バイオインフォマティクスの分野における注目も集まってきている。

我々の対象とするデータはシークエンスからネットワーク構造にまで渡っているが、平成 27 年度は主に RNA-seq データを対象とした深層学習に基づくデータ解析技術を開発した。具体的には、RNA-seq 実験から得られる各リード配列中に観測される、塩基配列パターンの出現頻度バイアスの補正である。このバイアスは、遺伝子発現値の推定において影響を与えることが知られており、その補正はより正確な発現値推定のために重要である。一方これらのバイアス自体は、実験時の様々な要素（ライブラリ作成、RT-PCR による cDNA 断片増幅、等）の影響において生成されると考えられ、そのバイアス生成過程を陽にモデル化することは困難である。そこで我々は、深層学習モデル

の一つである Recurrent Neural Network (RNN)に基づく、シークエンス構造の先見情報を用いない塩基配列パターン出現頻度バイアス自動検出および補正手法を開発した。

より詳細には、我々の提案した解析手法は深層学習ライブラリである“Torch7”を用いて実装を行った。図 43 にパイプラインの概要を示す。まず RNA-seq リード配列を、STAR aligner を用いてアライメントする。次に seqbias により foreground 配列および background 配列を抽出する。それらを元に RNN に基づくバイアス予測モデルを学習し、各リードのバイアスの確率を推定する。最後に各リードのバイアスに基づき、リードの重みを計算し、発現値を計算することでバイアス補正を行う。RNN の本体としては、二種類の先進的なモデル、Long Short Term Memory (LSTM)および Gated Recurrent Unit (GRU)を実装し評価した。同手法を、公開データである MAQC-III データに適用した結果、我々が提案した RNN に基づく塩基配列モデルの学習は効率的であった。特に 4-10-4 GRU (入力層ノード数 4、隠れ層ノード数 10、出力層ノード数 4)によるバイアス補正後の遺伝子発現値は、バイアス補正無しの遺伝子発現値および既報のベイジアンネットワークに基づく先進的なバイアス補正手法“seqbias”による遺伝子発現値の双方に比べて、ゴールデンスタンダードとして用いた RT-PCR による遺伝子発現値に対する Pearson 相関係数および r2 値において、より良好な結果を得た (表 2)。 更に、提案手法は GRU および LTM 双方において、発現値の log フォールドチェンジ値の分布の正負のゆがみが、seqbias 法に比べて 小さいという結果も得た (図 44)。

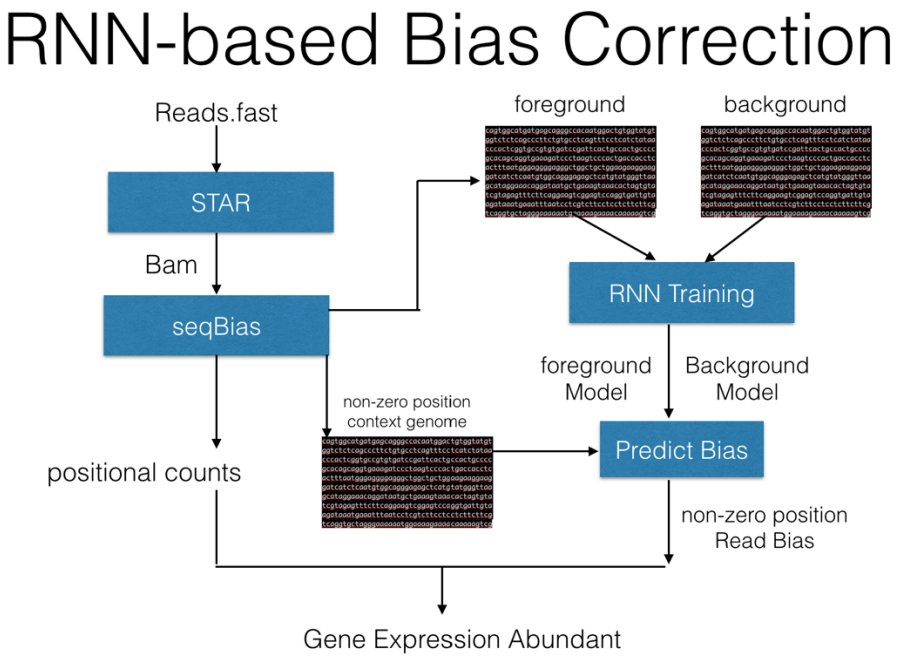


図 43　深層学習によるバイアス補正パイプラインの概要

表 2　RT-PCR 遺伝子発現値に対する相関係数値および r2 値の比較

Table 1. Correlation results on Sample A replicate 1. 922 genes are evaluated.

Method	Pearson	Spearman	r^2
original count	0.8644	0.8708	0.7471
seqbias	0.8658	**0.8751**	0.7496
GRU(4-10-4)	**0.8674**	0.8749	**0.7523**
LSTM(4-10-4)	0.8661	0.8736	0.7460
GRU(4-20-4)	0.8661	0.8729	0.7500
LSTM(4-20-4)	0.8669	0.8744	0.7515

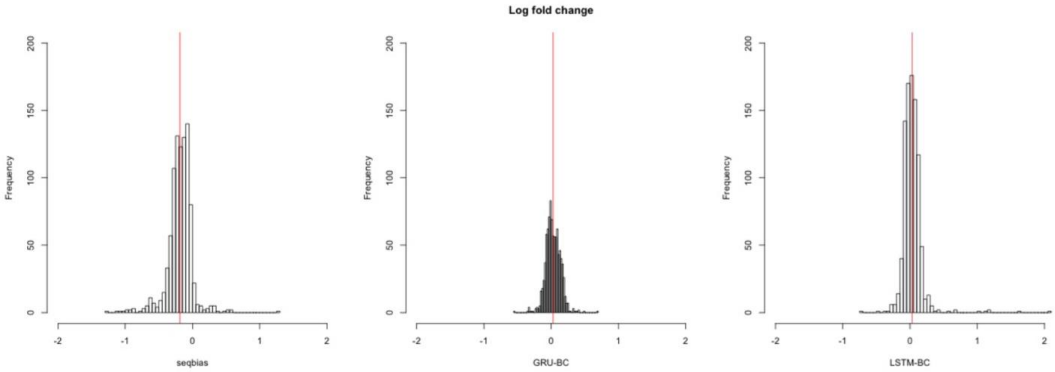


図 44　補正値分布比較

(4-2)大規模生体分子ネットワーク解析による脂肪細胞組織の刺激応答の網羅的解析とその応用

（松田秀雄・大阪大学）

【実施内容】

平成 27 年度は平成 26 年度までに得られた成果をさらに深化させ、micro RNA を含めた、脂肪細胞の褐色化と熱産生の制御機構の解明のための重要な知見を得た。具体的には、網羅的なネットワーク解析により褐色化の制御機構を解明することで、脂肪細胞が状態を変化させエネルギー消費に向けて働くときに重要な因子を明らかにした。

【成果】

ヒトを含めた哺乳類の脂肪細胞はその機能と組織学的な特性により、白色脂肪細胞と褐色脂肪細胞の２種類に大きく分類される。白色脂肪細胞は余剰のエネルギーを中性脂肪として蓄積するが、その肥大化は肥満につながり、糖尿病など生活習慣病の大きなリスク要因となっている。一方、褐色脂肪細胞は逆に脂肪を燃焼させることで熱産生を行う細胞であり、通常熱産生を行っている骨格筋と比べて約 70～100 倍の熱産生能を持っていると考えられている。この高い熱産生能力は、ミトコンドリアに存在する脱共役タンパク質 UCP1 に起因することが明らかとなっている。褐色脂肪細胞は、ヒトでは新生児の頃に数多く存在しており、体温維持の役割を果たすが、成長とともに主に骨格筋が基礎代謝の役割を担うため、褐色脂肪細胞は退縮していく。このため、少し前まで褐色脂肪細胞は成人になると消失すると考えられていたが、最近の研究で、ヒト成人での存在が PET 検査で確認され、褐色脂肪細胞の減少や機能不全が糖尿病の発症や肥満度との間に疫学的な相関が見られることから、その重要性が認識されている(New Eng J 2009, Diabetes 2009)。

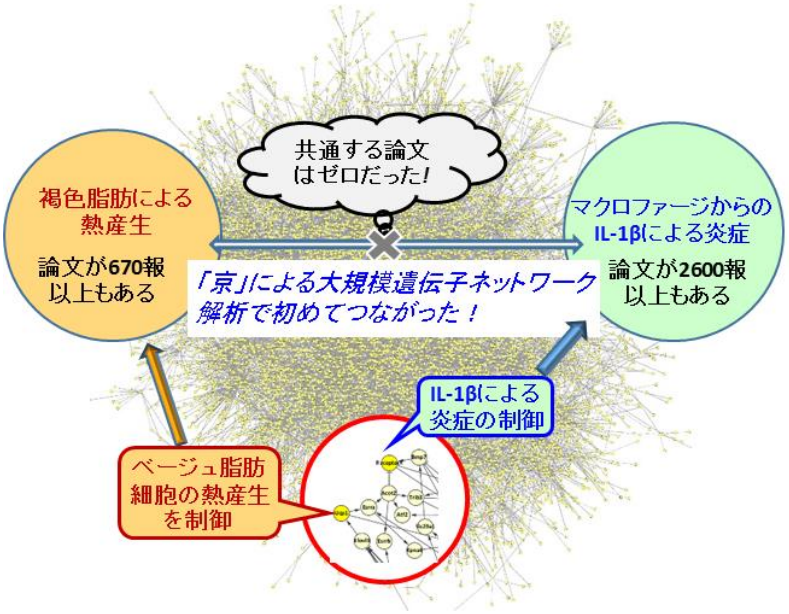


図 45　「京」により構築された白色脂肪細胞のベージュ脂肪細胞への変化過程で働く約 10,000 個の遺伝子のネットワークとそれにより得られた知見

近年、マウスを 4℃の低温環境で飼育したときに生じる、交感神経によるアドレナリンの分泌が刺激となり、一部の白色脂肪細胞が褐色化し褐色脂肪細胞に似た形に変化することが報告され、ベージュ脂肪細胞と名付けられた。さらに、ヒト成人でもベージュ脂肪細胞の存在が確認された(Nature Medicine, 2013)。ベージュ脂肪細胞は、蓄積した脂肪を分解して熱に変えることから、新しい視点からの肥満の是正につながるアンチメタボ細胞として注目を集めている。

従来の研究では、なぜ一部の白色脂肪細胞だけがベージュ細胞に変化するかは未解明であった。そこで、本研究では、京都大学農学研究科の河田教授のグループの協力を得て、4℃の低温環境で飼育したマウス個体において白色脂肪細胞からベージュ脂肪細胞へと変化する過程中の 10 時点（寒冷刺激なしと、刺激後 1，2，4，8，12，24，48，192，384 時間後）について RNA 検体を取得して、マイクロアレイにより時系列発現プロファイルを取得した。この発現プロファイルから、細胞変化過程で発現量が変化した約 10,000 個の遺伝子間の制御関係を表すネットワークを、本研究で開発したネットワーク解析ソフトウェア BENIGN により構築した（図 45)。遺伝子ネットワークの構築は、遺伝子数の２乗に比例する計算時間がかかるため、このように大規模な遺伝子ネットワークの構築は「京」クラスのスパコンがあって初めて実現できたことである。

得られた遺伝子ネットワークを調べたところ、ベージュ脂肪細胞への変化に重要な遺伝子である UCP1 の周囲に、炎症に関連した生理活性物質である IL-1β の働きを抑制する遺伝子が位置していた。IL-1β は免疫細胞の一種であるマクロファージが分泌する物質で様々な生体組織で炎症を引き起こすことが知られており、その作用について現在までに多数の報告がされていたが、ベージュ脂肪細胞を含む褐色脂肪細胞による熱産生との関連はこれまで全く知られていなかった。すなわち、褐色脂肪細胞による熱産生と IL-1β による炎症との関係は、「京」がなくては得られなかった知見であるといえる。

この結果は、京都大学農学研究科の河田教授のグループによりマウスの実験で次のように検証された。まず、肥満のモデルマウス（db/db マウス）および高脂肪食を与えて人工的に肥満化したマウスで脂肪組織における IL-1β の量を測定したところ、それぞれ通常のマウス(lean)および通常食を与えたマウス(ND)と比べて有意に増加していることが示された（図 46)。次に、低温環境で飼育したマウスに IL-1β を投与した時に、ベージュ脂肪細胞への変化に重要な遺伝子 UCP1 の発現量を計測した。室温で飼育したマウスと比べると、低温環境では約 2,000 倍にまで発現が上昇するが、IL-1β を投与すると急激に減少し、ベージュ脂肪細胞への変化が抑えられていることがわかった（図 47)。

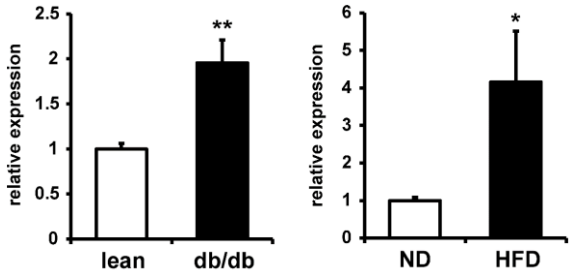


図 46　肥満モデルマウス(db/db)と高脂肪食飼育マウス(HFD)の脂肪組織での IL-1β の発現量

（それぞれ通常マウス(lean)と通常食飼育マウス(ND)との相対比（* P < 0.05, ** P < 0.01)）

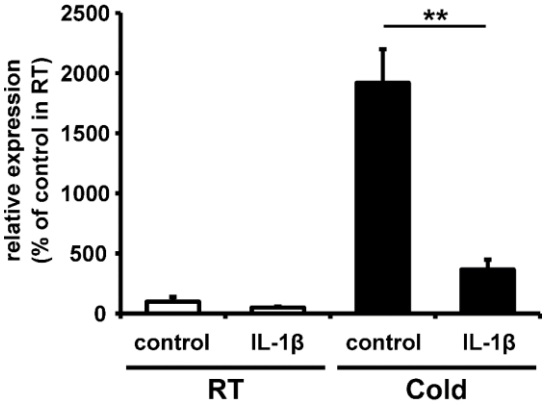


図 47　室温(RT)と低温環境(Cold)のマウスに IL-1β を投与したときの UCP1 の発現量の変化

（室温での発現量との相対比, ** P < 0.01）

以上のことから、低温環境に置かれたマウスで一部の白色脂肪細胞しかベージュ脂肪細胞へ変化しなかったのは、マクロファージが分泌する IL-1βによる抑制作用が原因であることが示唆された（図 48）。実際に IL-1βの抗体によりその作用を抑制すると、UCP1 の発現量の減少が回復することを確認した。すなわち、アンチメタボ細胞であるベージュ脂肪細胞が働くには低温環境に起因するアドレナリン刺激の他に、炎症作用の抑制が必要であるという肥満是正に向けての重要な知見が得られたが、これは「京」がなければ得られなかったものである。以上の結果は、2015 年に生理活性物質の専門学術誌 Cytokine にて発表している。

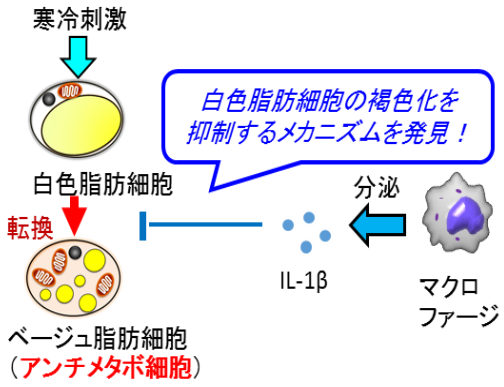


図 48 「京」を使って得られた、白色脂肪細胞からベージュ脂肪細胞への変化のメカニズムの全体像

また、最近の総説（M. Trajkovski et al. 2013, J. Y. Zhou et al. 2014 など）によると、褐色脂肪細胞における熱産生によるエネルギー消費において micro RNA の関与が報告されているが、その数は 10 個程度しかなく、しかも褐色脂肪細胞とベージュ脂肪細胞での microRNA の機能の違いがほとんど明らかになっていなかった。

そこで、本研究では、マウス個体の褐色・白色・ベージュの 3 種類の脂肪細胞について、マウス遺伝子について取得したのと共通の 6 時点（寒冷刺激前と寒冷刺激後 0, 1, 2, 4, 12, 24 時間後）について各 3 サンプルずつ micro RNA の時系列発現プロファイルを、マイクロアレイ（Agilent Expression Array Mouse miRNA 8x60k Rel.19.0）を用いて取得したものを発現プロファイルとして使用した。

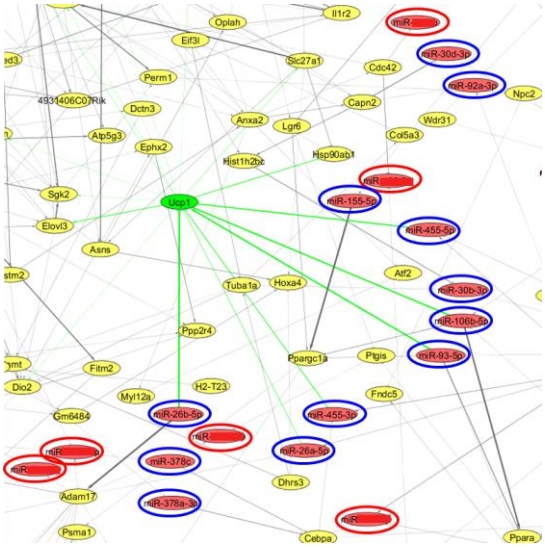


図 49 ベージュ脂肪細胞において mirco RNA を加えた遺伝子ネットワーク。UCP1（緑で表示）の周辺部分のみを拡大して表示。ベージュ脂肪細胞に関与していることが既知の micro RNA は青い丸で、未報告の micro RNA は赤い丸で囲っている。

褐色・白色・ベージュの脂肪細胞ごとに、寒冷刺激下において発現が検出された micro RNA の個数を調べたところ、それぞれ 291 個、232 個、351 個と多くの microRNA の関与が示唆された。特に、白色脂肪細胞でのみ発現している microRNA の個数が 8 個にとどまっているのに対して、褐色脂肪細胞では 11 個に増え、ベージュ脂肪細胞になると 56 個にまで増大している。このことからベージュ脂肪細胞でのエネルギー消費に向けての転換に micro RNA が大きく関与していることが示唆された。そこでベージュ脂肪細胞で遺伝子に micro RNA を加えた時系列発現プロファイルからネットワークを構成した（図 49）。図からベージュ脂肪細胞に関与していると推測される新規 micro RNA が示唆された。

(4-3)次世代シークエンサデータ解析のための情報処理システムの開発（秋山泰・東京工業大学）

【実施内容】

平成26年度までに、次世代シークエンサから産出される大量のゲノム配列情報の超高速解析ソフトウェアとして、配列相同性解析のための大規模並列ソフトウェア（GHOST-MP）の開発、エクソーム解析パイプラインを「京」上に構築してきた。平成27年度は、ヒトロ腔内菌叢のメタゲノム解析の結果をとりまとめた。さらに開発したGHOST-MPを活用して、東京歯科大学と連携して歯肉縁下のメタゲノム解析を行った。健康な歯周組織を有する対象者間でも細菌叢の構成が大きく異なり、歯周炎に関わるとされる細菌の含有率が異なることが分かった。また、東大医科学研究所と連携して腸内細菌叢のメタゲノム解析を実施した。通常下痢を引き起こす病原細菌を有する健常者が存在することが分かり、免疫がその病原細菌を抑え込む機構についての研究の端緒を開いた。

【成果】

塩基配列のシークエンシング技術の発達により、大量の塩基配列データが短期間で得られるようになった。現在もシークエンシング技術は急速な発展を続けており、得られたデータを高速に情報解析する方法が求められている。本研究開発項目では、塩基配列の高速解析を実現するためのアルゴリズムの開発とその大規模並列化を行い、開発した手法を医学分野でのメタゲノム解析の実データに応用した。基礎研究面での科学的な成果として前者のソフトウェア開発が、実用的成果として後者の解析への実応用が挙げられる。

塩基配列のシークエンシング技術の発達は、大量の塩基配列データをもたらしたが、判明した塩基配列のみから直接に得られる生物学的知見は少ない。塩基配列のどの部分がどのような生物学的機能を有するか（どのような働きをするか）を調べることで、その塩基配列、あるいは、その塩基配列を有する生物についての理解が深まる。新規に判明した塩基配列に対して、生物学的な実験によって調査することは多大な時間と資金を要するため、計算機による推定手法がよく用いられている。本研究開発項目では、塩基配列の機能推定に最も広く用いられている手法の１つである、配列相同性検索（類似配列検索）の検索アルゴリズムの開発と並列化を行った。

配列相同性検索は、クエリとなる配列の類似配列を参照配列セットから検索する。進化的類縁関係にある配列同士では類似性が高いほど機能が保存されていることが期待される。このため、参照配列セットとして、機能既知の配列を用いることで、クエリ配列の機能を推定することが可能となる。

本研究開発項目では、あらかじめ参照配列セットに対して索引を作成することで、配列相同性検索を高速化した。配列間の類似性はSmith-Waterman アルゴリズムを用いた配列アラインメントによって、O(MN)の時間計算量（M：クエリ配列長、N：参照配列長）で計算できるが、参照配列セットのすべてに適用すると多大な時間を要してしまう。索引を用いてクエリ配列の部分文字列と類似する領域を発見することで、配列アラインメントの回数を減らし、検索を高速化できる。研究開発期間中に段階的に検索アルゴリズムの改良を行ったが、いずれも索引部分が重要な役割を果たし、それぞれ、ハッシュ表(Suzuki et al. *PLoS ONE*, 7(5), e36060, 2012)、接尾辞配列(Suzuki et al. *PLoS ONE*, 9(8), e103833, 2014)、ハッシュ表と参照配列セット内でのクラスタリング(Suzuki et al. *Bioinformatics*, 31(8), 1183-1190, 2015)を索引に利用することで高速化を実現した。ハッシュ表と参照配列セット内でのクラスタリングを索引に利用する手法（GHOSTZ）では、クラスタ代表との距離が既知ならばメンバとの距離の上界が三角不等式から求められることを利用して（図50)、ハッシュ表のサイズと部分文字列の比較回数を削減することで、配列相同性検索として広く

用いられている BLASTX に対して 261 倍の検索速度を実現し、高速な手法として知られる BLAT や RAPSearch と比較して高速であることが確認された。(表 3)。

さらに、開発した検索アルゴリズムを Message Passing Interface を利用してノード間で配列データを分配して並列に検索を行い、「京」の大規模並列環境を最大限に活用することで、高速化を実現するプログラム GHOST-MP を開発した。このプログラムは、「京」の全計算ノード (82,944 ノード) の使用に対しても優れたウィークスケーリングを示した (図 51)。

また、本研究開発項目で開発したプログラムをメタゲノム解析の実データに応用した。メタゲノム解析とは、環境中の微生物群由来のゲノムを分離・培養を経ずに集団のままでシークエンシング・解析する手法のことであり、環境中に存在する新規微生物・遺伝子の発見や、環境と微生物、また、その遺伝子との関係性の調査などを行うものである。

ヒトの様々な部位 (口腔、腸内、皮膚など) における微生物についての研究プロジェクトである Human Microbiome Project (HMP、http://hmpdacc.org/) が公開する口腔内細菌叢のメタゲノムデータに適用し、GHOST-MP が大規模プロジェクトの産出するデータ解析に利用可能であることを示した。この解析において、口腔と口腔前提、歯垢では、サンプル中の遺伝子組成が明らかに異なること、組成が大きく異なる遺伝子群 (Bacterial chemotaxis や Flagellar assembly、Lipopolysaccharide biosynthesis などの機能に関わる遺伝子群) を明らかにした。このように GHOST-MP を評価した。

口腔内細菌叢のメタゲノム解析については、東京工業大学と東京歯科大学石原和幸教授の共同研究として歯周炎病巣細菌叢のメタゲノム解析を行っている。両大学の倫理審査を経て、健常者と歯周炎患者の歯肉縁から得たサンプルのシークエンシングを行い、そのデータ解析に GHOST-MP を適用した。現在、その結果の解析が進行している段階である。

表 3 GHOSTZ の検索速度

	Computation time (sec.)	Acceleration ratio
GHOSTZ	460.8	261.3
RAPSearch	1285.5	93.7
BLAT	2514.9	47.9
BLASTX	120395.2	1.0

SRA Accession Number: SRR407548 から無作為に抽出した 10,000 配列をクエリに、KEGG GENES のアミノ酸配列をデータベース配列として検索を行った結果。Acceleration ratio は BLASTX の結果を基準とした。(表の出典、Suzuki et al. Bioinformatics (2015) 31 (8): 1183-1190 Table 2.)

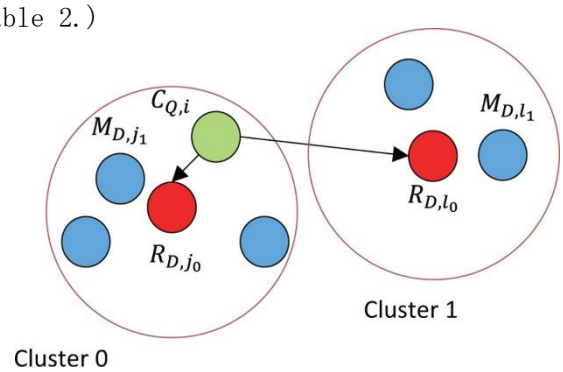


図 50　GHOSTZ のアラインメント候補探索

それぞれ、緑：クエリの部分文字列、赤：クラスタ代表である部分文字列、青：クラスタメンバである部分文字列を表す。クラスタ代表から一定距離内にある部分文字列がクラスタメンバとしてクラスタリングされる。三角不等式を利用することで、クラスタメンバと直接比較することなく類似部分文字列の判定が可能になる（偽陽性を含むが、偽陰性はない）。図では、クエリと２つのクラスタ代表、Cluster 0 と Cluster 1 の部分文字列を比較することで、Cluster 0 のメンバをアラインメント候補部位として列挙できる。（図の出典、Suzuki et al. Bioinformatics (2015) 31 (8): 1183-1190 Fig. 2.）

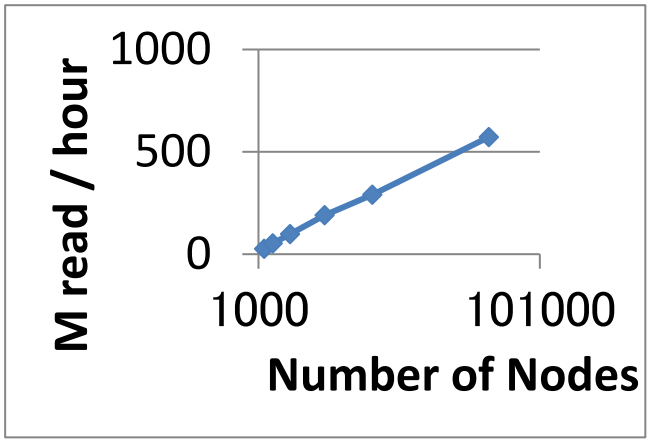


図 51　GHOST-MP の並列性能（ウィークスケーリング）

ヒトロ腔内細菌叢の次世代シークエンサのリード配列をクエリとして、KEGG genes 配列データベースに対し検索を行った。縦軸は時間当たりの処理リード数、横軸は使用ノード数を示す。

もはや、腸内細菌叢を調べるメタゲノム解析は世界中どこでもやっている。しかし、メタゲノムを解析し、その比較を行っても dysbiosis（細菌種の構成異常）が見えないという声が多い。秋山、宮野、植松智教授の連携で開発した高速・大規模に世界最深度（遠いホモロジー検索）のアミノ酸配列探索法を組み込み「京」に実装したメタゲノム解析パイプラインは、腸内細菌などのメタゲノムデータ解析をアミノ酸配列に翻訳されたレベルで（即ち、機能に関する解析が可能）10 分以内に実現し、腸管粘膜免疫の研究に革新をもたらしつつある。

東京大学医科学研究所では、コレラに対する交差抗原を誘導出来る常在菌の同定が急速進展し、メタゲノムを基盤とした革新的ワクチン開発の着想を得た。「お腹の強い人」がいることは知られている。旅行者下痢、コレラ、０１５７など、かかる人、かからない人がいる。腸内細菌が違うのではという仮説があるが、メタゲノム解析パイプラインによるメタゲノムデータ解析によって、腸内常在菌の中に通常では激しい下痢を起こす“病原細菌”を持っている健常人を発見した。コレラの流行地域への旅行や感染経験のない人だ。そこで免疫が病原細菌を抑え込み常在菌化させる機構を調べることを開始した。これにより、常在菌化を可能とする免疫応答をワクチンで再現する研究が始まっている。まさに「京」でメタゲノムを基盤とした革新的で安全なワクチンデザインが生れようとしている。