

第8章

ライフサイエンスへの計算科学活用

《HPCI計算生命科学推進プログラム》

前章で詳述したように、2006（平成18）年からスタートした第3期科学技術基本計画において、次世代スーパーコンピューティング技術は、国家的な大規模プロジェクトとして本計画期間中に集中的に投資すべき基幹技術であると位置付けられた。それを受けて文部科学省では、同年度から「最先端・高性能汎用スーパーコンピュータの開発利用」プロジェクトを開始し、2012年の完成を目指した次世代スーパーコンピュータの開発、および次世代スーパーコンピュータを最大限に活用するためのソフトウェアの開発を進めることになった。それが「京」スーパーコンピュータと、「京」のためのアプリケーション開発である。後者のソフトウェアはグランドチャレンジ・アプリケーションとよばれ、ナノテクノロジーとライフサイエンスの2分野における研究開発であった。

理化学研究所はライフサイエンス分野の中核拠点として、2006年10月より「次世代生命体統合シミュレーションソフトウェアの研究開発」プログラム（具体的な内容等は第1節に記載）を開始した。その目的は、次世代スーパーコンピュータの性能を十分に引き出して革新的なシミュレーションを実現し、ライフサイエンス分野における計算科学という新たな方法論を確立することであった。

2008年には、文部科学省に「次世代スーパーコンピュータ戦略委員会」が設置され、次世代スーパーコンピュータの計算資源を必要とし、かつ、社会的・学術的に大きなブレイクスルーが期待できる戦略分野が決定された。この戦略委員会の役割は、「次世代スーパーコンピュータプロジェクト」の目標達成のために研究開発を着実に進めるとともに、次世代スパコンの完成後をにらみ、その利活用の在り方を検討することであった。こうして「HPCI戦略プログラム」5分野が決定された。

「HPCI戦略プログラム」（2011-2015年度）

「京」を中心に、大学などのスパコンをネットワークでつないだHPCI（ハイパフォーマンス・コンピューティング・インフラ：高性能の計算環境）を最大限に活用して、世界最高水準の研究成果を創出するため、下記5分野を重点的に研究開発する文部科学省のプログラムである。

- 分野1 予測する生命科学・医療および創薬基盤
- 分野2 新物質・エネルギー創成
- 分野3 防災・減災に資する地球変動予測
- 分野4 次世代ものづくり
- 分野5 物質と宇宙の起源と構造

こうした動向を踏まえ、理研は、「ライフサイエンス分野における計算科学研

究」分野にさらなるブレークスルーをもたらすべく、2009年6月、「計算生命科学センター設立準備室」を設置し、文部科学省の「HPCI戦略プログラム」に対する提案書を提出した。その後、実施可能性調査のヒアリング審査会を経て、2010年7月に理研が戦略分野1の戦略機関に選定されたのである。

その後の準備研究と並行して、戦略分野1の中村春木分野マネージャー、柳田敏雄統括責任者を中心に戦略プログラム実施体制が検討され、2011年4月、理研組織として「HPCI計算生命科学推進プログラム」が設立されたのであった（具体的な内容・進展等については第2節に記載）。

以上が「HPCI計算生命科学推進プログラム」が設立されるまでの背景である。

第1節 グランドチャレンジ・アプリケーション (2006-2012年度)

生命体統合シミュレーションとは

「京」のためのアプリ開発の第1弾である「グランドチャレンジ」の一つとして選ばれた「次世代生命体統合シミュレーションソフトウェアの研究開発」とは、いかなる試みだったのか。

2006（平成18）年ごろ、天文や素粒子、材料科学、工学的な設計、気象予測など、身近な幅広い分野ですでにスーパーコンピュータを活用した研究開発が進展していた。それとは対照的に、ライフサイエンス分野における計算科学的なアプローチは限定的で、タンパク質の挙動や限られた生体内の反応などに使われていたに過ぎなかった。そうした中、理研は、今後はライフサイエンス分野でもスーパーコンピュータによる研究が重要となるため、積極的に取り組んでいくべきだ、とする提言をまとめた。

文部科学省では、スーパーコンピュータの開発とともに、それを最大限に活用するためのソフトウェア開発（グランドチャレンジ・アプリケーションと命名）を進めることになり、ライフサイエンス分野における課題として、「次世代生命体統合シミュレーションソフトウェアの研究開発」の公募が行われた。理研は15機関が集結した研究開発体制を提案し、2006年6月に文部科学省の審査を経て、10月に発足した（図1）。

発足時点では、ライフサイエンス分野におけるシミュレーションソフトウェアの開発研究に対して、うまく研究開発を進めることができるのかどうか、疑問視する向きも多かった。その理由は、生命体といえば分子レベルから全身スケール、さらにはバイオインフォマティクスまで、扱う領域が非常に広く、まさに階層が複数にまたがる（マルチスケールである）上に、多細胞レベルの現象では基本となる確固たる理論が十分に整備されていない状況だったからである。例えば一つの細胞における全分子レベルのシミュレーションを行うためには、分子の数が巨大である上、細胞質は単なる濃度の高い溶液としては取り扱えず、計測もできていない極めて複雑な構造を持っている。非常に難しいテーマであり、これを正確

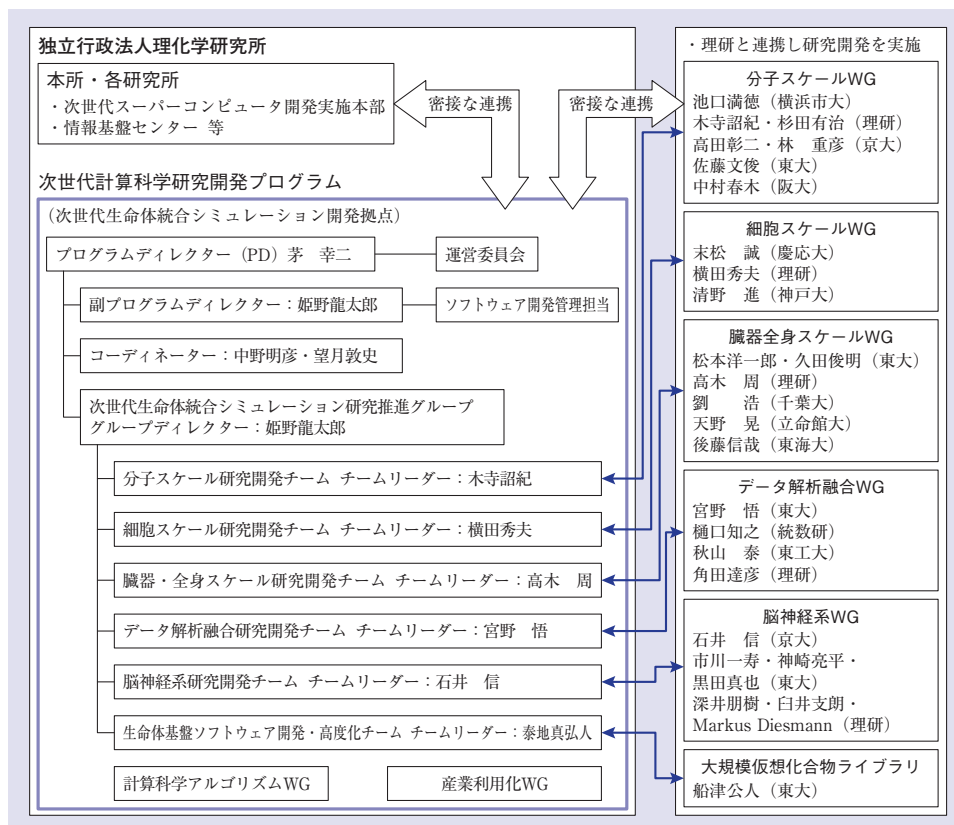


図1 研究開発体制 (2010年12月時点)

に再現し、機能を計算で予測できるのかどうか、疑問を感じる人も多かった。

しかしながら、その後、さまざまなタンパク質の機能が分子レベルで理解・再現できるようになり、細胞レベルでもシミュレーション研究が着実に進展し、研究開発の発展は目覚ましいものがあった。なお脳研究の分野でも、「京」を使うことによって、その機能に迫る研究が進展している。

また、当初は、次世代シーケンサーで得られた巨大なデータや、疾患に関連した膨大な遺伝子データを、ペタフロップス・クラスのスーパーコンピュータで扱うことができるのかどうか、疑問視する向きもあった。しかし、より大規模なゲノム解析が高速に実行できる環境が整い、膨大な量のデータを処理することが求められる時代となって、高速で大量の情報処理ができるスーパーコンピュータを用いた研究が有用であることが実証されてきた。

ライフサイエンス分野のシミュレーション科学が大きく成長する中で、この分野に興味を抱いた若手研究者が数多く参加し、垣根を超えた幅広い分野の連携が進んだことも、グランドチャレンジのもう一つの大きな成果であった。新しいライフサイエンス分野を作り上げてきたと言ってもよい。グランドチャレンジは2013年3月まで7年間の長期プログラムであったが、その成果は、2011年より本格的にスタートした「HPCI計算生命科学推進プログラム」に直接活用され、さらなる成果が創出されてきた。また、開発ソフトウェアについては、すでにホームページで公開が行われており、HPCI戦略プログラム分野1と理研の情報基盤センターが協力して、引き続き利用者への提供、利用支援を進めてきた。

実際の取り組み

次世代生命体統合シミュレーションソフトウェアの研究開発にあたっては、大きく二つの方向から取り組みを進めた。一つが「第一原理的アプローチ」で、分子から全身まで、生体内で起こるさまざまな現象を、基礎原理に基づいて統合的に理解することを目指して、シミュレーションソフトウェアの研究開発を進めたものである。もう一つは、大量の実験データから未知の法則に迫る「実験データの解析的アプローチ」であった。それぞれの研究開発を統合的に進めたことは言うまでもない。

ライフサイエンス分野への貢献を考え、目標を次の二つに定めた。

- 1) 次世代スーパーコンピュータの完成時での利用を目指したアプリケーション・ソフトウェアの開発
- 2) 将来のライフサイエンス分野の基盤構築に向けた長期的なグランドチャレンジ

特に後者は、実験とコンピュータシミュレーションが一体となって初めて可能となる、新しい知見の獲得に向けた取り組みであった。

階層間をつなぐ

分子・細胞・臓器全身・データ解析・脳神経系などの分野におけるシミュレーション技術は、このプロジェクト以前には、それぞれ別の研究領域だった。しかし、次世代スーパーコンピュータ「京」のソフトウェア開発という旗印の下、これらの研究領域を統合して一つのプロジェクトとして実施できたことは、複雑な現象が多階層で関連する生命体の本質に迫る上で、極めて意義のある試みであった。世界的に見ても、これだけの領域が一つにまとまったプロジェクトは過去に例がなく、「京」によって開発された計算技術で世界をリードしていく上でも、重要なプロジェクトとなった。2012年12月の国際シンポジウムでも、海外の研究者から高く評価され、特に米国立衛生研究所（NIH）のペン（Grace Peng）博士からは「成功物語を学びたい」とNIHとの連携を申し込まれるほどであった。

階層間の連携は困難であったが、それでも、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（タンパク質・心筋細胞・心臓）、多剤排出トランスポーターの機構解明（原子・分子・粗視化）、血栓シミュレーション（細胞・臓器全身）などにおいて、階層間、チーム間の連携を進めることができたのは意義ある成果であった。

それまでの生命科学分野では、現象の複雑さゆえに、基礎方程式から出発して計算で到達できる範囲では、生命現象に十分に迫ることはできなかった。また、モデル化が十分に行われていない細胞レベルでの現象では、スーパーコンピュータの活用はなされてこなかった。しかし、次世代スーパーコンピュータの開発によって、計算技術でこれらの問題が解決されるのではないかという期待が高まり、本プロジェクトで、生命科学分野の計算科学技術の研究開発を進めた結果、他の

科学技術分野を凌駕する速度で、必要なソフトウェア群を開発することができた。これは、ソフトウェアの高度化チームが各研究者と密接な連携を取って初めてなし得たことといえる。このようなソフトウェア群は、スーパーコンピュータを生命科学分野で活用する上で重要な資産として、活用され続けている。

世界トップクラスの成果

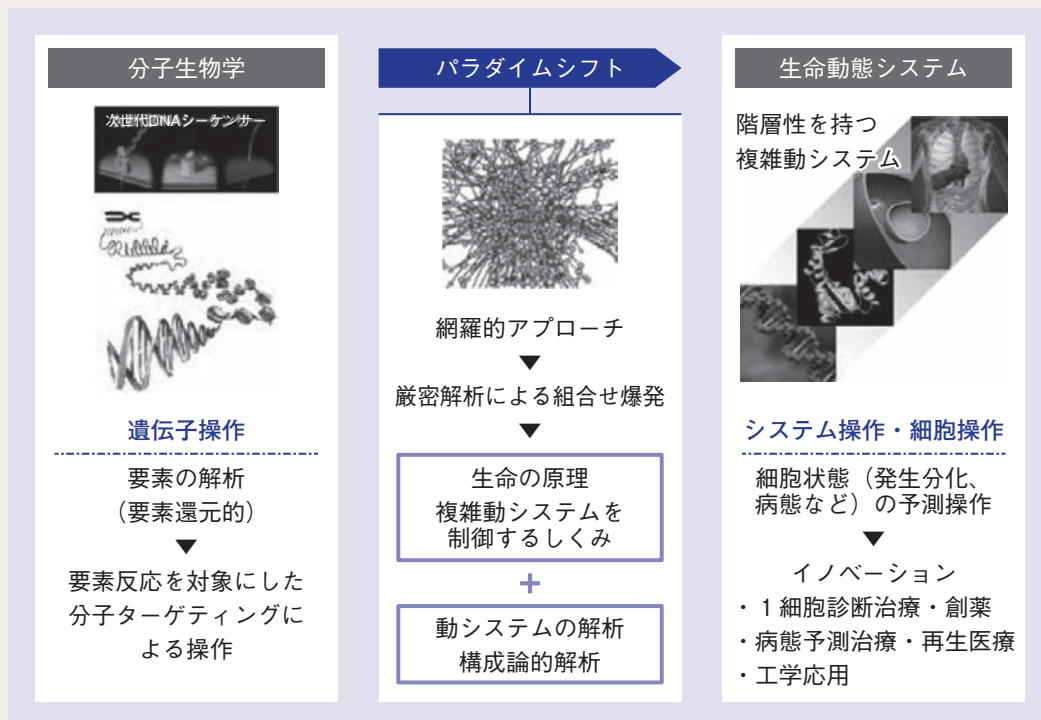
「京」の性能をフルに発揮するソフトウェアの開発では、全チームで合計31本の開発に取り組み、そのうちの27本が「京」上で稼働することを確認した。そ

Column

計算生命科学とは

20世紀後半の生命科学は、「生命現象を分子のレベルまで掘り下げて解析すれば、その機能をあますところなく理解できる」という仮説の下に進められてきた。特にDNAシーケンサー（塩基配列解読装置）、質量分析器、一分子計測などの計測技術の進歩は目覚ましく、生命体の分子レベルでの理解は急速に進んだ。しかし、生命体は分子レベルから個体レベルに至るまでさまざまな階層性を示しており、しかも個々の階層は質においても動的特徴においても異なっている。このような複雑で動的な多階層システムとしての生命体を理解するために、今、新たな概念的枠組み（パラダイム）が必要となっている。

計算生命科学は、新しい計測技術から生み出される多様なビッグデータを効率的に解析し、動的な多階層システムを整合的につなぎ、生命システムを総体として理解することを目指している。スーパーコンピュータを積極的に活用してこの目的が達成できれば、生命科学において初めて、予測可能性と制御可能性が獲得できることになる。



生命科学のパラダイムシフト

の27本中16本のソフトウェアが1万ノード以上の高並列化を果たしており、生命科学分野のソフトウェアとしては今までに類を見ないほど速く、ハイパフォーマンス・コンピューティング（HPC）に対応できたと言ってよい。

特に構造流体連成ソフトウェアZZ-EFSIで4.5PFLOPSを達成し、これは世界最速の構造流体連成計算となった。構造流体連成計算というのは、流体の流れによって構造物が変形し、その変形によって流れが影響を受ける過程を、正確に計算していくものである。生物では当たり前の現象であるが、柔らかな生物のダイナミックな動きは、新たな計算手法を考案することにより初めて計算が可能となった。その他、分子動力学ソフトウェアcppmdで3.8PFLOPS、心臓シミュレータUT-Heartで2.9PFLOPSの実行性能を達成、いずれも世界トップクラスの性能が得られた。ソフトウェア開発に関しては、当初想定していた目標よりはるかに高い成果が得られたと評価された。

また、生体现象の深い理解と新たな発見を目指し、従来の計算機性能では不可能であった規模で計算することで、次のような医薬品や医療機器、診断や手術方法の開発につながる成果を達成することができた。

分子スケール研究開発チームでは、共通テーマとして多剤排出トランスポーターを設定し、量子化学計算、分子動力学計算、粗視化分子動力学計算を連携させて機構の解明を進めることができた。

臓器全身スケール研究開発チームでは、マルチスケール・マルチフィジックス心臓シミュレータにおいて、心筋細胞内のタンパク質の挙動から心筋細胞、そして心臓全体まで、3階層にわたる階層統合に世界で初めて成功した。そして具体例として、肥大型心筋症の病態を分子レベルの変異から捉えて解析することに成功した。さらに、「京」に至る途中段階で開発した全身血管網シミュレータの解析結果を活用して、動脈硬化を調べる新しいタイプの電子血圧計が商品化された。国産初の超音波治療器の開発に向けて、乳がん治療器設計のシミュレーションも実施された。

データ解析融合研究開発チームでは、大規模な生命データに対応した汎用性のある独創的なデータ解析手法が考案された。そして、それを実装した高並列ソフトウェアにより、がんの理解と薬との関係に新たな知見を得ることに成功した。

脳神経系研究開発チームは、他のチームに2年遅れてスタートし、必ずしも十分な期間があったわけではないが、興味深い成果をいくつも達成することができた。例えば、神経形態形成シミュレーションの成功、視覚系シミュレーションによるロボットヘッドのオンライン制御、理研の新しいスーパーコンピュータ（RICC）とインターネット接続の成功がある。また、昆虫嗅覚系シミュレーションでは、詳細な細胞構造の抽出、標準脳作成、標準脳マッピングなどのソフトウェアにより、前運動中枢の5領域モデルに基づく結合の推定を行い、嗅覚系中枢回路の高精細シミュレーションを達成することができた。




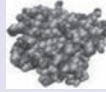

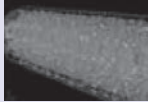




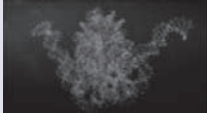
細胞スケール研究開発チームでは、非常に難しい階層のテーマに取り組み、慶應義塾大学医学部等と共同で研究を実施することで、肝臓シミュレーションなどで有用な成果を得ることができた。

Column

スパコンで生命現象を理解する

私たちの体は、臓器、組織、細胞、分子という階層から成る複雑な体系とみなすことができる。生命現象は血液系や呼吸器系や神経系なども含め、このシステムが一体となって働くことで起こる。

HPCI計算生命科学推進プログラムでは、このシステムをスパコンの中で再現し、機能させることによって研究を進めている。これによって生命現象を理解し、予測するとともに、新しい薬や病気のよりよい治療法につなげることを目標としている。

1665年	フック：「ミクログラフィア」（生物は細胞からなる）		フックが描いたコロクの細胞 (実際には細胞壁だった)
1687年	ニュートン：運動方程式		
1859年	ダーウィン：「種の起源」（生物は進化する）		
1865年	メンデル：遺伝の法則		メンデルが調べたエンドウマメの形質の例
1953年	ワトソン、クリック： DNAの二重らせんモデル		DNAの二重らせん構造 (PDB ID：3bse)
1958年	ケンドル：タンパク質のX線結晶構造解析		ミオグロビンの構造 (PDB ID：1mbn)
1977年	サンガー：DNA塩基配列決定法		
2005年	次世代シーケンサー		
2006年	「次世代生命体統合シミュレーションソフトウェアの開発」 分子から全身まで体内で起こるさまざまな現象を統合的に理解するためのシミュレーションソフトウェアの研究開発		
2011年	HPCI戦略プログラム分野1 「予測する生命科学・医療および創薬基盤」 スーパーコンピュータ「京」を中心としたHPCIを最大限に活用することによって、戦略的に生命科学の研究分野において画期的な成果を生み出す	 タンパク質の動きのシミュレーション	 血流のシミュレーション
		 がん再発の鍵となる遺伝子を明らかに	 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
2012年	スーパーコンピュータ「京」		
2016年	フラッグシップ2020プロジェクト ポスト「京」重点課題1 生体分子システムの機能制御による革新的創薬基盤の構築	 HPCI戦略プログラム分野1より	 HPCI戦略プログラム分野1より

生命科学の新たな時代を切り拓いた計算生命科学

第2節 HPCI計算生命科学推進プログラム (2011-2015年度)

研究課題は4テーマ

理研の「HPCI計算生命科学推進プログラム」は、2011（平成23）年4月に推進組織として発足した。文部科学省のHPCI戦略プログラムのうち、特に、社会的・学術的に大きなブレークスルーが期待されている「戦略分野1 予測する生命科学・医療および創薬基盤」において、理研は研究活動等をけん引する機関（戦略機関）と規定され、大学や研究機関なども参加して、共同で先端的なシ

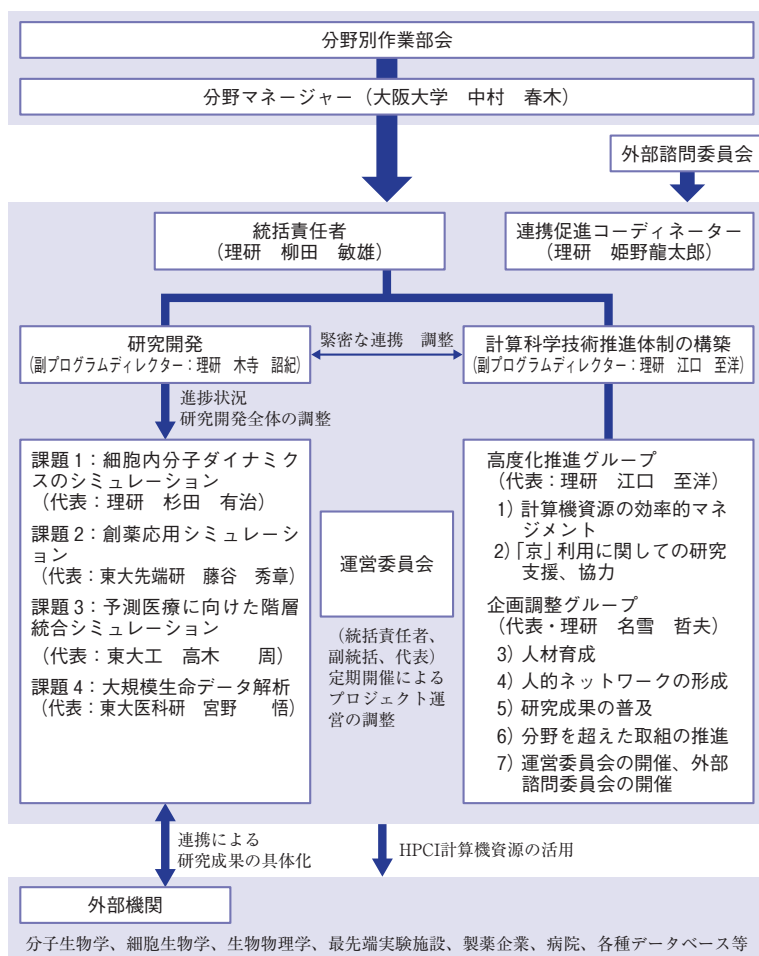


図2 実施体制（2015年12月時点）

ミュレーション研究や人材育成プログラムなどを推進してきた（図2）。参加機関は、東京大学、日本原子力研究開発機構（当時）、沖縄科学技術大学院大学、東京工業大学、産業技術総合研究所、九州大学、岡山大学、大阪大学、横浜市立大学、京都大学、東海大学であった。

研究課題は、①細胞内分子ダイナミクスのシミュレーション、②創薬応用シミュレーション、③予測医療に向けた階層統合シミュレーション、④大規模生命データ解析の四つを設定した。いずれの研究課題でも、「京」で動作するアプリケーションの開発を進め、ゲノム・タンパク質から細胞・臓器・全身にわたる生命現象を統合的に理解することを通して、疾病メカニズムの解明、その予測、創薬研究などを進めてきた。それらの成果は医療や創薬プロセスに寄与することが期待されている。

また、計算科学技術分野の発展や人材育成を目的に、「京」と互換性の高い独自の計算機システムを運用して、大学、公的機関、民間の計算生命科学者や技術者がHPCIを積極的に活用できるよう支援してきた。さらに、生命科学分野における高度な計算機環境の重要性を理解してもらうため、高校生などを対象にアウトリーチ活動も積極的に進めてきた。

研究の目標

「京」コンピュータを用いた大規模な計算により生命科学分野での予測可能性

を拡大させること、そして日本を支える基盤科学を創出し、医療開発・創薬支援などの健康医療開発を劇的に加速・効率化すること、それがHPCI戦略プログラムの分野1の目的であった。

2016年現在、生命科学分野において、計測技術の進展は目覚ましく発展しており、生命科学で得られる実験結果の情報量は、増大の一途をたどっている。これらの計測情報を基盤として、生命システムを物理・化学的な系として定量的に記述し、その振る舞いを計算機で予測することも、今や可能となりつつある。したがって、このプロジェクトでは、「より精密で説明能力の高い予測が可能な生命のモデル化」を目標に研究を進めた。ペタフロップス級のスーパーコンピュータの演算能力を使って、初めて可能となるような大規模シミュレーションと、最先端の計測器から産出される大量データの計算科学的解析とを通して、生命システムの定量的な記述、それによる生命現象の理解や予測、さらに、その予測を創薬や医療などに活かすことを目標として掲げた。

具体的な課題としては、創薬と予測医療という二つの重要な応用と、生命科学で重要性が拡大し続けている計測データ（細胞内の動態、次世代シーケンサー）とを意識し、四つの研究課題を設置したのである。以下、個々の研究課題について少し詳しく記していきたい。

①細胞内分子ダイナミクスのシミュレーション

ここでは、マルチスケール分子ダイナミクス・シミュレーションと一分子粒度シミュレーションを高度化し、細胞内環境下における生体分子の挙動をシミュレーションすることによって、生体膜を介した物質輸送、タンパク質/DNA相互作用、シグナル伝達機構を解明し、細胞機能の理解や薬剤設計に貢献することを目指した（図3）。

実際、「京」を活用するシミュレーションソフトウェアを複数開発することで計算基盤を構築し、アカデミアのみならず、創薬など産業界へも広がる大きな貢献をすることができた。さらに、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を利用して、創薬応用計算が進められている。こうしたソフトウェア開発や「京」なしには不可能な大規模計算を通して、HPCが生命科学において重要な役割を果たすことを示すことができた。

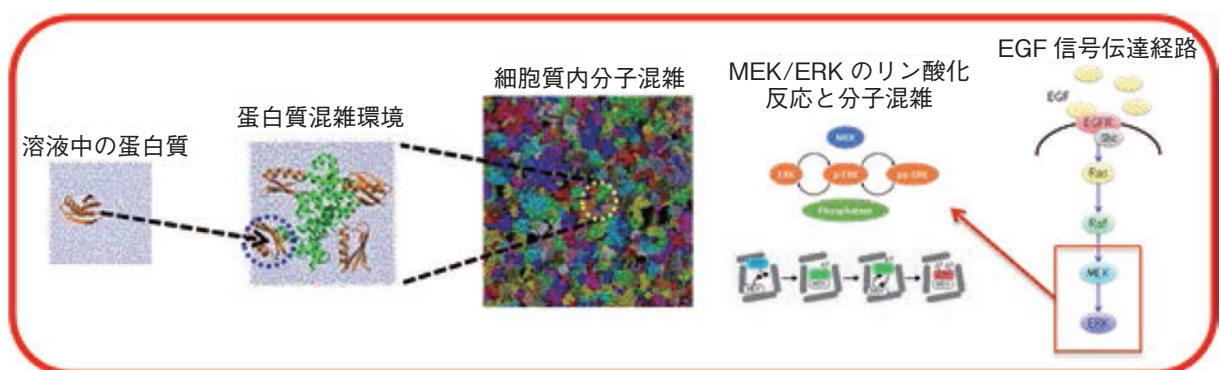


図3 細胞内環境を考慮した信号伝達経路のモデリング

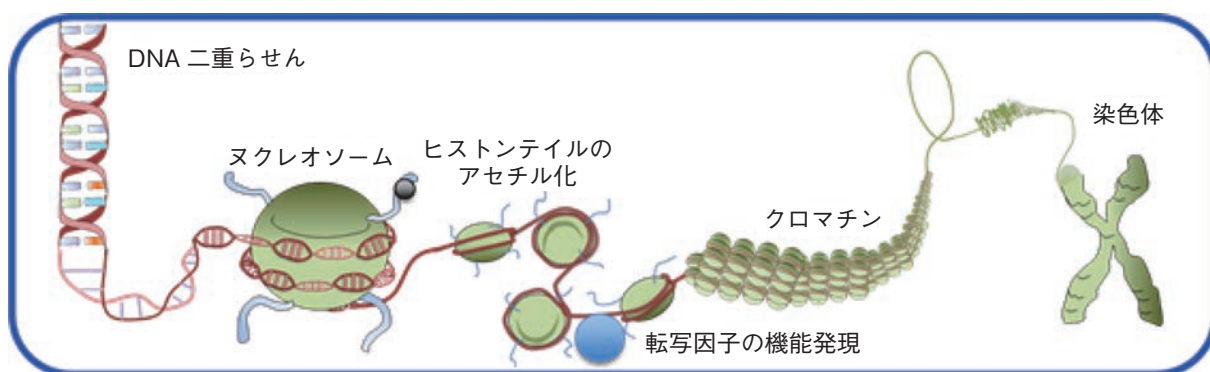


図4 ナクレオソーム、クロマチンの機能発現機構

主な研究成果としては、大規模並列分子動力学シミュレーションを実施し、分子混雑やヌクレオソーム中のDNAなど、細胞内環境を取り込んだ生体分子のモデリングとシミュレーションに成功した（図4）。

バクテリア細胞質内の分子混雑系の全原子分子動力学は、1億個もの原子から成り立っており、生体分子のシミュレーションとして世界最大規模の計算である。ヌクレオソームがどのようにして分子を認識しているかを分子レベルで明らかにする自由エネルギー計算も、レプリカの並列計算によって初めて実現した超大規模計算であった。

一方で、粗視化分子モデルや一分子粒度モデルというマルチスケール・マルチフィジックスなシミュレーションを併用することで、規模的にも時間的にも、研究対象の振る舞いを拡大させることができ、その解明に成功した。さらに、実験情報とシミュレーションの密接な連携を進めるために、MD/SAXS法を開発した。これにより、X線溶液散乱データを再現するシミュレーション結果から、溶液中のヌクレオソームの構造を正確に決定できるようになった。

また、それぞれの課題について実験研究者と共同研究を行い、「京」を用いた

分子動力学計算の大規模化、マルチスケールシミュレーション、実験と計算の密接な連携を重ねることで大規模シミュレーション結果を検証し、生体分子シミュレーションの可能性を広げることができた。

②創薬応用シミュレーション

ここでは、分子動力学計算を用いた結合自由エネルギーシミュレーションを高度化し、薬剤候補化合物の設計方法の構築、および化合物とタンパク質の結合部位、結合強度の新たな同定法を提案した。これにより、スーパーコン

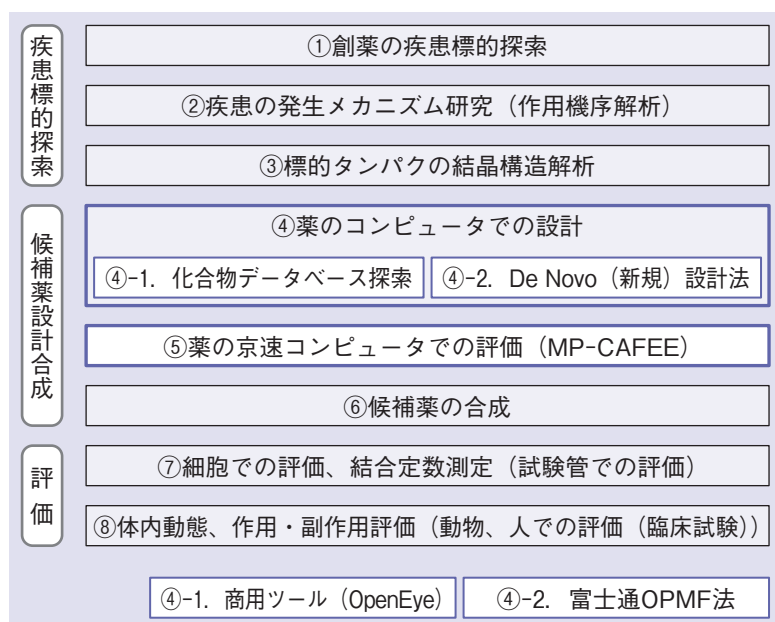


図5 創薬プロセス全体と計算創薬で取り組む部分の関係

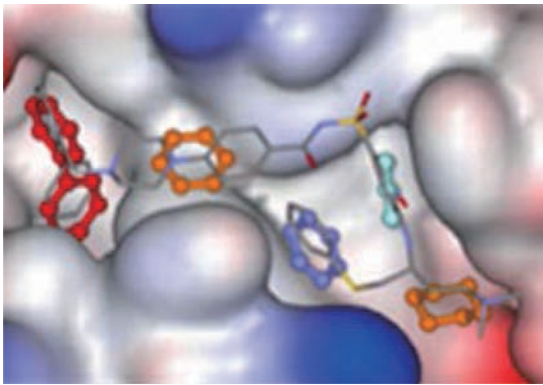
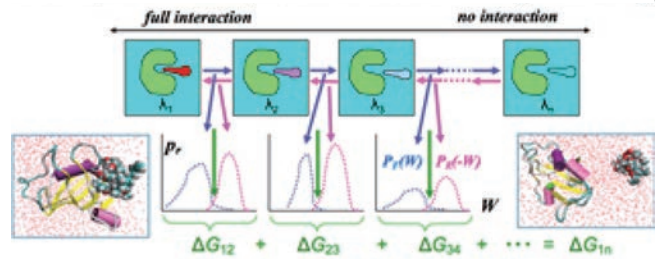


図6 フラグメント最適化 (OPMF法)

Massively Parallel Computation of Absolute binding Free Energy



$$\Delta G = -RT \ln \left\langle \exp\left(-\frac{W}{RT}\right) \right\rangle \quad \therefore K_d = \exp\left(\frac{\Delta G}{RT}\right)$$

図7 超並列結合自由エネルギー (MP-CAFEE) 計算法

コンピュータを用いたシミュレーションによる創薬設計手法を開発、創薬プロセスにつながる成果を上げた (図5)。

具体例を挙げると、新しい治療医薬品が待望されている標的タンパク質に十分な強度で結合する薬剤候補化合物をフラグメント最適化法で設計し、従来の計算法では5kcal/mol以上の誤差であった化合物と標的タンパク質の結合自由エネルギーを、MP-CAFEE法を用いて1kcal/mol以下の精度で求めることに成功した。これによって、新規の薬剤候補化合物のスーパーコンピュータによる設計と活性予測が可能となった。さらに実験グループと連携して、設計した化合物を実際に合成し、化学・細胞アッセイを行い、薬剤開発における「京」コンピュータの有用性を実証することができた。

得られた主な研究成果として、がん治療標的タンパク質に結合する薬剤候補化合物をフラグメント最適化法 (図6) で設計し、MP-CAFEE法で1kcal/molの精度で結合自由エネルギー (図7) を求めた後、ウェット実験と連携してがん治療標的キナーゼで薬理活性が高い化合物を発見することができた。製薬企業との共同研究により、この化合物に対する動物実験を2014年10月より実施している。また、バイオ医薬品では、マウス抗体のヒト型化と結合力を増強するためのアミノ酸置換を設計し、2014年末より東京大学先端科学技術研究センターで動物実験を進めている。このように二つの創薬標的の薬剤設計で、スーパーコンピュータによる分子動力学計算の有用性を示すことができた。また応用面だけではなく、抗原抗体反応で測定されるエントロピーとエンタルピー変化の起源を明らかにして、どのような要素が抗原と抗体の結合強度に影響しているかを明らかにした。

③予測医療に向けた階層統合シミュレーション

ここでは階層統合シミュレータを構築し、分子、細胞レベルから組織、器官、臓器の挙動のシミュレーションをすることによって、複雑な生命現象と幅広い疾患の理解を目指す研究を進めた。これにより、わずかな兆候から将来の病態を予測したり、負担の少ない治療法を検討したり、さらに薬効を評価したりすることなどへの道が開かれた。

心臓については、心筋細胞内のサルコメア (筋原繊維) レベルの分子運動から心臓全体の拍動までの3階層統合を達成した。そして、分子生物学の知見と巨視

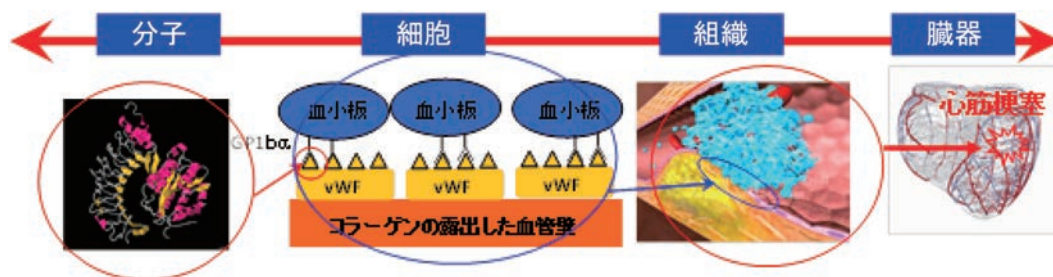


図8 心筋梗塞の階層統合モデリング

的な病態を結び付けた階層統合シミュレーションを実施し、疾患メカニズムの解明、治療法の検討などを行うための開発を行った(図8)。運動機能障害については、パーキンソン病のような脳神経系が関わるモデル構築が困難な問題に対して、神経細胞レベルからの物質輸送・シグナル伝達の影響を考慮した脳神経系シミュレータを開発し、それと筋骨格系シミュレータとを連成した全身統合シミュレーションを実施して、運動機能障害のメカニズムの解明と治療法の検討を進めた。最終目的は、個々の疾患の事例を介して、循環器系、脳神経系、筋骨格系の階層統合を実現することにある。短期的には臨床応用、長期的には生命そのものの理解につながるような階層統合シミュレータを作り上げるのが目標であり、そのための研究を実施した。

ここでの研究成果としては、特に顕著な成果を上げたマルチスケール・マルチフィジックス心臓シミュレータUT-Heartがある。サルコメアレベルから心筋細胞、心臓全体までの3階層統合シミュレーションに世界で初めて成功し、計算科学分野に大きなインパクトを与えることができた。得られた成果を基に作成された動画は、CG分野で権威のある国際会議SIGGRAPH(2015)において、最優秀可視化/シミュレーション賞に選ばれた。スミソニアン博物館の科学サイトや海外の科学系メディアでも高い評価を受けた。また、このシミュレータはすでに臨床データを用いた病態予測の段階に入っており、小児先天性心疾患の手術後の

予測が精度よくできるなどの成果が得られた。2016年現在、UT-Heartの開発者である久田俊明らは、研究成果をもとに(株)UT-Heart研究所を設立し、シミュレータの臨床応用に向けて、精力的に研究開発を行っている。

脳神経系シミュレータNESTでは、本課題の中で、「京」上で17億3000万個の神経細胞、10兆4000億個のシナプス結合を持つ世界最大の神経ネットワークシミュレーションに成功した。これは、マームセットクラスの小型霊長類の脳サイズに匹敵する計算ができることを意味し、ヒトの脳機能解明に向けた第一歩といえるものである。さらにそのNESTにより、大脳基底核モデルと視床-大脳基底核モデルを構築し、筋骨格系のシミュレータと統合し、全身

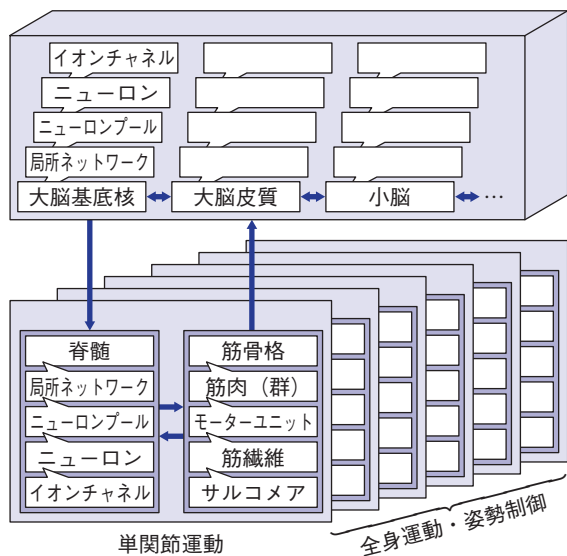


図9 パーキンソン病の再現に向けた脳神経系-筋骨格系階層統合モデリング

の脳神経-筋骨格系統合モデルを構築し、パーキンソン病の脳モデルが振戦や固縮などの病態を引き起こすかどうかシミュレーションを実施した（図9）。その結果、パーキンソン病のサルの実験で観測されている脳波の振動（15ヘルツ： β バンド）を再現することに成功した。そのシグナルが視床で約半分の周波数になり、大脳皮質、脊髄から筋線維へと伝わって、パーキンソン病特有の手の震え（振戦）につながるということが明らかになった。

④大規模生命データ解析

ここでは次世代シーケンサーから得られる大規模データを解析するための計算手法を開発し、未知の遺伝子が多いメタゲノム解析、がんの原因となる機能的突然変異の探索、脂肪細胞の機能分化のネットワーク解析などを進めた。これにより、個別化医療やゲノム情報の産業応用などに貢献することができた。

具体的には、「京」を中核とするHPCIに最適化した最先端・大規模シーケンズ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の解析研究を行うことにした。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した。

主な研究成果としては、600以上のさまざまながんの遺伝子発現データと100以上の薬剤に対する感受性・耐性データから、世界最大規模の遺伝子ネットワーク解析を実施し、世界最高精度の個別化抗がん剤投薬基盤を構築した（*PLoS One*, 2014等に発表）。また、426例という過去最大規模の成人T細胞白血病・リンパ腫（ATL）の症例を用いて、大規模オミックスデータ解析を行い、ATLのシステム異常の全体像を解明（図10）するとともに、新規治療薬剤の開発に向けた標的を発見した（*Nature Genetics*, 2015）。このように、がんという極めて

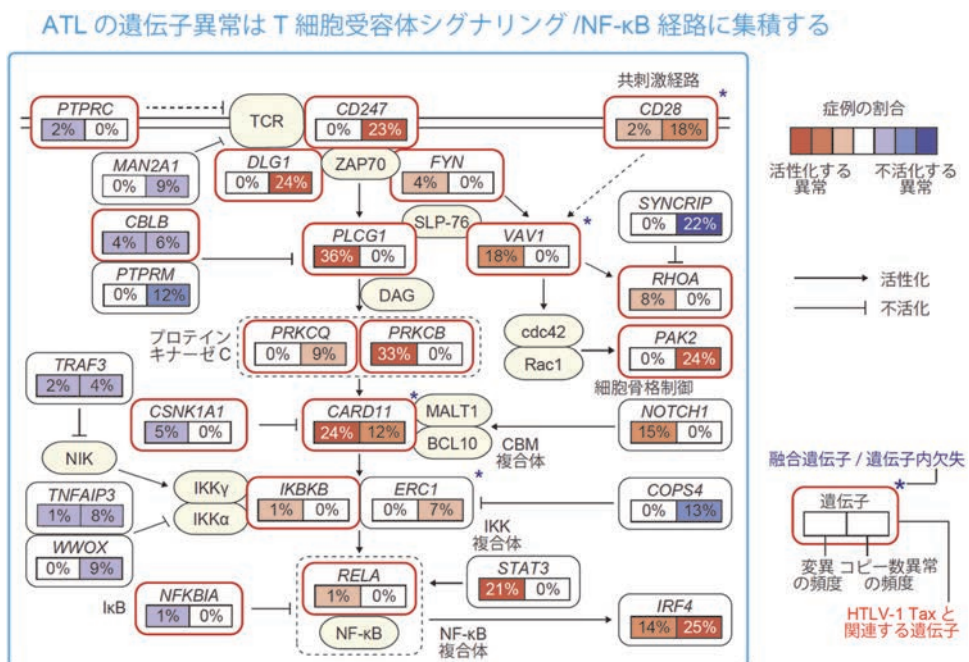


図10 成人T細胞白血病・リンパ腫（ATL）のシステム異常の全貌解明

複雑なシステム異常を背景とした病気に挑戦し、目標を超える成果を得た。

また、大規模遺伝子ネットワーク解析により、IL-1 β という炎症に関与する生理活性物質と熱産生のメカニズムが初めてつながり、白色脂肪細胞がアンチメタボ細胞へ変身する分子メカニズムの全容を解明した (*Cytokine*, 2016)。この成果は、生命プログラムの解明に対して、方法論も含めてブレークスルーを起こした。

さらに、33種類の主要ながん種を含む1万例を超えるがん試料のゲノム解析データを研究対象として、スーパーコンピュータを用いた大規模な遺伝子解析を実行した。それにより、がん細胞が免疫監視機構を回避する新たなメカニズムを解明することに成功した (*Nature*, 2016)。

加えて、ヒト糞便メタゲノム解析が10分以内に可能となり、免疫研究者との共同研究でコレラに対する交差抗原を誘導できる常在菌の同定が急速に進んだ。ここから、「京」でワクチンを開発するという可能性が生まれている。このように、「京」を用いた大規模生命データ解析は、医学分野にブレークスルーをもたらしている。

推進体制の構築

計算生命科学を推進するにあたっては支援部門もさまざまな活動を展開してきた。全体としては、多くの研究者が「京」コンピュータを中核とするHPCIを効果的に利用できるよう、研究開発担当者との連携を密にして、ソフトウェア環境および実行環境の整備を進めてきた。他の戦略分野、計算科学研究機構およびHPCIコンソーシアムと連携しつつ、HPCIの利用普及、情報発信、理解増進、人材育成なども進めた。

具体的な活動としては、まず第1に、計算資源の効率的な管理がある。画期的な成果が創出されるよう、研究開発課題の進捗評価や見直しによる「京」の計算資源配分を実施し、HPCI環境を効果的に利用するための管理システムを構築した。さらに、「京」互換の計算機システムを整備して、大学や製薬関連企業の研究者にもHPCI環境を体験してもらうこととした。

第2は、高度化の推進とユーザー支援である。具体的には、「京」ヘルプデスクと連携し、各分野の研究者が開発または使用するソフトウェアについて、「京」の性能を引き出して高い処理性能を発揮するよう高度化の支援を行った。また、高度なHPCI環境を使いこなせる人材育成の一環として、SCLS（「京」互換機）計算機システムを使用した開発ソフトウェアの講習会を開催した。研究機関、教育機関、民間企業を対象に、13回開催し、合計105名の参加者を得た。

第3が人材育成で、拠点となる大学等研究機関を核に、学生、大学院生、製薬企業研究者などの社会人を対象に、継続的に人材育成を実施した。具体的には、大阪大学大学院基礎工学研究科では数年にわたって教育プログラムを実施した。また全国の大学で、生命科学を専攻する学生の講義に協力してきた。さらに産業技術総合研究所において創薬基盤研究部門HPCI人材養成プログラムをスタートさせ、セミナー・講習会を開催した。加えて、神戸大学計算科学教育センターと

連携した遠隔インタラクティブ講義「計算生命科学の基礎Ⅰ」、「計算生命科学の基礎Ⅱ」（各連続15回）は、多くの研究者の協力を得て、標準的な計算生命科学のカリキュラムとして確立することができた。この講義には社会人を含む幅広い層が受講し、「現場での実践に役立った」、「最新の生命ビッグデータについての状況がよく分かった」などの好反響を受けた。

若い年代に向けては、高等学校、大学、大学院において計算生命科学への理解と興味と関心を促すために、特別講義や出張授業を実施した。

計算生命科学を推進するために重ねてきた試みの第4は、人的ネットワークの形成である。分子生物学、細胞生物学、生物物理学、薬学、医用工学、バイオインフォマティクスなど生命科学のコミュニティにHPCI環境での計算生命科学を理解してもらうため、セミナーやシンポジウムで講演やポスター発表を行った。場所は日本分子生物学会、日本生物物理学会、日本蛋白質科学会、日本バイオインフォマティクス学会、情報計算化学生物学会、そして全国の大学に及んでいる。地域ごとに計算生命科学のシンポジウム、セミナー等も開催した。

ほかに、2012年度に導入したSCLS（「京」互換機）計算機システムの利用公募を、全国の大学、企業の研究者、生命科学研究者、技術者を対象に年3回の割合で実施し、利用者に対しては講習会を開いて計算機利用の情報提供、意見交換などの手厚い支援を行って、HPCIの活用を促進させることができた。ここから一部の利用者が「京」の利用を始めるという事例につながっていく。

第5は研究成果の普及である。国内外の研究者、産業界、国民に対するウェブサイト、研究者向けと一般向けの紹介冊子、ニューズレターを通じて、研究成果の紹介、および情報の発信を行った。その他、“グランドチャレンジ”で開発された「次世代生命体統合シミュレーションソフトウェア」の一部をSCLS（「京」互換機）計算機システムに移植したソフトウェアと、戦略分野1で進めているソフトウェアをウェブサイトで公開し、技術面での利用支援を行い、生命科学の研究者・技術者への活用を図った。

社会的な理解を促進するため、シンポジウム、セミナー、展示等の開催活動におけるツールとして、研究紹介コンテンツや可視化コンテンツを制作した。中でも心臓シミュレータ「UT-Heart」の映像は、YouTubeで26万回以上の視聴回数を記録し、全世界で大きな反響を得ることができた。

推進体制として取り組んだ6番目は、分野を超えた取り組みであった。高度な計算科学技術環境を使いこなせる人材を育成するため、計算科学研究機構や他の戦略分野と連携し、並列化プログラミングやソフトウェアの利用法について講習会等を開催した。また、他の戦略分野や計算科学研究機構との連携の下、相互の技術交流を図るためのシンポジウムやHPCI戦略プログラム合同研究会を毎年開催した。さらに、計算科学研究機構、他の戦略分野等と協力し、「京」コンピュータシンポジウムを2010年から毎年開催して、研究成果の普及を効果的に実施した。

第3節 ポスト「京」に向けて

ポスト「京」の重点課題

「世界最高水準のハイパフォーマンス・コンピューティング（HPC）技術」は、科学技術の振興、産業競争力の強化、国際貢献、安心・安全の国づくりなどの実現に不可欠な国家の基幹技術であることから、2011年からの第4期科学技術基本計画においても、国家安全保障・基幹技術に位置付けられた。

国際的にもさまざまなスーパーコンピュータの開発利用が進められており、2020-2022年ごろに、欧米中でエクサスケール・コンピューティングの実現を目指した研究開発が、国家主導の形で活発に推進されている。

そのような中、文部科学省では、スーパーコンピュータ「京」の後継機の開発を2014年度より着手した。このポスト「京」においては、国家として解決を目指す社会的・科学的課題に戦略的に取り組み、わが国の発展に寄与し、世界を先導する成果を創出することが期待されている。2014年8月20日に産学の幅広い有識者から構成された「ポスト「京」で重点的に取り組むべき社会的・科学的課題についての検討委員会」において、九つの重点課題が選定された。

選定された各重点課題について、2014年10月に実施機関の公募が行われ、同年12月に実施機関が選定された。その結果、ライフサイエンス関係重点課題のうち、次の二つがそれぞれ、理研、東京大学医科学研究所が代表して実施することになった。

重点課題1「生体分子システムの機能制御による革新的創薬基盤の構築」

（代表機関：理化学研究所生命システム研究センター

課題責任者：理化学研究所生命システム研究センター 奥野恭史）

重点課題2「個別化・予防医療を支援する統合計算生命科学」

（代表機関：東京大学医科学研究所

課題責任者：東京大学医科学研究所 宮野悟）

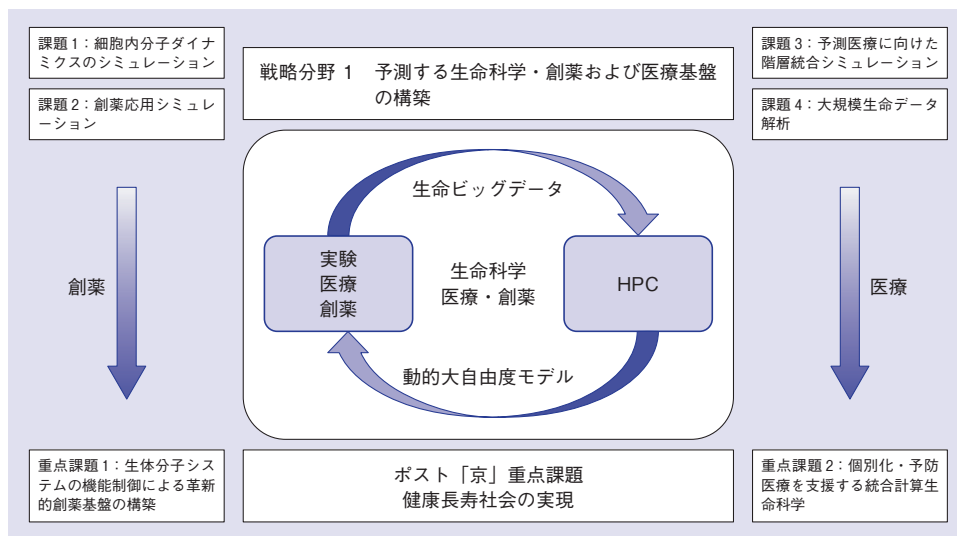


図11 「HPCI戦略プログラム」戦略分野1とポスト「京」重点課題の関係

ライフサイエンス分野では、「HPCI戦略プログラム」の戦略分野1で取り組んできた研究課題によって、実験、医療、創薬から得られた生命ビッグデータをモデル化し、「京」などのHPCによるシミュレーションが可能となっている。また、その計算結果を実験、医療、創薬にフィードバックする連携体制もでき上がった。こうした状況において、ポスト「京」では、二つの重点課題の発展を通して、「健康長寿社会の実現」を目指している（図11）。

そこで次に、HPCI計算生命科学推進プログラムで取り組んできた四つの課題の成果を、ポスト「京」でどのような形に発展させようとしているか、その概略を紹介する。

課題1からポスト「京」へ

まず、課題1の細胞内分子ダイナミクスのシミュレーション成果をどう発展させるかである。

従来行われていた分子シミュレーションは、さまざまな実験結果と比較するためには、計算の規模や精度も不十分であり、実験結果のごく一部をサポートするにとどまっている。それでも、この研究開発で可能となった計算規模の拡大、マルチスケール化、実験データを繰りこんだシミュレーションなどにより、計算と実験を共同して進める共通基盤が初めて構築された。これを受けてポスト「京」では、さまざまな連携研究の在り方が議論されている。

特に、近年急速に発展しつつあるクライオ（低温）電子顕微鏡を用いた立体構造の高精度化や、単粒子解析での構造アンサンブルの精密化などに対して、課題1で開発された技術を適用することが計画されている。また、In-cell NMRや溶液NMRからタンパク質間相互作用の情報をきちんと取り出すには、課題1がカバーしてきた分子動力学計算による情報が必要になる。この点で、緊密な連携が実現すれば、従来の「静的な」構造生物学から「動的な」構造生物学へと大きな展開がなされ、結果として創薬などへの応用も新展開する可能性がある。そのためにも、シミュレーションと実験、あるいはデータ解析の二つ以上の技術を有する新しい人材の開拓・育成が特に重要で、「動的な構造生物学」を構築できるかどうか、そのカギを握っていると言える。

「京」のために開発された分子動力学ソフトウェア（GENESISなど）は、世界的に見ても高いパフォーマンスを示しており、先行するNAMDやGROMACSに追いつき追い越すことも現実的な目標となりつつある。ポスト「京」においては、ソフトウェアとハードウェアの開発を連携させたコデザインが進められており、さらなるパフォーマンスの向上が期待できる。実際、ソフトウェアの性能は計算機アーキテクチャーに大きく依存することが明らかであり、分子動力学のような大規模計算を高速化するために最もふさわしいハードウェアとは何なのか、汎用化と専用化のいずれが使いやすいのか、といった議論は継続されるべきである。つまり、汎用機であるポスト「京」を超えた議論が、この分野では必要となっていることは間違いない。

課題2からポスト「京」へ

課題2の創薬応用シミュレーションにおいても、扱える対象は限られていたのが現実である。キナーゼのような水溶性の単一ドメインの標的タンパク質に対する低分子化合物の結合自由エネルギー計算では、薬剤設計に必要な1kcal/molの計算精度が実現できた。しかし、膜タンパク質や核内受容体など、脂質や核酸と複合している創薬標的に対しては、計算精度が不足していた。このため、ポスト「京」では、DNAやRNAなどの核酸や脂質分子の高精度力場の開発が必須となる。DNAに関しては、藤谷秀章（東京大学）が開発した力場パラメータ割り当て法であるFUJI力場の拡張がほぼ完成しており、脂質分子の力場についても重要なブレイクスルーが得られている。したがって、これらの研究成果をつなげて、タンパク質複合体に対する計算創薬法の確立を目指すことが具体的目標となる。

以上、課題1、課題2の成果を踏まえて、ポスト「京」の重点課題1の「生体分子システムの機能制御による革新的創薬基盤の構築」が導かれる（図12）。ポスト「京」の演算能力を最大限に活かすような分子シミュレーション技術を開発し、生体分子システムの時間的・空間的な機能解析を実現する新たな構造生命科学を開拓し、それらを通じて次世代の創薬アプローチとなる計算技術の開発を目指している。

課題3からポスト「京」へ

課題3の予測医療に向けた階層統合シミュレーションでは、タンパク質、細胞

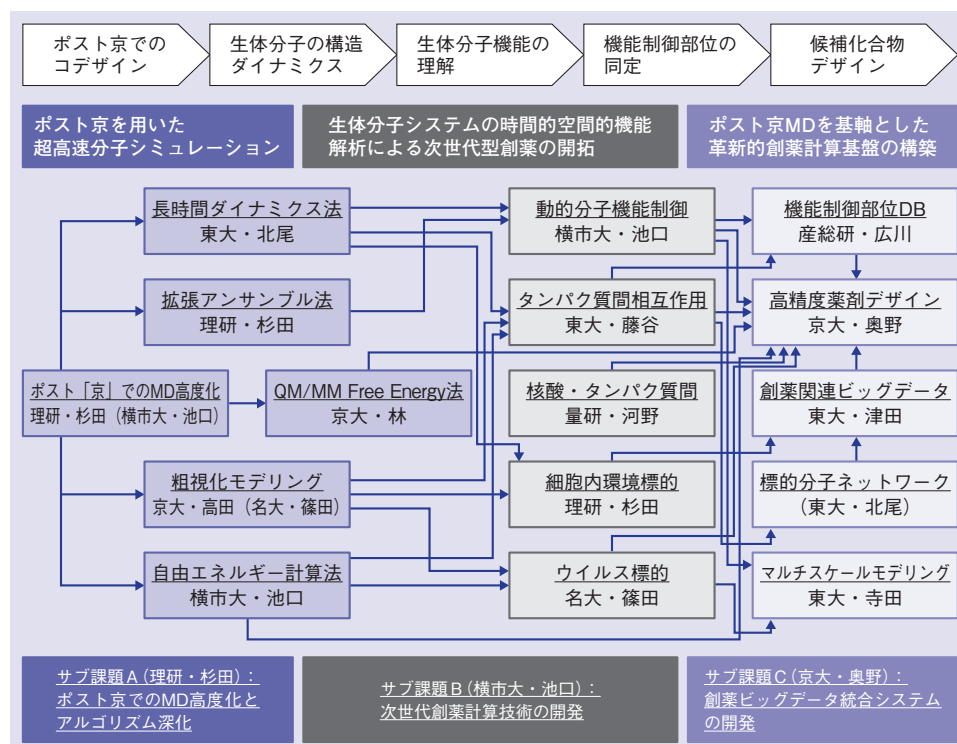


図12 ポスト「京」重点課題1「生体分子システムの機能制御による革新的創薬基盤の構築」の研究体制

から組織・器官の働きまでを統合するソフトウェアの開発を進めてきた。2016年現在、ソフトウェアがようやくそろった段階であり、今後の医療応用の観点からは、個々の患者のデータを用いた臨床応用が重要な意味を持つ。ポスト「京」では、複雑な病態を再現することも重要となるが、それ以上に臨床の現場にソフトウェアを持ち込み、疾患の早期発見、早期治療に役立てることが期待される。これまで作ってきたソフトウェアを社会還元できるかどうかである。

研究テーマとしては、課題4で培ったバイオインフォマティクスを生体力学シミュレーションに活かして、個人の身体的特性だけでなく、ゲノム情報と身体的特性を合わせて疾患を予測したり治療法を検討したりできるツールの開発がある。さらにその先には、シミュレーションを介して、生命の謎そのものを解明していくという大きなテーマもある。

課題4からポスト「京」へ

課題4の大規模生命データ解析では、全ゲノムシーケンス、全RNAシーケンス、コピー数解析、エピゲノムデータといったビッグデータの解析が戦略的に重要であることを明らかにした。特に、がん研究ではスーパーコンピュータが不可欠であり、数理統計、並列アルゴリズムの設計、臨床サンプルを使った検証実験などを一体として進める必要性が強く認識された。単にデータをスーパーコンピュータに投入すれば成果が出るわけではなく、大規模なサンプルにより検証実験を行うことが不可欠であった。その結果、HPCを使ったがんの医学研究のためのプロトタイプ（人材・ソフトウェア・計算リソースから成る）ができた。

2016年現在、シーケンスの技術は急速に一般化しており、現時点でヒトゲノムの解析コスト（30コピー）は日本において商業ベースで20万円を下回っているところがある。今後、ナノ技術を応用したシーケンス技術が登場すれば、1万円以下、1時間以内でヒトゲノムのシーケンスが可能となる。一方で、ATLも含め、がんの病態の本態は、タンパク質をコードしているエクソンだけを調べても捉えられないことが分かってきた。一人の患者に対して、全ゲノム、全RNA、全エピゲノムを身体の複数の場所から採取して、また可能な場合は時系列で解析する時代が来ている。具体的には、600以上のさまざまながんデータ解析ソフトウェア「Genomon」による高精度化・大量処理化が待ったなしの状態であり、この対応を今後進めていく。コスト面からは、スーパーコンピュータの超省電力化が不可欠であり（シーケンスのコストより電気代のコストが大きくなる）、「京」以上に省電力のスーパーコンピュータが必要となる。

以上、課題3と課題4の成果を踏まえて、ポスト「京」の重点課題2「個別化・予防医療を支援する統合計算生命科学」のうち、がん研究（大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明）が展開されることになる。

また、ノンコーディングRNAを含む遺伝子発現データを対象とした大規模遺伝子ネットワーク解析ソフトウェアSiGNシリーズ（およびその改良版）を發展させ、新学術領域「システム癌新次元」（領域代表：宮野悟、期間：2015-2019年）の中で基礎研究を展開していくことになる。

