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
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」

BioSupercomputing Newsletter

2013.3 Vol.8

4th Biosupercomputing Symposium
2012年12月3日-5日　東京国際フォーラム

CONTENTS

●SPECIAL INTERVIEW
○革新的なアプローチでライフサイエンス分野の未来を切り拓いてきたグランドチャレンジ　2-3
理化学研究所　次世代計算科学研究開発プログラム プログラムディレクター　茅 幸二
○ライフサイエンス分野の研究開発に革新をもたらした画期的なプロジェクト　4-5
理化学研究所　次世代計算科学研究開発プログラム 副プログラムディレクター　姫野 龍太郎

●研究報告
○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　6
東京大学新領域創成科学研究科　久田 俊明、杉浦 清了、鷲尾 巧、岡田 純一、高橋 彰仁
（臓器全身ケールWG）
○膵臓β細胞内インスリン顆粒動態シミュレーション・モデル　7
神戸大学大学院システム情報学研究科　玉置 久（細胞スケールWG）
○京による全脳シミュレーションへの道のり　8
理化学研究所　脳科学総合研究センター　ユーリッヒ研究センター 神経科学・医療研究院（INM-6）
アーヘン工科大学医学部　マーカス・ディースマン（脳神経系WG）
○大規模並列用MDコアプログラムの開発　9
理化学研究所　次世代計算科学研究開発プログラム　大野 洋介（開発・高度化T）

●SPECIAL INTERVIEW
○循環器系および筋骨格系・脳神経系における階層統合シミュレーションの実現をめざす　10-11
東京大学大学院工学系研究科機械工学専攻 教授　高木 周（課題3 代表）
○最先端シークエンサーによる大規模データを「京」で解析し生命プログラムとその多様性の理解を進める　12-13
東京大学医科学研究所ヒトゲノム解析センター 教授　宮野 悟（課題4代表）

●報告
○「4th Biosupercomputing Symposium」の開催報告　14
理化学研究所　次世代計算科学研究開発プログラム　田村 栄悦
○京互換機：SCLS計算機システムの導入　15
理化学研究所　HPCI計算生命科学推進プログラム チーム員　木戸 善之

●イベント情報＆ニュース　16

「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6年半をふりかえる

革新的なアプローチでライフサイエンス分野の未来を切り拓いてきたグランドチャレンジ

理化学研究所　次世代計算科学研究開発プログラム プログラムディレクター
茅 幸二

●ハードとソフトの両面から「京」を見守ってきた6年半

理化学研究所が、今日の「京」の開発主体となり、多くの公的研究機関や大学とともに、その開発と利用環境の整備に取り組むことになるにあたっては、内部の科学者会議で何度も議論を重ねました。当時、私はその会議の議長をしていました。物質科学分野をはじめ幅広い分野で、既にスーパーコンピュータを活用した研究が進展していましたが、ライフサイエンス分野ではまだまだという状況でした。そうしたなか、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非常に重要であり、理化学研究所は積極的に取り組んでいくべきだとする提言をまとめるに至りました。プロジェクトのイニシアティブをとる文部科学省をはじめ、理事の方々もさまざまな議論をされていたと思いますが、理化学研究所が中核機関に決まる経緯には、こうした私たちの提言もひとつの契機になったのではないかと考えています。

スーパーコンピュータの開発とともに、それを最大限に利活用していくためのソフトウェア開発として、グランドチャレンジ・アプリケーションの研究開発が、プロジェクトの大きな目標のひとつになっていました。「次世代ナノ統合シミュレーションソフトウェアの研究開発」に関しては、ちょうど私が自然科学研究機構分子科学研究所で所長を務めていた2000年代初頭に、すでにスーパーコンピュータが導入され、ナノテクノロジーに関する理論家たちが集まって計算科学の共同研究が始まっていました。そのため、分子科学研究所を中心に、電子・原子・分子レベルからの精緻な大規模計算によってナノ物質の特性や現象を理解して予測につなげるグランドチャレンジのシミュレーションソフトウェア開発がスタートしました。しかし、ライフサイエンス分野に関しては、それまで計算科学的なアプローチはほとんど未開拓に近い領域でした。それでも、この分野で研究開発を進めていかなければいけないと強く主張され、グランドチャレンジ「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」の推進役を果たしたのが姫野龍太郎さん、そして泰地真弘人さんたちでした。その見識は正しくて、これはやらなければいけないということになり、13機関（最終的には15機関）が集まって研究開発体制が確立され、2006年6月に公募による文部科学省の審査を受けて、正式にスタートしたのはこの年の9月でした。私は、ライフサイエンスの専門ではなかったので、正直なところプログラムディレクターへの就任を依頼されたときは、少し悩みました。それでも、ナノテクノロジー分野の計算科学には関わったことがあったので、組織管理の面では多少なりともお手伝いができるのではないかと考えて、お引き受けすることにしました。2008年には次世代スーパーコンピュータ開発実施本部の副本部長も兼務することになり、結果的にはハードとソフトの両方から「京」の開発を見守ることになりました。

●確かな成果を達成したグランドチャレンジ

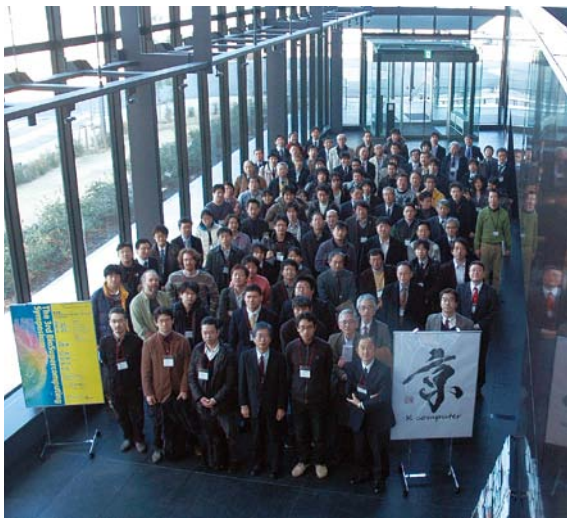
物質科学の分野では、誰もシミュレーションソフトウェアの開発研究に疑問を抱いていませんでしたが、一方のライフサイエンス分野に関しては、当初、理解を得ることもなかなか難しい状況でした。何といっても分子レベルから全身スケールまで扱う領域が非常に広く、まさにマルチスケールである上に、それぞれの階層においても、ベースになる確固たる理論がほとんどありませんでした。例えば細胞に関しては、もちろん理化学研究所内でも実験による研究は進められていましたが、とにかく細胞は複雑です。タンパク質などいろいろなものが含まれており、水のなかにこれほどまでものが溶け込むのかというくらい細胞内は濃度の高い溶液になっていて、これを正しく再現し、機能を計算することなどできるのだろうかと、誰もが疑問を感じるほどの非常に難しいテーマでした。しかし、結果的にはこの6年間で、ある程度の見通しがつくまでに研究が進みました。もちろん、実際に再現するためには、「京」の計算能力をもってしても追いつかない問題であることも、また事実ですが。

グランドチャレンジのプログラムディレクターに就任した当初、姫野さんたちと一緒に、たくさんのライフサイエンス分野のリーダーシップをとっておられる先生方にお会いしました。ところが、多くの先生方は素っ気ないというか、私たちがやろうとしていることに積極的な理解と支持を示していただけませんでした。わずか6、7年前ですが、当時はそんな状況だったのです。しかし、今や細胞レベルでもシミュレーション研究による理解が進み、分子レベルでもさまざまなタンパク質の機能が再現できるまでになり、開発研究は飛躍的な進歩を遂げています。脳研究の分野でも、「京」を使うことによって、世界で初めて本当の意味で脳らしい脳の機能がコンピュータ上で再現されるだろうといわれるまでに研究が

建築中の計算科学研究機構の建屋（2009年4月）

報道関係者に公開された建設途中の計算機室（2010年2月）

「京」完成記念写真



The 3rd BioSupercomputing Symposium（2011年3月）

進展しています。さらに、臓器・全身スケールでは、非常に優れた心臓シミュレーションが可能になり、血流シミュレーションでも全く新しい流体構造連成解析手法の開発によって、血栓ができる過程の再現をめざすまでになり、また、超音波治療機器の開発に結び付く生体の音響的シミュレーション研究も成果を挙げています。これらは医療や医療工学への貢献という意味でも、非常に重要な成果といえます。そして、データ解析融合研究の分野でも、当初はバイオインフォマティクス（生物情報科学）の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を疑問視する声もあったようですが、より手軽にゲノム解析ができる環境が整い、膨大な量のデータを処理することが求められるオーダーメード医療の時代が近づくなか、さまざまな処理の問題を含め、高い計算性能をもつスーパーコンピュータが役立つことが実証されました。

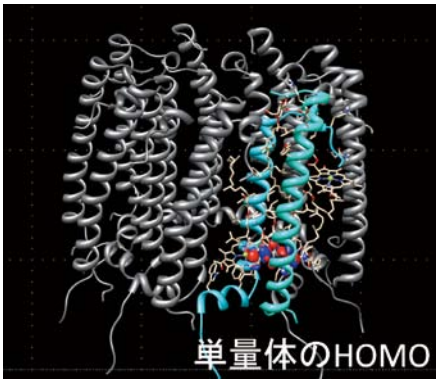
当初は、非常に複雑で幅広いライフサイエンス分野において、計算科学的なアプローチによって十分な結果が出せるのか、疑問を感じていた人が多かったと思います。そんななか、この6、7年間で、それなりの手ごたえと十分な可能性が得られたことは間違いありません。また、このような段階に至るまでには、ソフト開発の技術開発の中心として本プロジェクトを支えた、泰地さんをリーダーとする生命体基盤ソフトウェア開発・高度化チームの果たした重要な役割を、忘れることはできません。もちろん、複雑な生命現象や生命のメカニズムを明らかにするまでには、まだほど遠いわけですが、それでもこれだけの確かな成果が得られことは事実であり、グランドチャレンジのインパクトは大きかったと思います。シミュレーションという新しい技術が、これまでの生物学や医療に対して、まさに“一石を投じる”ことができたのではないかと感じています。

●若い力でライフサイエンスの未来を開拓してほしい

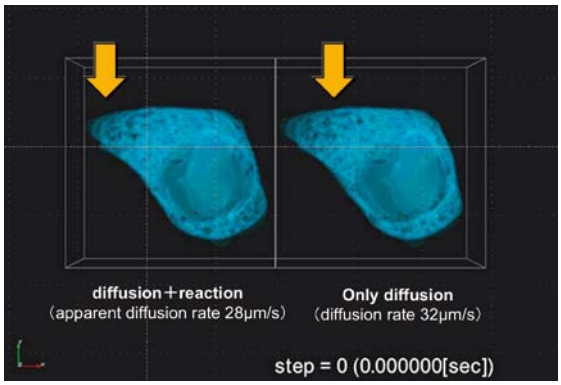
グランドチャレンジのもうひとつの大きな成果は、ライフサイエンス分野のシミュレーション科学が大きく成長するなかで、この分野に興味を抱いた若手の研究者が数多く参加し、幅広い研究分野との連携が進んだことです。こうした若い研究者たちは、今後の研究を支えていく非常に重要な人材になると、私たちは考えています。しかし、問題はこれらの若い研究者が取り組む計算科学を用いた新しい研究が、まだ十分に認知されていないことです。理解が進んでいる米国などでは、計算科学に取り組む研究者にもそれなりのポストがありますが、日本のライフサイエンス分野ではまだそこまで達していません。ポストに空きがあると、実験の人を入れてしまうという傾向は今も多く見られます。シミュレーション研究が、今後ライフサイエンス分野で重要な役割を担っていくことは間違いありませんが、これまでこの分野の研究をけん引してきた先生方には、まだ確信が持てないでいるように思えてなりません。また、研究費の面でも非常に厳しい時代ですから、そう簡単にポストを増やせないという事情もあると思います。それでも、少し無理をしてでも新しい分野を拓く若い研究者たちにポストをつくっていただきたいと、私は願っています。

2013年3月をもって、グランドチャレンジのプロジェクトは終了します。それぞれのチームで優れた成果を挙げながら、これまで築き上げてきた研究開発の流れをいかにして継続し、拡大していくか、それがこれからの大きな課題になると思います。また、ライフサイエンス分野というのは、これまで他の分野との交流や連携が非常に限られていました。しかし、グランドチャレンジでは、そうした垣根を越えて、さまざまな分野の研究者がひとつのテーブルについて、新しいライフサイエンスをつくりあげてきました。若い研究者たちには、これからもどんどん垣根を越えて、いろいろな分野の知識や情報を吸収しながら、新しい世界を切り拓いていってほしいと願っています。

いつの時代も、新しい分野を拓いていくのは苦しいものですし、たいへんなことも多いものです。しかし、今、確かにいえることは、ライフサイエンスは大きな変革の時代を迎えようとしているということです。これからたくさんのイノベーションが生まれていくはずです。グランドチャレンジはそのための滑走路のひとつであると思っています。一人でも多くの若い研究者たちが、この滑走路を活かして、新しい道を切り拓いていく努力を続けていってほしいと考えています。



Protein DF ータンパク質全電子計算



Ca+ kinetics simulation (HepG2)

「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6年半をふりかえる

ライフサイエンス分野の研究開発に革新をもたらした画期的なプロジェクト

理化学研究所　次世代計算科学研究開発プログラム 副プログラムディレクター
姫野 龍太郎

●プロジェクトから生まれた驚異的な成果

私たちが2006年10月から取り組んできた「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」は、ひとことでいえば、「次世代スーパーコンピュータプロジェクト」の一環として、ライフサイエンス分野のシミュレーションソフトウェアの“ショーケース”をつくろうという取り組みでした。10PFLOPS級のスーパーコンピュータの性能をもってすれば「こんな新しいことができます、画期的なことができます」ということを示すために、このプロジェクトが組まれました。

この6年半を振り返ってみると、いろいろな意味でできたことと、できなかったことがありますが、シミュレーションソフトウェアの研究開発という面では、当初想定していたよりも、「京」に最適化されたソフトウェアがしっかりとでき上がり、理論性能に対する実効性能の高さをアピールできるものを、数多く生み出すことができたと思います。

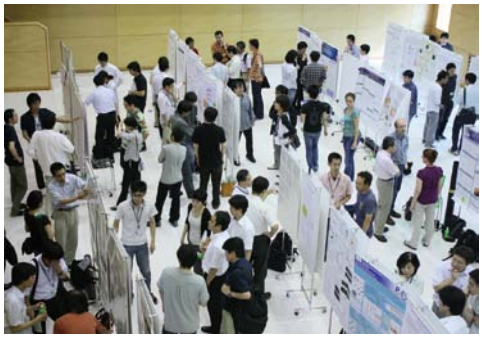
「京」に最適化した分子動力学ソフト「大規模並列用MDコアプログラム（cppmd）」が、期待されていたとおり40%近い性能を出したことはもちろんですが、「ボクセル流体構造連成解析プログラム（ZZ-EFSI）」も、43%という驚異的な実効性能をたたき出してくれました。さらに「マルチスケール・マルチフィジックス心臓シミュレーション（UT-Heart）」も30%近い性能を出していますし、「低侵襲治療シミュレーション（HIFU）」も20%を超える性能を出しています。もともと私たちは、「京」全体を使用して実効性能1PFLOPSを超えるシミュレーションソフトウェアの開発を目標にしていました。そうしたソフトウェアを、2本くらいは開発したいと思っていたのですが、この4本は、実際に「京」全体を使えば、間違いなく20%を超える性能を出してくれるはずです。つまり、当初の目標を上回る成果を達成することができた――というとちょっと大袈裟かもしれませんが、とにかく十分に性能を発揮するシミュレーションソフトウェアができました。もちろんこれらだけでなく、「京」1万ノードを超えた時点で確かな実績をあげたソフトウェアが、この4本を含めて13本あります。全体約30本のソフトウェアのなかの3分の1以上が、十分な実績を達成したというのは、実は私たちの当初の予想を超える驚異的な成果です。「わずか3割で?」と思われるかもしれませんが、2006年にプロジェクトが始まったころ、スーパーコンピュータで使えるライフサイエンス分野のシミュレーションソフトウェアなど、本当に開発できるのかといった、開発そのものを疑問視する厳しい声が相当ありました。それを考えれば、実際に使えるソフトウェアが13本も開発できたこと自体が、インパクトのある素晴らしい成果といえると思います。

振り返ってみて、よかったと思うのは、毎年、きちんとした成果を出し続けることができたことです。年々、確かな実績を示してきたことで、少しずつまわりも認めてくれるようになりました。何より、実際に開発をしている研究者たちも、努力して並列性能が向上していくことによって、成果として自分たちにフィードバックされることを理解し、今では「もう『京』なしにはできない、いや『京』でも足りないくらいだ」というまでになっています。当初、「10PFLOPSのマシンなんて、使いこなせるだろうか」といった不安の声があったことなど、今では信じられないくらいです。

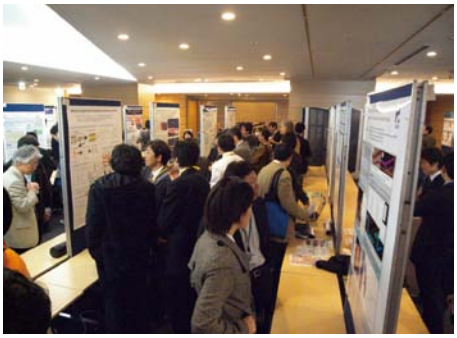
特に2012年度は、私たちがHPCI戦略プログラムの5分野とともに優先的に10PFLOPSを使えるようになったのが、7月からわずか2カ月間だけだったこともあり、計算時間が足りないというみんなの悩みをどう処理したらよいのか、こちらも苦しい思いでした。当初の状況から考えれば、まさに嬉しい誤算ともいえますが、優先的に使える期間が少なかったために、サイエンティフィックな成果が十分に出せていないものが少なからず残されている点が、とても残念です。例えば、脳の発達や学習の再現をめざして脳全体レベルでの入出力関係を再現するためのシミュレーションソフトウェア「局所回路シミュレータ（NEST）」は、「京」全てを使うと、世界最大の脳科学の計算ができるところまでソフトウェアはでき上がっていますが、十分な計算時間が取れませんでした。その意味では、あと1年あったら、もっといろいろなことが達成できるのではないかという思いもあります。

このように、やり遂げられなかったものもありますが、全体的には、プロジェクトとして確かな成果をあげることができたと思っています。もしこのプロジェクトがなかったとしたら、恐らく今でも「ライフサイエンス分野では『京』なんか必要ない」という研究者が、数多くいたのではないでしょうか。

「次世代生命体統合シミュレーション研究開発プロジェクトシンポジウム2007」で公演する姫野副プログラムディレクター



Joint Computational Science Workshop 2009（2009年7月）のポスターセッションの様子



第2回バイオスーパーコンピューティングシンポジウム（2010年3月）のポスターセッションの様子

●サマースクールから生まれた新たな融合

「京」については、ネットワークも相対的に速く、非常に優れたハードウェアに仕上がっていると思いますが、使いやすくなるまでには、研究者が使う前の段階でバグを取ったり、性能をフルに発揮させるためのさまざまな取り組みを行ってきた、生命体基盤ソフトウェア開発・高度化チームの存在を見落とすことはできません。本当によくやってくれたと思います。彼らの地道な努力のおかげで、「京」が使いやすいマシンに仕上がったとともに、実効効率の高いソフトウェアが数多くできたのも、こうしたチームがあったからこそだと思います。特に、高速化のノウハウというものは、ソフトウェアを開発する研究者全体で共有することが大切です。どうすればより高い性能が出せるか、その細かい情報の積み重ねを共有する上でも、彼らが果たした役割は大きかったと思います。

共有ということで思い出すのは、プロジェクトが始まったばかりのころです。多くの研究機関や大学からスケールが異なる幅広い研究分野の方たちがプロジェクトに参加してくださったのはよかったのですが、最初のうちはシンポジウムを開催しても、各研究開発チームごとにそれぞれの専門用語や略語が飛び交って、何をどう話しているのかお互いに理解できないという状況もありました（笑）。分かりやすく説明し合って、少しずつ理解は深まっても、なかなか融合が生まれませんでした。そうしたなかで、実施してよかったと思うのが、若手を集めたサマースクールでした。それぞれの研究者が新しいモデルに取り組んでいるとき、研究分野が少し異なる人たちが、同じようにコンピュータを使って、自分と違うやり方でアプローチしている様子を目にすると、それが直接的な刺激になるわけです。手法やアルゴリズムだけでなく、取り組む姿勢も含めて、全てが刺激になります。サマースクールで、自分と違う研究分野の人たちが、お互いに何をやっているのかを理解し始めたことで、研究分野や専門用語の壁を乗り越えることが可能になり、相互の融合が進んだと思います。流体構造連成解析手法による血栓のシミュレーションなども、こうした新たな融合によって生まれた成果といってよいでしょう。

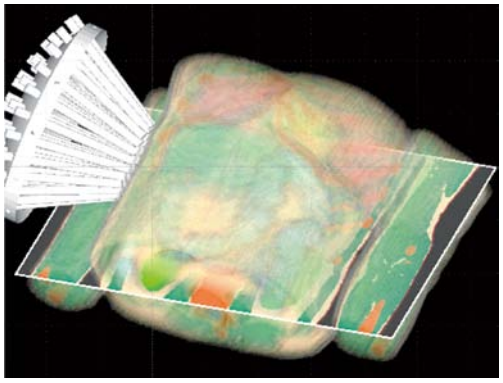
新たな融合によって、それぞれの分野に新たな広がりが生まれつつあり、このプロジェクトがそのきっかけになったことは、私にとって非常にうれしい成果です。このこと自体は論文にもなりませんし、成果ではないかもしれません。それでも、長いスパンで見たときに、このプロジェクトがターニングポイントとなって、新たな広がりが生まれ、研究開発が大きく進展したといわれる日がくることを願っています。

●プロジェクトの成果をさらに発展させていくために

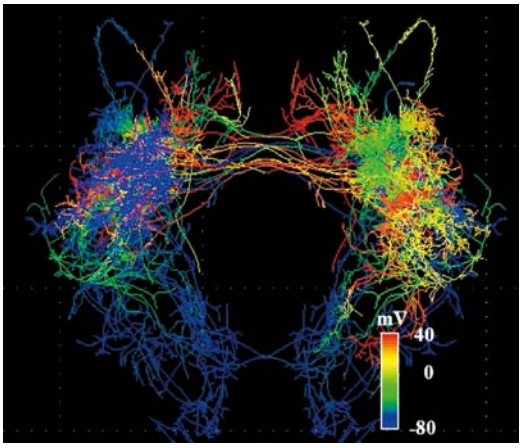
生命現象は、まさに多体系かつ多階層で、こうした現象を対象にシミュレーションを行っていくためには、新たな融合によって新しいフェーズを開拓していくことが必要です。その意味では、今回のプロジェクトでは、それぞれの研究開発チームで新たな融合が生まれ、それぞれに興味深い成果が得られようとしています。プロジェクトそのものはソフトウェアの開発に主眼が置かれていましたが、ここへきて、開発したソフトウェアを活用した研究成果の方へシフトしつつあり、非常に面白い成果も生まれつつあります。また、どちらかといえば私たちはサイエンス・ドリブンで、研究者が自分たちのシーズとするためにソフトウェア開発を進めてきましたが、一方では、実際の創薬や医療分野で使えるといった実用化へ向けた可能性が開けているものも、いくつか生まれています。

プロジェクトは2012年度末で終了し、それぞれの研究開発チームも解散することになりますが、それぞれの成果のなかで、ニーズに結び付く可能性が高いものは、HPCI戦略プログラム（分野1）に継承されていきます。それ以外にも、このまま終わらせてしまうのでなく、継承・発展させていくべきと思われるソフトウェアについては、理化学研究所情報基盤センターで引き受け、継続的な開発や更新、サポートなどを進めていく考えです。理研のさまざまな研究センターなどで、実際に使ってもらえるようになればよいと思っています。そうした仕組みを用意して、一部とはいえ使える環境を維持していかないと、ただ開発して終わりということになってしまいますから。そして、それを理化学研究所の研究テーマのなかで進めていくことで、このプロジェクトを理化学研究所がやってよかったということを、多くの人たちに実感してもらいたいと願っています。

また、先ほどサマースクールに参加して意識が変わった若手研究者の話をしましたが、こうした研究者たちを継続的にサポートしていくために、バイオスーパーコンピューティング研究会を立ち上げましたので、今後もこの研究会を発展させていきたいと考えています。せっかく育ち始めた若い世代の研究者、新しい視点を持った研究者たちを、引き続き育てていくには、非常によい場になると思っています。



超音波治療シミュレータ：ZZ-HIFU



神経ネットワーク活動のシミュレーション

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

東京大学新領域創成科学研究科
（上から）久田 俊明、杉浦 清了、鷲尾 巧、岡田 純一、高橋 彰仁
（臓器全身スケールWG）

ヒトの心臓はATP（アデノシン三リン酸）加水分解による生化学反応をエネルギー源として、電気(イオン電流、興奮伝播、心電図など)・化学(物質輸送、反応、エネルギー変換など)・力学（心筋張力、血圧、血流など）の諸現象に広く派生するマルチフィジックス問題を構成する。また空間尺度としては、タンパク分子(～10nm) から細胞(～100μm)、組織(～mm)、臓器(～cm) を経て血液拍出に至るマルチスケール問題を構成している。

ATPは、一つの心筋細胞内に5, 60本存在する筋原線維を構成する沢山のアクチンフィラメントとミオシンフィラメントの間の相対的滑り運動のためのエネルギーとして消費される。フィラメントの相対滑りはミオシンフィラメントから延びるミオシン分子のヘッドがアクチンフィラメントに結合して牽引力を与えることにより生じるが、その具体的な機構に関する仮説は多様性に富み未だ議論に終止符は打たれていない。図1右はミオシンヘッドの首振り説を示す。一方、心肥大は弁膜症や高血圧など心臓が高い血圧を発生しなければならない状況に適応するために壁厚（心筋の断面積）を増やす現象と理解されているが、肥大型心筋症と呼ばれる疾患では血圧は正常で肥大する理由がないのに壁が厚くなる。初期の検討ではミオシン分子が点突然変異（代表的なものとしてアクチン分子との結合部位に当る403番目のアミノ酸の変異がある）し心機能が落ちるのでこれを代償するために肥大すると理解されたが、その後の研究では変異ミオシンはむしろ機能が亢進していることが報告された[1]。では機能の亢進したミオシンが混ざると心臓に何が起きるのか。

私達はこれまで超並列計算機による心臓のマルチスケール解析技術を開発してきたが[2]、現在は“京”を用いて本格的にこの問題に取り組んでいる。 即ち、従来は計算機パワーの限界からアクチン分子とミオシン分子の確率的結合状態は平均的な一分子を考え、これを状態遷移を表す方程式に適用して解かざるを得なかったが[3]、超並列計算機の特徴を活かせば図1に示すように、分子内の弾性要素をばねでモデル化した一つひとつのミオシンヘッドの確率的運動を具体的にシミュレートするところから、細胞の収縮を経て、組織、臓器の運動までをシームレスに解くことが出来るようになった。ミオシンヘッドの運動には、例えば一つのミオシンヘッドがアクチンフィラメントと結合すると近隣のミオシヘッドも結合し易くなる“協調性（Cooperativity)”と呼ばれる特徴が知られるが、本計算モデルではこのような特徴もありのままに導入することが出来、また前記の機能の亢進したミオシンが混じった場合についての心拍動の様子も調べることが可能となる。具体的手順は次のようになる。(1) ミクロスケールにおけるミオシン分子の協調性をともなう確率的な振る舞いを直接モンテカルロ法でシミュレートする。(2)細胞間隙などを含む細胞のメゾスケール構造体モデル内の筋肉線維部に上記モンテカルロモデルを埋め込み、スケール間の力学的相互作用を物理的に正しく取り扱う。(3)上記2次構造体モデルとマクロスケール臓器モデルの運動を均質化法により結び付ける。

以上のような解析法の実現により、分子レベルでの状態変化の法則および 細胞組織レベルでの各構造体が心臓の拍動性能やエネルギー効率にどのような影響を与えているのか、逆にマクロ的な筋肉の収縮弛緩がフィードバックされて分子レベルでの状態変化にどのような影響を及ぼしているのか、などをシミュレーションを通じて分析することが可能となった。このようなシミュレーションプラットフォームは今後、分子生物学者が自分の仮説を新たな観点から検証する手段にもなると考えられる。

UT-Heartの研究開発・検証は、富士通（株）次世代テクニカルコンピューティング本部（門岡良昌アプリケーション統括部長のグループ）、東京大学医学部附属病院（永井良三客員教授のグループ）と共同で行われている。また臨床研究は東京大学医学系研究科倫理委員会の承認のもとになされている。

【参考文献】

[1] Belus A, Piroddi N, Scellini B, Tesi C, Amati GD, Girolami F, Yacoub M, Cecchi F, Olivotto I, Poggesi C, The familial hypertrophic cardiomyopathy-associated myosin mutation R403Q accelerates tension generation and relaxation of human cardiac myofibrils, J Physiol, 586(Pt 15), 3639-3644, 2008
[2] Hosoi A, Washio T, Okada J, Kadooka Y, Nakajima K, Hisada T, A Multi-Scale Heart Simulation on Massively Parallel Computers, SC10, NewOrleans, proceedings 1-11, 2010
[3] Washio T, Okada J, Sugiura S, Hisada T, Approximation for Cooperative Interactions of a Spatially-detailed Cardiac Sarcomere Model, Cellular and Molecular Bioengineering, 5, 113-126, 2012

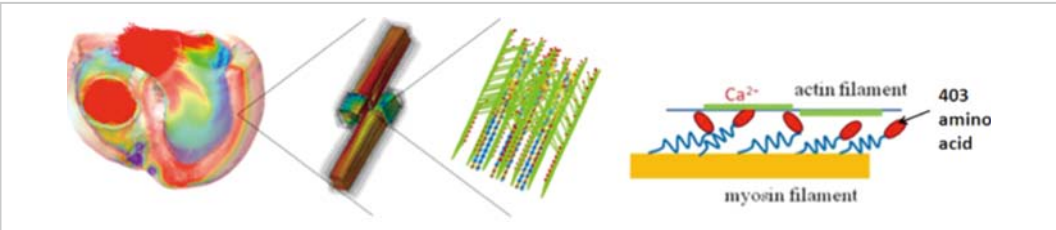


図1：京によるサルコメア動力学からの心拍動マルチスケールシミュレーション

膵臓β細胞内インスリン顆粒動態シミュレーション・モデル

神戸大学大学院システム情報学研究科
玉置 久
（細胞スケールWG）

今日、世界的に糖尿病が激増しており、特にアジア人においては増加が著しくなっています。ここで、アジア人に見られる糖尿病は、インスリンの分泌不全によるいわゆる2型の糖尿病がほとんどです。健常者においては、膵臓細胞内のインスリン顆粒がゴルジ体で生成され、細胞膜に輸送された後、細胞外に分泌されますが、2型糖尿病では細胞内にインスリン顆粒は存在するものの、分泌が何らかの要因により阻害されてしまいます。この要因を明らかにし、適切に対処するためにも、膵臓細胞におけるインスリン分泌機構の解明が求められているところです。

これまでにも、細胞膜におけるインスリン分泌機構について多くの研究がなされ、その解明が進んでいます。しかしながら、細胞内部での顆粒動態メカニズムについてはほとんど未解明であるというのが現状です。我々は、近年観察可能になったグルコース刺激後のインスリン顆粒動態の部分的な観察結果と生物学的知見から、膵臓細胞内インスリンの顆粒動態シミュレーション・モデルを作成し、インスリン顆粒分泌において見られる典型的な現象を再現することによって、インスリン顆粒動態に関する新たな知見獲得へのアプローチを試みています。

インスリン顆粒の動態、すなわち膵臓細胞内部で生成されたインスリン顆粒が細胞膜まで移動して分泌されるといった過程で、顆粒が細胞骨格に沿って運動すると考えられています。そこで、シミュレーション・モデルの構築に際しては、膵臓細胞の骨格・顆粒の動態原理を踏まえ、4つの段階、すなわち顆粒生成部、内部層、外部層および顆粒分泌部に分割して考えています（図1）。

再現を目指す現象については、インスリン顆粒分泌の過程を大きく、（1）第Ⅰ相…グルコース刺激に伴って現れる急峻な分泌、（2）第Ⅱ相…グルコース刺激の後に現れる穏やかな分泌、（3）第Ⅲ相…基礎分泌と呼ばれる持続的でわずかな分泌、からなるものと捉えた上で、健常者、糖尿病患者、投薬時に見られる分泌の典型的特徴として、（a）健常者…第Ⅰ相、第Ⅱ相および第Ⅲ相が現れる、（b）糖尿病患者…第Ⅰ相が現れない、（c）投薬時…第Ⅰ相と第Ⅱ相がともに強く現れる、といった分泌パターンに注目し、これらの典型的パターンを有する分泌過程の再現を目標としています。

シミュレーションの一例として、図2に、健常者と糖尿病患者を想定した二つのケースに対する結果を示します。ここで、ケース1とケース2におけるパラメータ設定の差異は、グルコース刺激下での顆粒分泌部における顆粒分泌数のみですが、実際に観測されるような2相性を持つインスリン分泌（ケース1）および第Ⅰ相が生じないインスリン分泌（ケース2）を再現できていることが確認されます。

現在、投薬時に見られるインスリン顆粒分泌の様子を再現するパラメータ設定を模索中です。これに加えて、典型的な分泌パターンを再現するようなシミュレーション結果を拡充し、各々のパラメータ設定が動態と分泌に及ぼす影響について、より詳細な検討を進めることも肝要であると考えています。さらには、膵臓β細胞内での顆粒動態の実測データ（かなり制限的・断片的なものになるでしょうが）に基づくパラメータ設定、シミュレーション・シナリオの充実、代謝系のシミュレーションとの連係によるシミュレーションの精緻化などについても検討を進めていきたいと思っています。

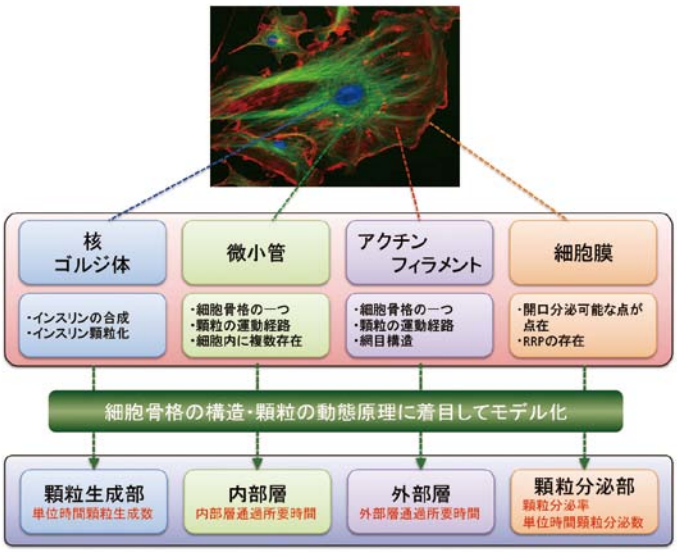


図1：インスリン顆粒動態モデルの概略

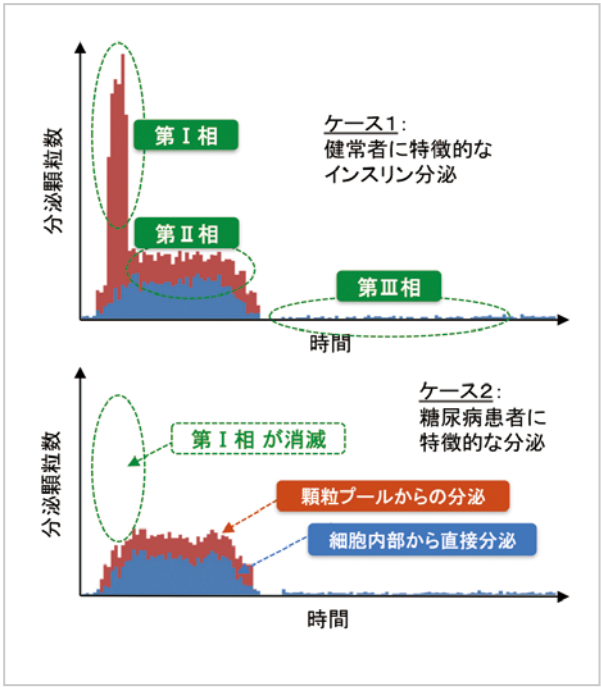


図2：シミュレーション結果

京による全脳シミュレーションへの道のり

理化学研究所　脳科学総合研究センター
ユーリッヒ研究センター　神経科学・医療研究院（INM-6）、アーヘン工科大学医学部
マーカス・ディースマン
（脳神経系WG）

脳神経系研究開発チーム（BNT）が、京の計算能力を計算論的神経科学の分野において利用可能となるようにソフトウェア開発を進めてきた経緯をここで紹介します。本レポートの拡張版は［1］で見ることができます。

ヒトの脳は約10^{11}個の神経細胞で構成されており、各々が10000個の別の神経細胞と結合しています。計算論的神経科学では、神経回路のダイナミクスを解明するためのボトムアップアプローチとして、しばしば、神経細胞およびそれらの相互関係をモデルとして記述することから始めます[2]。NESTシミュレーター[3]はこうしたアプローチに向いたツールです。神経細胞は、微分方程式からなる小さなシステムとして記述され、δインパルス[4, 5]によって相互に情報交換することで、自然な規模と複雑さをもつ回路を構成します。

トップダウンアプローチでは、特定の脳機能を大まかに記述することから始め、この機能がどのように神経細胞レベルで実行されているかを調べます（例：[6]）。脳の機能的回路は、通常多くの脳領野により成り立ちます。しかし、これまでのシミュレーションでは、主に計算機のメモリの必要性から[7]、局所回路（例：[8, 9]）またはスパイクベースでない巨視的モデルに制限されていました。京のようなスーパーコンピュータを用いると、脳のようにマルチスケールな結合をもつ巨大システムのシミュレーションも可能となります。

第1回BNT会議（2008年11月）の時点では、NEST（第2世代カーネル、2g）は、10^5個の神経細胞からなる回路シミュレーションを1024プロセッサにより実施していました。第3回会議（2009年10月）では、ユーリッヒ研究所のJUGENEコンピュータを用いることで、10^6個の神経細胞によるシミュレーションが可能になりました。これを受けて私たちは、2010年の2月までに、京の上で動作させる回路における神経細胞数を10^6から10^8個以上へと拡大することを、マイルストーンとして設定しました。NESTは2010年11月に京のプロトタイプで、2011年5月に試験システムで、そして同年9月には京で、コンパイルすることに成功しました。

3gカーネルでは、神経細胞の局所結合が疎であるという性質を考慮したデータ構造となり、あわせて、神経細胞の非局所結合に関する情報を効率的に保存することができるようになりました[7]。2012年3月の第8回BNT会議では、次のマイルストーンを視野に入れることが示され、5月[10]には、京の利用ノード数として12288に到達しました。近々公開されるNEST 2.2は、このカーネルが基礎となっています。

神経細胞間の結合（シナプス）には、異なるタイプのダイナミクスと可塑性を持つものがあります[11,12]。しかし、一つの神経細胞が、各計算機コア上で、シナプスを持たない、または一個のシナプスだけを持つ場合に制限すると、このシナプス構造の非一様性は無視でき、単一種のシナプスのみからなるようなものと等価に扱うことができます。2012年9月の第9回会議で発表された4gカーネルは、この特性を利用しています。図は、横軸のノード数を用いることでシミュレーション可能な最大回路サイズ（三角）およびその実行時間（点）を示しています。4gカーネルにより、京上で10^9個の神経細胞からなる回路のシミュレーションが可能となることが分かります。NESTのユーザインタフェースは同じままで、カーネルの一般性および機能性はどちらも損なわれていません[13,14]。

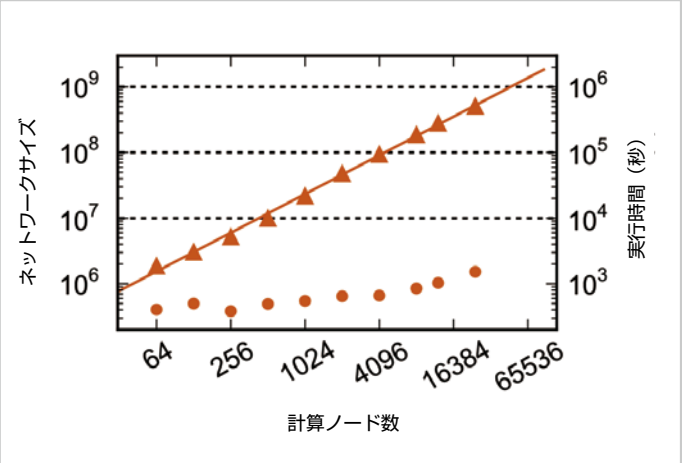
私たちは、神経科学者が、脳の解剖学およびダイナミクスに関するデータをモデルとして統合することで脳機能に関する洞察を得る際に、シミュレーション技術が役立つようになったと考えています。

今後は、4gで新たに開発された技術をNESTの公開版として実装し、さらに、現在準備段階にあるエクサ規模の計算機のためのソフトウェアの開発に取り組みます。

www.csn.fz-juelich.de
www.nest-initiative.org

謝辞

本研究の一部は、理化学研究所 計算科学研究機構での京のアーリーアクセス、文部科学省の次世代スーパーコンピュータ戦略プログラム、ヘルムホルツ協会、HASBおよびポートフォリオテーマSMHB、JARA、並びにEU助成金269921（BrainScaleS）によるものです。



【参考文献】
[1] Diesmann (2012) p83-85 In Proc 4th BioSupercomputing Symposium, Tokyo
[2] Lansner and Diesmann (2012) Chap 10 In Computational Systems Biology, Le Novére ed., Springer
[3] Gewaltig and Diesmann (2007) Scholarpedia 2(4):1430
[4] Morrison et al. (2008) Biol Cybern 98:459-78
[5] Hanuschkin et al. (2010) Front Neuroinform 4:113
[6] Potjans et al. (2011) PLoS Comput Biol 7(5): e1001133
[7] Kunkel et al. (2012) Front Neuroinform 5:35
[8] Potjans et al. (2012) Cerebral Cortex doi:10.1093/cercor/bhs358
[9] Wagatsuma et al. (2011) Front Comput Neurosci 5:31
[10] Helias et al. (2012) Front.Neuroinform 6:26
[11] Potjans et al. (2010) Front Comput Neurosci 4:141
[12] Pfeil et al. (2012) Front Neuromorph Eng 6:90
[13] Eppler et al. (2009) Front Neuroinform 2:12
[14] Gewaltig et al. (2012) Chap 18 In Computational Systems Biology, Le Novére ed., Springer

大規模並列用MDコアプログラムの開発

理化学研究所　次世代計算科学研究開発プログラム
大野 洋介
（開発・高度化T）

我々のチームでは、「京」用に高速化されたコアプログラム・ライブラリの提供および「京」でのアプリケーション高速化技術の習得・蓄積を目的として、大規模並列用MDコアプログラムを開発してきました。

分子動力学（Molecular Dynamics, MD）シミュレーションは分子の運動・構造変化等を計算で求める手法で、生命科学においては、タンパク質等の生命現象の基盤となる生体分子を計算対象としています。シミュレーションでは原子間に働く力を計算し運動方程式を数値積分することで全ての原子の運動を計算します。運動方程式の数値積分と古典力場で近似した共有結合による力の計算は原子の個数に比例した計算量ですが、ファンデルワールス力（分子間力）とクーロン力は原子の組み合わせの個数に比例、つまり、原子の個数の二乗に比例した計算量が必要です。ファンデルワールス力は1.4nm以遠では非常に小さいので距離の離れた原子のペアの計算は省略するカットオフ法を使いますが、クーロン力は距離の二乗に反比例するので単純に無視することはできません。このため、FFTを使うことで遠距離クーロン力の計算量を$O(N \log N)$に抑えるPME（Particle Mesh Ewald）法が主流の計算手法となっています。しかし、FFTは並列化した場合の通信が多く大規模並列では効率が低下するため、大規模並列ではFMM（高速多重極展開法）等のFFTよりも通信の少ない手法の採用が進みつつあります。また、PME法は周期境界条件で厳密に正しい手法で、現実の細胞内のような周期性のない対象では擬似的な周期性が問題になる場合もあります。そのため、電荷・双極子モーメントが中性等の条件下で補正したカットオフ法が使われることもあります[1,2]。カットオフ法は通信はほぼ近接ノード間なので大規模並列にも適しています。

いずれの手法においても、計算時間の多くを何らかの形のカットオフ法計算が占めているので、本研究ではカットオフ法の「京」での最適化を重点的に行なってきました。最もよい条件ではカットオフ法計算部分の単体CPUでの計算効率は60%に達しました。2012年10月に行なった「京」全ノード評価(注1)では、5.4億原子・カットオフ距離2.8nm・ポテンシャル計算4ステップに1回の計算で実効計算速度4PFlops・効率38%を達成しました。ポテンシャルを毎回計算する場合では、79872ノードでの5.2億原子の計算で実効計算速度4.4PFlops・効率43%になりました。

並列性能をまとめたのが図1です。ノードあたりの計算規模（原子数）を一定にした場合の並列性（weak scaling）が緑の破線で、上から順にノードあたりの原子数が52338、6542、818、102の場合です。いずれの場合もほとんど一定の計算時間になっており、高い並列性が達成できています。全体の計算規模を固定した場合の並列性（strong scaling）が赤い線で、左から順に418707、3349656、7536726、26797248、214377984原子の場合です。並列性はノードあたりの原子数に依存しており、6000原子以上ではほぼノード数比例で性能が向上しましたが、ノード数を8、64倍にしノードあたり800、100原子になると、並列効率が50%、20%に低下しており、ノードあたり100原子程度で1ステップの計算時間10ミリ秒程度が実用上の限界だと考えられます。FMMによる遠距離クーロン力計算も実装中の試験評価ですが、カットオフ計算と比べて並列性の低下は少なく、「京」全ノード規模でも効率の低下は少ないことを確認しました。

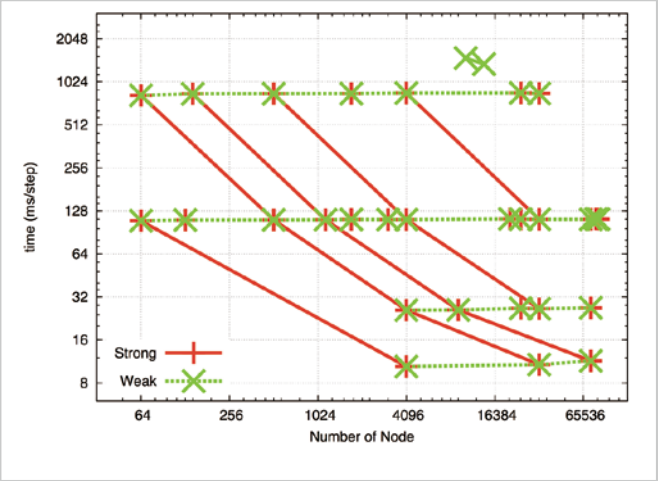


図1：カットオフ法の並列性能。カットオフ距離2.8nm。横軸はノード数、縦軸は1ステップの計算時間(ミリ秒)。緑破線がノードあたりの原子数固定(weak scaling)、赤線が全原子数固定(strong scaling)

本研究で得られた「京」での最適化の知見は他チームのアプリケーションの開発・最適化支援にもいかされています。また、このMDコアプログラムは、「京」での最適化の実例・参照コードとして、あるいは最適化済みの再利用可能MDコードとして、公開される予定です。

本研究は高度化チーム小山洋*、舛本現、長谷川亜樹、森本元太郎、沖本憲明、平野秀典との共同研究です。

* 現職 独立行政法人 物質・材料研究機構（NIMS）

表1：カットオフ法の「京」全ノード規模計算性能。カットオフ距離2.8nm

ノード形状	48x52x32	48x52x32	48x48x36*
ノード数	79872	79872	82944
原子数	522546336	522546336	542644272
1000ステップの計算時間(秒)	116.357	112.414	112.085
ポテンシャル計算頻度	毎回	4ステップに1回	4ステップに1回
実効計算速度(PFlops)	4.387	3.871	4.031
計算効率	0.429	0.379	0.38

表注*「京」のネットワーク形状48x54x32とは異なりますが今回は大きな影響はありませんでした。

注1
HPCIシステム利用研究課題として理化学研究所のスーパーコンピュータ「京」を利用（課題番号:hp120068）

【参考文献】

[1] Wolf, D. and Keblinski, P. and Phillpot, SR and Eggebrecht, J., J. Chem. Phys. 110 8254 (1999)
[2] Fukuda, I. and Yonezawa, Y. and Nakamura, H., J. Chem. Phys. 134, 164107 (2011)

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
課題3　予測医療に向けた階層統合シミュレーション

循環器系および筋骨格系・脳神経系における階層統合シミュレーションの実現をめざす

東京大学大学院工学系研究科機械工学専攻 教授
高木 周
（課題3 代表）

●トップダウン的アプローチで生命システムを明らかに

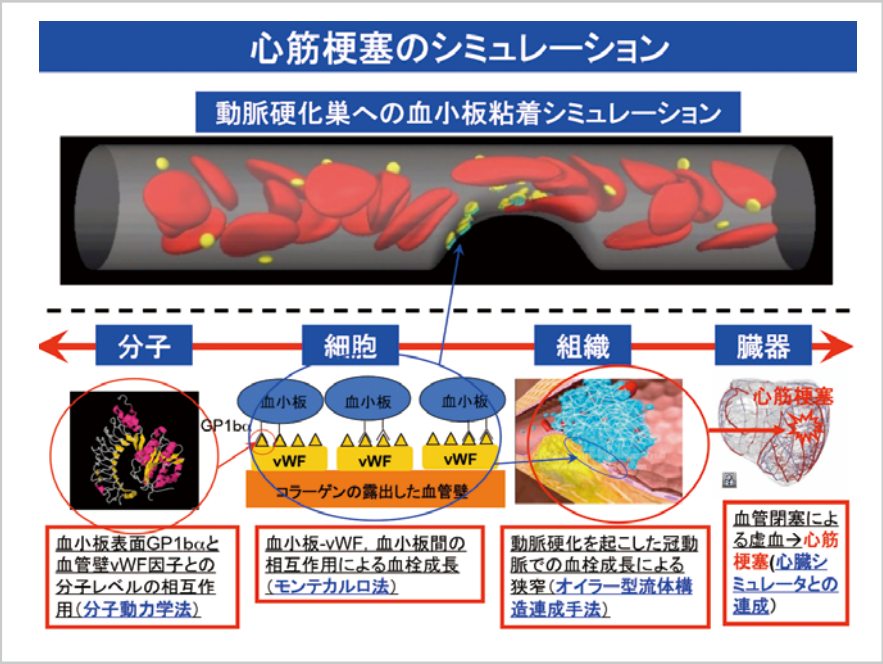
課題3の究極の目標は、「“生きている”とは何ぞや」を明らかにすることです。そのために、統合されたシステムとしての生命を理解することをめざして、階層性を持った人体をつくり上げていく階層統合シミュレーションに取り組みます。

私たちの生命が維持されているというのはどういう状態かというと、栄養となるものを口から取り込んで、その栄養分が血流を介して体全体に送られる。また一方で、空気を吸うことにより、肺で得た酸素がやはり血流を介して全身に送られ、体を構成する細胞は届けられた栄養と酸素を得ることによってそれぞれの機能を果たす。さらに、細胞が機能した結果、人体全体が機能し、やがて栄養が必要だという指令が脳に送られて、脳は空腹を感じ、私たちはまた食事をしたりするわけです。当然ながら生命の維持には、より細かいさまざまなプロセスがありますが、そこにも熱や物質や情報をいろいろなかたちで輸送するシステムが存在しています。そこを原点として、体のなかでいったい何が起きているのかを明らかにしていく、分かりやすくいえば、そうしたアプローチで生命システムを解き明かそうというのが課題3です。

例えば、薬が効いて病気が治ったとき、分子レベルで「薬がタンパク質のある機能を阻害し、細胞が正常に機能するようになって治りました」という説明は、間違っていません。しかし治らなかった場合、なぜ治らなかったのかは、分子レベルの局所的な話だけでは説明がつかないケースが数多くあります。実際に、分子レベルではうまくいくはずなのに、効かない薬もたくさんあります。病気の症状が消えて、脳が「治った」と感じるのは、タンパク質や細胞ひとつひとつの話ではなく、タンパク質の機能が変化して細胞が正常に働き、そうした細胞の集団である組織や臓器が正常に活動し、最終的には神経系を介して脳のなかで「治った」ことを感じるわけです。つまり、局所的なイベントだけでは決まらないことが多いのです。ですから、薬が効かない場合は、他のいろいろな因子が関わっていることを考えなければなりません。“生きている”ことを理解するためには、分子レベルから全身レベルまでのいろいろな要素を捉えていくことが必要なのです。とはいえ、分子から体全体まで、すべてを精緻にシミュレーションすることは難しく、どうしても限界があります。どのように階層を統合していけばいいのか、そのための方法論が重要です。

さまざまなイベントが、スケール的にも時間的にも複雑に相互作用している生命現象を理解するためのアプローチの方法には、ボトムアップ的に小さな要素から組み上げていくことで大きな現象の理解につなげるやり方と、トップダウン的にまず体全体が維持できていることから下りていき、小さな要素がどう機能しているのか、何が効いているのかを見ていくという2つがあります。私たち課題3は、後者のアプローチで、生命現象を明らかにしていこうとしています。

私たちの体には、わずかな擾乱が起きても元の正常な状態に戻そうとするホメオスタシス（恒常性維持）という機能が備わっていて、多少調子が悪くなっても、すぐに悪化するわけではありません。それがある一線を越えたときに、がらりと状態が変化して病態になります。何が効いて、どの段階で大きな変化が起きるのかを定量的に捉えるのは難しいかもしれませんが、どの因子がどの程度まで進んだら大きな変化が起きるのかを見つけ出すには、トップダウン的なアプローチの方がやりやすいと考えています。逆にボトムアップ的なアプローチで組み上げていくと、最後に残ってくる因子が何なのかが非常に見えにくくなります。階層統合は、ある意味では粗視化であり、何かを切り落とすことによって、異な



血栓症シミュレータと心臓シミュレータの統合により、心筋梗塞へと至るシナリオの再現をめざす。これまで開発を進めてきた、タンパク質分子レベルの相互作用から、流体力学レベルの血流現象までを取り入れたマルチスケール血栓シミュレータをさらに拡張し、心臓の拍動に連成して収縮・膨張する冠状動脈に対する血管閉塞を再現する。

る階層を見ていこうという試みになります。気を付けなければいけないのは、単純に平均化してしまうと、次の階層で現れる機能や個性が、かき消されてしまうということです。何を残せばいいのか、それは結局のところ、上の階層に現れる機能や個性で決まるわけです。下の階層の理論をきちんと踏まえた上で、心臓なら心臓、血流なら血流の機能を残すにはどこをどのようにつなげればよいのか――私たちが手がけてきた研究課題の多くが、まさにこの点を取り扱うことにあったといっても過言ではありません。

●グランドチャレンジの成果を活かしてさらなる発展を

5年間行ってきたグランドチャレンジプログラムにおいて、血栓シミュレータ、心臓シミュレータ、筋骨格系シミュレータ、脳神経系シミュレータなどの優れた性能を持つ各種生体シミュレータが開発されています。課題3では、その成果を積極的に利用し、「京」を使って、これまで個別に開発が進められてきたこれらのシミュレータを統合することにより、さまざまな疾患に関する複雑なプロセスを再現するための基盤ツールを整備し、将来的に階層性を持った人体全身に統合していくためのフレームづくりにも結び付けていきたいと考えています。さらに疾病に関するシミュレーション成果を、病態予測や治療支援に役立てることにも取り組んでいきます。

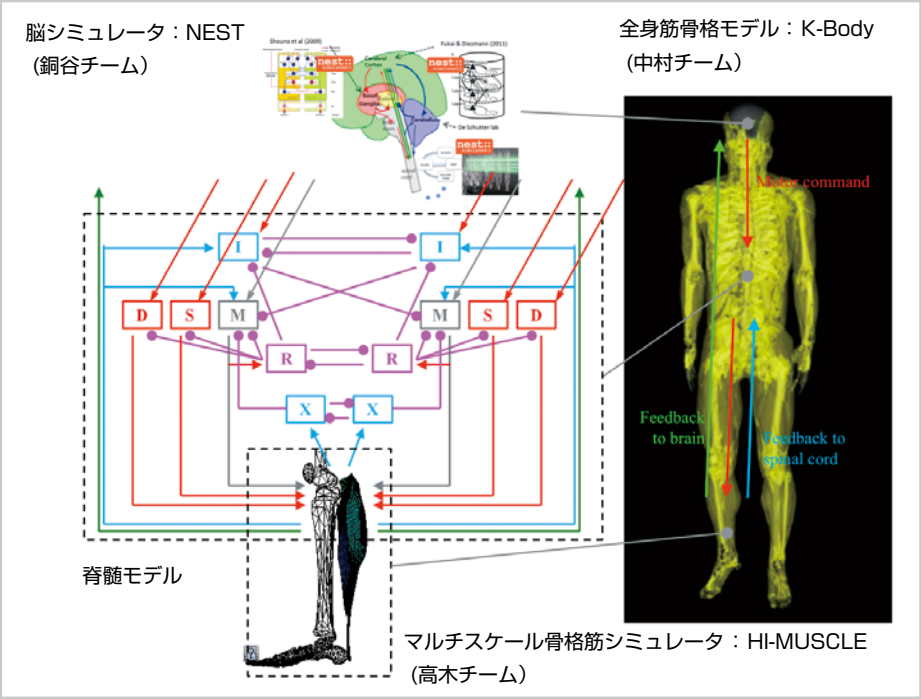
具体的なシミュレーションのターゲットに選んだのは、心筋梗塞とパーキンソン病という2つの疾患です。先に申し上げた通り、人体において循環器系と神経系は、生命維持のために熱や物質や情報を輸送するための不可欠な物質輸送網・情報伝達網であり、いわば人体の幹線網として機能しています。将来的に全身統合シミュレータを構築していくことを考えた場合も、循環器系と神経系は非常に重要です。こうしたことも考慮して、心筋梗塞とパーキンソン病をターゲットにしました。

心筋梗塞のシミュレーションでは、グランドチャレンジプログラムで開発した血栓シミュレータを拡張し、血栓形成の初期段階で血小板が粘着・凝集する一次血栓（血小板血栓）だけでなく、血液凝固機能の発動による二次血栓（フィブリン血栓）の成長から血管閉塞に至るまでのシナリオを再現することをめざしています。さらに、久田俊明（東京大学）らが開発している心臓シミュレータとの連成を行い、その効果を含めて、心臓を取り巻く冠状動脈における動脈硬化巣への血栓粘着から始まって血栓が成長し心筋梗塞に至る過程をシミュレーションしていきます。また、血栓に関連する薬の効果を評価することにより、新たな薬剤の開発に対して重要となる知見を与えることができればと思っています。

一方のパーキンソン病は、筋骨格系シミュレータと脳神経系シミュレータの統合パートとして、その再現をめざしています。パーキンソン病は、体（筋骨格系）が正常でも、脳からのシグナルの異常によって手足の震えや体の固縮が起きると考えられています。脳からのスパイクシグナルが、運動ニューロンを介して筋繊維に伝わり、震えが生じる振戦と呼ばれる病態や、筋肉がこわばり姿勢が固まる固縮と呼ばれる病態を再現することにより、脳神経疾患による運動機能障害のひとつであるパーキンソン病の病態予測や治療支援に役立てたいと思っています。

●全身統合シミュレータの構築を見据えて研究を推進

現在進めている心筋梗塞およびパーキンソン病のシミュレーションに加えて、今後は、神経系を介して筋骨格系と循環器系のシミュレータを統合することにより、幅広い疾患に展開できる統合シミュレータの基盤ソフトを構築することにも取り組んでいきたいと考えています。また、心臓部分については、心筋梗塞をはじめ、狭心症、拡張型心筋症、肥大型心筋症、微小血管障害による心機能異常、ストレスや興奮による影響まで含めて、さまざまな疾患のシミュレーションを行い、治療法の検討、薬効の評価も実施していきたいと考えています。また、パーキンソン病に関しては、神経細胞レベルからの物質輸送やシグナル伝達の影響を考慮した脳神経系を伴う全身統合シミュレーションを実施することにより、そのメカニズムの解明と治療法の検討を行ない、例えば、これまでの電気刺激による対症療法とは異なる、個々の患者の症状や運動テストのデータなどをもとにした投薬と電気刺激のコンビネーションなど、最適な処方を予測することにもつなげていきたいと思っています。さらに開発されたシミュレータを活用することにより、転倒時の体勢の変化や損傷の予測、筋肉・脳の疲労の影響の違いなど、広範囲の現象の解析が将来的には可能になると考えています。



脳神経系シミュレータ、筋骨格系シミュレータの統合によって、脳神経疾患による運動機能障害のひとつであるパーキンソン病を再現することをめざす。

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

課題4　大規模生命データ解析

最先端シークエンサーによる大規模データを「京」で解析し生命プログラムとその多様性の理解を進める

東京大学医科学研究所ヒトゲノム解析センター 教授
宮野 悟
（課題4 代表）

●大規模データ解析で生命システムを明らかに

課題4「大規模生命データ解析」は、「京」を中核とした高性能計算施設（HPCI）に最適化した最先端・大規模シークエンスデータの解析基盤を整備した上で、がんや細胞分化のプロセスに現れる生命プログラムをシステムとして捉え、その複雑性や多様性をゲノムによって理解する研究を進めるとともに、ゲノムを基軸とした生体分子ネットワーク解析研究を実施しています。そして、こうした研究によって、薬効・副作用の予測、毒性の原因の推定、オーダーメード投薬、予後予測などへの応用に貢献することをめざしています。

課題の名称にある「生命データ」は、いうまでもなくゲノム、エピゲノム（DNA修飾の総体）、トランスクリプトーム（転写産物の総体）など、ゲノムを基軸とした大量データをさしています。これらは、生命プログラムを理解するための重要なデータです。

その重要性を理解してもらうための一例として、私たちはなぜがんになるのかを考えてみましょう。親から受けた遺伝的要因、環境的要因によるDNAの修飾（エピゲノム）、腫瘍細胞に蓄積した遺伝子変異――これらが複雑に組み合わさって引き起こされる生命システムの何らかの異常、それががんです。システム異常が起こると、本来異常が起きないように自滅する仕組み（アポトーシス）が機能しなかったり、増殖を止める外部からの命令が機能せず、自分自身で増殖命令を出したりといったことが起きます。さらにがん細胞は、血管内皮細胞や免疫炎症細胞などの正常細胞とシェークハンドしながら薬剤耐性を獲得したり、浸潤と転移でどこへでも広がっていくなど、まさに時空間で進化する複雑で異常な細胞集団なのです。こうしたがんの悪性度や治療応答性、副作用の出やすさなどを規定しているのがゲノムであり、システム異常を明らかにするためには、ゲノム、エピゲノム、トランスクリプトーム、さらにはプロテオーム（タンパク質の総体）、メタボローム（代謝産物の総体）、インタラクトーム（相互作用の総体）などを見ていかなければなりません。しかし、それだけで治療法や病気の状態が理解できるというわけではありません。がんの多様性・複雑性やダイナミズムを解き明かすためには、数学とスーパーコンピュータを駆使した大規模データ解析と数理モデリングによって、システムを理解することが重要なのです。

●「京」による数理解析で生命システム理解にブレークスルーを

2003年、13年の歳月と1千億円という費用をつぎ込んだ「ヒトゲノム計画」によって、ヒトのゲノムが解読されました。この年に、米国国立衛生研究所（NIH）は、ヒトゲノム解読後のロードマップを出しましたが、そのなかに「Biology is changing fast into a Science of Information Management.」という一文があります。当時、多くの日本の研究者たちは、驚きとともに「言い過ぎでは」と思ったものですが、10年過ぎた今、誰もがそうなりつつあることを実感しています。さらに米国を中心に、2003年から個人の違いをDNAレベルで解明する国際ハプマップ計画が行われ、ヒトの病気や薬剤応答に関わる遺伝子を効率よく見つけるための基盤ができました。2008年からは主要ながんのゲノム異常カタログを作成する国際がんゲノム計画が始まり、2万5千人のがんサンプルと正常細胞、合わせて5万人分の全ゲノム情報が解析されています。一方で、2008年に大統領に就任したバラク・オバマ氏は、上院議員時代に「ゲノムと個別化医療法案」を提出し、2009年には「遺伝子差別禁止法」が成立するなど、米国のゲノム情報を基盤とする医療ヘルスケア戦略は、着々と進められてきたわけです。また、NIHは早くも2004年から「1千ドルゲノム計画」（1千ドルでヒト一人のゲノムを読めるように解析のコストダウンをめざす計画）に予算を付け始め、シークエンス技術の発展とその実用化が進み、今まさにバイオロジーは、大量のゲノムデータをどう解析し、解釈していくかというサイエンスになろうとしています。

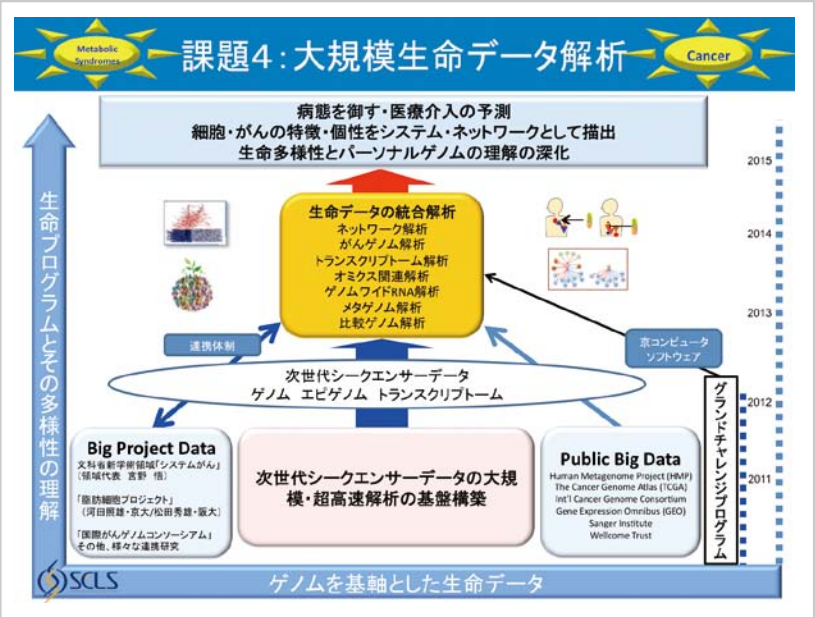
シークエンス技術については、今日その高度化が急速に進んでいます。これまでの光シークエンス技術を用いた次世代シークエンサーは、装置そのものが高価である上、非常に高価な蛍光試薬が必要であることやリード長も限られているといった課題がありました。しかし現在、コストの安いシリコンチップを使う半導体シークエンサーのような最先端シークエンサーの開発が進み、2013年には「1千ドル（10万円）、数時間」で個人のDNAシークエンス情報が得られる装置が実用化を迎え、2、3年後くらいには、安定的に利用できるようになる見込みです。さらに「100ドル、1時間以内」というシークエンサーの実現も間近かといわれています。こうした超安価・高速・高精度のシークエンサーの普及に伴って、自分のDNAを調べる臨床シークエンスに基づいて病気の診断、治療の方針や薬の種類・量などを決める個別化医療は急速に広がり、誰もが自分のDNAシークエンスを持つことができる時代も、それほど遠くないと思っています。そして、パーソナルゲノム時代が本格化すれば、膨大なゲノムデータが生まれ、生命システムの理解にも新たなブレークスルーがもたらされることでしょう。

そのさきがけともいえる成果は、すでに生まれています。それが「骨髄異形成症候群（MDS）の原因遺伝子の発見」です。骨髄異形成症候群は、骨髄で正常な血液がつくれなくなることにより、急性骨髄性白血病へと移行していくケースが多い病気です。これまでその原因は不明でしたが、東京大学医学部附属病院キャンサーボードの小川誠司らの研究チームは、東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピュータをフルに用い、次世代シークエンサーを使って、検体のエクソン（RNAスプライシング反応で残る部位）の解析を徹底的に行うことにより、RNAスプライシングの経路の4つの遺伝子に現れる変異がその原因であることを明らかにしました。初めてこの病気の原因遺伝子を発見したこととともに、RNAスプライシングの異常ががんの発症に関わることを世界で初めて示した研究成果であり、がん研究の歴史に残る成果ともいわれています。これは、まさにシークエンス技術とスーパーコンピュータ、統計的数理解析チームが一緒になってなし遂げた成果のひとつといえます。現在、私たちが取り組んでいるHPCI戦略プログラム（戦略分野1）の「大規模生命データ解析」においても、こうしたブレークスルーの成功例を生み出していきたいと考えています。

●課題4が取り組む研究内容

課題4「大規模生命データ解析」の大きな目的は、生命プログラムとその多様性の理解です。この目的を達成するためにまず必要となるのが、次世代シークエンサーデータの大量高速処理です。そのために2011 ～ 12年度は、「京」に最適化した次世代シークエンサーデータの大規模・超高速解析の基盤構築に力を注いでいます。この解析システムでは、毎時1,000万リードの性能を達成することにより、大量のゲノム配列情報の高速解析を実現させ、より難度の高い、微小な類似性を感知できる世界最深度の高速配列探索を行い、世界に対して圧倒的な優位性を確保する体制づくりをめざしています。

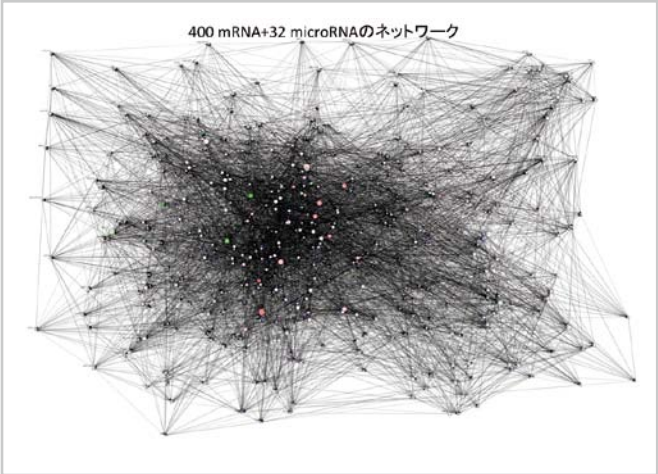
解析の基盤となるデータに関しては、パブリックな公開データベースに加えて、がんのTCGAプロジェクト、メタゲノムプロジェクト、さらに課題4の研究者が参加する国際がんゲノムコンソーシアム、新学術領域・システムがん、脂肪細胞プロジェクトなど、さまざまな研究プロジェクトと連携し、実験による検証もこうしたプロジェクトのなかで行うことになります。



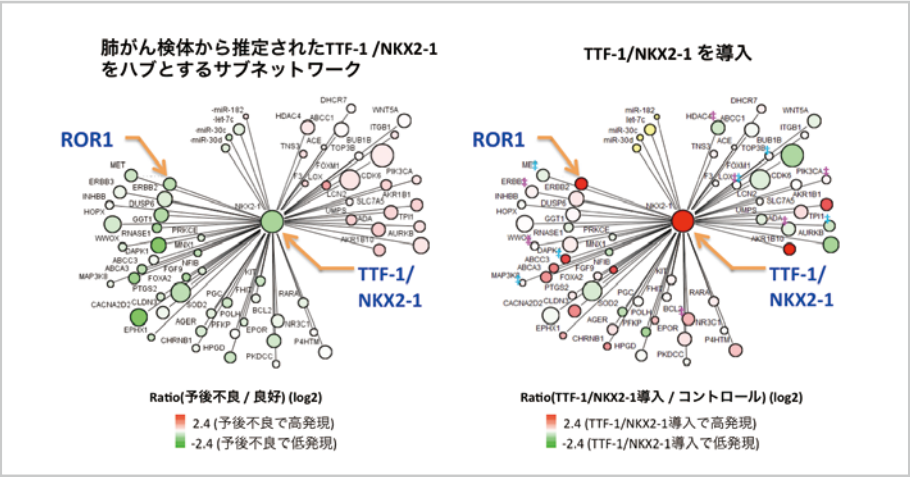
課題4の研究推進計画

具体的な研究内容は、「がんのリネッジと多様性の理解」、「パーソナルゲノム理解の深化」、「個別化医療介入予測」、「細胞分化ネットワーク」、「薬剤応答ネットワーク」、「パーソナルがんネットワーク」など幅広い研究内容が展開されますが、その到達点として3つを挙げることができます。1つ目は、「生命の多様性とパーソナルゲノムの理解の深化」です。なかでもメタゲノムとがんの多様性に重きを置き、さらにパーソナルゲノムの理解を進化させていくことで、生命の多様性の理解を深めていけるだろうと考えています。2つ目は、「細胞やがんの特徴・個性をシステムやネットワークとして抽出」していくことです。そして、3つ目は、医療介入の予測につなげていくことによって「病態を御す」ということです。これらの3つをめざして、研究を進めていこうとしています。

ライフサイエンス分野におけるシミュレーション研究が盛んに行われていますが、現在は、まだ大腸菌ひとつとっても、完全にシミュレーションできるまで生命システムが明らかにされているとはいえません。ましてヒトの場合は、理解は進んでいるものの、数理モデルを構築してシミュレーションすることにより、新たなバイオロジカルな知見を導き出したり、誰にでも効く薬を開発するというのは、とても難しいのが実情ではないでしょうか。課題4では、まずデータからスタートして、データドリブンのサイエンスを行っていこうと考えています。データを統合的に解析し、そこから見えてくるものを明らかにすることによって、個人個人の病気の原因の糸口を探し出していくための手法を開発していく、それが私たちの取り組み方です。



肺線がんのマイクロRNA/mRNA遺伝子ネットワーク。遺伝子間の因果関係をネットワークでとらえた全体図。
（名古屋大学医学研究科・高橋　隆教授提供）



TTF-1/NKX2-1遺伝子を導入すると有意な関係を示すまわりのネットワークの発現が大きく変化し、肺線がんが生き残るかどうかのスイッチが変わったことが明らかになった。
（名古屋大学医学研究科・高橋　隆教授提供）

「4th Biosupercomputing Symposium」の開催報告

理化学研究所　次世代計算科学研究開発プログラム
田村 栄悦

「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(略称ISLiM)」では、2006年10月のプロジェクト開始以来スーパーコンピュータ「京」をターゲットにしたライフサイエンス分野のソフトウェアの研究開発―ライフサイエンスのグランドチャレンジ―への取り組みを、分子スケール、細胞スケール、臓器全身スケール、脳神経系、データ解析融合、および生命体基盤ソフトウェアに渡って包括的におこなってまいりましたが、2012年度末にプログラム終了を迎えます。

そこで、これまでの研究開発のトピックスを紹介すると共に、国際的な研究交流をさらに深めるため海外から7名のモデリングおよびシミュレーションで国際的に活躍している著名な研究者：

医薬分野 - Sandra R. B. Allerheiligen（Merck）
脳科学分野 - Sten Grillner（Karolinska Institutet）
細胞分野 - Ion I. Moraru（University of Connecticut Health Center）
医学分野 - Grace Peng（National Institute of Health）
臓器分野 - Aleksander Popel（Johns Hopkins University）
分子分野 - Karissa Sanbonmatsu（Los Alamos NationalLaboratory）
分子分野 - Ruhong Zhou（IBM T.J. Watson Research Center

を講演者にお招きし、茅幸二プログラム・ディレクターのオープニング・スピーチを皮切りに、キーノート講演者2名、招待講演者7名、ISLiM講演者14名による、「ISLiM 国際シンポジウム: 4th Biosupercomputing Symposium」が2012年12月3日～5日の三日間の日程で東京国際フォーラム ホールD7を会場に開催されました。

生命体の多岐にわたるマルチスケール・マルチフィジックス・シミュレーションを対象にした英語によるシンポジウムのため、参加へのハードルがやや高いかもしれないという懸念もありましたが、結果は145名(事前登録者数：134名)という大勢のみなさまの参加を得て好評裏に終えることができました。

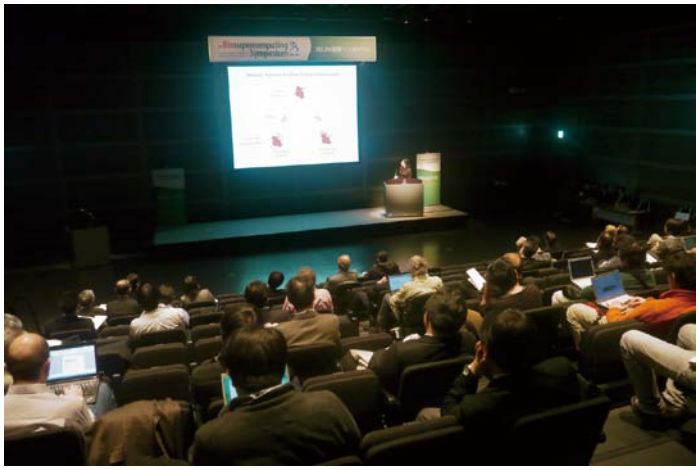
シンポジウム後、海外からの招待講演者からは、

- シンポジウムはたいへん広範囲な分野にわたり、私の専門性はその一部のみではあるが、非常に刺激的かつ多くのことを学べた楽しい経験だった
- 茅プログラム・ディレクターと姫野副プログラム・ディレクターの講演はISLiMプロジェクトの巨大な像を紹介するのにたいへん役立ったし、他のすべての講演もたいへんよく練られていた。欲を言えば、複雑なスライドよりは1スライドあたり一点にハイライトした簡潔なスライドにすると聴衆はもっと理解しやすいと思う
- ISLiMプロジェクトは明らかにサクセス・ストーリーである。ISLiMプロジェクトで研究開発されたソフトウェアとその経験を日本以外の国にも知らせることは極めて大事なことだと思う。ぜひNIH のIMAG Wiki で紹介して欲しいし、チームリーダーにもweb セミナーで発表していただきたい
(IMAG wiki: http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Japan)
- シンポジウムの進行はたいへんスムーズで、非常に温かい雰囲気の中で進められていた

といったシンポジウムを高く評価する多くのフィードバックをいただきました。

さらに、三日間の講演の合間、前後などには、招待講演者とISLiMチームのメンバーとのコミュニケーションも行われ、ISLiMの膨大な活動とその成果をこの分野の海外の有力研究者に示し、国際的な評価につなげるための良い機会ともなりました。

なおシンポジウムの講演のProceedings はWebからダウンロードできます（URL:http://www.csrp.riken.jp/4thbscs/program.html）



京互換機：SCLS計算機システムの導入

理化学研究所　HPCI計算生命科学推進プログラム チーム員
木戸 善之

背景と目的

HPCI戦略分野１では「京」を中心としたスパコンを利用した研究活動を行う一方で、研究成果の普及、人材育成、人的ネットワークの形成など研究推進体制の構築の役割を担っています。その一環として我々は「京互換機」によるSCLS計算機システムの導入を決定しました。図１はSCLS計算機システムの搬入の現場です。本システムの役割としては、

図1：SCLS計算機システム搬入現場

1．生命科学研究における「京」などスパコン利用の促進
2．戦略分野１、ISLiMプロジェクトにおける研究成果のテストベット環境の提供
3．「京」および「京互換機」における開発者、利用者コミュニティの形成

といった狙いがあり、多くの生命科学者らにSCLS計算機システムの利用、参加を広く呼びかけています[1]。また導入するSCLS計算機システムの仕様を表１にまとめ、比較のため「京」の仕様も参考資料として掲載しました。

生命科学研究における「京」などスパコン利用の促進

創薬の現場では、計算機を利用したバーチャルスクリーニングやドッキングシミュレーションによる候補となる化合物の絞り込みを行っており、効率化を図っています。さらに分子間の結合自由エネルギーを計算で求め、精密な分子動力学のシミュレーションを行う手法も現実的になりつつあります。より精密な分子動力学シミュレーションでは、因子が増加すると急激に計算量が増加し、スパコンでなければ実質不可能な計算時間となります。しかしスパコンは高額なため、中小規模の企業や研究所では導入することは困難です。そこで我々は生命科学者らに対し小規模でかつ気軽にスパコン利用できる環境としてのSCLS計算機システムを整備しました。SCLS計算機システムでは利用に際し公募制を敷いており、公募に通った課題は無料でシステムを利用することができます。SCLS計算機システムは小規模なノード数ですが「京」と互換性を持ったシステムであり、SCLS計算機システムで得られたノウハウが、「京」を中心としたHPCI利用課題申請への切掛になると期待しています。

研究成果のテストベット環境の提供

戦略分野1およびISLiMプロジェクトでは「京」など超並列な計算機で動作するアプリケーションの開発も行なっています。研究成果の普及という意味で、アプリケーションの利用を呼びかけるとともに、SCLS計算機システムは、そのアプリケーションらのテストベット環境としても整備します。ISLiMプロジェクトが開発するアプリケーションは多岐に渡り、６のカテゴリ、計31のアプリケーションが開発されています[2]。それらをSCLS計算機システムにインストールし、ユーザらが利用し評価できる環境を整えます。

表1：SCLS計算機システムと「京」仕様対比

		SCLS計算機システム	京
CPU	名　前	SPARC64™ IXfx	SPARC64™ VIIIfx
	理論性能	211GFLOPS(1.65GHz)	128 GFLOPS（2GHz）
	コア数	16	8
システム全体	ノード数	48	88,128
	理論性能	10.1 TFLOPS	11.28 PFLOPS
	ノード当たりのメモリ	32GB（システム全体1.5TB）	16GB（システム全体1.5PB）

開発者、利用者コミュニティの形成

「京」および「京互換機」で動作するアプリケーションの多くは、大規模なプロジェクトで開発されており、プロジェクトが終了したのちはメンテナンスを行うことが困難になります。そうした状況は大きな損失であり、「京」および「京互換機」で動作するアプリケーションをメンテナンスし続けていくための開発者コミュニティ育成を、我々は目指しています。SCLS計算機システムでは計算機資源を提供すると同時に、開発者コミュニティの闊達な議論の場を提供します。具体的には開発者らが自由に読み書きできるWebベースシステム（図２参照）とメーリングリストを提供します。編集者にはシステム運営側の技術者も含まれており、システムに関するトラブルシューティングや、並列化手法の開発におけるノウハウなど、様々な情報を包含するWebベースシステムとして運営する予定です。

まとめと今後の課題

生命科学、医療、創薬はデータ量や計算量など、人間では処理しきれない問題を抱えています。しかしそれは計算機の進化によって解決でき、言い換えれば生命科学には計算機は必要不可欠な状況にある、と言えます。「京」は2011年で世界一であったスパコンであり、現在でも世界トップクラスの計算機です[3]。しかし「京」を利用するにはアプリケーションのスケーラビリティなどのスパコン利用実績を示す必要があり、スパコンの知識が必要不可欠になります。そうした計算機の専門性を持たない生命科学者らが「京」のような大規模な計算機を利用するにはあまりにも敷居が高いので、SCLS計算機システムのような小規模だが気軽に扱えるスパコンが生命科学者らに必要であると我々は考えています。そしてSCLS計算機システムがスパコン利用の導入、切掛になるものであると我々は期待しております。

【参考文献】
[1] HPCI戦略分野１, http://www.kobe.riken.jp/stpr1-life/
[2] 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
http://www.csrp.riken.jp
[3] Top 500, http://www.top500.org

図２：SCLS計算機システムの試験運用中のポータルサイト

イベント情報＆ニュース

●公開シンポジウムのお知らせ

文部科学省「革新的ハイパフォーマンス・コンピューティング・インフラ（HPCI）の構築」
HPCI戦略プログラム
「グランドチャレンジ・アプリケーションの研究開発」公開シンポジウム

- 開催日：３月11日（月）10：00～18：20（終了後、懇親会）
- 場　所：東京大学山上会館（東京都文京区）
- 参加費：無料（懇親会は有料）

詳しくはWebページをご覧ください（http://www.csrp.riken.jp/）

● SCLSでは、ツイッターやFacebookでイベントや研究成果などのさまざまな情報をいち早く発信中です。ツイッターは“@HPCI_Senryaku1”、Facebookは“HPCI戦略分野１”でご検索ください。フォローやいいね！コメントをお待ちしております。

● SCLS計算機システム（京コンピュータ互換スーパーコンピュータシステム）利用の公募情報をSCLSホームページにて公開中です。 http://www.kobe.riken.jp/stpr1-life/
2013年度の公募は年3回（4月，7月，10月）を予定しています。皆さまのご応募をお待ちしております。

● 産業技術総合研究所　生命情報工学研究センター（CBRC）では、「HPCI戦略プログラム　分野１における人材養成プログラム」を実施しています。広く一般を対象にしたセミナー、ワークショップ、チュートリアル、e－ラーニングを実施し、計算科学技術と生命科学をより高度に融合させ、バイオインフォマティクス入門者からスパコン「京」を活用する人材の育成まで、幅広い活動を実施しています。平成25年度の予定は決まり次第、ホームページに掲載します。
http://hpci.cbrc.jp

文部科学省委託事業

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

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
Next-Generation Integrated Simulation of Living Matter

「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」は、文部科学省の委託事業で、スーパーコンピュータ「京」の性能を最大限活用するペタスケールのシミュレーションを実現するため、分子から全身まで生体内で起こる様々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を進めています。

文部科学省高性能汎用計算機高度利用事業

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
Supercomputational Life Science

HPCI戦略プログラムは、スーパーコンピュータ「京」を中心としたHPCI（High Performance Computing Infrastructure）を最大限に活用することによって、戦略的に取り組むべき5つの研究分野において画期的な成果を産み出し、計算科学技術の飛躍的な発展を目指す文部科学省のプログラムです。
「予測する生命科学・医療および創薬基盤」は、理化学研究所を代表機関として、大規模シミュレーション・高度なデータ解析に基づく生命現象の理解と予測、およびそれを通じた薬剤・医療のデザインの実現を目指して研究を実施しています。

BioSupercomputing Newsletter Vol.8 2013.3
発行：平成25年3月

独立行政法人理化学研究所
社会知創成事業
次世代計算科学研究開発プログラム
〒351-0198 埼玉県和光市広沢2-1
TEL: 048-462-1488　FAX: 048-462-1220
http://www.csrp.riken.jp

HPCI計算生命科学推進プログラム
〒650-0047 兵庫県神戸市中央区港島南町7-1-26
理化学研究所計算科学研究機構研究棟3階R301
TEL: 078-940-5692　FAX: 078-304-8785
http://www.kobe.riken.jp/stpr1-life

RIKEN 2013-027