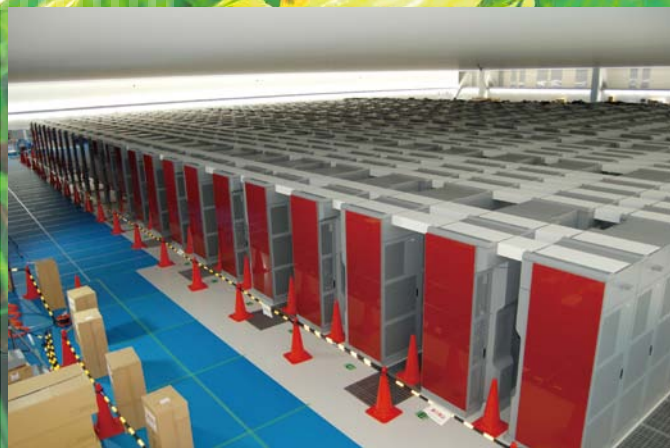


BioSupercomputing Newsletter

2011.11

Vol.5

2011年6月 京速コンピュータ「京」、演算速度世界一位を獲得



2011年6月の「京」の整備状況

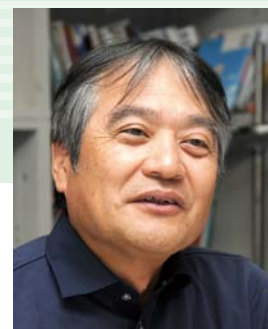


TOP500ランキング1位の証明書

CONTENTS

● SPECIAL INTERVIEW	○「予測する生物学」をめざすバイオスーパーコンピューティングの挑戦はいよいよ世界一の「京」で成果を出すフェーズに入った 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	2-3
	○高性能計算機資源および開発アプリケーションの産業利用促進を図るために何をすべきか 計算科学振興財団 チーフコーディネーター 福田 正大 都市活力研究所 主席研究員 バイオグリッドセンター 関西 理事・事務局長 志水 隆一	4-5
● 研究報告	○QM/MM 自由エネルギー法による酵素反応分子機構の解析 京都大学大学院理学研究科 林 重彦 (分子スケールWG)	6
	○アクチン細胞骨格の計算メカノバイオロジー 京都大学再生医科学研究所 井上 康博 (細胞スケールWG)	7
	○血栓シミュレーションに向けた血流解析手法の開発 東京大学工学系研究科 伊井 仁志 (臓器全身スケールWG)	8
	○生命体シミュレーションのためのデータ同化技術の開発 統計数理研究所 樋口 知之 (データ解析融合WG)	9
	● SPECIAL INTERVIEW	○複雑な生命現象の理解と予測に向けて計算生命科学の明日を拓く 理化学研究所 HPCI計算生命科学推進プログラム プログラムディレクター 柳田 敏雄 理化学研究所 HPCI計算生命科学推進プログラム 副プログラムディレクター 木寺 詔紀 理化学研究所 HPCI計算生命科学推進プログラム 副プログラムディレクター 江口 至洋
● 研究報告	○創薬応用シミュレーション 東京大学先端科学技術研究センター 藤谷 秀章 (分野1-課題2)	12
	○次世代DNAシーケンサデータの超高速解析 東京工業大学大学院情報理工学系研究科 秋山 泰 / 石田 貴士 / 角田 将典 / 鈴木 脩司 (分野1-課題4)	13
● 報告	○バイオスーパーコンピューティングサマースクール2011 理化学研究所 次世代計算科学研究開発プログラム 石峯 康浩 (臓器全身スケールWG) 統計数理研究所 データ同化研究開発センター 斎藤 正也 (データ解析融合WG) 新潟国際情報大学 近山 英輔 (細胞スケールWG) 東海大学医学部内科学系循環器内科 七澤 洋平 (細胞スケール/臓器全身スケールWG)	14
	理化学研究所 次世代計算科学研究開発プログラム 半田 高史 (脳神経系WG) 理化学研究所 次世代計算科学研究開発プログラム 舩本 現 (開発・高度化T) 理化学研究所 次世代計算科学研究開発プログラム 森次 圭 (分子スケールWG)	
	○「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 (ISLiM)」 開発アプリケーション紹介ページ、オープン 次世代計算科学研究開発プログラム 次世代生命体統合シミュレーション研究推進グループ	15
	● イベント情報	

“予測する生物学”をめざすバイオスーパーコンピューティングの挑戦はいよいよ世界の「京」で成果を出すフェーズに入った



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

● 開発ソフトウェアの3分の1が「京」環境でのテストを開始

非常に複雑な“超多体系多階層問題”ともいべき生命現象を、ベタスケールという桁違いの性能を有するスーパーコンピュータ「京」をフル活用することによって理解し、創薬や医療に貢献するためのソフトウェアを開発する——こうした目的で2006年に動き出した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）も5年が経過しようとしており、ソフトウェアの開発は、いよいよ「京」を活用しながらの超大規模並列化と実機チューニングを進める、新たなフェーズを迎えています。

整備途中ながら、「京」は、今年（2011年）6月に発表された第37回TOP500において、LINPACKベンチマークで8.162PFLOPSを達成（実行効率93.0%）し、第1位を獲得しました。そして、この「京」で実行することを目標に、現在31本のソフトウェア開発が進行中です。すでに、そのうちの11本が大規模並列化を進める開発フェーズ（8,192並列）をクリアし、「京」環境でのテストを行っています。その1つである「大規模並列用MDコアプログラム（cppmd）」は、3月末の時点で実効効率30%を超える性能を発揮して、1.3PFLOPSを達成しました。これらのソフトウェア開発には、「京」の性能をフルに引き出すことが求められており、私たちとしても、マシン全体（10PFLOPS）を使ったときに、少なくとも1PFLOPSを超えるソフトウェアをできる限り多く揃えたいと考えていますが、とりあえず1本は目標となる性能を達成したことになります。cppmdを含めて現在（2011年10月）までに4本（cppmd、ZZ-EFSI、CafeMol、ZZ-HIFU）が実効効率20%を超え、さらにそのうちの2本（cppmd、ZZ-EFSI）は約40%に達しています。「京」

が使えるようになったのが今年度に入ってからですから、それを考えるとかかなり早いペースで性能チューニングが進み、ある程度満足ゆく性能が出ていると言えるでしょう。ただ残念なことに、いくつかのソフトウェアでは、まだ並列性能や単体性能がよくなかったり、Platypus-MM/CGのように並列性能は出ているのに単体性能があまりよくないといったものもあり、さらなる改良を進めているところです。もちろん、「京」そのものも未だ整備途中のマシンですから、今後、コンパイラや通信用のライブラリなどが充実していくことによってソフトウェアの性能も向上していくと思いますが、それを待っている余裕はありません。私たちが試験利用を通じて「京」を優先的に使えるのは2012年10月末までです。それまでに、まとまった計算を行って、サイエンティフィックに意味のあるしっかりとした結果を出さなければいけません。そのためにはコードの最適化など、今できることをすべて行いながら、なるべく短い時間で最大限の性能を出すことが求められています。

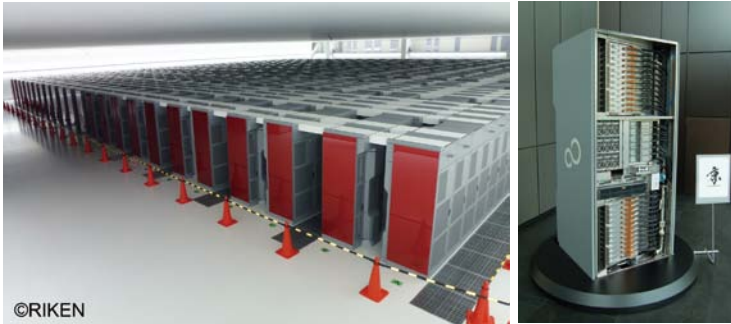
もちろん私たちとしては、開発中の31本のソフトウェアすべてを「京」で走らせて、結果を出す計画です。しかしながら期間は限られていますので、計算規模や時間をどのように割り振っていくかも、これから具体的に決めていかなければいけません。システム全体を使って計算できるものは10本以下、いやもっと少ないかもしれません。さらに、計算規模を縮小して長い時間をかけるなど、問題設定の検討も必要です。限られた時間内に計算資源を効率よく活用して最大の成果を出すにはどうすればよいのか、それも今後の大きな課題の1つだと思っています。

● ライフサイエンスの新たな時代を拓くために種をまいたグランドチャレンジ

グランドチャレンジと並行して、今年度から「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」」が始まりました。グランドチャレンジがめざしてきたのは、基本的にソフトウェアの開発です。それもニーズ志向ではなく、シーズ志向、つまりサイエンスオリエンテッドで開発に取り組んできました。ライフサイエンス全体をコンピュータシミュレーションやデータ解析の方向へ導いていきたいという思いで、遺伝子、生体分子から臓器、全身まで、いろいろなスケールの幅広い研究を含めたソフトウェアのパッケージをつくってきました。さらにその結果を出し、それらを公開することによって、この分野の敷居を低くし、新たに研究したいと思っている人たちに私たちの成果を役立ててもらおうことが狙いです。ですから、必ずしも製薬・医療分野が望んでいるソフトウェアばかりではありません。一方、HPCI戦略プログラムは、「京」などの高度計算機資源を最大限に活用して、社会的に意味のある成果を生み出していくことをめざしています。ですから、同じ分子動力学の計算ソフトでも、グランドチャレンジにおいては「京」に向けて開発することが目的であり、戦略プログラムではそれを利用して創薬につなげていくことが期待されているのです。同じものを扱っていても、「開発」と「利用」というように、フェーズは全く違います。私たちは、いってみれば種をまいてきたわけで、きれいな花が咲くように全力で努力してきました。さらにそこから実を上手に育てて刈り取るのが、戦略プロ

グラムの仕事といえるのではないのでしょうか。

私たちは、ある意味で社会のニーズとは無関係に設定したゴールに向かって走ってきました。しかし今後は、必ずしもそれだけではなく、実際の創薬に向けて活用されたり、医療機器の開発に役立てられたりという応用の道を開いていくために、新たな機能をつけ加えたり、使いたいと望む人たちと共同研究をしたり、講習会を開いたりすることが求められるでしょう。そうして、グランドチャレンジの成果が、戦略プログラムに引き継がれていくのだと思っています。産業利用、特に創薬の面では、すでに分子スケールの研究チームやデータ解析融合の研究チームの成果に大きな期待が寄せられています。すぐにでも活用できるソフトウェアとして、注目されているわけです。臓器・全身チームの取り組みも、少し時間がかかるかもしれませんが、創薬や医療に貢献するものとして期待されています。私としては、これらのソフトウェアをベースに新たな機能を付加したり、組み合わせたりすることで、さらに私たちが予想し得なかった新しい方向へ伸びてほしいと願っています。とはいえ、私たちがまずやらなければいけないことは、これまで開発してきたソフトウェアにさらに磨きをかけることです。そして、今までにない、世界が驚くような研究成果を出すこと、すなわち素晴らしいサイエンスの花を咲かせることであるの言うまでもありません。



計算機室に並び京速コンピュータ「京」(左)。1筐体に24枚のシステムボードが搭載されている(右)。



「京」の設置が進められている計算科学研究機構の建物。

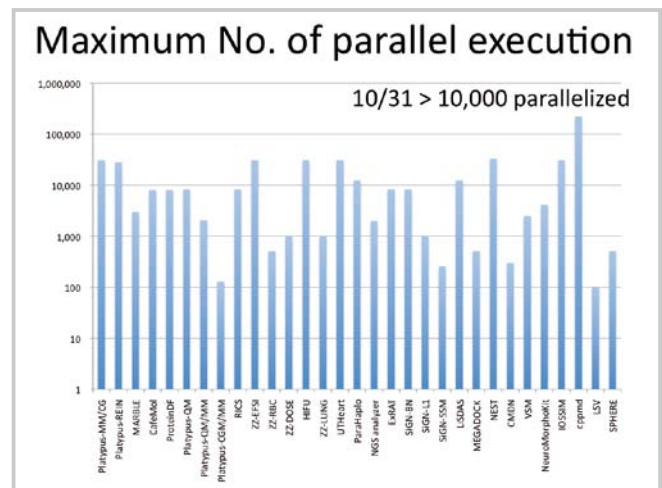
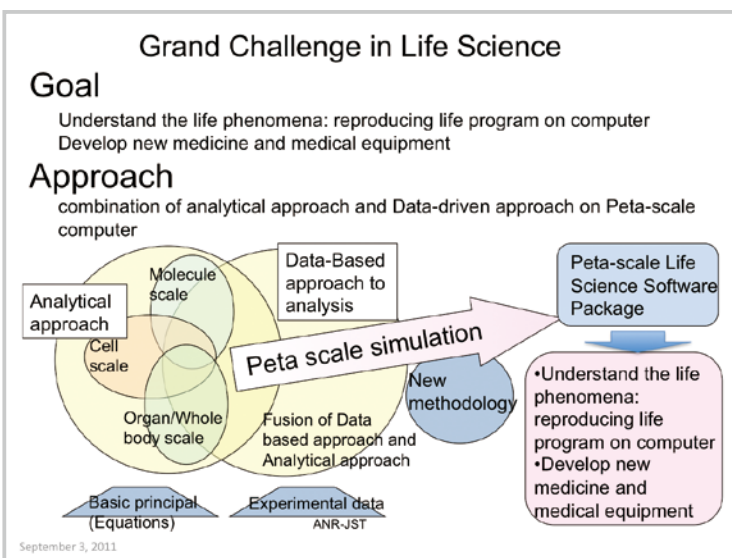
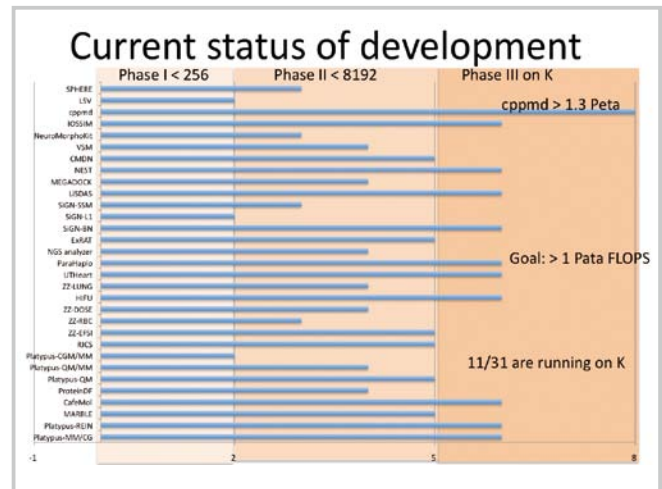
● 「京」が使えるようになってより深まった研究者の一体感

そもそもグランドチャレンジが計画されたころ、私たちのまわりの圧倒的多数は、ライフサイエンスの研究を加速したいのであれば、コンピュータシミュレーションにお金をかけるより、実験装置を充実させる方がよほど効果があるという考えでした。ライフサイエンスのソフトウェアで、スーパーコンピュータでちゃんと成果が出せるものが本当にできるのか、そんな疑問の声があがる状況のもとで、グランドチャレンジはスタートしたのでした。ところが、今の段階で、すでに開発中のソフトウェアの3分の1が1万並列を超えるレベルで「京」でのテストを開始し、実効効率10%を超える性能を出すものが次々に生まれています。もちろん、これは単にプログラムの性能の話であり、科学的に意味のある結果を出していくのはこれからですが、確かに言えることは、このプロジェクトを取り巻く人々の見方が、確実に変化してきたということです。コンピュータシミュレーションの可能性に期待する声は、どんどん大きくなっています。

同時にプロジェクトのなかにいる人たちも変わってきました。もともと非常に広い研究分野の人々の集まりだったこともあり、当初は自分の研究分野にしか関心がなく、共通の言語がない状態でスタートしました。同じテーブルにつくようになって、ようやくお互いの顔も分かるようになり少しずつ理解が進むようになりましたが、劇的な変化が生まれたのは、「京」が使えるようになってからでした。「京」で性能を出すためにはどうしたらいいのか、これはプロジェクトに関わるすべての研究者の共通の関心事です。そして、共通体験、共有言語を通して、「京」を中心に1つのプロジェクトを推進しているという一体感が研究者たちのなかに生まれ、お互いの絆がますます深まっていくことを実感しています。その一例が、最近できた「京」を使う若手研究者たちだけのメーリング

リストです。そこには、「『京』がハードウェアで世界一になったのだから、次は私たちが成果を出す番だ」、「私たちが頑張らなければいけない」といった熱意あふれる意見なども書き込まれていて、目的意識の高さと、一緒になって1つのプロジェクトを推進しているという実感をみんなが持っていることを、あらためて知ることができました。

「京」が動き出すと同時に結果を出す、それが私たちに与えられた使命であり、そのために残された時間はあと1年です。見事な花を咲かせ、次に続く「利用」というフェーズにつなげていくためにも、ラストスパートを迎える最後の1年は、最も重要な年になるはず。そんな来年に向けて、今いちばんの懸念は、実は来年度の予算です。これをしっかりと確保すること、これもまた私たちの大きな課題になっています。



京速コンピュータ「京」の性能をフルに活用して、生命現象の理解と医療に貢献するためのソフトウェア開発をめざす。

開発が進む31本のソフトウェアのうち、すでに11本が予定の大規模並列化を達成し、「京」を使用したテストに入っている(上)。1万並列以上を達成したのも10本を数える(下)。(2011年10月現在)

高性能計算機資源および開発アプリケーションの産業利用促進を図るために何をすべきか

財団法人 計算科学振興財団 チーフコーディネーター
福田 正大 (写真左)

公益財団法人 都市活力研究所 主席研究員
 特定非営利活動法人 バイオグリッドセンター関西 理事・事務局長
志水 隆一 (写真右)



● 民間企業にコンピュータシミュレーションが普及しにくい理由

— お二人は、HPCIコンソーシアムの産業利用促進検討ワーキンググループのメンバーでもあります。最初に、お二人が現在どのような活動をなさっておられるかをお話してください。ちなみに福田さんは、かつて航空宇宙技術研究所で世界最速を達成したスーパーコンピュータ「数値風洞 (NWT)」の開発・運用に携わり、その後も一貫して計算科学の世界を歩み、その発展に貢献してこられました。

福田 (敬称略) 現在は、計算科学やシミュレーション技術の普及によって日本の産業を活性化させていこうと設立された計算科学振興財団で、いろいろな企業のみなさんにスーパーコンピュータを活用した研究・開発のお手伝いさせていただいています。財団にも産業界向けの計算機「FOCUSスパコン」があり、これを活用して企業の技術高度化支援なども行っておりまして、みなさんに大いに活用していただきたいと思っています。

志水 私が関わっていますバイオグリッドセンター関西の話をごささせていただきます。もともとは大阪大学が推進してきた「バイオグリッドプロジェクト」がありまして、これはグリッドコンピューティング環境でバイオ・医療向けの開発アプリケーションを動かしていこうというのですが、その技術や研究成果を産業界など広く世の中で活用してもらうことをめざして設立されたのが、NPO法人のバイオグリッドセンター関西です。2004年にテストベッドを開始して、「ぜひ使ってください」と呼びかけたのですが、当初、製薬会社は全く動いてくれませんでした。大阪大学のスーパーコンピュータを確保し、アプリケーションを移植して動くようにしていたのですが、「データを出せない」、「自社から外部へネットワークをつなぐことなど許されない」と、どこにも使ってもらえませんでした。ならば、自分たちで成果を出してみようということで、ちょうど動き出そうとしていた阪大発ベンチャーと手を組んで知的クラスター創成事業を、翌年には医薬基盤研究所からも「インシリコ創薬をやりたい」という話がありまして、コンピュータシミュレーションに基づいて医薬品候補化合物を創製する「創薬バリューチェーンプロジェクト」が始まりました。時期的にも、ちょうどインシリコ創薬が流行になりつつあったところで、それも幸いだったと言えるかもしれません。

福田 バイオグリッドセンター関西で使えるようにしたアプリケーションは、大阪大学の先生が開発されたものですか。

志水 いろいろ使いますが、基本的には大阪大学蛋白質研究所の中村春木教授や産業技術総合研究所の福西快文氏らがつくられたドッキングシミュレーションなど (myPresto) です。

福田 プロジェクトが始まってから、実施の計算は製薬会社の人たちがやったのですか。

志水 企業に外注しました。そもそも製薬会社は入っていないんです。まずは製薬会社の色をつけずに開発しようということでした。

福田 そうですか。なぜそんなことを聞いたかということ、アプリケーション

を開発しても、民間企業の人たちにはなかなか使いこなせない、使うにはサポーターがついて手とり足とり指導しないといけない。だから民間企業に普及しないんだという話を、よく聞くものですから。

志水 製薬業界について言うと、いくつかレベルがあって、日本の大手10社くらいは計算科学をやっている人がいて、自分たちで使える力を持っておられます。中身をしっかりとっておきたいということでアプリケーションの説明は求められますが、後は簡単な手ほどきだけで十分。さらに私たちと一緒に、大学のスーパーコンピュータをどんどん使って計算するところまでやっているのは、そのうちの3社くらいではないでしょうか。

福田 すでに数社はそこまでやっておられるんですね。

志水 どのくらいの計算資源をお持ちかは分かりませんが、大手企業は自社で化合物ライブラリーを持っていて、自分たちで反応試験をやり、化合物を見つけられる体制が整っていると思います。ただ、「一から教えてほしい」という企業もたくさんあります。また、「自分たちも、今後は大手のようにスーパーコンピュータを使った創薬に取り組んでいきたいので、勉強させてほしい」と私たちの研究プロジェクトにオブザーバーとして参加している企業もあります。もちろん、秘密保持契約を結んだ上での話ですが。

福田 中堅レベルの製薬会社のなかにも、積極的に取り組んでいこうという意欲をお持ちのところがあるということですね。

志水 ええ。ただ、製薬業界全体に広がっているかということ、どうでしょうか。まだメインストリームにはなっていないくて、実験で行き詰ったときに、ちょっとシミュレーションでやってみようというところが多いように感じます。

福田 まずインシリコチームがシミュレーションでやってみて、仕上げに実験をやるという流れにはなっていないわけですね。

志水 時間的な問題もあるようです。計算結果を半月待つなら、実際にやってしまった方が早いと……。そうしたタイムラグの問題などもあって、合成チームとインシリコチームがうまくコラボレーションできる体制になっていないのではないかと思います。あとは、これまで自社の計算機リソースでできることしかやってこなかったために、「大規模な計算なんてできない」と、最初から諦めているところもあると思います。それこそ「京」のような桁外れのリソースを使ってがんシミュレーションをやっていくということが、もともとの発想として出てこない……。

● シミュレーションへの製薬会社の認識も変化しつつある

— 先ほどお話しにあった、社外にデータを持ち出せない、ネットワークにつなげないといった状況は、今も変わっていないのでしょうか。

志水 少しずつ意識は変わっていると思います。自社の研究員に計算ばかりやらせてられない、そんな余裕はないといった事情もあると思いますが、最近は組織同士の秘密保持契約さえきちんできるとなれば、作業的な計算は外注に出してもよいという動きが広がっているように感じます。

福田 重要な産業に対しては、計算を引き受けるしっかりとした受け皿を用意すべきだと、私は思いますね。例えば、製薬業界の計算は理化学研究所が引き受けるといった体制を、国策として整備したらいい。そうすれば、責任の所在もはっきりするし、コストパフォーマンス面はもちろん、いろいろなノウハウも集まるはず。うちの財団のFOCUSスパコンも、まさに産業界で使っていたことをコンセプトに導入したわけ。25TFLOPSほどだからそれほどのパワーではありませんが、多くの企業を回ってお話をさせていただいて、つくづく分かったのは、十分にニーズがあるということです。私たちのところはCAE（Computer Aided Engineering）系が多くて、製薬会社はまだあまり回っていませんが、志水さんのお話をうかがっていると、ニーズは十分あると思いますね。

志水 製薬業界は、今ちょうど過渡期にあつて、これまで頑なだったものが、外のリソースを使ってもいいのではないかという流れに向かいつつあります。一方で、世界一の「京」が産業利用に力を入れるということで、これは非常にいいタイミングだと思います。あとは成功事例といいますが、成果が目に見えるかたちで出てくると、「やってみようか」と乗り出してくる製薬企業も出てくるのではないのでしょうか。

福田 その意味でも、理化学研究所がこれまで取り組んできたグラウンドチャレンジの成果は、非常に重要になってきますね。

志水 海外と比較しても、日本ほどバイオシミュレーションのソフト開発に力を入れている国はないと思います。日本が世界のトップを走っていると言ってもよいのではないですか。ここまで日本の英知を集結しているのですから、これを使わない手はないですよ、もったいない。そういうことを、もっとみんなで宣伝していかないといいけませんね。

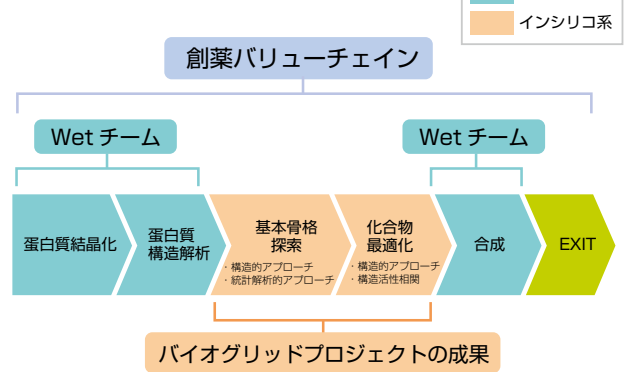
● 種をまいておかなければ、上から水をかけても芽は出ない

— なかなか新しい薬が生まれにくい時代ですが、製薬業界において、候補化合物を見つける手法としてコンピュータシミュレーションの評価、期待度はどうなんでしょうか。

志水 それほど詳しいわけではありませんが、コンピュータを使ってみようかという意識は育ちつつあると思います。ただ、そのためのやり方が分からないので、踏み込むのに躊躇しているというところはあるでしょうね。あとは、企業のトップに「そんなやり方で薬ができるのか」といった、ある種の不信感があるのではないのでしょうか。その意識を変えていくことも課題だと思います。

福田 私が航空宇宙技術研究所にいたころの話ですが、スーパーコンピュータを入れたときに、「これからの航空機開発にはシミュレーションをもっと活用していかなければいけない」ということを業界の主要企業に話したのですが、当時、中間管理職の人たちはあまり乗り気ではありませんでした。そのときにどうしたかという、まず新しい技術を吸収することに意欲と熱意を持つ若い現場の担当者とお話をしました。私たちが開発したアプリケーションをどんどん使ってもらったのです。その次に重役のところへ行き、「今、取り組まなければ、日本の技術に将来はない」と話しましたが、なかなか重い腰を上げようとしません。そこで、「おたくの若い元気な技術者たちはやりたいと言っていますよ」と言うと、「そういう

創薬バリューチェーンとは



出展：NPO 法人バイオグリッドセンター関西

う若い者がいるなら、一度話を聞いてみよう」ということになり、話が進み始めたという経験があります。企業に、どのように情報をインプットしていくかということも大事です。事を急いで、ただ上から攻めていけばよいというものではない。土に種をまいておかなければ、いくら上から水をかけても芽は出ません。

志水 実はある製薬会社で福田さんが今お話しになったことと同じようなケースが進みつつあるんです。大学で開発されたアプリケーションを使った共同研究を一緒にやりながら、自社の課題を持ち込み、計算結果をフィードバックしようとしています。

福田 成果が出たところで、誰かが上から経営陣の背中を押してくれればいい（笑）。

— アプリケーションを開発している研究者の方々に、実用化といいますが、産業利用をめざすために知っておいてほしいことはありますか。

福田 計算例として、こうした入力でこういう結果が出るといったものが用意されていると分かりやすいです。あと大事なのは、企業から「このプログラムで、ちょっと違うこうした計算ができないでしょうか」といったいろいろな要望に応えられる機能が付け加えられているといいですね。そうした要望に抵抗感なく応じていただけると、「このチームと一緒にやっていきたい」という信頼関係もできてくるのではないのでしょうか。もちろん、何でも企業の言う通りにするというではありませんよ。

志水 まずはしっかりアプリケーションをつくっていただいて、あとは簡単なものでいいのでマニュアルを用意してもらえば、その先のユーザーインターフェイスなどは研究者の領域ではありませんからね。できれば汎用性を意識したものにさせていただくとよいと思います。

福田 このグループの化合物には使えるけど、あっちのグループは計算できないというものは困りますからね。

志水 とにかく私たちとしては、コンピュータシミュレーションの理解と普及をめざして、製薬業界を中心に「こういうものが使えますよ」、「こんなすごいアプリがありますよ」、ということを広くお伝えしていきたいと思っています。

— 難しい部分もありますが、それでも今後が楽しみという気持ちになってきました。今日はありがとうございました。

QM/MM 自由エネルギー法による 酵素反応分子機構の解析



京都大学大学院理学研究科

林 重彦

(分子スケールWG)

我々のグループは、酵素反応分子機構の解析のための QM/MM 自由エネルギー法の開発を行っています。生体分子の分子機能は、多くの場合に酵素触媒活性を持つ化学反応と共役しています。従って、酵素化学反応機構の理解は、生体分子機能の制御及び設計に重要となります。その目的のために、現在、主流となっている手法が、量子化学的 (QM) 手法、及び分子力学的 (MM) 手法を組み合わせるハイブリッド QM/MM 法です。この手法では、局所的な酵素触媒活性部位を QM 法、及びその他の広大なタンパク質環境を計算コストの低い MM 法で記述することにより、非常に高い計算効率を得ています。これまでに我々のグループは、QM/MM 法を用いて光受容や分子モータータンパク質等の様々な酵素反応の分子機構を解明してきました。

QM/MM 法の有用性の一方便、分子機能に共役する酵素化学反応の計算は、依然困難でした。多くの場合、分子機能発現には生体分子の大きな構造変化が伴います。このような構造変化は、通常 MM 法に基づく分子動力学 (MD) シミュレーションにより解析されますが、その構造変化に共役する酵素化学反応の計算には QM/MM 法に基づく記述が必要となります。しかしながら、QM/MM 法では、局所的な活性部位のみの適用に制限した QM 計算であっても MM 計算に比べて圧倒的に計算コストが高く、生体分子系の遅い緩和を適切に考慮できるだけの MD サンプリング時間を得ることができません。

この問題を解決するために、我々は、新規な QM/MM 自由エネルギー法 (QM/MM-RWFE-SCF 法) を開発しました。QM/MM 自由エ

ネルギー法では、MD シミュレーションによりサンプルされた MM 部分の構造分布により定義される自由エネルギー曲面上で、QM 法によって取り扱われる活性部位分子の最適自由エネルギー構造が決定されます。我々は、平均場近似と統計的 reweighting の手法を組み合わせ、さらに QM-MM 間の長距離クーロン相互作用を Ewald 法により適切に考慮することにより、非常に精度の高い効率的な手法を開発しました。特に、本手法のスキームでは、QM/MM 法部分の計算と MM 構造サンプリングのための MD シミュレーション部分が完全に分離されます。従って、MD シミュレーションに既存の洗練された MD プログラムを使用することが可能になり、計算遂行上の大きな柔軟性を達成しています。

開発された手法のテストとして、図1の反応スキームに示される α -amylase のグリコシド結合の加水分解に適用しました。周期境界条件で表される水中のタンパク質系は 68,000 以上の原子で構成され、活性部位の QM 分子は 600 基底以上の豊富な基底関数を用いて記述されています。本手法を用いて、この反応の始状態及び終状態の自由エネルギー最適構造を決定しました。その結果、図2に示されるように、これらの最適構造は、活性部位周辺のタンパク質ループ部分のそれぞれ 90ns 及び 21nsにわたる大きな構造緩和を伴い決定されました。このような、サブマイクロ秒に迫る QM/MM 自由エネルギー最適化計算は前例のないものです。すなわち、本手法で行う構造最適化計算は、直接的な QM/MM MD 計算に比べ、1万倍の長さのタンパク質構造緩和過程を追うことができ、また、他の同様な QM/MM 自由エネルギー計算に比べても 100 ~ 1000 倍程度の長時間緩和を記述しています。さらに、この計算により、反応前後において、近接ループが異なる構造を取る可能性が示唆されています (図2)。このような、酵素反応に相関するタンパク質構造の大きな変化は、これまでの手法では得ることのできない知見であり、酵素反応の新たな分子機構の解明や、制御・設計への道を拓くものと考えています。

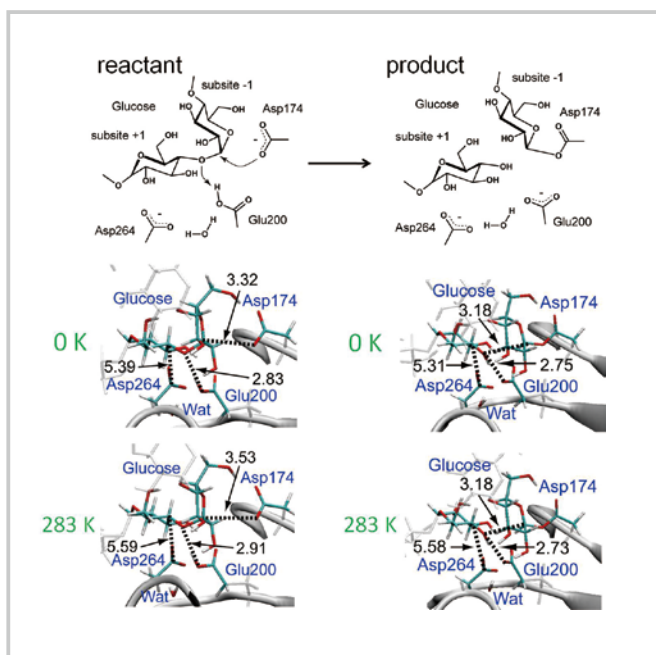


図1: α -amylase の酵素反応スキーム (上図) と反応始状態 (reactant) と終状態 (product) の構造変化 (中・下図)。自由エネルギー曲面上の構造最適化構造はポテンシャルエネルギー曲面上 (0 K) でのそれから変化している部分も多いが、終状態で強い電子的相互作用をしている原子間距離は、熱ゆらぎの効果を取り込んでも保持されていることがわかる。

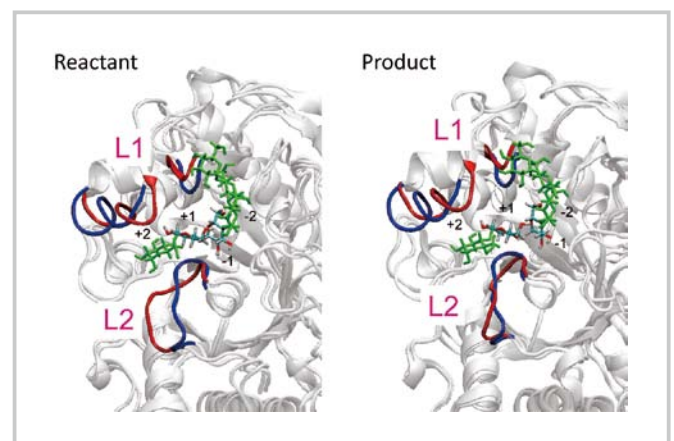


図2: 反応始状態 (reactant) と終状態 (product) でのタンパク質構造の変化。自由エネルギー最適化により得られた活性部位近傍のループ (L1 及び L2) の構造 (赤) が、構造最適化前 (青) の構造から大きく変化している。また、L2 ループの構造は、始状態と終状態で大きく異なる。

アクチン細胞骨格の 計算メカノバイオロジー

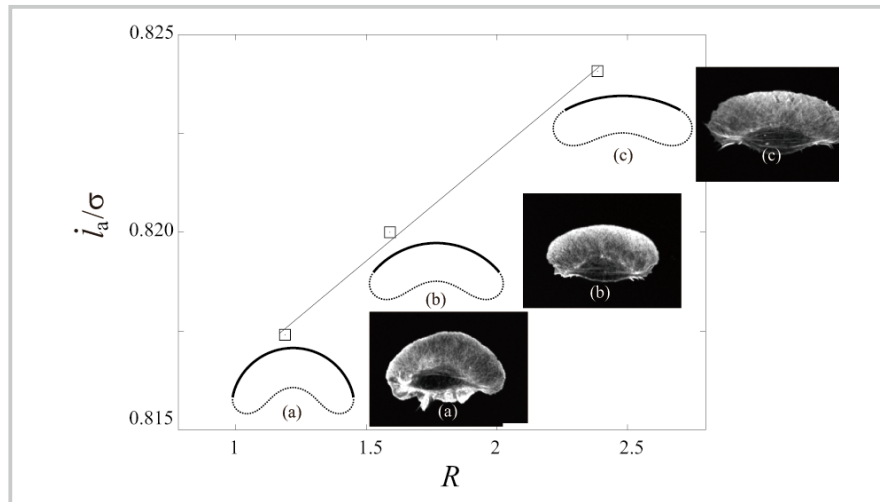


京都大学再生医科学研究所

井上 康博
(細胞スケールWG)

アクチン細胞骨格は、およそ5nm程度の大きさのアクチン単量体が互いに繊維状に連結したアクチンフィラメントを基本構造に、様々な補助タンパク質との相互作用を介して、網目構造や束構造などの高次構造を作り、細胞の形態や運動を支える重要な力学的・化学的基盤を成しています。例えば、細胞表面のアクチンは、糸状仮足や葉状仮足などの運動性を担う構造を作り、これらは、複雑に枝分かれしたニューロン構造の形成や創傷治癒における細胞運動などに必要です。また、アクチンとミオシンからなる複合繊維(アクトミオシン)構造では、ミオシンのモーター運動により、力が発生し、形態形成における細胞変形や収縮環による細胞質分裂、細胞接着のターンオーバーなどにおいて重要となります。これら構造は、アクチンを中心とした分子システムの自己組織化によって現れ、また、常に力学的・化学的に準安定な状態を経巡ることで、ダイナミックにその構造や特性を変化させることができます。私たちは、このようなアクチン細胞骨格のダイナミクスと細胞機能との関連について、計算メカノ

バイオロジーの手法を用いて研究しています。最近の私たちの研究により、細胞運動における細胞形状とアクチン重合が熟ゆらぎを介して達成すること^[1]や、アクトミオシンネットワークに発生する力とミオシン濃度との関係がネットワーク構造の変化に対応して双線形になること^[2]を明らかにしました。これらは、細胞スケールの機能発現と分子スケールのダイナミクスとの関係について、重要な知見を与えています。また、計算機を用いてアクチンメカノバイオロジーシミュレーションができるようになってきたことで、細胞をまるごと1個、計算機上に再現することも全く手の届かない夢ではなくなってきました。もちろん、まだ、解決しなければならない課題は山積みです。これらに関しては、次世代生命体統合シミュレーション研究グループ内外の研究者と連携する必要があり、現在、そのような場として、計算科学アルゴリズム研究会が定期的に開催され、分野を超えて数理モデル・計算手法の知見を共有し、また、関連する境界領域の研究者との連携が進められています。



【参考文献】

- [1] Inoue, Y. and Adachi, T. Biomech. Model. Mechanobiol. 2011 10:495-503.
- [2] Inoue, Y. Tsuda, S., Nakagawa K., Hojo M., Adachi T. J Theor Biol. 2011 21;281(1):65-73.

図1：移動性細胞の計算機シミュレーション：シミュレーションによって得られた細胞形状（実線図）は、実際に観察される細胞形状（蛍光画像）と一致しました。また、シミュレーションにより、細胞先端のアクチン重合速度（縦軸）は、細胞先端の曲率半径（横軸）に比例することがわかりました。

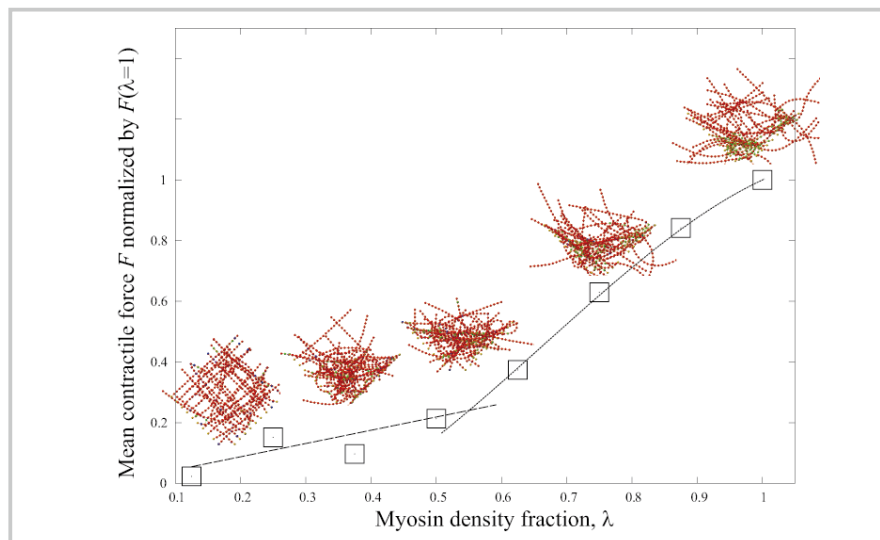


図2：アクトミオシンネットワークの自律的構造変化：アクトミオシンネットワークに存在するミオシンのモーター運動によって、アクトミオシンネットワークの構造は自律的に変化します。シミュレーションにより、構造再編に伴い、ネットワークに発生する力は、ミオシン分率に対して双線形となることわかりました。

血栓シミュレーションに向けた 血流解析手法の開発



東京大学工学系研究科
伊井 仁志
(臓器全身スケールWG)

血管内壁が損傷を起こすと、そこで血栓が形成されます。通常、軽い怪我をしてもある程度時間が経てば出血が止まるのは血栓のおかげであり、この止血こそが血栓の主な役割です。形成された血栓は止血後に溶解されますが、何らかの要因で形成された血栓が消えず肥大化し血管を防いでしまったり、肥大化血栓が剥がれて末端の血管で詰まり血流を塞いでしまうこともあります。このような血栓をもとに、脳こうそく、心筋こうそくなどが引き起こされ、またエコノミークラス症候群(ロングフライト血栓症)も血栓が原因で発症します。こうした理由から、血栓形成を阻害する薬剤が臨床の場で用いられています。しかしながら、止血に携わる正常血栓に対しても作用する可能性があるため慎重に薬剤を投与する必要がありますが、人体内の血流中において薬剤の効果を前もって見積もるのは非常に難しいとされています。このような背景から、我々のグループでは血栓形成をシミュレーションで再現することを目指しており、例えば、新薬開発の際に血流中での薬効を把握するのに役立てたいと考えています。

血栓形成過程においては、損傷血管壁への血小板の接着、血小板同士の凝集、また赤血球をも取り込んだ凝固などが、血流や細胞代謝など様々な要因を伴い段階的に起こっています。ここでは、その初期過程である血小板が損傷血管壁に付着する血小板一次凝集を対象にします。この過程においては、血管内皮細胞下のコラーゲンに付着したvWF(von Willebrand factor)と血小板表面に分布するGPIIb α (glycoprotein IIb α)が結合することで、血小板が損傷血管壁に接着します。このような現象

を解析するため、赤血球・血小板を含んだ血流現象と、血小板表面・損傷血管壁でのGPIIb α -vWF結合現象を、統合的に取り扱う手法を開発しています(図1)。

血流中において赤血球の容積率は45%程度と非常に割合が大きいため、血管径がマイクロメートルサイズになると、血流の流動特性は赤血球の影響を受けるため血小板の凝集にも強く関わってきます。また赤血球は非常に軟らかく変形しやすいため、流動場とあわせた相互作用現象となります。これらの現象を非常に多くの赤血球が含まれる系で解析する必要があるため、我々のグループでは計算効率に優れかつ血流解析と親和性の良い新たな解析手法を提案しています^[1,2]。実験的に得られている赤血球の物理特性を用いて解析を行ったところ、図2に示すように、流動場における実際の赤血球の変形挙動と比べて妥当な結果が得られました。

血小板一次凝集の解析においては、血小板表面のGPIIb α と損傷血管壁のvWFとの結合を考慮する必要があります。これらは分子スケールでの現象ですが、現状では血流とあわせた相互作用的な解析は不可能なので、そのような結合現象を統計的に取り扱うことができる手法を導入し、血流解析と組み合わせています。図3に示すのは解析事例の一例で、血小板表面のGPIIb α と壁面底部(損傷血管壁を想定)のvWFが結合を起こし、血小板が血管壁に接着している様子を再現しています。

今後は、血小板細胞の代謝反応をモデリングすることで血小板同士の凝集を取り扱い、血栓形成過程を段階的に再現していく予定です。

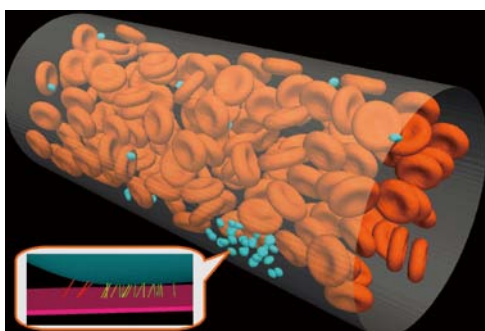


図1：血小板一次凝集に向けた解析の模式図

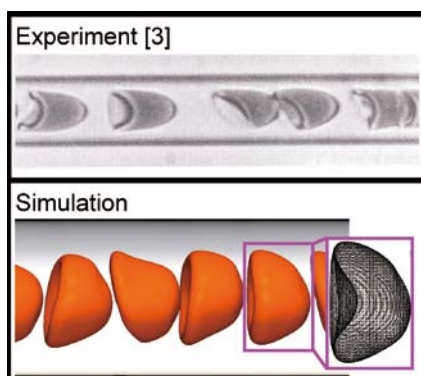


図2：流動場中の赤血球の変形挙動^[3]とシミュレーション結果

【参考文献】

- [1] K. Sugiyama, S. li, S. Takeuchi, S. Takagi and Y. Matsumoto, A full Eulerian finite difference approach for solving fluid-structure coupling problems, *J. Comput. Phys.*, 230 (2011) 596-627.
- [2] S. li, X. Gong, K. Sugiyama, J. Wu, H. Huang and S. Takagi, A full Eulerian fluid-membrane coupling method with a smoothed volume-of-fluid approach, *Commun. Comput. Phys.* (2011) accepted.
- [3] P. Gaehtgens, C. Dührssen and K.H. Albrecht, Motion, deformation, and interaction of blood cells and plasma during flow through narrow capillary tubes. *Blood Cells*, 6 (1980) 799-817.

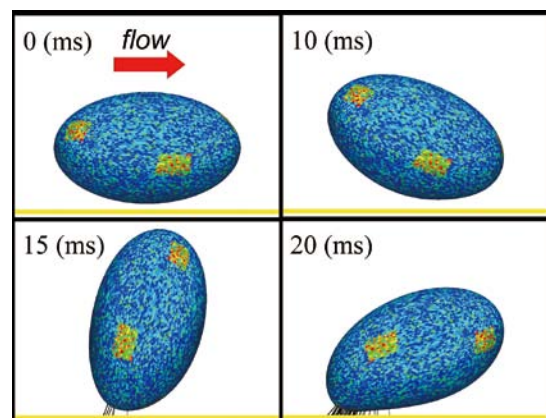


図3：GPIIb α -vWF結合を考慮した流体解析における血小板接着の様子

生命体シミュレーションのための データ同化技術の開発



統計数理研究所
樋口 知之
(データ解析融合WG)

今回の東日本大震災によって私たち研究者は、複雑なシステムを理解し防御あるいは制御することの困難さと、その問題克服に向けて英知を傾ける永続的な努力の必要性を改めて認識させられました。地球に限らず生命体のような複雑なシステムの理解と制御においては、対象に関する知識は常に不完全であることを前提に、現象の予測能力でもって研究の進め方を評価し修正する方策が有効です。このことは統計学が教えるところであり、そもそも人間が地球上に繁栄してきた大きな理由の一つでもあります。

予測能力は大きく分けて二つの機能に関する有効性の総合指標となっています。一つはフォワード（前向き）計算モデルの記述力、もう一つは対象の現状（現況）を捉える認識力です。フォワード計算とは、分かりやすく言えば繰り返し代入計算のことで、右辺に値を入れたら左辺の値が出、その値をまた右辺に入れば次のステップの左辺の値が出るような計算方式を指します。時間発展を陽に求めるシミュレーション計算の多くはこの方式をとり、また長期予測はこのフォワード計算の反復で実現されます。一方、後者は計測手法のイノベーションと直結しています。画期的な計測法の発明により従来よりも圧倒的な規模と精密さで情報を得ることのできる新しい装置は、どの分野の研究者にとっても大きな魅力であり、特に生命科学においては、その発展の大きな牽引力となってきました。

しかしながら、計測装置の研究開発のみに傾斜することは、予測能力の観点からすると得策ではありません。それは、対象を直接的にまるごと計測する方向性には理論上限界があることもありますが、予測能力を上げるためには、フォワード計算モデルの記述力を強化することも極めて効果的だからです。地球・宇宙科学や物性物理のようなシミュレーションを用いた研究に長い歴史がある分野では、フォワード計算の基礎となる支配方程式がたいてい確立しており、支配方程式にもとづく計算をスーパーコンピュータ上でどう近似的に上手に実現するかが成功の鍵です。この近似計算の改善がフォワード計算モデルの改良に相当しています。バイオスーパーコンピューティングの大きな目標の一つも、このフォワード計算モデルの記述力を、計算ハードのスケールメリットを最大限に生かして劇的に向上させることです。ただし残念ながら生命科学においては、支配方程式に相当する原理がないと言っても過言でなく、従って、フォワード計算モデル自体も多種多様な考え方にもとづいた一般性の低いものとならざるを得ない状況です。

では生命科学においてはフォワード計算モデルの系統だった改良は困難なのでしょうか？ 前述したように、計測法の発展とフォワード計算モデルの改良が両輪となって、現象の予測能力の向上は実現されます。そうすると、フォワード計算モデルを何某かの独自の評価基準で改善するのではなく、予測能力が向上するようにモデルに手を入れることは自然に思えます。つまり計測データから学習する機能をフォワード計算モデルに付加するわけです。事実、シミュレーション研究の最先端領域である気象・海洋分野では、全世界から時々刻々と集められる大量の時空間データと、スーパーコンピュータ上での世界最大規模のシミュレーション計算結果を、ベイズ統計を用いて情報統合し、シミュレーションモデルをリアルタイムで改良することで予測性能をあげる予報業務が普通になりつつあります。また、

最近話題になった放射線の影響を予想するシミュレータSPEEDIが、その力を十分に発揮できなかった理由の一つに、実観測データをリアルタイムでシミュレータに反映させる機能がなかったことが指摘されています。

このような観測データとモデル計算結果の情報統合はデータ同化と呼ばれ、近年シミュレーション科学の分野で大きな注目を浴びています。データ同化の考え方を生命体シミュレーションに適用すれば、少なくとも予測性能の着実なアップに繋がり、ひいては複雑システムの理解と制御に役立つことでしょう。私どもはこの強い思いのもと、データ解析融合チームメンバーとして生命体シミュレーションのためのデータ同化技術の研究開発に日々取り組んできました。現在の生命体シミュレーションモデルは服に例えて言えば既製服のようなものです。バラエティがあったとしても、せいぜいS・M・Lのサイズの違いがあるくらいでしょう。片や人の体内システムは一人一人異なっています。薬や治療法の副作用に苦しむ患者さんの診療データから、その人にあったオーダーメイド、せめてセミオーダーメイドの生命体シミュレータが自動的に構築できる、そんな時代が早く到来することを夢見て頑張っています。

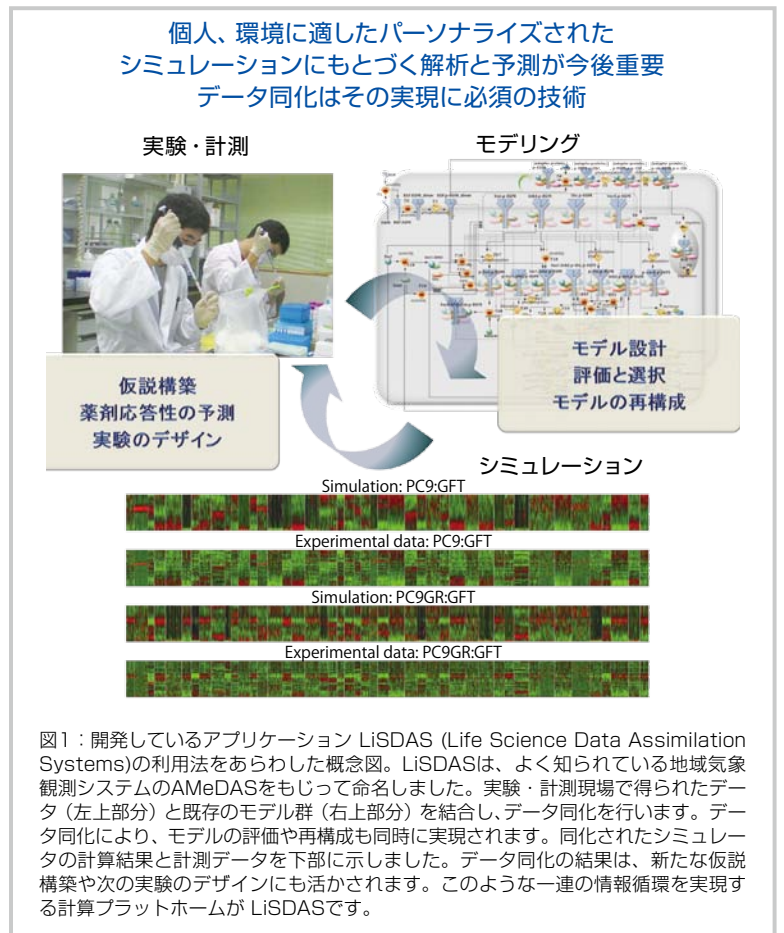


図1：開発しているアプリケーション LiSDAS (Life Science Data Assimilation Systems)の利用法をあらわした概念図。LiSDASは、よく知られている地域気象観測システムのAMeDASをもじって命名しました。実験・計測現場で得られたデータ(左上部分)と既存のモデル群(右上部分)を結合し、データ同化を行います。データ同化により、モデルの評価や再構成も同時に実現されます。同化されたシミュレータの計算結果と計測データを下部に示しました。データ同化の結果は、新たな仮説構築や次の実験のデザインにも活かされます。このような一連の情報循環を実現する計算プラットフォームが LiSDASです。

複雑な生命現象の理解と予測に向けて 計算生命科学の明日を拓く

理化学研究所 HPCI計算生命科学推進プログラム プログラムディレクター 柳田 敏雄
理化学研究所 HPCI計算生命科学推進プログラム 副プログラムディレクター 木寺 詔紀
理化学研究所 HPCI計算生命科学推進プログラム 副プログラムディレクター 江口 至洋



左より、木寺、江口、柳田

「京」を中核とする日本の高性能計算基盤（HPCI）を活用して、幅広い研究分野で世界最高水準の研究成果を生み出し、社会への還元をめざすHPCI戦略プログラムが、今年度から本格的にスタートした。分野1「予測する生命科学・医療および創薬基盤」は、生命科学にどのようなブレークスルーをもたらすのか。統括責任者・柳田敏雄氏、副統括・木寺詔紀氏、江口至洋氏にお話しいただいた（今号・次号の2回に分けて掲載）。

●生命現象をコンピュータ上に再現することが目標

柳田（敬称略）今、私たちがどのような夢を持っているかを一言でいうと、非常に複雑で、非線形かつ動的な系の代表ともいえる生命現象をコンピュータ上に再現できるようにする、これが最終的な目標です。例えば細胞のいろいろな動きを再現するためには、コンピュータに載せるデータがしっかりと揃っていないといけない。当然ながらデータだけでシステムが分かるわけではなく、データを処理・解析するためのコンセプトが必要で、それらをきちんと用意して、それをベースにモデルをつくり、細胞の動きをコンピュータのなかで再現するわけです。もちろん、細胞だけでなく、組織、器官まで再現したいのですが、それは簡単なことではありませんから、まずはタンパク質の機能といった部分的な機能をしっかりと理解して、コンピュータ上に再現したいですね。さらにそこから創薬につながっていけば素晴らしいと思います。もちろん、タンパク質の複雑な動きがどのような仕組みで機能につながっているかも興味はありますが、まずタンパク質の動きが操作できるようにすれば、創薬に結び付けることが期待できますからね。



木寺 私は、これまでずっとタンパク質のシミュレーションの研究をやってきたわけですが、HPCI戦略プログラムにおける目標も、最終的には柳田先生がいわれたように、タンパク質を計算機のなかでシミュレーションし、その振る舞いを一目瞭然という形で再現したいと思っています。しかし、材料分野などのように計算科学が進んでいる他の分野と比べると、生命科学はまだまだ未熟で、誤解を恐れずにいうならば、そもそも生命現象をシミュレーションすることは、限りなく不可能に近いことだというべきかもしれません。ならばどこまで可能なのか、私たちのチャレンジは、そこから始まった



のだと思っています。これまでは、細胞の振る舞いを再現することは難しい、ならばもっと限定してタンパク質1つならどうかというように、できるところまで話を限定して、そこをきちんとシミュレーションしたときに何が分かるかというストーリーを作り出してきました。分かるところまで問題をどんどん小さくしていったわけです。その結果、何がおきたかという、生物学でなく物理学の問題になってしまいました。最初は物理の問題だったのです。しかし、多くの“生物屋さん”は、そこに見向きもしてくれませんでした。では、生命現象と物理現象はどこが違うかということになります。例えばタンパク質1つを計算機のなかに置いてシミュレーションしても、何も起こらないんです。何か起きるためには相手が必要です。その相手が動くためには、また別の相手が必要で、別の相手もさらに別の相手が……、というように延々と続く世界があって、それが恐ろしく複雑で不均質な細胞という環境のなかで繰り返されているのです。かつては計算機の能力の限界のために、そこまで見ていくことができませんでした。ところが、だんだん計算機の能力が高くなるにつれて、再現できる範囲が空間的にも時間的にも大きくなって、小さく限定しなくてもよくなりました。そこでようやく“生物屋さん”たちも関心を持ってもらえるようになってきました。その段階のさきに、「京」を使ってもっと大きなもの、もっとリアルなものを見ようということになったわけですね。ようやく実験をやっている生物学の研究者たちが認めてくれるレベルの研究になってきたということです。さらに、これまで「計算なんて関係ない」と思っていた医療や創薬の分野でも、「役に立つのではないか」、「京」を使えば面白いことができるかもしれない」と言ってもらえるようになってきました。かつて不可能と思われていたことが、可能性が見えてきたというか蓋然的といえるくらいまで……。いや、これは言い過ぎかな（笑）。

柳田 「5倍よくしましょう」と言う現実問題に縛られて悩んでしまうけれど、「100倍」と言えば、かえって軽い気持ちになれて、研究は進むそうです（笑）。

木寺 私たちがこれまで使っていた計算機のレベルが10TFLOPSでしたから、「京」の10PFLOPSは1,000倍ですから、より多くの研究者が乗ってくれるかもしれませんね（笑）。

柳田 冗談はさておき、確かに研究者が「面白い」と思うことは非常に大事なことだと思います。シミュレーションって面白いんだ、役に立つんだと思えば、それをベースにものを考えるようになり、研究が新しい方向に進んでいく可能性が生まれますからね。

●生命科学が情報を活用したサイエンスとして成立することを示す

江口 柳田先生が「夢」といわれましたが、私はこのプロジェクトを通して、これまで試行錯誤を繰り返しながらたくさんの実験をやってきた研究者が、「この実験はコンピュータシミュレーションで結果が分かるから、シミュレーションに任せよう」と言える時代になればいいなと思っています。例えば、自動車の安全性を調べるための衝突実験がありますよね。昔は実際にたくさんの自動車を使って、角度や速度を変えて実験したそうですが、今はかなりの部分をシミュレーションでやっているそうです。それだけコンピュータシミュレーションの信頼性が高まっているということです。生命科学においても、同じようになってほしいと願っています。



柳田 おっしゃる通り、ものづくり分野などでは、今や実験の半分はシミュレーションです。実験といえばシミュレーションを活用するのが当たり前になっています。その点では、生命科学はコンピュータサイエンスが遅れているといえるかもしれません。しかし、これからは生命科学もそうなっていくと思います。すべて置き換えることは無理ですが、できる部分は実験をせずにパラメータを決めてシミュレーションでやり、どうしてもできない部分を実験するというかたちで効率を上げていかないと、特に創薬ではアプローチできません、試行錯誤では組み合わせが多すぎますからね。

木寺 江口先生のおっしゃられたことに加えてもうひとつの大切なことは、生命科学が情報を活用したサイエンスとして成立し得ること

を示すことだと考えます。それは「京」の利用によって、いままで限定されていた情報処理のあり方を革新し、情報のイメージそのものも変えて、いままでできないと思っていたことをできるようになることです。その方向をよりはっきりと示して、可能性を大きく広げていくことが、私たちが取り組んでいるHPCI戦略プログラムのひとつの役割だろうと思います。「京」のような巨大な計算機資源を使ったデータ解析のあり方、プロトタイプを示して、その可能性と結果を大いに外に示していくことが重要ですね。それから、柳田先生が言われたものづくり分野などに比べて遅れているという点について、少しだけ弁護させていただくと、実は同じレベルの計算は私たちもやっているんです。ただそれだけでは生命の問題には間に合わない。だから小さな問題しか扱ってないように見える。例えば全身スケールのシミュレーションがありますが、人体も基本的には自動車と一緒に見えます。ところが、鉄板や鉄骨の組み合わせでできている自動車と違い、臓器1つとってもたくさんの種類の細胞がやまほど集まっていて、しかも、熱や物質が出たり入ったりしながら化学反応している。何を見るかによって、計算内容も変えなければならない、とにかく非常に難しい問題です。

柳田 他の分野との決定的な違いは、基本方程式がないことです。パラメータを探して、ある傾向をみつける、状態をみつけるということしかない。また、おこっていることが多要因であるということも問題です。生命科学は、対象が非常に複雑で難しいため、まだまだ時間がかかります。しかも、組み合わせが多いので、高い計算能力が必要です。でも、そこをクリアして、他の分野がやっているようなかたちまで研究が進めば、創薬などに貢献するいろいろなことができるようになるはずですよ。

●楽しまないといい研究はできない!?

江口 タンパク質の解析だけでも大変ですし、細胞はさらにいろいろなデータが含まれている、複雑で多様で非線形、そうした生命現象の難しさを熟知している柳田先生がプロジェクトのトップにおられ、どんどんメッセージを発信していかれることは、非常に重要なことだと思います。そうすることによって、新しいデータから新しい発想が生まれ、研究が進展していくと思います。さらに今後、分野1を盛り上げていくためには、それぞれの研究開発チームがお互いに問題を提起したり、新しい人たちが参加できるようなパスをつくり変えることなども必要になってくるでしょう。さらに、このプロジェクトは5年間ですが、6年目以降にさらに新しい動きが生まれて、この分野に参加したいという若い人たちが育つ、そういう仕組みづくりをしていきたいと思っています。研究の難しさはあると思いますが、新しい計算機パワーを活用して新しい成果をあげ、みなさんが活躍しているところを、生命科学のコミュニティや若いコミュニティに伝えて、研究のすそ野を広げていきたいですね。

柳田 計算機科学をやる人だけがこの分野を支えているのではなく、多くの生物学の研究者が計算機を物理的に活用している時代ですから、さらにそれを進めていくことがサイエンスそのものを向上させていくことにつながると考えています。しかし、いくら計算機は大事ですよと言ってもだめで、それをけん引していくもの、拠点となるものが重要です。その象徴的なものが、世界一の計算能力を示した「京」ではないでしょうか。そして、実際に「京」を活用したプロジェクトが動き出し、そこで素晴らしい研究が行われ、世界のトップを走る研究者が日本にいる、それが重要なことだと思います。

江口 「京」が世界一になった今が大きなチャンスですね。それだけに、先生方にはいい研究成果を出していただきたい。

柳田 研究資源が1位になったことで、計算科学的なアプローチ、日本の科学研究そのものが世界1位になれるのかということも当然、注目

されるわけですからね。そのためには、やはりインターディシプリナリーな研究のすそ野を広げていくことが大切であると思います。例えば理学と工学がうまく融合していくこともその1つです。

江口 世界一の「京」を使って研究ができるなんて、すごいことです。本当にいろいろな研究者が参加して、お互いが刺激し合い、高め合える環境が生まれてほしいですね。

木寺 計算機の能力が高まることによって、私たちの分野の研究が進み、さらにもっといい研究ができる、それは確かです。「京」が使えるということは、研究にとって非常に大きなメリットです。だからこそ、使う人がプレッシャーを感じるのではなく、いい緊張感を持って、わくわくしながら研究ができる、そんな体制づくりも、これからの課題かもしれません。まあ、「結果を出せ、結果を出せ」と尻を叩くのが私の立場ではあるのですが(笑)。

江口 「面白いから、一緒にやろう!」という雰囲気も、ぜひつくりたい(笑)。

木寺 実際に使ってみると、確かに「京」はすごいです。その速さを体感してしまうと、これはものすごい計算機だなと実感させられます。

柳田 今までと三桁違うと、量的な変化だけでなく、質的な変化を感じますよね。

木寺 これを使えばすごい結果が出るという感触は、みなさん持っています。だからこそ、逆にプレッシャーも感じているようです。私も、できるだけ「早く結果を出せ」と言わないようにしたいと思っているのですが、つつい……(笑)。

柳田 そういえば、先日審査委員会でも、研究者の発表に対して「研究は楽しくやらないといけない。楽しみながら取り組んでいますか」という話が出ました。

木寺 責任はもちろんありますが、結果的に、楽しまないといい仕事はできませんからね。

(次号に続く)

創薬応用シミュレーション

東京大学先端科学技術研究センター

藤谷 秀章
(分野1-課題2)



タンパク質とこれに特異的に結合する化合物(リガンド)の相互作用を調べる事は、生命現象の基礎的研究の大きな課題であると同時に応用の観点からも極めて重要である。特にタンパク質と薬分子との結合自由エネルギーを精確に求める事は創薬分野における永年の課題である。ドッキング・シミュレーションを初めとして様々な方法が試みられて来たが、薬分子の設計に有効な精度でタンパク質との結合自由エネルギーを予測する方法はなかった。タンパク質とリガンドが水中(細胞液)で結合した状態と、それが別々に水中に溶けている二つの異なった熱平衡状態の間の自由エネルギー差を求めるには、タンパク質の構造変化や分子の運動によるエントロピー変化を精確に取り入れる方法が必要になる。このような計算には分子動力学法が使われるがその精度は用いる力場パラメータに強く依存する。

タンパク質主鎖のラマチャンドラ角に対する振れ相互作用はタンパク質立体構造を決める最も重要な力場パラメータであり、図1でこのエネルギー・プロファイルをグリシンペプチドで比較した^[1]。黒い実線が今日最も精度の高い分子軌道計算(LCCSD)から求めたプロファイルで、AMBER(ff99, ff03)やOPLS-AAなど従来の分子力場と比較した。ff99sbは2006年に発表されたff99の主鎖の振れ相互作用だけを修正した力場で、幾つかの角度で高精度分子軌道計算のエネルギーに一致する様に決められた。この為にφやψが0°でのエネルギー値は黒い線に近いが、φが80°から180°間では黒い線のエネルギーと比較して1kcal/mol以上高いエネルギー障壁が見られる。我々はAMBERの電荷とvan der Waalsパラメータを使って、タンパク質やDNA・RNA、任意の有機化合物に対して統一的に力場パラメータを割当てる方法を開発してFUJIL力場と名付けた^[2]。

自由エネルギー差ΔGと非平衡仕事量Wの関係式を用いて^[3]、リガンドの他分子に対する相互作用が完全に存在する状態から、相互作用がなくなるまでの複数の中間状態で分子動力学計算を独立に実行して^[4]、隣の状態に移行する為に必要な仕事量Wから結合自由エネルギーを求める(図2)。このMP-CAFEE法とFUJIL力場を用いてタンパク質とリガンドの結合自由エネルギーを正確に求める事が可能になる。

【参考文献】

- [1] H. Fujitani, A. Matsuura, S. Sakai, H. Sato, and Y. Tanida : *J. Chem. Theory Comput.*, 5, 1155 (2009).
- [2] H. Fujitani, Y. Tanida, and Azuma Matsuura : *Phys. Rev. E*, 79, 021914 (2009).
- [3] C. Jarzynski : *Phys. Rev. Lett.*, 78, 2690-2693 (1997).
- [4] H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande : *J. Chem. Phys.*, 123, 084108 (2005).

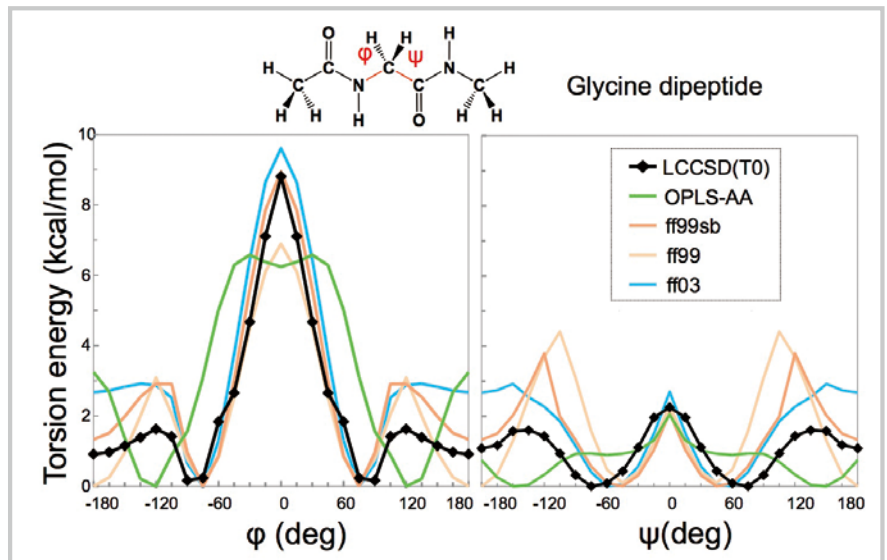


図1: タンパク質主鎖の振れ相互作用の比較

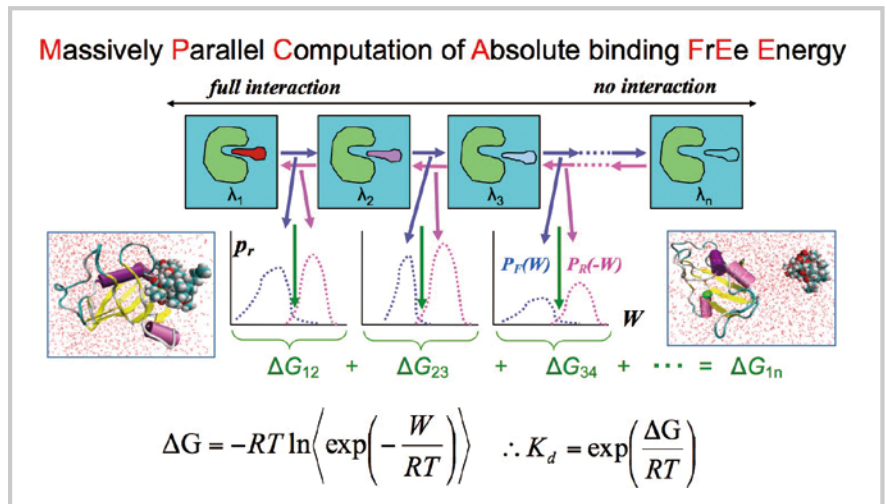


図2: 超並列結合自由エネルギー計算法 MP-CAFEE

次世代DNAシーケンサデータの 超高速解析



東京工業大学大学院情報理工学研究科

(上から) 秋山 泰、石田 貴士、角田 将典、鈴木 脩司
(分野1-課題4)



次世代DNAシーケンサは、飛躍的な技術革新により1回の走査で約1兆文字ものDNA配列読み取りを可能にしている。いわゆる旧型シーケンサの時代でも「1ドル当たりで読み取れる文字数」は約19ヶ月で2倍という驚異的な伸び率であったが、次世代シーケンサの登場以降は約5ヶ月で2倍の伸び率と言われ、また世界的に圧倒的な台数が普及しつつある。現在では読み取られた配列の意味解釈の計算が最大のボトルネックである。遺伝子ネットワークの推定のような総合的な高度計算にスパコンが必要な事は周知の事実であるが、そればかりか配列断片群に対する基本的処理にすら大規模計算が必要となった。

ヒトの遺伝子のわずかな変異などの研究では、既知配列との完全一致または強く類似した文字列の探索に適した高速アルゴリズムを使うことができ、大型のPCクラスタを用いて何とか対処されてきた。しかし一方、ヒトの腸内等の微生物のゲノム断片を網羅的に調べ上げる研究や、深刻な配列の変異を扱う研究では、遠縁あるいは大きく変異した配列間での配列相同性を検知しなければマッピング(既知配列との対応付け)が実行できない。これには桁違いに大きな計算が必要となり、PCクラスタ等では歯が立たない。

ヒトの健康は体内に共生する多数の微生物のバランスの取れた働きにより維持されている。またそのバランスには幾つかの局所安定解があると考えられている。米国のHuman Microbiome Project (HMP)では約500名、欧州のMetaHITプロジェクトでは約300名のヒトの口腔内、消化管内、皮膚、泌尿生殖系等の環境での微生物群のゲノム断片の網羅的な読み取りが実行されているが、このような貴重なデータを分析する際に、文字列の強い類似性を仮定した感度の悪い手法だけでは、解釈不能として読み捨てられる断片配列の比率が大きく増加してしまう。

我々は、ゲノム解析の専門家である黒川顕教授(東工大)の協力の下に、まず既存のBLASTXソフトウェアの超並列実行環境の構築を行った。東工大TSUBAME2.0上では16000コア超までのスケーラビリティを確認した(図1)。各計算ノードへのデータベースの初期配送では、ネットワークのバンド幅を活かした二分木転送を用いた。我々はDNA断片配列を可能な6通りの読み枠でアミノ酸配列に変換してから、既知アミノ酸配列との柔軟な比較を行う。DNA配列レベルでは大きく変異していても、アミノ酸レベルでは残基が一致する、あるいは似た物理化学的性質を持つ残基に変異していることが多いため、文字間の距離行列を用いて比較する柔軟な手法を用いることにより、同じ機能をもつ遠縁の遺伝子を探

す検出感度が高まるためである。

次に、我々はアルゴリズム上の大幅な高速化を行い、BLASTXを改良したGHOSTXと称するソフトウェアを作成した^[1,2]。GHOSTXでは、メモリに格納された区分データベースをSuffix Arrayで表現し、またクエリ配列側もSuffix Arrayに変換し、この両者を効率良く比較することで、同じ部分文字列の一致を多対多で一気に解決できる(図2)。またBLASTXでは固定長の文字列比較しか行わないのに対し、GHOSTXでは一致スコアを満たすまで可変長で文字列を伸ばせるため、計算コストあたりの感度が向上する。BLASTXと同等の感度が必要な場合でも速度は約20倍。実用上支障がない程度に感度を下げると100倍超の高速化が達成できた。高速だが感度が粗すぎて利用できなかったBLAT法より感度は大幅に良く、速度も2~3倍速いという結果も得ている。我々は「京」の上にGHOSTXアルゴリズムを実装し、OpenMPによるスレッド並列化と、MPIによるノード間のデータ並列性を組み合わせて、「GHOST-MP」と呼称する超並列解析システムを構築した。

相同性検索の主要部では、「京」で1536ノード(12288コア)までの良好なスケーラビリティを実測した。断片配列1本ごとの実行時間は千差万別だが、負荷分散機構を有するため、数万ノードまでのスケーラビリティが期待できる。GHOST-MPでは、読み込むデータベースとクエリ配列群の長さ同士の積のオーダーの計算を行うので、I/Oに終始する処理ではないものの、注意深いI/O設計は必要

である。「京」ではZ軸方向の通信でI/Oノードとの入出力転送が実施されるため、代表ノードによるデータ入力とブロードキャストの機構を調整してI/O負荷のさらなる分散と低減を図り、数万ノード級までの高効率実行を目指す。

【参考文献】

- [1] 鈴木, 石田, 秋山, 情処研報 2010-BIO-23(21), 1-6 (2010).
- [2] 鈴木, 石田, 秋山, 情処研報 2011-BIO-25(32), 1-8 (2011).

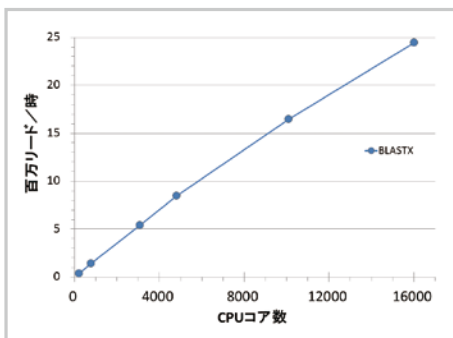


図1: BLASTXベースのシステムでの並列性能

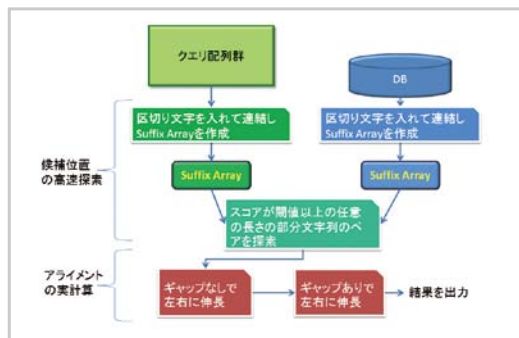


図2: GHOSTXアルゴリズムにおける処理の流れ

バイオスーパーコンピューティングサマースクール 2011

理化学研究所 次世代計算科学研究開発プログラム 石峯 康浩 (臓器全身スケールWG)

統計数理研究所 データ同化研究開発センター 斎藤 正也 (データ解析融合WG)

新潟国際情報大学 近山 英輔 (細胞スケールWG)

東海大学医学部内科学系循環器内科 七澤 洋平 (細胞スケール/臓器全身スケールWG)

理化学研究所 次世代計算科学研究開発プログラム 半田 高史 (脳神経系WG)

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

理化学研究所 次世代計算科学研究開発プログラム 森次 圭 (分子スケールWG)

(五十音順)

2011年9月26日27日の両日、兵庫県淡路市にある兵庫県立淡路夢舞台国際会議場において「バイオスーパーコンピューティングサマースクール2011」が開催されました。次世代生命体統合シミュレーション(ISLiM)のプロジェクトに参加する研究員が中心に企画・運営を行う若手の会としては昨年度のサマースクール・ウィンタースクールに引き続き3回目ですが、今回は「バイオスーパーコンピューティング研究会」(BSCRC)の後援のもと、名称を一新しました。

現在兵庫県神戸市において設置が進んでいる京速コンピュータ「京」が今年6月のTOP500で世界一になりました。ISLiMでは京速コンピュータ「京」の性能をフルに発揮するような生命科学の数値シミュレーションや大規模データ解析研究を行うためのソフトウェアを開発中ですが、このサマースクールでは、「京」で何ができるか、また、生命科学に対してどのような貢献ができるかをテーマに、「計算機・計算手法」と「アプリケーション」の二つのセッションを中心に講演方式で進められました。同プロジェクトに参加する研究員や関連分野の大学院生など総勢56人が参加し、活発な意見交換が行われました。

最初に「基調講演」として、次世代計算科学研究開発プログラムの茅幸二ディレクターから、ISLiMプロジェクトの道のりと今後の展望についてお話しがありました。次に、次世代計算科学研究開発プログラムの姫野龍太郎副ディレクターから、ISLiMで現在開発が進んでいるアプリケーションソフトウェアの現在の性能・開発状況と今後の「京」利用スケジュールについてご講演がありました。

次の「アプリケーション」のセッションでは、東京大学医科学研究所の上昌広先生による招待講演とISLiMの3チームの代表者からさまざまなスケールの生命体ソフトウェアに関する講演がありました。上先生からは研究者であると同時に臨床医でもある立場から、ゲノム解析に関連した個別化医療や医師不足の問題、また、東日本震災に見舞われた福島県での活動状況についてお話しがありました。計算科学者にとっては解析すべきデータがあるというのは前提ですが、実は人材ネットワークを広げつつデータを収集する作業が大事である、という主張が印象的でした。データ解析融合チームの斎藤正也さんからはLiSDASによる哺乳

類の概日周期転写ネットワーク解析の話とデータ同化ライブラリの紹介、細胞スケールチームの須永泰弘さんからは、細胞シミュレーション統合プラットフォーム(RICS)とその応用研究の紹介、また、脳神経系チームの加沢知毅さんからは、カイコガの嗅覚・運動神経系シミュレーションや神経回路のリアルタイムイメージングについてお話しがありました。

その後の懇親会・ポスターセッションでは35件のポスター発表があり、夜遅くまで議論が交わされました。参加された5人の大学院生に審査委員をお願いしたポスター賞の選定や、姫野先生のご提案による「エクサスケールコンピューティング」に向けた議論もそのなかで行なわれました。

二日目の「計算機・計算手法」のセッションではまず、富士通株式会社次世代TC開発本部PAプロジェクトの三吉郁夫先生、杉崎由典先生による招待講演ののち、粒子系・流体系シミュレーションとその高度化技法についてISLiMの3チームの代表者からご講演がありました。三吉先生からは、「京」のハードウェア構成やソフトウェアの開発状況の説明、杉崎先生からは、簡単な例を挙げながらのチューニング技法の説明があり、のちのQ&Aコーナーでは京開発者にしかわからない細かい点にまで教えていただきました。分子スケールチームの松永康佑さんからは分子動力学シミュレーション(MD)、特に長距離力の取り扱いの説明や開発中のPlatypus-MM/CGの紹介、高度化チームの大野洋介さんからは、特に短距離力におけるMDの高速化技法を「京」での実証例を挙げながら説明していただきました。臓器全身スケールチームの伊井仁志さんからは、オイラー型手法による血栓形成シミュレーションの話が、また、高度化チームの小山洋さんからは、微分方程式の数値解法における誤差とその原因についてお話しがありました。

ISLiM若手研究者の会合も第3回を迎えたこともあり、プロジェクトに関わるチーム間の交流を促すという当初の目的は達成されたのではと思います。今回は周知不足などによりうまくいきませんでした。今後は戦略分野や計算科学研究機構の研究者たちも参加者に取り込み、生命科学分野でのシミュレーション・大規模データ解析分野の若手コミュニティに向けた会合になることを期待しております。



「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 (ISLiM)」 開発アプリケーション紹介ページ、オープン

次世代計算科学研究開発プログラム
次世代生命体統合シミュレーション研究推進グループ

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2006年に始まり、「基礎方程式に基づく解析的アプローチと、大量の実験データから未知の法則に迫る実験データから解析へのアプローチ、さらには多階層を連結するアプローチにより、異なるスケールの研究と実験データを統合的かつ有機的に結びつけ、ペタスケールという桁違いの性能を持つスーパーコンピュータの性能をフルに発揮し、生体で起こる種々の現象を理解し医療に貢献するためのソフトウェアを開発する」を目標に、京速コンピュータ「京」でベタフロップス級の実効性能を実現するアプリケーション・プログラムの開発を行ってきました。この間、開発中に得られた成果は関係する学会や論文誌で発表を行うとともに、シンポジウムを開催し広く意見交換をしながら研究を進めてきました。

2011年度からは京速コンピュータ「京」での試験利用も始まり、開発しているプログラムの中から既にペタフロップス級の実効性能を達成したものが現れてきました。今後は、研究開発を行ってきたこれらのアプリケーション・プログラムを多くの方に知っていただき、さらには利用していただきたいと考えています。そこで、個々のプログラムを紹介するWebページをオープンしましたのでご覧下さい (図参照)。

このページでは現在開発を行っている全てのアプリケーション・プログラムを紹介しています。今後はプロジェクトの成果を社会に還元する一環として、ソース・コードなども順次公開していく予定です。そのため、これらのプログラムを実際に利用していただけるようダウンロードサイトも充実させていく予定です。

今回研究開発しているアプリケーション・プログラムに関しては「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 (ISLiM) 成果報告会 2011」(2011年12月21日 (水) - 22日 (木)) で研究開発状況が報告されます。ご興味のある方は是非ご参加下さい。

アプリケーション・プログラム紹介ページ

- 日本語ページ : http://www.csrp.riken.jp/application_j.html
- 英語ページ : http://www.csrp.riken.jp/application_e.html

「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 (ISLiM) 成果報告会 2011」案内ページ

http://www.csrp.riken.jp/2011/islim-houkokukai2011_j.html

「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」を紹介するビデオが下記で公開されています。こちらをご覧ください。

<http://www.riken.jp/r-navi/video/research.html>

ビデオ名: 生命の本質に迫る～次世代生命体統合シミュレーション研究開発～

The screenshot shows the ISLiM website interface. At the top, there's a header with the ISLiM logo and navigation links. Below that, the main content is organized into sections. The '開発アプリケーション' section lists various projects like '研究開発アプリケーション' and '分子スケール・アプリケーション'. The '細胞スケール・アプリケーション' section includes '細胞スケール研究の要となる 細胞シミュレーション統合プラットフォーム'. The 'データ駆動型アプリケーション' section features '脳がんの発生データに、生体におけるネットワーク構造と細胞シミュレーションモデルの検証 - データ駆動型アプローチ'. The '脳神経系アプリケーション' section includes '神経細胞内伝達-可塑性に関する分子ネットワークと人脳皮質構造情報に基づいた神経細胞シミュレーションプラットフォーム' and '脳神経系の全脳シミュレーション'. Each application entry has a small icon and a table with columns for code name, developer, and affiliation.

イベント情報

「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 (ISLiM)」 成果報告会2011

- 開催日：2011年12月21日(水)、22日(木)
- 場所：東京大学 武田ホール(東京都文京区本郷)

プログラム等詳細はwebページをご覧ください。

http://www.csrp.riken.jp/2011/islim-houkokukai2011_j.html

第4回バイオスーパーコンピューティング・シンポジウム

- 開催日：2012年3月5日(月)、6日(火)
- 場所：兵庫県神戸市内(予定)

詳細は決まり次第Webページ(<http://www.csrp.riken.jp>)に掲載します。

京速コンピュータ「京」が演算速度世界一を獲得

2011年6月に発表された世界のスーパーコンピュータ計算性能ランキング「第37回TOP500」で1位を獲得しました。整備途中段階(ピーク性能8.774ペタフロップス)の構成で、LINPACKベンチマークにおいて8.162ペタフロップス(1秒間に8,162兆回の浮動小数点演算性能)を記録しました。実行効率93%は、この規模のシステムでは驚異的な記録です。また、電力あたりの演算性能のランキングであるGreen500においても6位にランクインし、電力効率もトップクラスであることが認められています。

(表紙写真)

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

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

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、京速コンピュータ「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。

予測する生命科学・医療および創薬基盤

HPCI戦略プログラムは、京速コンピュータ「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。