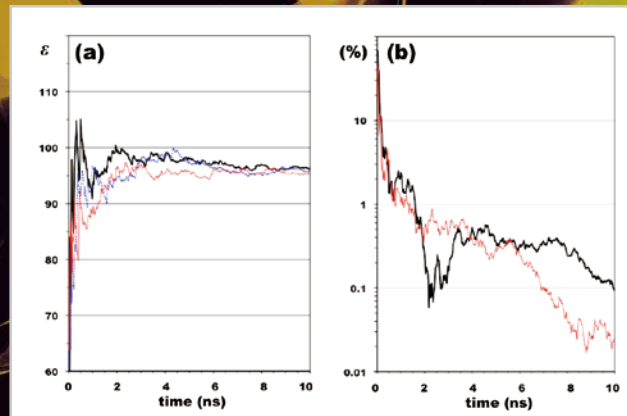
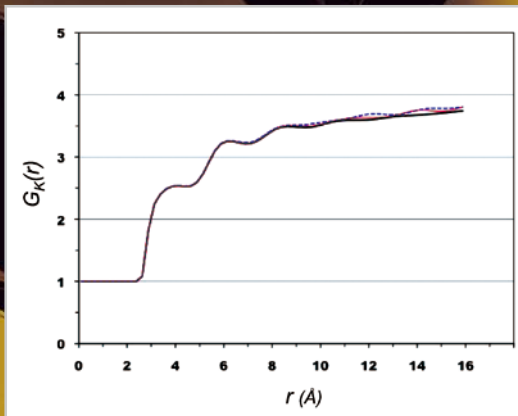


# BioSupercomputing Newsletter

2012.12 Vol.7



P6 図1：式[2]で定義される $G_K(r)$ 。黒い太線はPME法による値、赤い細線は $d_c=14 \text{ \AA}$ のZD法 ( $\alpha=0$ )、青い点線は $d_c=12 \text{ \AA}$ のZD法 ( $\alpha=0$ )による結果。パラメータ $\alpha$ については文献7)を参照。文献7)のFig. 5を改変。

P6 図2：(a) MD計算から得られる誘電率の累積平均の時間推移。(b) [1]式中括弧内第2項の第1項に対する比の時間推移を%として対数で表示したもの。線の意味は図1と同一。

## CONTENTS

●SPECIAL INTERVIEW	○「京」開発担当者に聞く今後のスパコン戦略とエクサスケールに向けた取り組み 富士通株式会社 テクニカルコンピューティングソリューション事業本部 エグゼクティブ・アーキテクト 奥田 基	2-3
	○「京」への展開で、実用に向けた最適化とさらなる規模の拡張を進める大規模仮想ライブラリ 東京大学大学院工学系研究科 化学システム工学専攻 教授 船津 公人	4-5
●研究報告	○水の誘電率計算から得られる古くて新しい問題 大阪大学蛋白質研究所 中村 春木 (分子スケールWG)	6
	○大規模並列計算用流体・構造連成解析プログラムの開発 理化学研究所 情報基盤センター 杉山 和靖 (臓器全身スケールWG)	7
	○スーパーコンピュータを用いた大規模遺伝子ネットワーク推定ソフトウェア SIGN 東京大学大学院情報理工学系研究科 玉田 嘉紀 (データ解析融合WG)	8
	○ISLIM研究開発ソフトウェアのソース・コード公開に向けた活動 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	9
●SPECIAL INTERVIEW	○「京」を用いた大規模シミュレーションによって細胞内分子ダイナミクスの理解と予測を実現する 理化学研究所 基幹研究所 杉田理論分子科学研究室 主任研究員 杉田 有治 (課題1 代表)	10-11
	○日本の優れたコンピュータ技術を活かして革新的な分子動力学創薬に挑戦 東京大学 先端科学技術研究センター 特任教授 藤谷 秀章 (課題2 代表)	12-13
●報告	○大学の新生に行った計算生命科学の講義 理化学研究所 HPCI計算生命科学推進プログラム 鎌田 知佐	14
●体制	○計算科学技術推進体制	15
●イベント情報		16

# 「京」開発担当者に聞く 今後のスパコン戦略と エクサスケールに向けた取り組み



富士通株式会社  
テクニカルコンピューティングソリューション事業本部 エグゼクティブ・アーキテクト

奥田 基

## ● 2015年ころには「京」クラスがオンサイトで 利用可能に!?

— 「京」の供用が開始されましたが、利用はプロポーザル制で誰もが自由に使える状況ではありません。10PFLOPSという計算性能を有するスパコンが、日本に「京」しかないの、プロポーザル制で行なわざるを得ないのですが、多くの研究者は自分が使いたいときに使える高度なスパコン環境を望んでいると思います。「京」の商用版である「FX10」の国内外での導入も進んでいるようですが、10ペタクラス導入の話は聞いていません。こうした状況のなか、研究者が10ペタクラスを自由に使える日はいつごろになるのか、また「京」を越えるスパコンの開発はどのように進められていくのか、課題はどこにあるのかなど、今後の動向についてお話をうかがわせてください。

奥田 これまでの世界のスパコン性能競争（TOP500第1位）の歴史を見ると、10年間で約1,000倍の性能向上がなされてきました。このペースで線を延ばしていくと、だいたい2018～20年ころに1EFLOPS（エクサフロップス）のマシンができるであろうと予測されており、日本もそれをめざしています。一方で、研究者が自由に使えるスパコンについて考えてみると、そのスパコンは各国に何台も導入されている必要があり、その性能はTOP性能のスパコンの1/10程度と考えられます。実際に、日本でも「京」の完成とほぼ同時に、東京大学情報基盤センターに「京」の約1/10のピーク性能1.13PFLOPSの新スパコンシステム（Oakleaf-FX）が導入されています。約3.3年で10倍の性能向上が見られますから、約3.3年経つと、現在のTOP500第1位と同等の性能を持つスパコンが各国に何力所か展開されると予想されます。すなわち、「京」と同等の性能を持つスパコンが大学や研究機関に入るようになるのは、2015年くらいではないかと予測されます。

— 逆にいうと、10ペタクラスのスパコンが何台か大学などに入る2015年くらいには、日本で100ペタクラスのスパコンが開発されるということですか。

奥田 ベンダー各社、それぞれ考え方や取り組みは異なりますが、私たち富士通がどんなことを考えているかというと、これまで「京」の開発を進めてきたことから、今後も国のプロジェクトに沿いながら、さまざまな製品の開発を進めていくという方向性を打ち出しています。これから本格的な運用が始まる「京」に対しても、運用支援や利用支援、アプリ最適化支援を行っていくと同時に、これから研究者のみなさんが「京」向けのアプリケーションをどんどん開発され、チューニングし最適化していくなかで蓄積されるソフトウェアの資産を、さらに発展させながら利用できる環境を継続的に提供していくことが、これからの私たちの役割であると思っています。商用機である「FX10」の提供もまさにそうですが、さらに2014～15年ころの製品化をめざして、エクサスケールへの繋ぎという意味で「Trans-Exa」と呼んでいます。「100PFLOPS級Trans-Exaシステム」を開発中です。これは、「京」や「FX10」のアプリケーションを、そのまま実行することができることを念頭に置いたマシンです。アーキテクチャも同様のコンセプトで、プログラミングモデルを変える必要もありません。もちろん、コンパイルし直してもらえれば、より高い性能を出すことができます。「FX10」は一つのCPUの性能が「京」の1.85倍（ピーク性能）上がっており、他にも運用性を向上できる機能が備わっていますが、「Trans-Exa」では、CPU性能、ネットワーク性能、

さらには実装密度、消費電力も大幅に改善したマシンを予定しています。もちろん100ペタまで拡張できるとはいえ、100ペタをそのまま導入できる場所は限られるでしょうが、それでも、数ペタクラスのスパコンがオンサイトで利用できる環境が、2015年ころには大学や研究機関に整うことになるはずですよ。

## ● エクサスケール実現への道り

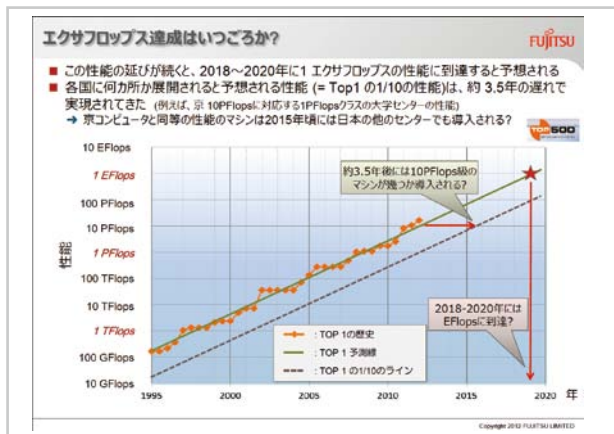
— スパコン開発における今後の技術動向については、どのようにお考えですか。

奥田 今年のTOP500で、非常に大きな技術的なジャンプがありました。それは第1位になった米国「Sequoia」の電力効率（電力あたりの性能）です。これまでのTOP1マシンのなかでは、「京」はいちばんよかったのですが、「Sequoia」は2,000GFLOPS/KW以上という圧倒的な差をつけて成績を塗り替えました。要するに、これまでにない非常に低消費電力のスパコンが誕生したということです。一方、CPUのなかにはコアと呼ばれる演算ユニットがありますが、TOP1マシンの1つのコアあたりの性能（LINPACK性能）は、この5年ほどの間、それほど大きく変わっていません。「京」は比較的高く、約15GFLOPSですが、「Sequoia」では逆に下がって、10GFLOPS程度です。つまり、現在はコアの性能を高めるというより、コアの数を増やすことによって性能を向上させる方向に動いているということです。

— 消費電力を下げる、コアの数を増やす、この2点がトレンドということですね。いい換えれば、それが今後のエクサスケールに向けた課題でもあるのでしょうか。

奥田 そうですね。先ほど「Trans-Exa」の話をしました。私たちは、これをエクサスケール実現に向けた技術開発の第1段階と位置付けて、さらに第2段階の研究開発を経てエクサスケールシステムを実現させたいと考えています。第1段階、つまり現在開発中の「Trans-Exa」では、1つのコアの性能を高めるとともに、さらなるマルチコア化を進めようとしています。「京」は8コアで、「FX10」は16コアですが、次はさらにコア数を増やしたい考えです。CPUの性能を高めることに伴い、当然ながらインターコネクタの高性能化も図ります。さらに低消費電力化も進める考えですが、実はこれがいちばん大きな課題です。それでも、「Sequoia」を越えることを目標に開発を続けているところです。あとは、実装密度、つまり面積あたりどれだけのCPUが搭載されているかということですが、これももっと高めて、高性能・高密度をめざしています。こうした第1段階を経て、さらにエクサスケールをめざすのですが、その間にはかなり大きな技術的なジャンプがあるだろうと考えています。先ほどの予測では、エクサ到達は2018～20年です。第1段階の開発からさらに3～5年あるわけです。コンピュータの世界で5年先のテクノロジーというのは、正直なところ、どうなるのか読めない部分があります。実際のところ、GPUの単なる演算性能の向上だけでは、方向性は見えていますが、低消費電力化に関しては、まだこれからという部分があります。開発中の100PFLOPS級のマシンではかなりメドがついているのですが、さらにその10倍となると、今のテクノロジーの延長では難しく、新しい技術の創出に期待しています。また、エクサスケールの実現には、さらなる高信頼化のための研究開発も必要です。アプリケーション開発においても、どのような形にしろ、エクサシステムではさらに並列度を高めざるを得ないわけですから、





**エクサスケール実現に向けた課題と取り組み**

- エクサスケールシステムの実現は現在の技術の延長では困難、少なくとも二段階の開発が必要
- 最大の課題は低消費電力化と高密度実装

■ 第一段階: 既に100ペタフロップス級システムを開発中

技術開発	■ さらなるCPUのマルチコア化やSIMD拡張
	■ さらなるインターコネクットの高性能、低消費電力化
	■ 高性能・高密度のメモリ技術

■ 第二段階: エクサスケールシステムの実現

研究開発	■ さらなる高性能、低消費電力化のための研究開発の推進
	■ 高信頼化のための研究開発の推進

Timeline: 2010 (No. 1 in Top500 (June, Nov, 2011) K computer), 2015 (100ペタフロップス級システム), 2020 (エクサスケールシステム)

エクサスケール実現に向けた課題と取り組み

エクサフロップス達成はいつごろか?

10年間で約1,000倍の性能向上というこれまでのスパコン性能の発達の流れが続くと、2018～20年には1EFLOPSに到達すると予測される。また、各国に数カ所展開される、TOP1の1/10程度の性能のスパコンは、約3.3年後にはその時代のTOP1の性能に到達する。

プログラミングモデルも変える必要が出てくるかもしれません。エクサスケールというのはまだまだ読み切れないところがあるわけです。ですから、100PFLOPS級システムが、次のエクサスケールに進むための準備をするためのプラットフォームになるのではないかと考えています。マルチコアが進み、SIMDが導入されたマシンをどうやって使いこなすかという研究開発を進めて、次に繋がっていただきたいと思っています。

● 世界に負けないスパコンの開発をめざす

「Trans-Exa」とされるマシンは、技術者にとっても研究者にとっても、次のエクサスケールマシンのテスト的な位置づけになるということですね。

奥田 「京」の場合も、プロジェクトの開始と同時にアプリケーションのプロジェクトが走り出しました。2011年度からHPCI戦略プログラムが動き出しましたが、後半で100PFLOPS級のマシンを使って、次に向けた準備を始めていただければ、エクサスケールで何をすればよいのかが見えてくるのではないのでしょうか。

かつてのように、ハードウェアが進化すれば、放っていても計算が早くなるという時代ではありませんからね。

奥田 「コ・デザイン」といわれるように、技術者と研究者が一緒になって、マシンとアプリを設計・開発していく時代だと思っています。「京」におけるグランドチャレンジアプリケーションの開発がそうであったように、事前の準備期間がないと、エクサスケールのマシンは完成したけれど、性能が活かせるアプリがないということになりかねません。「京」の場合は、4、5年前から準備が始まっていて、いろいろな分野で、これから続々とその成果が出ようとしている段階にきていますよね。2013、14年に優れた成果が出れば、次の研究開発への弾みになると思いますし、逆にいえば、そのときに次に進むためのマシンが、「京」のほかにも用意されているということが、とても重要なことだと思います。私たちとしても、そうした流れを描いて、100PFLOPS級マシンの開発を行っています。

ちょうど「京」の前に「FX1」が出て、それが「京」のアーキテクチャと似ているということで、いくつかの研究機関や大学にいち早く導入されたという経緯があったと思いますが、それと同じように、エクサスケールに向けた準備として「Trans-Exa」を活用してほしいということですね。

奥田 そのくらいの準備がないと、いくら性能の高いマシンができて、すぐには性能が出せないと思っています。

エクサスケール実現に向けた第2段階で、最大の課題は何でしょうか。

奥田 やはり、消費電力の低減がいちばんのポイントになると思います。CPU性能の向上に関しては、半導体テクノロジーが進んで、1つのチップのなかに演算回路を数多く入れる事は可能と考えられています。ただ、その高性能CPUの演算回路が全て動くと、消費電力がものすごくいいことになってしまい、効率的に動かすことができません。どうしても消費電力を下げるのが課題になってきます。電力をできるだけ喰わずに演算できる回路を考えることはとてもたいへんですが、そうしたことも取り組んでいかなければいけません。とにかく未知数な部分も含めて、さまざまな技術開発の積み重ねによってエクサスケールを実現させたい、実現させなければいけないと考えています。

開発者としては、より高いところをめざして進んでいく努力を止めることはできない。

奥田 もちろん、続けていかなければいけないと思っています。「京」を使い出した研究者の方から、一度「京」を使ったらもう元には戻れませんという話を聞いています。開発が始まったころは、「10ペタなんて必要な?」、「使えるアプリがあるの?」といった議論もあったと思いますが、5年たった状況は全く変わっている、そういうものではないでしょうか。それに、日本の研究者が他の国の人たちと一緒にプロジェクトをやろうとするときに、自分たちが「強み」を持っていなければ、対等に議論することもできませんよね。今、世界中の人々が、「京」に関心を持ち、「京」を使ってどんな研究成果が出るのかを見守っています。ですから、今後も世界に負けないマシンをつくり続けていかなければいけないと考えています。



富士通が商用機として開発した「PRIMEHPC FX10」

# 「京」への展開で、実利用に向けた最適化とさらなる規模の拡張を進める大規模仮想ライブラリ



東京大学大学院工学系研究科 化学システム工学専攻 教授

船津 公人

## ●創薬開発のカギとなる化合物ライブラリの現状

新薬の開発期間は十数年という長期間に及び、医療現場まで送り出される確率は数万分の1ともいわれています。そのため、開発コストも研究開発リスクも極めて高いのが実情です。こうした新薬の開発は、薬物標的の同定からスタートし、リード化合物を見つけ出し、それを最適化して活性をいいところまで持っていき、それからは臨床試験という流れになりますが、全体のなかで、成否のカギを握っているのは、化合物ライブラリからのリード化合物のスクリーニングに端を発する、初期段階の的確なリード化合物群の捕捉です。保有する化合物ライブラリの化学的な多様性と質、そして規模が、製薬会社の潜在的な開発能力を決定づけるといわれているのもそのためです。

では、化合物ライブラリの現状はどうなっているかという点、創薬研究の対象となる化合物の理論的な総数は10の60乗と推計されています。これに対して、製薬会社（メガファーマ）が保有する実在化合物のライブラリ数は、わずかに数百万化合物でしかなく、スクリーニングでヒットしなかったり、低活性の化合物しか得られなかったりと、漏れの大き

さが常に問題になっています。こうしたことから、スクリーニングのヒット率向上のためにも、化合物ライブラリの規模と多様性が強く求められ、より可能性のある化合物を網羅的に探索するという見地から、計算機上に構築されたバーチャルライブラリ（仮想化合物群）の活用には大きな期待が寄せられています。しかし、すでに存在するバーチャルライブラリを調べてみると、その備蓄件数は多くても数千万化合物程度であり、理論的総数と比較すれば、微々たるものでしかありません。これでは、いくらスクリーニングしても、ケミカルスペースのほんの一部しか見ていないことになり、大きな助けになっていないというのが実情です。さらに、バーチャルであるがゆえに、高スコアのリード化合物群を絞り込めたととしても、それらの合成検討に大きなコストが必要になるという問題も生じています。また、仮想化合物構造の創生を、原子種の組み合わせと各原子が取り得る結合回数に基づく理論操作だけに頼ると、せっかくのライブラリも、合成不可能な不安定化合物などを多く含むものになってしまう可能性があります。

## ●大規模仮想ライブラリの特徴とその概要

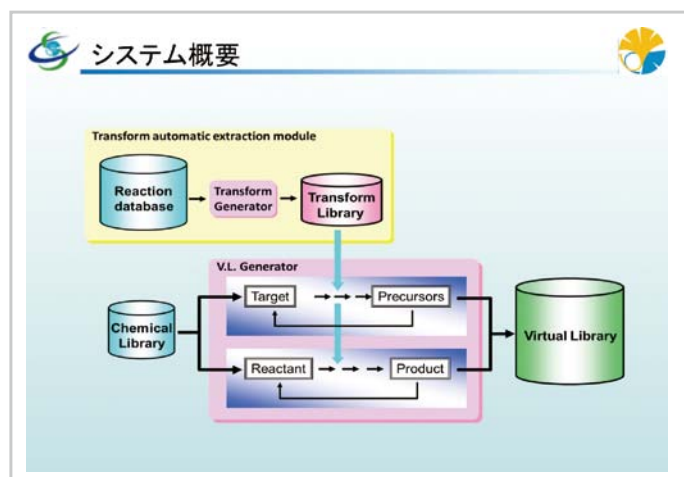
こうした問題を解決して、仮想化合物群の質や多様性、規模などの面で満足のゆく、これまでない新しいバーチャルライブラリを構築しようというのが、私たちが取り組んでいる「合成可能な化学構造および反応スキームからなる大規模仮想ライブラリの構築」です。ただ単に「大規模仮想ライブラリ」をつくるのではなく、「合成可能な化学構造および反応スキーム」によって構成されるライブラリであることが非常に重要です。化合物をつくるための合成経路も、一緒に含まれているわけです。もちろん、規模も大切です。私たちはドラッグライクネスと多様性を確保しながら、既存の数千万程度のライブラリを遥かに超える、10億から20億規模のバーチャルライブラリ構築をめざしています。

そのために開発したのは、42万種に及び既存化合物ライブラリから、

種構造を順次構造創出システムに投入し、トランスフォームという構造変換情報を適用して、新しい構造を作り出すシステムです。トランスフォームとは、反応データベースから抽出された、反応前後の反応部位の構造変化情報です。もっと分かりやすくいうと、既存の反応データベースから、各反応スキームの反応物と生成物の反応部位における結合回数の変化や構造環境変化などの情報、言い換えるなら「反応のエッセンス」を取り出して、それらを蓄積したデータベースです。

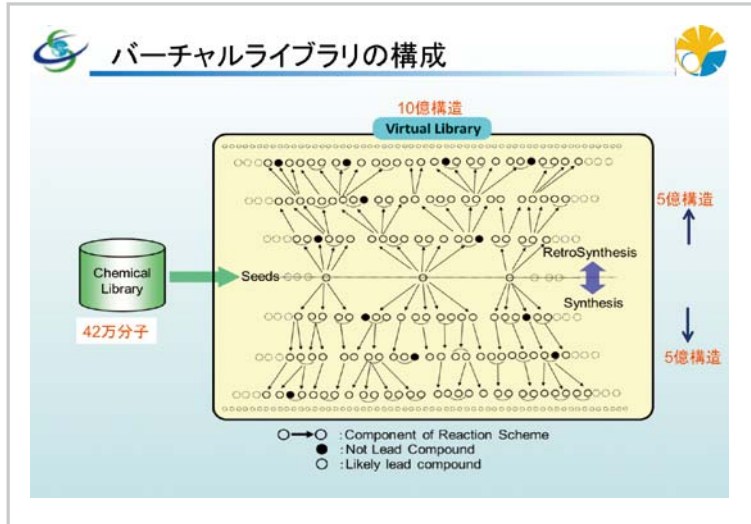
このトランスフォーム情報を適用し、反応物構造に対する生成物構造を反応スキームとして提示する順合成反応創出システムを連続運用することで、順合成ルートが付与されたバーチャルライブラリが構築されます。このバーチャルライブラリに含まれる化合物構造は、反応物構造と生成物構造との関連を保持した順合成ツリー構造を形成します。逆に、標的化合物構造に対する前駆体構造を反応スキームとして提示する逆合成反応創出システムを連続運用することで、逆合成ルートが付与されたバーチャルライブラリが構築されます。このバーチャルライブラリに含まれる化合物構造は、合成前駆体構造と反応生成物の関連を保持した逆合成ツリー構造を形成します。順合成のツリー領域であれば、反応物から生成物予測のスキームを示し、逆合成のツリー領域では、標的構造から反応部位を持つような合成前駆体を提案します。つまり、順反応方向だけでなく、何からつくればよいかという逆合成方向からの仮想化合物構造も、バーチャルライブラリに含まれることになります。

実際に既存の化合物ライブラリの種構造を投入すると、トランスフォームを適用することによって、いくつもの生成物構造の候補が出てきます。続いてそれらの生成物構造の候補を反応物構造としてトランスフォームを適用すると、さらに次の段階の生成物構造の候補が出てくる。一方では、これをつくるためには何からつくればよいかをたどっていくこともできる。また、なかには分子サイズが小さいもの、リード化合物の候補にならな



大規模仮想ライブラリのシステム概要





## 大規模仮想ライブラリの構成

矢印は仮想的な反応スキーム構成情報、白丸はリード化合物になり得る化合物、黒丸はリード化合物としてふさわしくない化合物を示す。ただ、合成経路をつなぐために、検索対象にはならないが、ライブラリを構成する構造となっている。順合成のツリー領域では反応物から生成物予測のスキームを表し、逆合成のツリー領域ではターゲット化合物から合成前駆体の提案を表している。

いと判断される、ドラッグライクネスを持たないものも出てきます。それは仮想ライブラリの検索対象にはなりません、合成ルートをつなぐという意味では必要な情報ですのでライブラリに含まれています。

今年度も、出力構造を再帰的に入力構造に置き換え、多段スキーム発生による規模の拡大を継続しており、バーチャルライブラリ全体として、

順方向と逆方向の反応スキーム例を含めて、重複しないユニークな化合物10億件を納めることを目標にしています。さらに、その先へ延ばしたり、最初の種となる構造を追加することで、20億件のバーチャルライブラリも可能ではないかと考えています。

## ● ライブラリ創出エンジンの評価

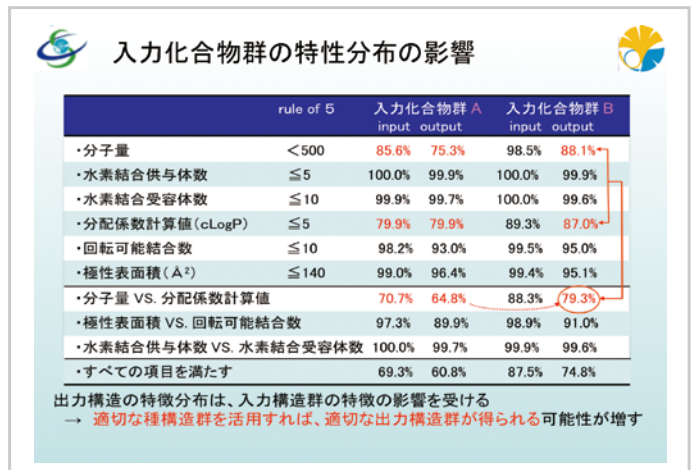
さらなる開発を続ける一方で、このバーチャルライブラリ創出エンジンの特性や出力された化合物群を把握するための評価も行っています。全てではありませんが、種構造群として利用した42万構造から1500万化合物を発生させた段階で、重複を排除したユニークな化学構造は630万種を数えました。出力化合物の重複率は58%で、半分をやや上回る程度でした。新規性については、この630万種の発生化合物群を、1500万件を含む購入可能な既存の化合物ライブラリと対比した結果、市販化合物との重複は、わずかに1.33%でしかありませんでした。したがって、このシステムによって出力された構造は、そのほとんどが新規化合物であり、新規性は十分に確保されていると判断しています。

さらに、入力化合物群の特性分布の影響の検討も行っています。バーチャルライブラリは入力した種構造から発生するわけですが、その発生した化学構造が薬物として意味があるかどうかを判断する基準に、ADMIT特性(吸収、分布、代謝、排泄、毒性)があります。この特性予測を算出することによって、その化学構造を事前に評価できるわけです。例えば、経口投与された場合、体内に吸収されなければ、薬として働きません。薬は有機物で、しかも結構大きな分子量ですから、基本的に水に溶けません。したがって、当然ながら吸収されにくいわけです。それでは具合が悪いので、水に溶ける特性をある程度保持している必要があります。溶ける、溶けないだけでなく、極性表面積が大きければ溶けやすいとか、水素結合の供与体数や受容体数が多ければ溶けやすさを助長するだろうといった、いろいろな特性で評価できるわけです。そうした吸収に関して予測するための経験則が「リビンスキーの規則(ルール・オブ・ファイブ)」です。今回、この「リビンスキーの規則」に示される各特性値の分布の検討を行いました。その結果、本システムが出力する仮想化学構造は、入力化合物群の特徴を引き継ぎ、その特性値の分布を拡張しつつ、各特性値指標の適合率の高い化合物群を種とすると、指標適合率の高い化学構造群を出力することが確認できました。つまり、種構造群を適切に選択すれば、医薬品としての適性がある仮想化合物を、高い確率で出力できることを示しています。

今後は、実際に「京」の上にこの大規模仮想ライブラリを載せて、一

般の利用者も含めて活用していくこととなります。スクリーニングのためのソフトウェアは別のグループで開発していますので、私たちは大規模仮想ライブラリを提供していくわけです。また、ライブラリの要素となる化学構造と反応スキームを創出するためのライブラリ創出エンジンを「京」に載せておけば、今後は「京」のユーザーになる製薬会社の方々が、自分たちの所有している化合物ライブラリからバーチャルライブラリを構築させることも可能になります。このライブラリ創出エンジンそのものは、本プロジェクト以前に船津研究室で開発済みですが、希望も多いことから、今後この創出エンジンについても提供の準備をしていくことになると思います。

私たちとしては、この道具立てはほぼできたと思っています。これからはユーザーの方たちの具体的な希望なども聞きながら、実際の創薬に向けた実利用を進めていくために、今後、大規模仮想ライブラリをどのように活用していけばよいのかを考える、そうした新しいフェーズに入ったといえるでしょう。



リビンスキー規則(ルール・オブ・ファイブ)を指標とした入力化合物群の特性分布の影響の評価の一例。

# 水の誘電率計算から得られる古くて新しい問題



大阪大学蛋白質研究所

中村 春木

(分子スケールWG)

「水は真空に比べて80倍ほどの大きな誘電率を持ち、大きな分極効果が観測される」と教科書には書かれています。この純水の誘電率の分子シミュレーションによる計算には長い歴史があり、最近になって問題点がクリアになったこともあります。この古くて新しい問題を振り返ることにより、分子だけでなく、シミュレーション計算全般において我々が注意すべきことがあるようにも思い、以下に紹介をいたします。

まず、ある純水の系の誘電率 $\epsilon$ は、その系にある各水分子 $\{i\}$ の電気双極子モーメント $\{\vec{\mu}_i\}$ の総和の統計平均により、 $\epsilon = 1 + 4\pi N\mu_0^2 G_K / 3k_B T V$ として算出されます。ここで、 $\mu_0$ は一つの水分子がもつ電気双極子モーメントであり $G_K$ はKirkwood因子と呼ばれ、

$$G_K = \frac{1}{N} \left[ \left\langle \sum_i \frac{\vec{\mu}_i}{\mu_0} \sum_j \frac{\vec{\mu}_j}{\mu_0} \right\rangle_{ensemble} - \left\langle \sum_i \frac{\vec{\mu}_i}{\mu_0} \right\rangle_{ensemble}^2 \right] \quad [1]$$

で定義されるスカラー量です。一方、Distance dependent Kirkwood因子と呼ばれる

$$G_K(r) = \frac{1}{N} \left\langle \sum_i \left( \frac{\vec{\mu}_i}{\mu_0} \sum_{\substack{j \\ r_{ij} < r}} \frac{\vec{\mu}_j}{\mu_0} \right) \right\rangle_{ensemble} \quad [2]$$

は、動径分布関数に比べ極めて敏感に水分子の配向と構造を反映する重要な指針を与えます(図1<表紙>)。ある水分子を取り巻く第1層目、第2層目の水分子の電気双極子モーメントとの相関、さらに離れた層との相関が見て取れます。式[1]第2項は十分大きな統計をとるとゼロとなる部分ですので、 $r$ を無限大にした時の式[2]の $G_K(r)$ が、誘電率に対応する $G_K$ に一致します。すなわち、図1の右端の十分遠方の部分での値が誘電率に対応し、標準的な計算法である周期境界条件を使ったPME (particle-mesh Ewald)法での1気圧、300Kでの誘電率の値は96程度になります。80からのずれは、用いたTIP3Pの水分子モデル<sup>1)</sup>に依存すると考えられます。

ところで、液体の純水では水素結合ネットワークができており、そのダイナミクスとしての緩和は遅いため、誘電率の値が収束するには図2(a)<表紙>に見られるように1~2ns程度の短い計算では全く不足、少なくとも6ns程度以上の長いシミュレーションが必要です。しかし、この現象は2011年のGereben & Pusztai論文<sup>2)</sup>により初めて系統的に指摘されたもので、それ以前の他の論文における短い計算時間での値は信用なりません。私どもも自ら計算していて、数nsのトラジェクトリでは同一の計算手法であっても誘電率の値が異なることが多く、悩んでいた問題でもあったため、この論文の指摘によって極めて明快に問題がクリアできました。さらに、式[1]の $G_K$ の第2項についても図2(b)<表紙>のように1~2ns程度では時間平均としての値が無視できるほど小さくなっていないことがわかりました。以上は、計算資源が豊富になったために明らかにされた点です。

一方、もう一つの問題として、純水における遠距離的な静電力の取り扱いがあげられます。今ではPME法が標準的な手法となっていますが、以前は計算資源の問題から、あるカットオフ距離 $d_c$ 内の水分子同士の相互作用のみを考慮する方法がよく使われていました。その際、単純に考え

るとなるべく大きな $d_c$ をとれば良いと思われそうですが、Yonetaniの論文<sup>3)</sup>により、たとえ $d_c=18\text{\AA}$ と大きくしても $G_K(r)$ は正負に一桁ほど大きく振れてすれてしまい、単に定量的に誘電率が異なるだけでなく、定性的な水の構造も正しく再現できないことが示されています。

歴史的には、Neumann<sup>4)</sup>によって、 $d_c$ を半径とする球の外側に誘電率 $\epsilon_{RF}$ の誘電体があるとし、そこからFröhlichによる反作用場(Reaction field)<sup>5)</sup>を受けるとしたReaction field法によって、上記のカットオフによるartefactを取り去る手法が提案されています。この手法はその後多くの研究者によって試みられましたが、蛋白質水溶液のような均一でない系では、パラメータとして与えるべき $\epsilon_{RF}$ の見積もりが困難なこともあり、近年は盛んではありません。一方、最近、理研の福田育夫博士らによりNon-Ewald法の一つとして提案されたZero-dipole summation (ZD)法<sup>6-8)</sup>は、Wolfによる電荷の中性条件<sup>9)</sup>だけでなく電気双極子モーメントの中性条件も課すことによって遠距離力の効果を繰り込み、簡潔なアルゴリズムでありながら高い精度の計算を実現する優れた手法です。この手法を用いると、 $d_c=12 \sim 14\text{\AA}$ という常識的には近距離的な相互作用しか考えていないような短い距離で相互作用をカットオフしても、図1および図2に示されるように、PME法とほとんど同一の誘電率や $G_K(r)$ が得られます。興味深いことに、このZD法は、ある条件ではReaction field法における $\epsilon_{RF} \rightarrow \infty$ の場合と全く同じ式を与えることが示されており、また最近提案されているその他の種々のNon-Ewald法とも共通の性質があることがわかっています<sup>9)</sup>。周期境界条件を課さない手法には、基盤となる共通の物理があるものと思われま

ところで、計算科学では、既に定まったアルゴリズムを基にどれだけ高速に計算するかが競われることも多いと思います。もちろんそれは必要なことですが、アルゴリズムあるいはモデルそのものから考え直すことによって、全く新たな世界が開ける可能性もあると思います。現在、私たちは、上記したZD法を、周期境界系としてではなく3次元トラス系としての蛋白質やDNAの水溶液などのヘテロな系に対して応用する研究に取り組んでおり、それなりの良い結果が得られ始めています。

新しいアルゴリズムやモデルを用いる計算の研究では、「天動説」に凝り固まったレビューアとの戦いが必ずおきて、論文を出版する際には苦労します。しかし、研究が成功した場合の波及効果も大きく、実際、周期境界系を使わずに計算ができることは、多くの生体超分子の高速シミュレーションを、より容易にかつ計算資源もより少なく具現化できることにつながります。

最後に、ここで紹介した研究は、福田育夫博士(理研)、神谷成敏博士(阪大蛋白質研)、米澤康滋博士(近畿大)との共同研究によるものです。皆様

## 【参考文献】

- 1) W. L. Jorgensen et al., *J. Chem. Phys.* 79, 926 (1983);
- 2) O. Gereben, L. Pusztai, *Chem. Phys. Lett.* 507, 80 (2011);
- 3) Y. Yonetani, *J. Chem. Phys.* 124, 204501 (2006);
- 4) M. Neumann, *Mol. Phys.* 50, 841 (1983);
- 5) H. Fröhlich, "Theory of Dielectrics" Clarendon Press (1958);
- 6) I. Fukuda et al., *J. Chem. Phys.* 134, 164107 (2011);
- 7) I. Fukuda et al., *J. Chem. Phys.* 137, 054314 (2012);
- 8) I. Fukuda, H. Nakamura, *Biophys. Rev.* 4, 161 (2012);
- 9) D. Wolf et al., *J. Chem. Phys.* 110, 8254 (1999)



# 大規模並列計算用流体・構造連成解析プログラムの開発



理化学研究所 情報基盤センター  
 東京大学 大学院工学系研究科 (9月30日まで)

杉山 和靖  
 (臓器全身スケールWG)

血流は、健康な状態を維持する機能(止血、物質輸送、異物除去、体温調整)を担っています。例えば、血管壁が傷つくと、血小板の付着がきっかけとなって血栓が形成し、血管が修復されます。一方、何らかの要因で血栓が過度に成長し、血管が閉塞すると、後遺症や死に至る危険性の高い心疾患、脳血管疾患が引き起こされます。論理的根拠がしっかりと計算物理の方法論によって血液の正常・異常を正しく予測できれば、治療や創薬の進歩に活かせるでしょう。我々の研究グループでは、連続体レベルでの血流現象に注目して、流体・構造の力学的作用を結びつける連成解析プログラム(ZZ-EFSIコード)を開発しています。

血液には、柔軟に変形する血球が大量に存在します。径が数百 $\mu\text{m}$ 以下の細い血管内では、赤血球の変形能や、高濃度の粒子流れとしての性質が、血流の機能に大きく影響します。血液・血球の動力学を記述する原理原則(保存則、粘性・弾性を記述する構成式)は単純です。ただし、系には多様な時空間スケールの現象が内包されるため、複雑な挙動を示します。ZZ-EFSIコード開発の狙いは、単純な原理原則に基づいて、大規模計算を実現することです。その方策として、「京」の性能を引き出すことを念頭に置き、通常の連成解析コードをチューニングするのではなく、新たに実装すべき式を定め、計算スキーム、アルゴリズムを見直してきました [1, 2]。

「京」を始めとする近年のスカラー型スーパーコンピュータの特徴として、並列処理の階層性(計算機どうしの通信を伴うMPI並列、計算ノード内のコア間のスレッド並列、コア内の多重演算)が挙げられます。我々は、メッシュの生成・再構成を必要としないオイラー型(空間的に固定した点で全ての物理量を更新)の計算手法を開発しました。計算プログラムでは、矩形の計算領域をx, y, z方向に賽の目状の格子に分割することによって、式の記述やMPI領域分割を行なっています。これは、どの階層のハードウェア構成とも相性が良く、計算規模を拡張するのに優れています。一般的な流体のアプリケーションは、演算量に対するメモリアクセス頻度が高い傾向にあります。演算処理に比べてメモリの読み書きの遅いスカラー機では、そのため、演算待ちの状態が頻発し、「京」での実行効率(理論ピーク性能に対する実効計算速度)は、せいぜい10%程度に留まります。我々は、メモリアクセス頻度の低く済むアルゴリズムを開発して、計算の高速化をはかりました。「京」で実施した流体・構造連成計算の性能を図1に示します。単体ノードでの実行効率は46.6%であり、スカラー機での連続体力学計算としては、十分に高いと言えます。また、並列台数の増加に伴う効率の変化が小さいことから、線形拡張性が高いと言えます。82,944ノードを用いて、 $6.96 \times 10^{11}$ 格子点数で約500万個の分散体を含む系を計算したところ、実効計算速度4.54PetaFlopsの性能を達成しました。

これまで、赤血球、血小板を含む脳細動脈内流れを模擬した計算を実施しました(図2(a))。いくつかの血小板の軌跡

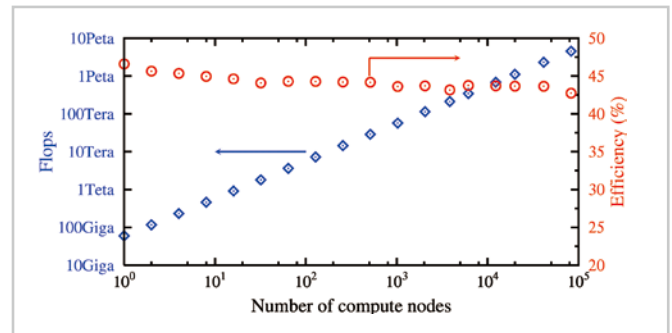


図1: 「京」でのZZ-EFSIコードのウィークスケーリング並列性能(左軸: 実効計算速度。右軸: 実行効率)。計算ノードあたりの格子点数は $512 \times 128 \times 128$ 。

を図2(b), (c)に示します。赤血球のない条件(図2(b))では、各々の血小板の半径座標の変化が小さく、血小板は管軸に沿ってほぼ真っすぐに進むのに対して、赤血球のある条件(図2(c))では、半径座標の変化が大きく、分散しやすいことが確認できます。この結果は、赤血球群が流体を攪乱すると、血小板が大きく揺らぎ、その血管壁へ近づく機会が増えることを表します。血小板血栓形成に対する赤血球の攪拌効果の重要性を示唆する実験と辻褄が合っています。

現在、損傷血管壁への血小板接着のモデル [2] を導入しており、血小板血栓形成に関する実験的知見の実証を目指しています。将来的な展開について、患者個別の情報に基づく薬効評価が可能となれば、魅力的な医療の創出につながるでしょう。その実現のため、物性値の変化、凝固・溶解過程、生化学反応などのモデル化を今後の課題と考えています。

【参考文献】

- [1] BioSupercomputing Newsletter, Vol. 2, p. 11.
- [2] BioSupercomputing Newsletter, Vol. 6, p. 2-3.

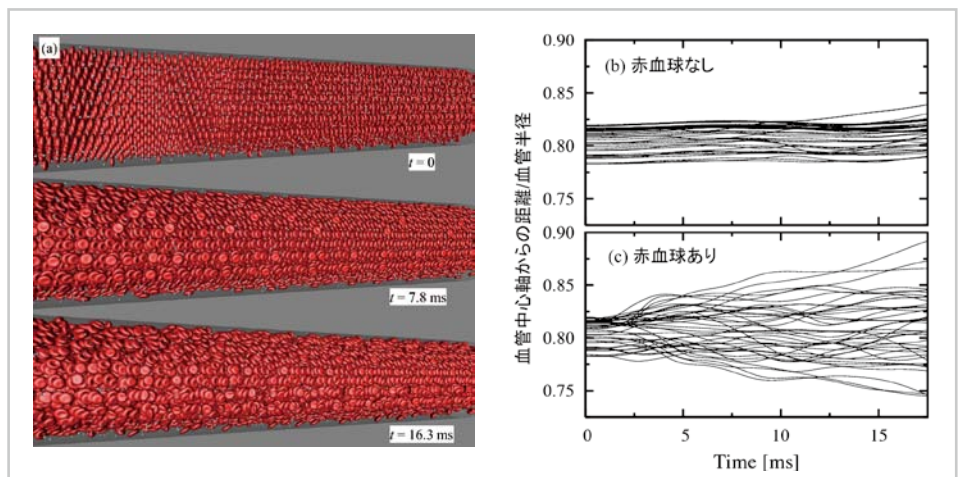


図2: 径約 $100 \mu\text{m}$ の血管内における多数の分散体の計算結果。(a): 血球分布のスナップショット(赤: 赤血球。白: 血小板)。流れは、左から右。(b), (c): 血小板の半径座標(血管中心軸からの距離)の時間変化。赤血球の有無による血小板の軌跡の違い。

# スーパーコンピュータを用いた大規模遺伝子ネットワーク推定ソフトウェア SiGN



東京大学大学院情報理工学系研究科

玉田 嘉紀

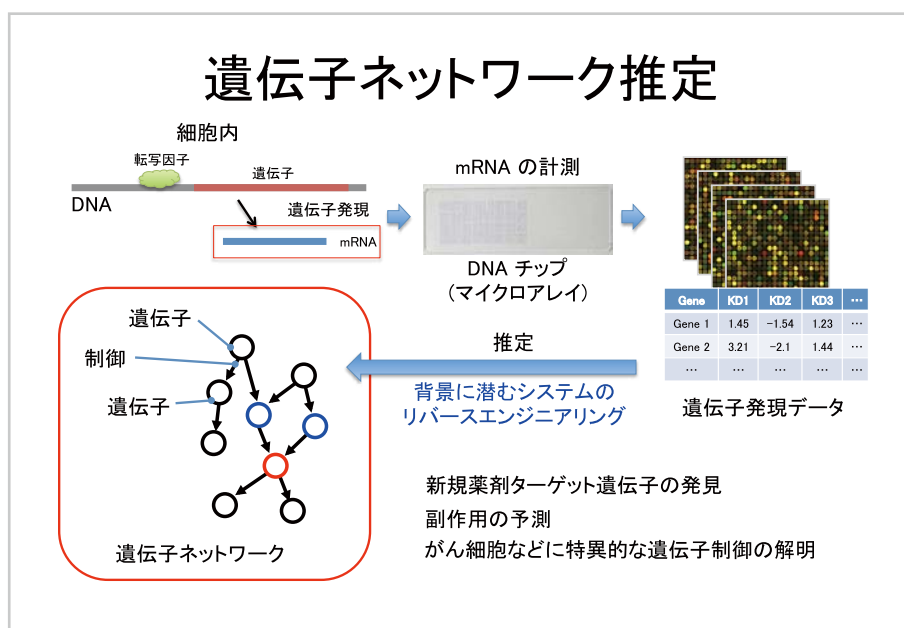
(データ解析融合WG)

ヒトの細胞にはおよそ2万から3万の遺伝子があるといわれています。ヒトの体はそのほとんどがタンパクでできていて、遺伝子は細胞内でのようなタンパクを作るのかという設計図に当たります。どのタンパクをいつどのくらい作ればいいのかという制御も専用の遺伝子が行っており、その遺伝子(=タンパク)もまた別の遺伝子によって制御されています。つまり遺伝子同士が複雑な制御のネットワークを形成しています。このネットワークはまだほとんどが解明されていません。同じヒトの細胞でも臓器によってネットワークは違います。また薬によっても変化し、がん細胞ではネットワークが壊れてしまっています。このような遺伝子の制御のネットワーク(=遺伝子ネットワーク)を計測可能なデータから数学的・統計学的・情報科学的な方法によって予測・推定しようというのが遺伝子ネットワーク推定です。現在の技術では細胞内で生成されているタンパク全てを計測することはできませんが、タンパクが生成される前段階で合成されるmRNAならば全ての遺伝子についてその量を計測することができます。このように計測したデータを遺伝子発現データといいます。1回の計測で得られるデータは、細胞のある状態の一瞬を捕らえたスナップショットです。この1回の計測データからでは遺伝子間の制御を予測・推定することは不可能で大量のデータが必要です。従って細胞に様々な刺激を与えたり、特定の病気の患者の細胞を集めたり、あるいは一定時間ごとに時系列にデータを計測したり、といったことで推定に必要なデータを集めます。遺伝子ネットワークを予測・推定することにより、これまで時間を掛けて一つ一つ遺伝子を探し実験を繰り返してきた遺伝子間の制御関係の解明を、計算によって網羅的に行うことで、新しい薬の開発、がん特異的な遺伝子の同定やその機能の解明が効率良くできるようになることが期待されています。

SiGN(サイン)は遺伝子発現データからスーパーコンピュータを用いて遺伝子ネットワークを推定するソフトウェアです。遺伝子ネットワー

クとして様々なモデルが提案されていますが、それぞれ一長一短があり最良のものというのはありません。またモデルが決まってもデータからそのパラメータを推定する方法にも複数あり、これまたどれも一長一短です。SiGNではスーパーコンピュータでの計算を前提とした膨大な計算時間の必要な複数の遺伝子ネットワークモデルと複数の推定アルゴリズムを実装した遺伝子ネットワーク推定ソフトウェアになっています。具体的には、SiGNは、静的・動的ベイジアンネットワークを用いたSiGN-BN、状態空間モデル(State Space Model: SSM)を用いたSiGN-SSM、L1正則化法によるパラメータ推定法を実装したSiGN-L1の3つのサブプログラムから構成されています。SiGN-BNはNNSR法という新しいアルゴリズムを搭載しており、これまで1000遺伝子程度が限界だったベイジアンネットワークを用いた遺伝子ネットワーク推定を全ゲノム(全遺伝子)に対しておこなう事ができるようになっています。SiGN-SSMは時系列データからシミュレーション可能な動的遺伝子ネットワークを推定しますが、ネットワークの構造自体は得られず全遺伝子間の関係の強度が数値で得られます。スーパーコンピュータによってこれまで難しかったネットワーク構造の計算が信頼度付きで計算可能になりました。L1正則化は元々大規模な遺伝子ネットワークに適用可能な方法ですが、個人の遺伝子発現の違いを考慮したネットワークを推定しようとすると、従来の方法では計算時間が足りませんでした。「京」を用いることによりこれも現実的な時間で計算することが可能になりました。

SiGNは「京」とヒトゲノム解析センターのスーパーコンピュータShirokaneを主なターゲットとして開発していて、Shirokaneではすでにいくつかのサブプログラムがインストールされユーザが自由に使える状態になっています。詳細はSiGNのウェブサイト<http://sign.hgc.jp>までどうぞ。





# ISLiM研究開発ソフトウェアの ソース・コード公開に向けた活動



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

## 1. 世界にも類のないソフトウェア研究開発プロジェクト

ISLiMプロジェクトでは、「京」の能力が発揮できるソフトウェアを開発し、優れた学术论文を発表すると共に、「京」で利用できるようにすることが目標とされています。

ISLiMが研究開発したソフトウェアの特長は、分子スケールから全身スケールまで、そしてシミュレーションからデータ解析までを約30個のソフトウェアで包括的に構成し、京をターゲットにそれらが高度にチューニングされていることです。ライフサイエンス/ヘルスケア分野でこのように包括的にソフトウェアを構築したのは世界でも類がなく、日本発のソフトウェア資産として研究用だけでなく教育用にも高い利用価値をもたらします。

## 2. ソース・コード公開にむけての活動

2010年後半から当プロジェクトでは、「次世代スパコンの創産産業利用促進研究会」と協力して、ソフトウェアの完成後の利用を念頭に国内の医薬品産業界と情報交流を進めてきました。その議論の中で、このような先進的なソフトウェアを公開する場合は、利用実績も豊富で迅速なサポート体制を提供できる市販ソフトのようなバイナリー・コードを提供するのではなく、利用者が自分でプログラムのソース・コードを確認し修正できるソース・コード公開の重要性が再認識されました。また「京」版だけでなく、企業で一般的に使われている「クラスター・システム版」のニーズも再確認しました。

### ●ソフトウェア開発責任者会議で目標推進

ソース・コード公開にあたっては、まずソフトウェア開発者がソース・コード公開の意義、公開に必要なプロセスなどについて具体的に理解し、公開にあたっての疑問点・懸念を払拭することが重要です。ISLiMでは、ソフトウェア開発責任者会議を新設し、趣旨の説明と議論を2011年11月9日、2012年2月21日、同7月23日の三回開催するとともに、ソフトウェアの知的所有権に詳しい本間高弘電気通信大学産学官連携センター 特任教授と、アンダーソン・毛利・友常法律事務所 重森一輝弁護士から貴重なアドバイスをいただいています。公開のための標準プロセスをプログラム開発責任者がわかるようにプロジェクト推進サイドからISLiM開発ソフトウェア公開準備のフロー

チャート」として提供し、進捗状況を図1の形で共有しています。

### ●ダウンロード・サイトから順に公開

ソース・コードを学界のみならず産業界にも広く利用していただくために、2011年にダウンロード・サイト ([http://www.islim.org/islim-dl\\_j.html](http://www.islim.org/islim-dl_j.html))を新設し、図2に示すように準備が整ったソフトウェアから順に公開してきました。ソフトウェア開発責任者会議で共有している目標は、2012年4月に全体の50%のソフトウェアを公開し、プロジェクト終了半年前の2012年10月に100%のソフトウェア公開です。その後の半年で、成果報告会、講習会などの普及活動をする予定です。

## 3. ソース・コード公開の現状と今後の課題

2006年に研究開発に着手したときの34個の独立したソフトウェアも、いくつかは開発最終段階で一つのソフトウェアに統合されるなどし、最終的には30個程度になる予定です。これらの最新の公開状況はダウンロード・サイトに示されていますのでご覧ください。「京」版と「クラスター・システム版」はコンパイラの指定等で切り替えられるようになっています。

多くの資源を投入して6年間研究開発してきた貴重な公開ソフトウェア資産ですが、プロジェクト終了後にどういう形で展開していくかを検討し、効果的に次へとボタンタッチしていくことが、今後の課題として残されています。

チーム: 分子スケール研究開発チーム		回答日: _____ (回答締め切り: 2012/6/20(水))						
各開発コード(京版+クラスター版全体)についての公開目標日記入と内容更新をお願いします。								
開発コード名	コード名	Step 1		Step 2	Step 3		Step 4	Step 5
		コード名は京版に"-K"採用	著作権者の確認	確認済み著作権者名	所属機関への届出手続	他コードが含まれているか	他コードが含まれている時の対応方法	公開ライセンス
Platypus-MM/CG (mu2lib-K) 公開目標日 (7/1)	<input checked="" type="checkbox"/> 使用する <input type="checkbox"/> しない(理由)	<input checked="" type="checkbox"/> 確認中 <input type="checkbox"/> 同意処理中 <input checked="" type="checkbox"/> 確認済み	理研	<input checked="" type="checkbox"/> 確認中 <input type="checkbox"/> 確認済み	<input checked="" type="checkbox"/> 確認中 <input type="checkbox"/> 含まれない <input type="checkbox"/> 含まれる	<input type="checkbox"/> ライセンスに従う <input type="checkbox"/> 使用許可を取得 <input checked="" type="checkbox"/> 検討中	<input checked="" type="checkbox"/> GPL <input type="checkbox"/> MIT <input type="checkbox"/> その他 ( )	マニュアル等ドキュメント <input type="checkbox"/> ドラフト未着手 <input type="checkbox"/> ドラフト作成中 <input checked="" type="checkbox"/> ドラフト有 <input type="checkbox"/> 完成版有
質問・コメント ( )								

図1: 進捗状況の共有

分子スケール・アプリケーション			
 <p>量子科学計算(QM)、分子動力学計算(MM)、粗視化モデル計算(CG)の手法を結合したQM/MM、MM/CG法によって、マルチスケールシミュレーションを実現するためのプログラム。</p>			
番号	アプリケーション	コード名	ダウンロード
M-1	マルチコピー・マルチスケール分子シミュレーション法 開発の基盤となるクラスライブラリ	mu2lib-K (開発コード名 Platypus-MM/CG)	<a href="#">ダウンロード</a>
M-2	レプリカ交換分子動力学計算インターフェイス	Platypus-REIN	<a href="#">公開予定</a>
M-3	全原子分子動力学計算	MARBLE	<a href="#">公開予定</a>
M-4	粗視化モデル計算	CafeMol	<a href="#">ダウンロード</a>
M-5	密度汎関数法に基づくタンパク質全電子波動関数計算	ProteinDF	<a href="#">ダウンロード</a>

図2: ダウンロード・サイト ([http://www.islim.org/islim-dl\\_j.html](http://www.islim.org/islim-dl_j.html))の一部

# 「京」を用いた大規模シミュレーションによって 細胞内分子ダイナミクスの理解と予測を実現する



理化学研究所 基幹研究所 杉田理論分子科学研究室 主任研究員

杉田 有治

(課題1 代表)

## ● 研究開発課題の全体像

生命科学におけるシミュレーション研究は、その方法論やアルゴリズムの開発の急速な発展に伴い、現在、最もアクティブな研究分野のひとつになっています。ゲノム情報、立体構造情報、細胞内でのタンパク質発現情報などの実験データが次々に生産されているなか、今後、さらにその重要性は高まっていくはずで。また、これまで、生命科学はデータ重視の学問でしたが、これからはそのデータを統合して、生命体システムとしての理解を深める時期に来ています。その意味でも「京」の優れた計算能力によって実現するシミュレーションが、生命科学を予測や制御可能な新しい研究体系に変えていくために果たす役割は大きいといえます。

こうした状況のもと、HPCI戦略プログラム(戦略分野1)「予測する生命科学・医療および創薬基盤」研究がスタートしました。私たちが取り組む課題(細胞内分子ダイナミクスのシミュレーション)では、細胞環境を強く意識した分子・細胞スケールの大規模シミュレーションを行うことによって、細胞内分子ダイナミクスの理解と予測を実現することをめざしています。

これまで生体分子シミュレーションは数多く行われてきましたが、そのほとんどが、水溶液中あるいは脂質二重膜中におけるタンパク質やDNAなどの振る舞いを明らかにしようとするものでした。しかし、細胞内にあるタンパク質の数と細胞の大きさから計算すると細胞質内の環境が水溶液中の環境と大きく異なっていることは間違いなく、また、それは実験的にも明らかにされつつあります。ひとつのタンパク質が動く現場のすぐ近くには他のタンパク質などが複数存在しているわけ。そ

のような環境(細胞内分子混雑環境)が、タンパク質の構造やその安定性、そして機能にどのように影響しているのかはまだ完全に理解されているわけではありません。理論的には、タンパク質分子をひとつの粒子に近似したモデルを用いた計算や組み合わせ効果を統計力学的に調べた研究はあるものの、タンパク質とその周囲の環境を全原子レベルで取り扱った計算事例はほとんど存在していません。私たちは、今回のプロジェクトで、これまでにないスケールで細胞環境に近い条件でタンパク質の働く姿をとらえたいと考えています。

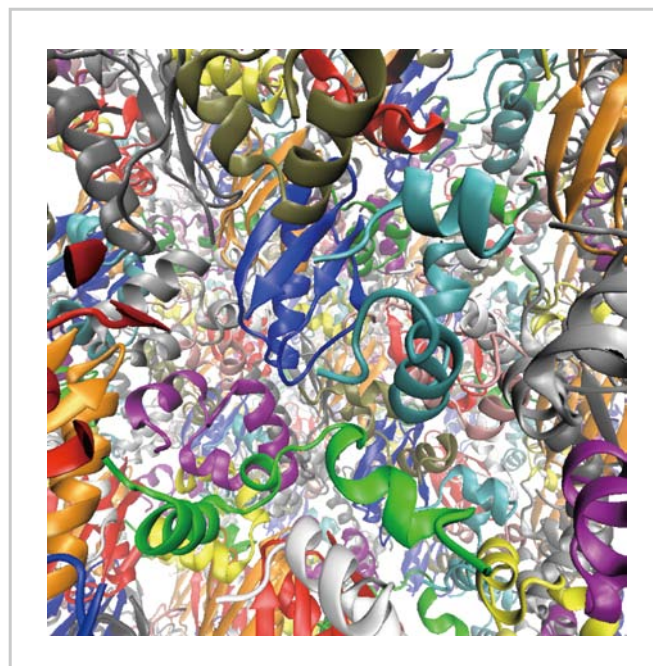
プロジェクトでは、次の3つのサブ課題を実施しています。ひとつは、一分子粒度シミュレーションと分子動力学計算の連携によって新たな研究分野を開拓する「細胞環境下での信号伝達経路のシミュレーション」、長時間の分子動力学計算に基づく自由エネルギー計算によって定量性・予測性を持つ分子シミュレーション達成をめざす「膜タンパク質による細胞膜を隔てた物質輸送」、さらに、核内DNAタンパク質の相互作用を全原子および粗視化分子動力学計算を用いて解析する「核内DNAタンパク質の動的構造と機能の解明」、この3つです。

生命現象を理解し、予測に結び付けていくためには、分子レベルの計算とシステムズバイオロジーを融合させることによって、「細胞まるごと」を理解することが欠かせません。そのため、プロジェクトの終盤では、研究開発によって得られた知見を活用して、「細胞まるごとシミュレーション」をめざした次の課題を明確にしたいと考えています。

## ● 分子シミュレーション研究の目標

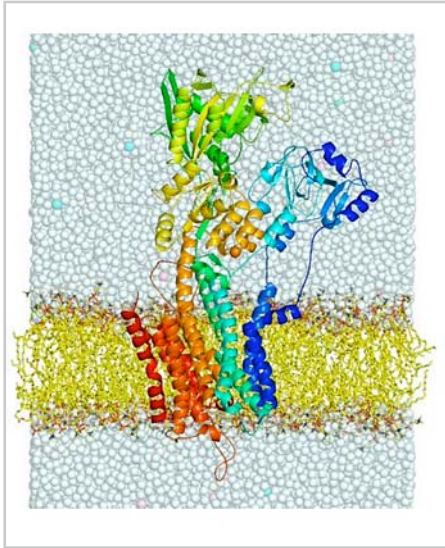
課題1「細胞内分子ダイナミクスの理解と予測」を実現するためには、二つの要素が必要です。ひとつは、タンパク質や核酸などの生体分子の“かなり遅い運動”のシミュレーションを実現させること。もうひとつは、分子スケールの研究(原子粒度)と細胞スケールの研究(分子粒度)とを接続し、細胞機能を分子の視点から明らかにしていくことです。

前者については、ミリ秒スケールの大規模な構造変化のシミュレーションに挑みたいと考えています。“遅い”といいましたが、それは私たち研究者の感覚で、実際にはものすごく速い分子運動です。しかし、現在行われている分子シミュレーション研究ではマイクロ秒程度が長時間ダイナミクスの時間スケールなので、その1,000倍も長い、ミリ秒の分子運動は遥かに“遅い”わけです。「京」が完成するまでに私たちが利用可能な計算環境は、約100TFLOPS程度でしたが、「京」の誕生によりおよそ100倍(「京」の演算性能は約10PFLOPS(10,000TFLOPS))強化されました。したがって、これまでと同じような計算機の利用方法で1,000倍長い計算を行うのは、実はかなり難しい数字です。そこで計算

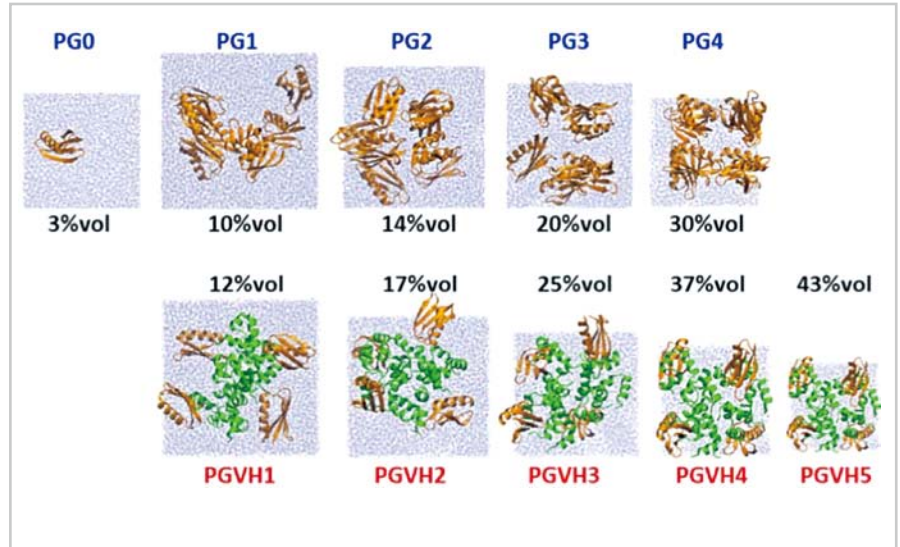


分子レベルの計算から「細胞まるごと」の理解への過程で、「細胞内分子ダイナミクス」の理解と予測は非常に重要な役割を持つ。





イオンポンプの生体膜中での分子ダイナミクスをシミュレーションによって見ていくためには、タンパク質、生体膜、水やイオンの約26万におよぶ原子数とともに、その分子間相互作用を含む膨大な計算が必要になる。



タンパク質濃度を変化させたときのシミュレーション。細胞質内でタンパク質分子などが混み合った状態を分子シミュレーションで再現し、水分子に与える影響が明らかにされた。

手法を何とか工夫して、ミリ秒相当のダイナミクスが見られるようにしようというわけです。

そのための高度な並列化技術の開発をすでに進めています。例えば、多次元レプリカ交換分子動力学計算法というアルゴリズムがあります。この手法ではレプリカと呼ばれる系のコピーについて異なる温度や異なるパラメタでの分子動力学計算を並列に実行し、ある頻度で温度やパラメタを交換することで計算を加速することができます。1個の分子動力学計算について数万CPUを用いた並列化を行うのは難しいのですが、レプリカ交換法では各レプリカの分子動力学計算を数百から数千CPUを用いて並列化し、レプリカを数十から数百用意することで、数万CPUを同時に効率よく利用することが可能になります。こうした方法をいろいろ積み重ねていかなければなりません。

これまでにない計算性能を持つ「京」を活用していくとはいえ、これまで説明したように、計算だけではどうしても限界があります。そのため、この研究課題を実施していくためには、実験とのコラボレーションが欠かせないと考えています。例えば、私たちは立体構造を扱っているので、X線結晶構造解析などによって立体構造が得られれば、それを分子動力学計算するというのがいちばん手取り早いわけです。また、NMR（核磁気共鳴）によるスペクトルは溶液中や細胞質中でのタンパク質のダイナミクスに関する情報を与えてくれます。私たちがミリ秒にこだわるのも、それが実現すれば、実験と直接比較ができるのではないかと考えているからです。実験の測定限界とシミュレーションがミリ秒でマッチし、長時間の実験と計算の時間スケールが一致するわけです。

## ● 分子ダイナミクスシミュレーションの成果

「京」の計算性能を効率よく活用していくためのソフトウェア開発や高度化を進める一方で、シミュレーション研究そのものも、すでにいくつも実施されており、成果も得られています。私たちが中心になって進めている細胞質内の分子混雑環境を考慮したシミュレーションもそのひとつです。水を露に含む全原子分子動力学計算を行うことによって、細胞質内と同じ程度に混み合う環境ではタンパク質の周囲に存在する水分子の性質が大きく異なっていることが明らかになっています。それにより、タンパク質間の相互作用も変化し、溶液中と異なるふるまいをしている可能性も示唆されました。細胞内環境での物理パラメタは、実験的に測定することは困難であるため、この計算結果は非常に有用であるといえます。また、今後は「京」を活用することによって、より大規模な系での分子混雑の解析が実現することになり、さらなる研究の進展が期待されます。

また、膜輸送体の分子シミュレーションなども進歩しています。従来はX線結晶構造が解明されてから、分子シミュレーションによるダイナミクスを解明することができるようになるまでしばらく時間がかかってい

ましたが、最近X線結晶構造が解けてすぐに脂質二重膜も含めた分子動力学計算が実行できるようになってきました。すでに膜輸送体の輸送サイクルの一部で起こる大規模構造変化がシミュレーションを行うことにより、解明されつつあります。「京」を本格的に用いることにより、膜輸送体が行う膜を隔てた物質輸送をシミュレーションにより明らかにできる日も遠くないと期待しています。

私たちが取り組んでいるプロジェクトには、基礎科学としてだけでなく、創薬や医療への貢献が期待されていることはいうまでもありません。ただ、そうした成果が得られるようになるまでには、まだ時間が必要でしょう。それをめざす一方で、今、私たちが「京」を用いた研究開発で大切にしなければいけないのは、これまでの計算技術では全く不可能であった生命現象のシミュレーションにチャレンジし、その成果を積み上げていくことであろうと思います。それにより、計算科学を用いた生命科学を志す研究者や学生を増やし、彼らが活躍できる場をつくっていく、それも長期的に見れば、このプロジェクトのひとつの役割なのだろうと考えます。

## 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2 創薬応用シミュレーション

# 日本の優れたコンピュータ技術を活かして 革新的な分子動力学創薬に挑戦



東京大学 先端科学技術研究センター 特任教授  
藤谷 秀章  
(課題2 代表)

## ●実際に薬を創り出すことをめざす

「京」の登場によるコンピュータ能力の向上は、病気標的タンパク質に薬分子が作用し、結合するまでの分子動力学計算を可能にしました。これにより、タンパク質の形状変化を原子レベルで明らかにして薬設計を行うIT創薬革命が始まろうとしています。

私たちが取り組んでいるHPCI戦略プログラム戦略分野1の課題2「創薬応用シミュレーション」では、「京」の計算能力をフルに活用して、創薬プロセスを革新する新しいComputer Aided Drug Design (CADD) 技術を確認するとともに、実際に薬を創り出すことをめざしています。同じ分子シミュレーション分野でも、課題1では、より生物学的に重要な幅広い現象が扱われるのに対して、課題2では、扱う対象を創薬がターゲットにしている病気標的タンパク質に絞っているのが特徴といえるでしょう。

これまで、国の研究機関や大学では、独自に薬を創り出した経験がありませんでした。というのも、臨床試験まで含めて薬を開発しようとすると、200億、300億といった膨大な費用がかかるからです。さらに臨床試験の前の段階で、数多くの薬候補化合物を合成する設備においても、

残念ながら、多少実験できる程度しか持っていないところがほとんどです。また、抗体医薬の場合は薬といってもタンパク質であるため、基本的にはすべて大学などでも合成でき、製薬会社に頼らなくてもよいのですが、低分子化合物の場合は、考えたものを合成するということに非常にお金がかかってしまいます。ですから、開発段階から臨床試験まで、すべて大学や研究機関だけで行って薬を創るというのは、まず不可能というのが現状です。

こうしたことから、私たちのプロジェクトでは、はじめから製薬会社とタイアップし、共同研究的なかたちで開発を進めています。実際に創薬に結び付けるためには、製薬会社を巻き込まざるを得ないわけです。国の研究費を企業のために使うようにお感じになる方がおられるかもしれませんが、実際は全く逆で、研究開発段階では、企業側が完全な持ち出しになってしまいます。それでも共同研究に参加してもらうため、こちらも製薬会社に納得してもらえただけのシミュレーション結果を出さなければいけません。本当に薬を創り出すためには、どちらもリスクを背負いながら、真剣に取り組んでいくことになります。

## ●IT創薬の先駆者となることが重要

実はこの20年ほどの間、コンピュータシミュレーションによって薬を創ろうとする試みは何度も繰り返されてきましたが、未だ実現していません。その最大の理由は、タンパク質そのものを計算するだけの計算パワーが、これまでなかったからです。溶液のなかのひとつのタンパク質に、薬となる化合物が付いて機能を阻害する——こうした現象をシミュレーションするためには、少なくとも5万から10万原子、大きなものでは20万原子、さらには数百万オーダーの計算が必要になってきます。ひとつのタンパク質だけでなく、全体をシミュレーションしないと作用するかどうか分からないことも、最近明らかになってきました。そうすると、100万原子を簡単に超えてしまうのです。日本では「京」ができたことで、ようやく課題になっていた計算パワーを手に入れることができました。これまで標的タンパク質に付くか付かないか、経験的に進めてきたものが、すべてコンピュータによる計算で明らかになり、論理的に薬を開発する環境が整ったわけです。

現在、私たちが行っているIT創薬(図1)の取り組みは、ほぼ同時進行で欧米でも始まっています。米国のデヴィット・E・ショー氏が、分子

動力学専用計算機(ANTON)をつくって、メガファーマ(巨大製薬会社)とともにIT創薬を進めていることをご存じの方々も多いことでしょう。こうしたことによって、今まで半信半疑だった日本の製薬会社も、IT創薬によく目を向けはじめています。しかし、世界的にもまだ成功例がないために、自ら投資して取り組むという段階には至っていません。そこを私たちが開拓していこうとしているわけです。

大切なのは、世界と同じスタートラインに立つ今、先駆者として走り出すことです。製薬会社の研究者とともに、実際の創薬に向けた取り組みを通して、何を計算し、どんな結果を取って、それをどのように化合物の設計に活かすかを一緒にやっていくなかで、IT創薬の研究開発の裾野を広げていくことができるはず。実際に薬ができることが理解できれば、恐らく5年後くらいには、「京」クラスのスパコンが大手の製薬会社に導入され、さらに研究が進むことでしょう。そうした時代を拓くことが、ある意味でこのプロジェクトの最終的なゴールでもあり、「京」としての重要なミッションのひとつであると考えています。

## ●「京」の高い計算能力で高精度に薬効を予測

多くの医薬品はタンパク質を標的としており、より効果的な薬を見出すためには、生体内の標的タンパク質と強く相互作用する化合物(リガンド)を探し出す必要があります。私たちは、スーパーコンピュータを用いた分子動力学計算により、標的タンパク質と薬の候補である化合物を含む系のシミュレーションを行い、タンパク質と化合物間の相互作用を調べ、標的タンパク質だけに強く作用する新しい化合物を設計することにより、短時間で効率よく低分子医薬品を開発することをめざして

います。

そのために私たちは、ジャルジンスキーが1997年に発見した自由エネルギー差と非平衡仕事量の関係式を用いて、結合自由エネルギーを計算するアルゴリズムであるMP-CAFE法を考案しました(図2)。これは、化合物の他分子に対する相互作用が存在する状態から、相互作用が全く消滅して離れる仮想状態までの複数の中間状態に関して分子動力学計算を実行し、隣の状態に移行するために必要な仕事量から結合自由エネルギー



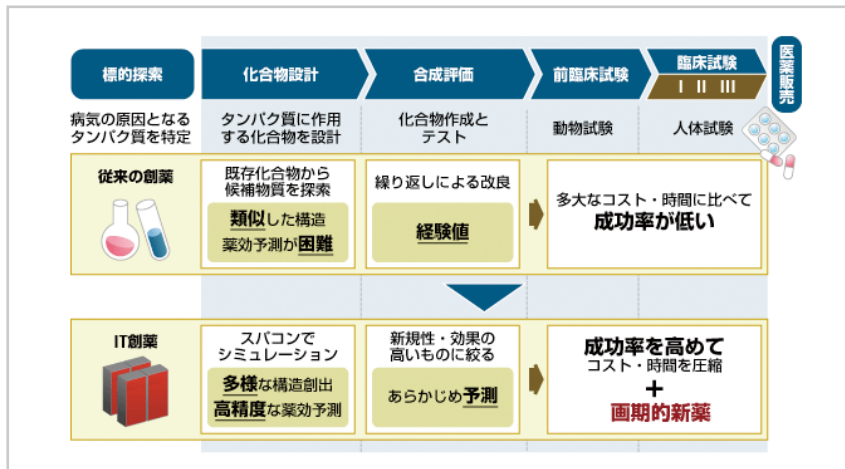


図1：従来の創薬とIT創薬の違い  
図表提供／富士通株式会社

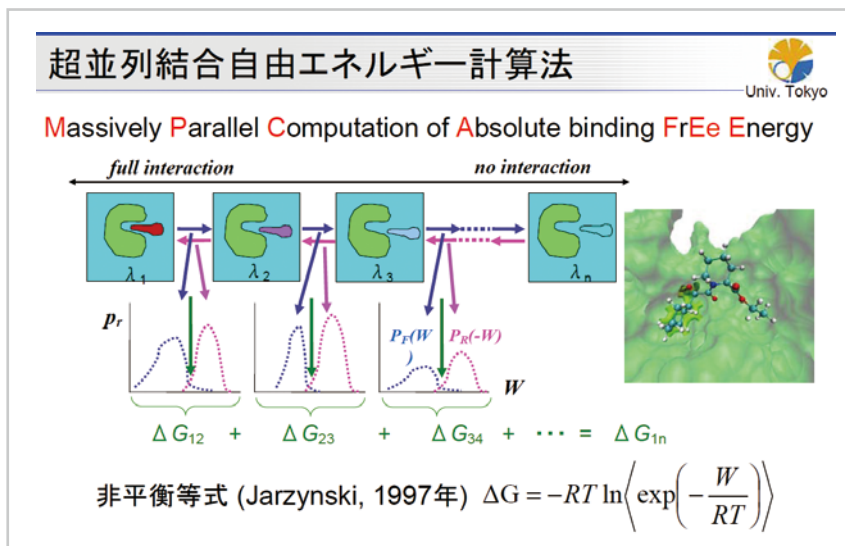


図2：超並列結合自由エネルギー計算法 MP-CAFEE法  
タンパク質と化合物(薬)が溶液中で熱運動している状態において、薬の相互作用パラメータを変化させながら大量の分子動力学計算を実行し、パラメータが変化するときの仕事量分布から標的タンパク質と薬の結合自由エネルギーを求める計算法。

ギーを求めるものです。これにより、溶液中の標的タンパク質と化合物の結合自由エネルギーを正確に計算することができます。MP-CAFEE法の特徴は、原子レベルから計算しているため、タンパク質のキャラクターに影響されず、あらゆるタンパク質に対して安定的に高い精度が出せる点です。その分大きな計算パワーが必要で、2005年に発表した時点では、そのような膨大な計算をいったいどの計算機でやるつもりなのかといわれていました。しかし、「京」ができたことにより、具体的な成果を出すための条件が整ったわけです。

2011年5月より、超並列結合自由エネルギー計算プログラムMP-CAFEEを「京」に移植する作業を開始し、これまでにMP-CAFEEが、「京」で正常に動作していることが確認できています。現在は、より高速に計算するための調整を行いつつ、8月からは、「京」向けに改造したMP-CAFEEを用いて、ターゲットとしている癌や白血病などの標的タンパク質に対する薬を実際に開発するための計算が始まっています。そのためには、数百万単位のさまざまな化合物のなかから、うまく作用しそうな候補化合物を、その構造を見ながら探し出す必要があります。そこで、コンピュータを使った薬設計ビジネスを進めている富士通株式会社バイオIT事業本部との共同研究などによって、まずはMP-CAFEE計算に持ち込むおよそ100のオーダーの候補化合物の選別が行われました。候補化合物は、既存の化合物だけでなく、より効果的に作用する新規化合物も設計されています。もちろんその際には、合成が可能かどうか、毒性(副作用)がないかどうかのチェックも行われます。IT創薬のメリットは、

コストや時間を圧縮できるということもありますが、何よりも大きいのは、既存の化合物にとらわれることなく、分子シミュレーションによって多様で新しい化合物を設計したり、候補化合物を予測したりできる点にあります。

今年度は、比較的小さな標的タンパク質に対して薬候補化合物がひとつのシリーズとして出せるのではないかと考えています。計算機時間の関係で、まずはいちばん確実なものを最優先に進め、あとは今後より大きなもの、難しいものを扱うための準備ということになると思います。実質的にはあと3年ですから、それまでに今年度出た成果をはじめ、臨床試験まで行く可能性のあるものを複数個創るとというのが、最大の目標です。

先に申し上げた通り、日本の製薬会社にこれまでなかなか成果が出なかったIT創薬に目を向けてもらい、一緒に取り組んでもらうということが、このプロジェクトのひとつの目的でもあるわけですが、世界の動向や「京」の完成ということも手伝って、製薬会社の反応は非常によく、オファーも数多く来ています。国のプロジェクトですから、こちらでもできるだけ多くの製薬会社に声をかけて、コラボレーションの話をさせてもらっていますが、すでに多くの会社からいろいろなご提案をいただき、「できれば今年度から始めたい」という声も出ています。もちろんすぐに始められるわけではないので、今はディスカッションを続けながら今後に向けて準備をしているところです。

# 大学の新生生に行った計算生命科学の講義

理化学研究所 HPCI計算生命科学推進プログラム

鎌田 知佐



今年2月に神戸国際展示場で開催された「第4回サイエンスフェア in 兵庫」(主催:兵庫「咲いてく(サイエンス&テクノロジー)」事業推進委員会)での、HPCI戦略プログラム戦略分野1(以下「分野1」)、江口至洋副プログラムディレクターの特別講演がきっかけとなり、私たちは、同委員会顧問をされている神戸大学大学院発達科学部人間環境学科、蛭名邦禎教授より神戸大学発達科学部での新生生を対象としたオムニバス形式の必修科目、「発達科学への招待ー学際性について」での特別講義の依頼をいただいた。

講義は6月29日、7月13日の2回にわたり行われ、293名(男子学生:140名、女子学生:153名)の学生が参加した。神戸大学発達科学部は、人間の発達とそれを取り巻く環境の問題に取り組む文系理系を包括した学部である。宇宙物理学や計算生物学を専攻する学生から、芸術学、心理学やスポーツ科学を専攻する学生まで在籍し、知的好奇心という点に於いて、日本の若者を代表するような多様性に富んだ集団である。

講義タイトルは「学際性についてースーパーコンピュータ「京」が仲介する生物学と物理学、化学、数学の融合を例にー」、講義の構成はスーパーコンピュータと「京」と「京」の持つ汎用性を軸に学際性の重要性へと展開された。

分野1では、高校生、大学生・大学院生や社会人向けの様々な教育とアウトリーチ活動をプロジェクト発足時から実施しているが、対象者は主に理数や科学に興味のある、または専攻している学生や、スーパーコンピュータそれ自身、あるいは「京」の話題性に関心のある人たちであった。今回の聴き手は前述とは異なり、スーパーコンピュータや「京」に興味のない学生も含まれているため、学生からの感想は実に様々なことが寄せられ、驚くこともあり、ゆえに講義後の学生たちの理解の変化の大きさも見る事ができた。

スーパーコンピュータ、そして「京」について、学生たちの反応で一番大きかったのは、「スーパーコンピュータが思っていたものとは、違っていった」という声だった。その中でも、今回の講義でスーパーコンピュータの社会や個人への有用性と多岐にわたるシミュレーションをこなす「京」の汎用性を知るまでは「スーパーコンピュータ、あるいは「京」の名前は知っているけれども、自分の生活には関係ないものと思っていた」、 「計算速度の速さを競ったところで何が変わるのだろう」と思っていたと書いた生徒が1割以上おり、「スパコンが税金の無駄遣いだといったニュースをよく耳にしていたので、悪いイメージばかりでした」というような意見も寄せられていた。そのような学生たちに「京」の社会での重要な役割を説明するいい機会であった。

次に多かった感想が、20年前のスーパーコンピュータの性能が現在の私たちの生活に活かされていて、今の「京」の性能に近い将来自分たちの日常生活にあるであろうこと、そしてスーパーコンピュータの歴史とその変遷を知って、科学技術の進歩の速さに驚き、関心を寄せたものだった。読み手まで、反応の変化が伝わる文を書いてくれた感想文を抜粋紹介しよう。

かつてのスーパーコンピュータが現代の僕たちが使っているパソコンや携帯へ姿を変えて行き、今後新鋭の「京」が20年後には身近な存在となっていく。その頃には新たなスーパーコンピュータが生まれ、今の技術では成しえない新たなステージが展開されていくのは、どこかロマンを感じる話である。ー抜粋ー

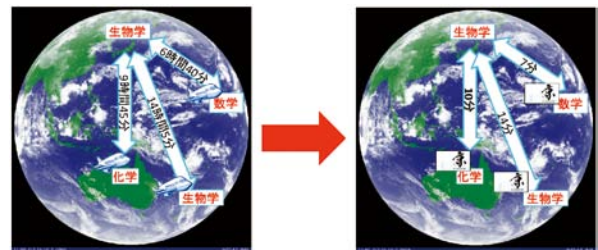
講義では、生物学と数学、化学、物理学の学問はそれぞれが孤島ではなく陸続きであること、「京」はその結びつきをさらに強くしていくことが述べられ、地球地図を使用して模式的にも示された。それについては、大学受験を終えたばかりの学生たちには非常に新鮮だったようで、一様に驚いた感想が寄せられていた。そして、講義終了後、文系の学生から「理系が世界を支えているのではないのか。しかし、文系・理系の融合が大事

ではないか。」という意見が寄せられた。それに対し、江口副プログラムディレクターから安心・安全な社会を創ることを例に、「現実起きている問題は理系だけでは解決できない。」という回答と学際性の重要性が説かれた。先述のように、発達科学部は学際性が特色の学部である。大学生活が始まり、他の学部と比較して履修科目が多様で不安を感じていた学生も少なからずいたようである。そのような学生たちも、また、そうでない学生たちも、自分たち自身の学部で広く学ぶことの意味を改めて理解し、感じいった様子が印象的であった。

今回、私たちは神戸大学発達科学部、白杉直子教授のご協力で講義の感想文をいただくことができ、飾りの無い学生たちの生の声を聴くことができた。学生たちの多くがコメントしていたスーパーコンピュータと社会との関わりの捉え方は、現実社会と重なるであろう。スーパーコンピュータの必要性への理解を得るためには、関心の無い人にもスーパーコンピュータについて「知ってもらおう」取り組みが計算科学への発展において重要である。そして、私たちの活動は眇眇たるものであるが、このような機会を継続して行くことも重要である。分野1計算科学推進体制構築グループでは人材育成として、スーパーコンピュータ「京」が生命科学に果たす役割についての出張授業を年に5回程度の割合で行っており、この活動が将来を担う若者の育成の寄与となれば幸いである。

最後に、今回の講義の機会をいただいた神戸大学・朴木発達科学部長、蛭名教授、白杉教授、そして293名の発達科学部の学生の方に感謝の意を深く表したい。

## 生物学と数学、化学、物理学は「陸続き」!



模式図:「京」は学問の結びつきを強める

気象庁地球地図 ([http://www.data.kishou.go.jp/obs-env/portal/chishiki\\_ondanka/p01.html](http://www.data.kishou.go.jp/obs-env/portal/chishiki_ondanka/p01.html)) を使用。

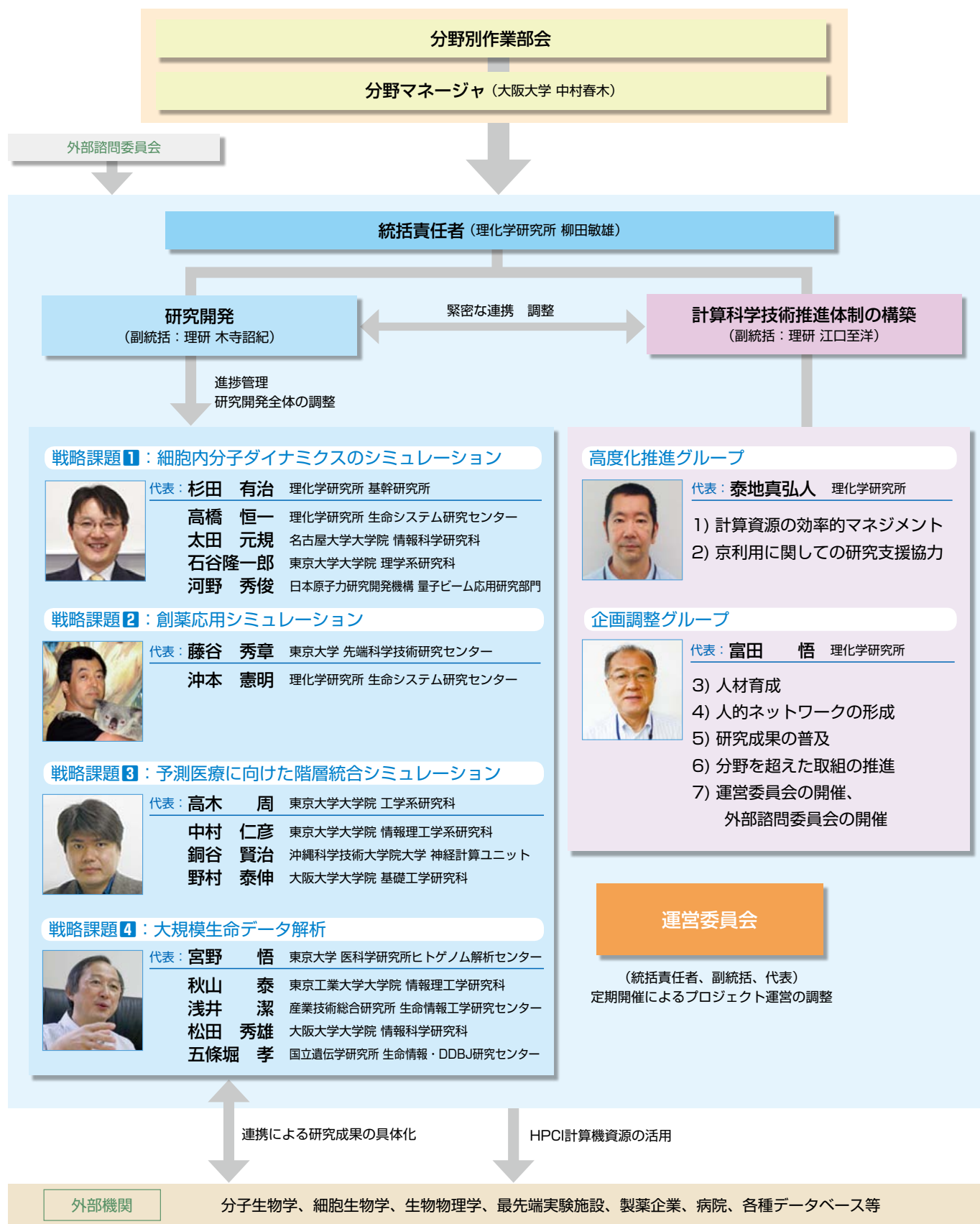
教育とアウトリーチ活動を強化していく一環として、同神戸大学発達科学部に於いて計算科学研究機構や他の戦略分野とも連携し、2012年度後期「自然環境科学特論 D」の枠でオムニバス形式の講義を企画、実施する

### 2012年度後期「自然環境科学特論 D」 ～スーパーコンピュータ「京」が拓く新しい科学～

- 11/17 「社会における計算機とスーパーコンピュータ「京」」  
前田 俊行 チームリーダー (理化学研究所計算科学研究機構高度化研究チーム)
- 「私たちの壮大な自然環境の歴史と変化」 <戦略分野5>  
大塚 孝治 教授 (東京大学大学院理学系研究科)
- 12/1 「エネルギー問題の深刻さ、いかにそれを乗り越えていくか」 <戦略分野2>  
杉野 修 准教授 (東京大学物性研究所物性理論研究部門)
- 「ヒトとともに存在する微生物の謎」 <戦略分野1>  
石田 貴士 助教 (東京工業大学大学院情報理工学研究所)
- 12/15 「人が作り出した環境、その安全性と健全性」 <戦略分野4>  
中島 憲宏 次長 (日本原子力研究開発機構システム計算科学センター)
- 「京コンピュータが担う海溝型巨大地震研究と減災の役割」 <戦略分野3>  
金田 義行 プロジェクトリーダー (海洋研究開発機構 地震津波・防災研究プロジェクト)



# 計算科学技術推進体制



## イベント情報

ISLiM国際シンポジウムのお知らせ

### 4th Biosupercomputing Symposium

- 開催日：2012年12月3日(月)～5日(水)
- 場 所：東京国際フォーラム(東京都千代田区)
- 参加費：無料(懇親会は有料)

詳細は、webページをご覧ください。<http://www.csrp.riken.jp/4thbscs/top.html>

### スーパーコンピューター「京」と創薬・医療との産学連携セミナー —HPCI計算生命科学推進プログラム—

- 参加費：無料

詳細は、webページをご覧ください。<http://hpci.me.es.osaka-u.ac.jp/>

第3回 ●開催日：2012年12月19日(水) 13:00-17:00

- 場 所：梅田センタービル(大阪市北区)

第4回 ●開催日：2013年1月25日(金) 13:00-17:00

- 場 所：フクラシア東京ステーション(東京都千代田区)

## ニュース

HPCI戦略プログラム  
分野1のロゴが完成しました



ロゴのコンセプトについて

「S」の文字を2つ組み合わせることでDNAの二重らせん構造をモチーフに“生命科学”を、背面の「輪」は“細胞=生命”、“輪=和・調和(連携)”を表現しています。知性を表す“紺”はスパコン、上昇志向・快活さを表す“オレンジ”は生命現象・生命科学をイメージしています。分野1の略称表記であるSupercomputational Life Scienceから、SCLSと表示しました。読み方は“エスクルス”です。当分野のホームページでもご覧いただけます。

[http://www.kobe.riken.jp/stpr1-life/outreach/pr/logo\\_stpr1.html](http://www.kobe.riken.jp/stpr1-life/outreach/pr/logo_stpr1.html)

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ーの  
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ー「京」の性能を最大限活用するベー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ー「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.7 2012.12

発行：平成24年12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>