

BioSupercomputing Newsletter

2012.3 Vol.6

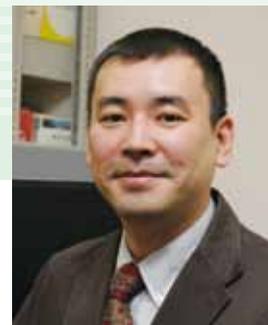


2011年12月 ISLiM成果報告会が開催されました。(本文、14ページ参照)

CONTENTS

● SPECIAL INTERVIEW	○新しい流体構造連成解析手法 (ZZ-EFSI) の開発によっていち早く高い演算性能を達成 東京大学大学院工学系研究科 特任准教授 杉山 和靖	2-3
	○開発・高度化チームに聞く「京」の実力と高い性能を引き出すために続くチューニングの取り組み 理化学研究所 生命システム研究センター 生命モデリングコア計算分子設計研究グループ グループディレクター 泰地 真弘人 理化学研究所 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 生命体基盤ソフトウェア開発・高度化チーム 上級研究員 大野 洋介 理化学研究所 HPCI計算生命科学推進プログラム 高度化推進グループ 高度化推進チーム 上級研究員 小山 洋 理化学研究所 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 生命体基盤ソフトウェア開発・高度化チーム 研究員 舩本 現 理化学研究所 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 生命体基盤ソフトウェア開発・高度化チーム リサーチアソシエイト 長谷川 亜樹	4-5
	● 研究報告	
	○全原子分子動力学シミュレーションによる多割排出トランスポーター AcrBの機能解析 横浜市立大学大学院生命ナノシステム科学研究科 山根 努 / 池口 満徳 (分子スケールWG)	6
	○ヒト循環器系のマルチスケールモデリング 理化学研究所 次世代計算科学研究開発プログラム 梁 夫友 (臓器全身スケールWG)	7
	○初期サッカーボール視覚運動系のスパイクニューロンレベルでのモデル化 京都大学 ジャン・モーレン 奈良先端科学技術大学院大学 柴田 智広 沖縄科学技術大学院大学 銅谷 賢治 (脳神経系WG)	8
	○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)	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	10-11
● 研究報告	○全原子モデルにもとづくヌクレオソームポジション変化の自由エネルギープロファイル計算 日本原子力研究開発機構 量子ビーム応用研究部門 河野 秀俊 / 石田 恒 / 米谷 佳晃 / 池部 仁善 (分野1-課題1)	12
	○骨格筋の活動の推定と脊髄反射の神経モデル 東京大学情報理工学系研究科 中村 仁彦 (分野1-課題3)	13
● 報告	○ISLiM 成果報告会2011 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	14
	○高等学校で行った計算生命科学の授業 理化学研究所 HPCI計算生命科学推進プログラム 鎌田 知佐 / 藤原 康広 / 江口 至洋	15
● イベント情報		16

新しい流体構造連成解析手法 (ZZ-EFSI) の開発によっていち早く高い演算性能を達成



東京大学大学院工学系研究科特任准教授
杉山 和靖

● アルゴリズムも見直し、新たな解析手法を確立

構造解析手法は、これまで主にもづくり分野で開発が進められてきました。しかし、機械部品などとは違い、生体にはもともと設計図が存在しません。そこで、CTやMRIによる医療画像と相性がよく、応力特性の数値的表現が異なる流体と固体を一緒に扱えるような、これまでにない流体構造連成解析を実現したい、それがアプリケーション開発の出発点でした。それによって、医療現場で使いやすく、生命現象の本質の理解、病気のメカニズム解明、さらには創薬にも貢献できるものにしたと考えてきました。

そのために取り組んだのが、メッシュの生成・再構築というプロセスを必要としないオイラー型の新しい連成解析手法です。ボクセルデータを利用して、固定メッシュ上ですべての物理量を更新できるように定式化することにより、複雑な境界を持つ問題や、多数の分散体を含む問題のシミュレーションを容易に実現することを目指しました。また、この方法は計算規模の拡張が容易であるため、大規模並列計算に適しています。一方で、時間積分のアルゴリズムもつくり直しました。最近、完全陽的な時間積分を実現する擬似圧縮性法が広く見直されています。それに動的パラメータを導入し、速度発散を最小化する最適化処理を行なうことによって、数値的に安定で、高い実行性能・並列性能を実現しています。これらの開発の背景には、ノードごとの演算量をできるだけ均一にするとともに、反復の多い処理を避けることで、「京」の性能を有効に活用したいという狙いがありました。

開発した計算手法は、従来の連成解析法と大きく異なりますので、その妥当性が検証されなければなりません。まず、これまでの十分に検証された流体構造連成問題や流体膜連成問題の計算結果との比較を行いました。非常によく再現できることを確認しました。また、界面で物理

量の跳躍を伴うような多媒質の系を数値的に扱おうとすると、保存則が破綻することがよくありますが、私たちが解きたい問題の運動エネルギー輸送の収支を調べると、その保存性がしっかりと満たされることを確認しました。さらに、実際の実験観測との比較も重要ですが、例えば細い血管のなかで赤血球の振る舞いがどうなっているかという、変形した赤血球群が血管の中心軸に集まろうとする軸集中という現象がおきます。私たちの計算結果でも、こうした赤血球群の振る舞い（管壁付近のセルフリーレイヤー）や形状（スリッパ型の赤血球）が、非常によく再現できていることを確認しました。

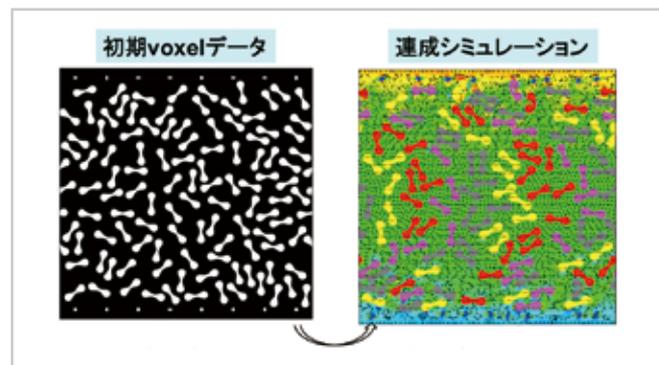
現在は、この新しい方法論を、血栓シミュレータに応用・拡張するための開発を行っています。最初に取り組んだのは、血小板血栓に至る過程、つまり血小板が傷ついた血管壁にどのように付着するのかを数値予測するという事です。すでに連続体レベルの血流解析シミュレーションの部分は開発できていますが、血小板の付着を捉えようとする、タンパク質同士の結合、言い換えれば揺らぎを持つ分子レベルの現象という、明らかにスケールが異なるものを考える必要があります。ただし、スケールの違いが大きすぎますので、分子スケールの現象を直接扱うのではなく、連続体スケールで見たとときの揺らぎの効果を確率論的に扱う方法を取り入れています。具体的には、付着するかどうか、リガンド・レセプター結合と呼ばれていますが、この分子スケールの効果が統計論的に反映されたモデルを使っています。こうしたマルチスケール、マルチフィジックスを考慮した解析手法を開発してきました。ほぼ準備は整いつつありますので、今後は規模を拡大して、実際に「京」で走らせながら、多数の赤血球が存在する血管の中で、血小板の付着に至るプロセスを解析していきたいと考えています。

● ピーク性能の約46%の計算性能を達成

今年度から、整備が進む「京」を使った計算が始まりました。スカラー機は流体解析には向いていないと言われていました。例えばベクトル機の「地球シミュレータ」で70%くらいの実効性能が出ていた流体のアプリケーションが、スカラー機では10分の1ほどの性能しか出ないといった話を聞いていました。そのため、最初は10%くらいの性能値が出れば……という気持ちでした。しかし、それでも、できるだけ「京」の機能を有効に活用して、高い計算性能を引き出すことを目指してきました。その結果、「京」環境で、ピーク性能の約46%の計算性能を出すことができました。比較的よい結果が得られたと思っています。スカラー機ではありますが、「京」はハード的にベクトル型と似たような発想がありますので、これくらい出ても不思議はないという思いもありますが、それでも、正直なところ、頑張ってきた甲斐があったという気持ちです。

「京」の優れている点のひとつは、通信のスピードが速いことです。特にハード的なリダクション処理に強みがあります。通信には隣接通信と大域通信があります。例えば場全体の誤差を評価しようとする、各ノードでの誤差を足し合わせるリダクション処理が必要で、その場合、大域通信が発生します。今回開発した疑似圧縮性法でも、リダクション処理がいくつか必要です。大域通信は、ノード数が増えるほど計算時間に対

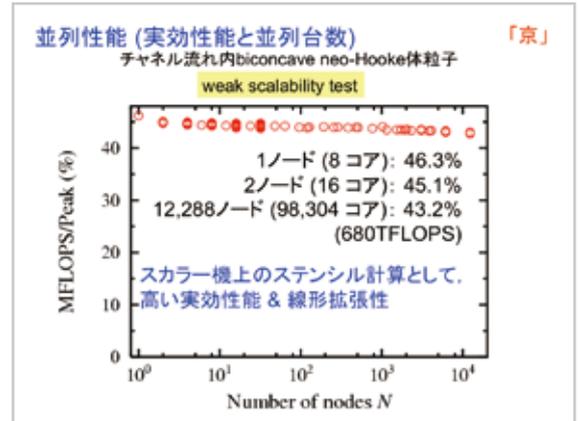
する割合が増えますが、当初、どれくらいのペースで増えるのか分からず、私たちのコードにとってボトルネックになると思っていました。しかし、実際にやってみると高速に処理されており、大域通信の占める時間の割



図：Euler法の特徴 固定メッシュ上で、全ての物理量を更新 複雑な境界形状を持つ/多数の分散体を含む 流体・構造連成問題の数値シミュレーションを容易に 境界適合メッシュの生成・再構成が不要



「京速コンピュータ「京」では、6次元メッシュ／トーラス構造ネットワーク(Tofu)という革新的なネットワーク構成を採用。ノード間通信の時間短縮に貢献している。写真はTofuの概念模型」



合は、現在やっている約10万コアで1%にも達していません。また、隣接通信は3～4%で収まっていて、予想以上に速いことが分かりました。

恐らく、8万ノード、64万コアを使っても、通信は8%程度に収まるのではないかと考えています。

●実際に「京」を使ってみて思うこと

私たちが取り組んでいる連続体力学というのは、原理原則が比較的単純で、コードで記述する部分も非常にシンプルです。研究分野によっては原理原則がとても複雑な場合もあるでしょうから、私たちの「京」を使った感想や、より高い性能を出すための取り組みは、必ずしも一般的に通用するとは限らないのですが、いくつか気付いたことをお話します。

問題点を見つけ出したり、性能を高めていくために、私たちが最初に注目したのは、プロファイラーによるハードウェアモニタ情報でした。これによって、ループごとの実効性能や、メモリーとのやり取りのスピードなどが示され、どこに問題点(ホットスポット)があるかを見つけることができます。「ホットスポットを見つけて対処する」ことはマニュアルにも記されている常套手段でして、まずはそこに手をつけました。これはループ毎に対する局所的な対処法ですが、一方で大局的にプログラム全体の流れを見て、いくつか分割化されている処理をまとめたり、逆に分割するといった見直しも行いました。これらはコードの流れを理解している開発者でないと対処が難しいように思いますが、大幅な性能向上をもたらす場合があり、非常に効果的でした。また、私たちのグループでは、別の方に1からコードを書いてもらい、クロスチェックを行ってきました。回り道のようなのですが、問題を効率的に解決するには重要で、迅速に開発を進めるのに有効だと思います。さらにチューニングに際しては、何種類もコードが存在することになりますので、プロファイラの結果を比較することで、見えなかった問題点が明らかになったり、逆によいところが見えてくるなど、いろいろな効果がありました。こうした

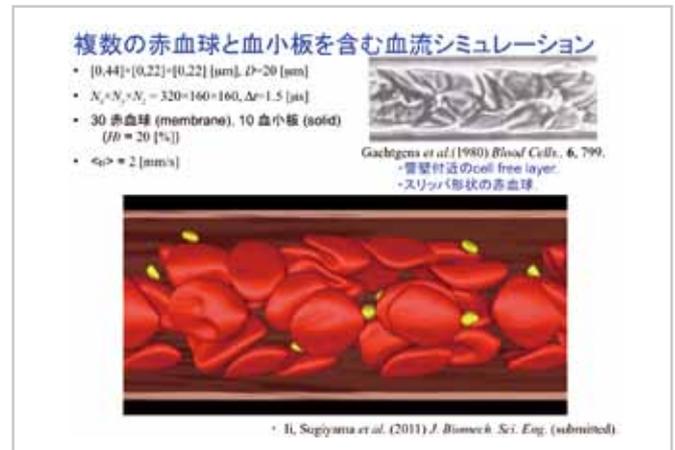
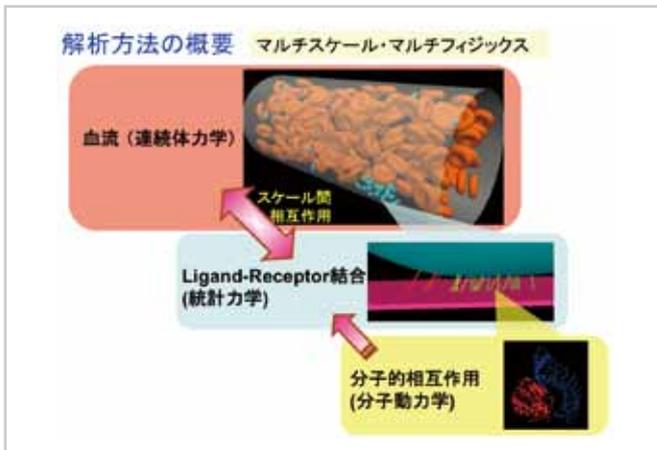
ことを地道にやり続けていくことによって、アプリケーションのパフォーマンスを高められたのだと思います。

私たちはFortranを使っていますが、結果的にそれがよかったという実感も持っています。Fortranは古いという人もいますが、メーカーがHPC用のコンパイラでまず何を準備するかというと、普通はFortranでしょう。CはFortranに比べてできることが多いでしょうが、それだけ整備するのは大変です。その点、Fortranはできることに制約がありますが、その範囲内で性能の高いコンパイラを完成させることは簡単なわけです。「京」が整備されていく初期段階では、CよりもFortranで書く方が、より最適化され、情報も多く得られたように思います。

性能を上げるために何よりも重要なことは、理詰めで考えていくことでしょうか。「京」の場合には、階層的な構造を持つハード構成を理解することです。ただし、理詰めで考えるだけでは、ベストな答えを見つけるのは困難です。論理的に「こうすればうまくいくはず」と対処しても、どこかにトレードオフがあって逆効果になることがよくあります。その場合には、試行錯誤しながら進めていくしかありません。そのためには、最初から完成度の高いもの一気に仕上げようとするのではなく、試行錯誤しやすいようにプログラムを書くことも、大事なポイントだと思います。

注：京は現在開発中であり、これらの数字は現状の値です。

謝辞 京での計算に関しては京速コンピュータ京の試験利用での結果です。



開発・高度化チームに聞く 「京」の実力と高い性能を引き出すために続く チューニングの取り組み



前列左から、大野洋介氏、泰地真弘人氏、舛本現氏、後列左から、小山洋氏、長谷川亜樹氏

理化学研究所 生命システム研究センター
生命モデリングコア計算分子設計研究グループ グループディレクター

泰地 真弘人

理化学研究所 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 生命体基盤ソフトウェア開発・高度化チーム 上級研究員

大野 洋介

理化学研究所 HPCI計算生命科学推進プログラム
高度化推進グループ 高度化推進チーム 上級研究員

小山 洋

理化学研究所 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 生命体基盤ソフトウェア開発・高度化チーム 研究員

舛本 現

理化学研究所 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 生命体基盤ソフトウェア開発・高度化チーム リサーチアシエント

長谷川 亜樹

● 予想以上に安定していた「京」

——スーパーコンピュータ「京」の整備が進み、試験利用ではありますが、グランドチャレンジのアプリケーションも次々と「京」を使ってのチューニング及び試験的な計算が行われています。みなさんが開発した「大規模並列用MDコアプログラム (cppmd)」は早々と1.3PFLOPSを記録し^{*}、他のアプリケーションも1PFLOPSを越えようとして開発が続けられているところです。そこで今日は、開発・高度化チームとして研究者らのサポートを担当しておられるみなさんに、実際に「京」を使った感想や、サポートのご苦労などについてお話を聞かせていただこうと思っています。

泰地 ハードウェア的には、最初から非常に安定していたという印象があります。急に止まったといった話もほとんど聞きませんでしたし、最初からわりと大きな規模で流れましたからね。はじめのうちは、いろいろ落ちたりして苦労するだろうと覚悟していたのですが、私たちが触れるようになった昨年の春には、とても安定していました。

大野 確かに、ハードウェアのトラブルがほとんどなかったのは意外といえば意外でしたね。いろいろと問題が出たのは、基本的にソフトウェア関係の方でした。特にコンパイラは開発途中ということもあって、CやC++はFortranに比べるとちょっと遅れていて、プログラムとして問題ないはずなのにコンパイルに失敗したり、結果が間違ったり、そういうことがありました。まあ、ソフトウェアまわりについては、早く使わせてもらう代わりに、そうしたバグ出しにも協力するというところで…… (笑)。

● Fortranユーザーが高い成果を達成

——Fortranの方がいい結果が出ているのですか。

大野 最初のうちは、性能面でもFortranの方が速いという状況がありましたし、開発途中のコンパイラでも最適化してくれるような書き方をするとということもやりました。ただ、最近は他のコンパイラも含めてだいぶ改善されてきています。

小山 僕自身がFortranユーザーだからというわけではありませんし、別にメーカーをよいしょするつもりもありませんが、コンパイラのFortranの自動並列に関しては、かなりよくできています。ユーザーは限定されるかもしれませんが、今までずっとFortranやってきた人は使いやすいだろうと思います。ただ、一方でメモリバンド幅の問題があります。地球シミュレータなどのベクトルマシンを使ってきたFortranユーザーの人たちの移植は楽かと思っただけですが、メモリバンド幅的に性能が出るか、出ないか、そこは完全に分かれてしまいます。出ない場合は、根本的にアルゴリズムを変えない限り、絶対に性能は出ないので、それはもうどうしようもないというところが、結構つらくもあり、チャレンジングでもあり……。



——確かに地球シミュレータと比べると、バンド幅は仕方ないですね。

小山 ただ、そこで暗明が分かれて、助かった人とそうでない人が出るというのは、厳しいですね。

——その点で、ZZ-EFSIIは非常にうまくいっているようですね。

大野 確かに、Fortranでループ文がきれいに書いてあると、自動並列で結構いい性能が出ます。

小山 もちろん開発努力を続けてこられた結果ではありますが、ある意味では、非常に幸運でもあった (笑)。しかし、「CやC++でプログラミングしたもの、ちゃんと最適化してほしい」というのが、多くの「京」ユーザーのニーズなので、そこはメーカーにも考えてもらわないといけませんね。

舛本 流体などをやってきた人たちは、もちろんベクトル型スーパーコンピュータの経験があると思いますが、バイオ分野ではスーパーコンピュータを使ってこなかった研究者も多いので、普通にCとかC++を使っているということを想定して、もう少し準備をしてほしかったという気はしますね。

小山 とはいえ、これはあくまで現状の話でして、本格運用のころには大きく状況が変わっていると思います。

大野 言語仕様の高機能になっているとコンパイラをつくるのは大変なので、後まわしになっている可能性もありますね。最初からできるだけ性能を出さなければいけないというプレッシャーもあったでしょうし。

舛本 まずは、やりやすいFortranから、というのは考えられる話ですね。

泰地 「京」には、超並列計算を効率的に実行するため、VISIMPACTと呼ばれる並列計算モデルが導入されていて、コア間の同期を高速にとる機能が付いています。それを使うと、ベクトル的な並列化がわりと簡単にできます。そのため、普通、コア間の並列をするときは上の方で並列化しますが、「京」の場合は下の方で並列化していて、そのループを分解するような形で並列化するのが得意という特徴があります。恐らく、そこで自動並列化できるかで、自動並列化がうまくいくかどうかが決まると思います。

小山 できるだけループ構造にして、たくさんデータを流すようなやり方をしていた人は、比較的スピードが出ています。メモリバンド幅の問題もありますが、最適化という意味では、ベクトルでやってきた人たちは、自動並列化が結構簡単に適用されそうな感じもします。

舛本 僕が担当したアプリケーションは、どれもCとC++だったので、最初のうちはコンパイルに苦労しました。あるアプリケーションは、すぐに性能を出さなければいけなかったこともあって、大野さんにも協力してもらい、できるだけのことをやり、最近すごくよくなってきています。ただ、別のアプリケーションでは、コンパイラが成熟するのを待つか、今のコンパイラに合わせて書き直すか、ちょっと考えています。時間が許すのであれば、トリッキーなことをしないで、コンパイラが整備されるのを待った方がいいんじゃないかと……。

——開発途上のマシンと、どのように向き合っていけばよいのかという悩みもあるわけですね。

舛本 C++限定の悩みですけどね。Cはまだいい方だと思います。だから、



小山さんの話を聞いて、Fortranはそんなにいいんだと思って……(笑)。

小山 確かにさくっと通ったし、Fortranはすごくよかったという印象があります(笑)。

● 本当の性能が試されるのはこれから

長谷川 私の担当しているデータ解析のアプリケーションは、移植自体はそれほど大変ではなくて、わりとすんなりいきました。ただ、まだハイブリッド並列などで性能を出していく余地はいろいろあると思うので、今後はそこに向けて取り組んでいこうと思っています。あとは、大量のI/Oとかがあるのですが、そこはシステム自体がまだ本番になっていないので、ちょっとペンディングといいますが、チューニング自体はもっと後になってしまうかもしれません。

舛本 他のアプリケーションと比べると、I/Oがネックになりそうな印象があります。

長谷川 それほど苦労していないというと語弊がありますが(笑)、もちろん苦労はしていますが、今できることが限られているという感じですね。もう少しI/Oが整備されたら、いろいろと考えることはあると思います。

小山 例えば流体の計算では、ダミーデータでも計算結果はあまり変わらないとか予測できるので、I/Oを後回しにできるのですが、バイオ分野ではデータありきといいますが、ヒット率が影響するようなベンチマークを取らないと意味がない場合があります。ダミーだと全然ヒットせずに空まわりで終わってしまい、それなりに本格的なデータを使ってチューニングしないと意味がなかったりするため、早い段階でI/Oが必要になるケースがあるのです。そのあたりは、苦労していることのひとつです。

— 「京」には6次元メッシュ/トラス構造のインターコネク (Tofuネットワーク) をはじめ、指定したデータをキャッシュに残せるセクタキャッシュなど、開発中から話題になっている新技術がいくつも導入されていますが、これらについては実際に使ってみて、いかがですか。

大野 cppmdのMD計算では、データ量に対して演算が比較的多く、あまりメモリネックにならない計算でしたから、実はキャッシュが足りないということがあまりなくて、セクタキャッシュ機能を使わなくても十分に性能が出てしまいました。今扱っているライフサイエンス分野のアプリケーションでも、セクタキャッシュを使うところまでいっているものは、まだないですね。ネットワークについても、MDの場合は比較的通信の比率は少ないので、ギリギリまでチューニングしなくても、結構いい性能が出ています。どちらも、まだフルには使っていないので、本当の性能が試されるのは、まだ先になるのではないのでしょうか。

● 開発途上であることの苦労

— 「京」がまだ開発中のころ、泰地さんから「これだけ大規模な計算機は、実際に使ってみないと何がおきるかわからない」というお話をうかがったことがありましたが、実際に使ってみて、予想していなかった発見や驚きはありましたか。

泰地 「使ってみなければわからない」という考え方は今も変わりませんが、いちばん意外だったのは、最初にお話したとおり、ハードウェアのトラブルが少なく、結構安定しているということでした(笑)。

— 嬉しい誤算というか……。

泰地 そうですね。悪い方では、先ほどから話に出ているように、コンパイラがFortran優先になっていることで、まだ開発途中とはいえ、そろそろC++もちゃんと動くようになってほしいですね。

小山 いちばんの難しさは、開発途中でまだ未実装であったり、バグを回避するために、どの程度アプリケーションに手を加えるべきかということですね。例えば、今、大がかりに手を入れて性能を上げなくても、バージョンアップされたら高い性能が出るかもしれないわけですから、そこをどうするか……。

舛本 今までにもずいぶんありましたよね(笑)。ただ、そういう努力は、

アプリケーションにとっては無駄かもしれませんが、われわれにとってはスキルアップにつながりますし、「京」の特性を理解する上で、全く無駄ということでもありません。

小山 確かに、経験を積む意味でも大事なことですが、個々のアプリを開発し、研究成果を出そうという人たちに、そこまで手間をかけていいのかを考えると、悩ましい気もします。

舛本 それでも、最初につくられたコードそのまま素晴らしい成果が出るということは、絶対にあり得ませんからね。いろいろと試すことは必要だと思います。もちろん、元に戻すところもたくさんあるでしょうけど(笑)。

— アプリケーションを開発しているユーザーからの質問で、多いのはどのようなことですか。

舛本 以前は、「コンパイルがうまく通りません」という問い合わせが多いぶんありましたね(笑)。

小山 クロスコンパイルで手間取るケースがいくつかありました。

— クロスコンパイルとは？

大野 「京」は、フロントエンドと計算ノードとが違うCPUのマシンになっているため、異なる環境で動かなくなってしまうケースがあるわけです。今は全部通りましたが、最初は苦労したのもありました。

小山 「京」がまだ開発中ということで、いろいろと機能制限もあるので、そこに引っ掛かってうまくいかないといった問い合わせもありました。もちろん、バグの場合も(笑)。

舛本 まだまだ初心者の人たちが常にいる状態ですから、初歩的な質問も多いですね。でも、マニュアルはよくできています。思ったよりも最初からちゃんとしています。

大野 むしろ、マシンの方がマニュアルに追いついていない部分もある(笑)。

● アプリ開発者の努力が必要

— これから「京」を使ってみたいという人もたくさんいると思います。そうした人たちに、先輩からアドバイスをお願いします。

小山 個別にはいろいろありますが、一般論としてのアドバイスは難しい。大野 「京」を使って大規模な計算をするのであれば、並列化が2階層のハイブリッド並列が必要ですね。このハイブリッド並列に手間取っているケースがそこそこ見受けられます。

舛本 ハイブリッド並列は、実施本部も推奨していますよね。

小山 「京」に限らず、大規模なマシンではノードあたりのコア数が増える方向に向かっていきますし、その意味でも、今後はハイブリッド並列でないとCPUパワーを引き出せないと思います、今のHPCの流れからすると。

大野 基本的なことですが、やっぱりアプリケーションを書く人の頑張りが必要ですね。どれだけHPCの性能を引き出せるかも、実はそこにかかっています。

— とはいえ、アプリケーションをつくる研究者たちは、自分たちの研究成果を出すために計算機を使うわけですから、なかなかそこまで要求するのは難しいかもしれませんね。

大野 そういう意味では、私たちのような開発・高度化チームを利用するのは、1つの手だと思います。極端なのは完全分業で、ひたすらコードをチューニングする人と、それを使って研究する人を分ける形ですが、その場合、アルゴリズムまで見直したいといふときに、研究者の計算したいものを壊さない変更できないといけませんから、ターゲットまでは理解していないにしても、計算手法とか、何を計算するのか、どういう計算をしたいのかということまでは理解してチューニングしないといけません。完全分業だと、意思疎通などで問題があります。ある程度は研究にも片足を突っ込んで、コードのチューニングを行う、そういうスタッフが必要だと思います。

— 「京」がLINPACK性能によるTOP500で2期連続世界一を獲得し、さらに「京」を用いたシリコン・ナノワイヤの材料に関する研究でゴードン・ベル賞を受賞しました。それに続く研究成果にも大きな期待が寄せられています。そのためには、これからのみなさんの仕事が非常に重要になってくると思います。素晴らしい成果を期待しています。今日は、お忙しいなかありがとうございます。

※2011年ゴードン・ベル賞エントリのための特別利用による成果

全原子分子動力学シミュレーションによる 多剤排出トランスポーター AcrBの機能解析

横浜市立大学大学院生命ナノシステム科学研究科

(上から) 山根 努、池口 満徳
(分子スケールWG)



近年、医療現場での薬物療法において、病原菌やがん細胞などに対して薬が効かなくなる現象である薬剤耐性化が大きな社会問題となっています。このような薬剤耐性化のメカニズムはいくつか知られていますが、その要因の一つとして、細菌の細胞膜に存在する多剤排出トランスポーターと呼ばれる膜蛋白質が大きな役割を担っています。多剤排出トランスポーターは、細胞にとって有害な様々な化学物質を能動的に細胞外へ排出する働きをもつ膜蛋白質です。

院内感染などの主要因である多剤耐性緑膿菌などで見られるRND型と呼ばれるタイプの多剤排出トランスポーターでは、村上聡氏（現東京工業大学教授）らにより大腸菌由来のAcrBについて原子レベルの結晶構造が2006年に決定されました。それによると、AcrBは約1000アミノ酸からなる巨大な蛋白質が3個集まった状態（ホモ3量体）で機能しています。そして、各蛋白質は、それぞれ薬剤の取り込み（取込型）、結合（結合型）、排出（排出型）を担う3つの異なる立体構造をとり、(図1A)、細胞内外のプロトン濃度(pH)の差を利用した細胞内へのプロトン移動により得られる駆動力で、各立体構造間を順に変化することで薬剤排出の機能を果たしています(機能的回転メカニズム、図1B)。また、膜貫通部位に存在する3つの荷電アミノ酸 (Asp407, Asp408, Lys940) が細胞内へのプロトン移動に寄与しており、排出型の状態のみLys940の側鎖の構造が異なります(図1C)。このことから、プロトン移動によるこれらの荷電アミノ酸のプロトン化状態の変化が引き金となって機能的回転メカニズムを生じていると考えられています。

AcrBは、私たち分子スケールチームの共通の研究ターゲットとして、様々な手法での分子シミュレーションが行われています。私たちの用いた全原子分子動力学法では、生体内の環境をより現実的に考慮するため、細胞膜を構成する脂質分子や水分子およびイオンなどについて水素原子までを含めた系を計算対象としています(図2中央)。そのため、全粒子数が約470,000と非常に大きな系に対する計算となりますが、系内の様々な相互作用を原子レベルで詳細に捉えることができます。そのような全原子分子動力学シミュレーションから、以下のことが明らかになりました。

- (1) 排出型の構造において、Asp407およびAsp408を同時に脱プロトン化した状態では、Lys940の側鎖が結合型および取込型で見られる構造に変化することが分かりました。(図2右パネル)。一方で、Asp408のみをプロトン化させた状態では、排出型の構造を安定に保持できることが分かりました。
- (2) 排出型のAsp407、Asp408を同時に脱プロトン化しLys940側鎖の構造が変化したシ

ミュレーションでは、排出型では通常閉じている薬剤の取り込み口が開き、取込型への構造変化の途中の様子が観察されました(図2左パネル)。

現在、排出型のLys940の側鎖の構造変化がどのようにして、薬剤の取り込み口が開く構造変化につながるのか、そのメカニズムの解明を目指して研究を進めています。

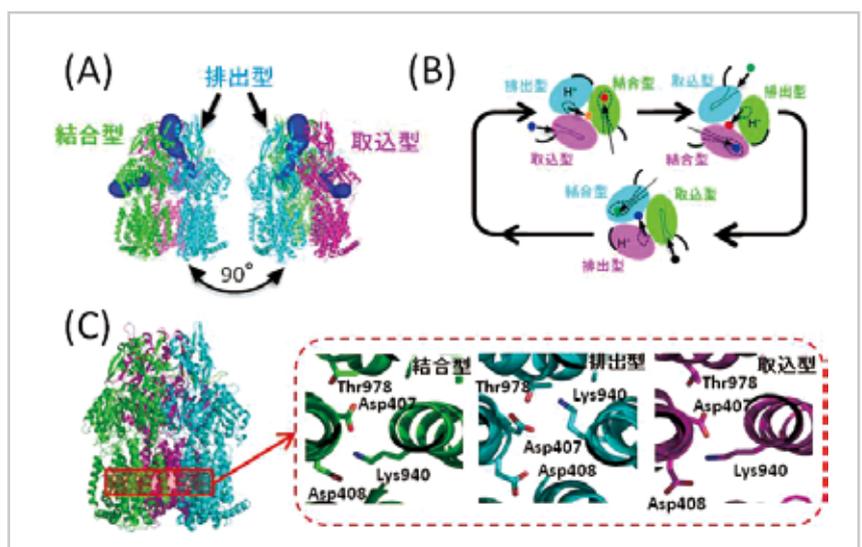


図1：多剤排出トランスポーター AcrBの構造と薬剤排出のメカニズム
(A) 各モノマーの構造と薬剤の経路(青)。(B) 薬剤排出のメカニズム(機能的回転メカニズム)。図中でプロトン: H⁺、薬剤: 丸印。(C) プロトンの細胞内への取り込みにかかわる膜貫通領域のアミノ酸。

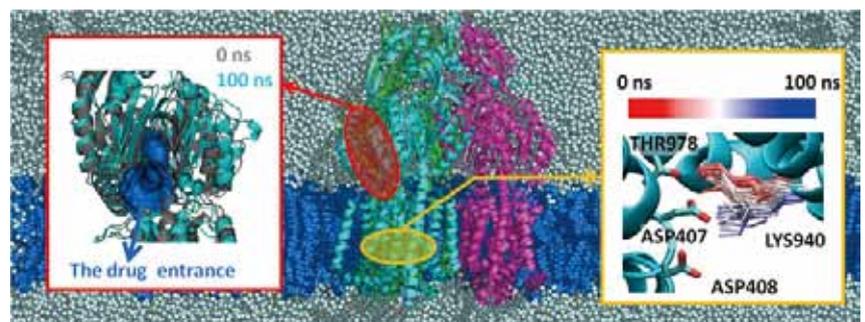


図2：全原子分子動力学シミュレーションに用いた系(中央、図中で水分子はグレー、脂質は青の球で示してある。)および結果(左右のパネル)。

ヒト循環器系のマルチスケールモデリング



理化学研究所 次世代計算科学研究開発プログラム

梁 夫友

(臓器全身スケールWG)

ヒトの循環器系は最も重要な生命維持システムですが、疾患の発生しやすいたちどころでもあります。世界的に見ると、循環器系疾患は現在、人類の最大死因となっています。循環器疾患の診断・治療を高度化あるいは支援するための研究や技術開発などが、数十年前から幅広い分野において活発に行われています。中でも血管病と血流との関連に関する研究は重要な課題の一つです。

特に、医用画像と流体解析技術を融合した研究は患者の3次元血流情報を詳細にとらえ、臨床応用に役立つと大きな期待を集めています。しかし、その一方で、幾つかの問題にも直面しているのも事実です。例えば、計算の境界条件を与えるための血流・血圧の測定については、精度や適応場所の制限など、様々な困難があります。また、患者が安静な状態でしか計測できない場合も多いため、多くの血流解析研究は安静状態での血流のみを反映するものとなってしまいます。しかし、血流循環の挙動は生理状態の遷移に応じて大幅に変わることがあります。

局部血流のほかに、より大きいスケールから見た血流挙動や脈波の伝播なども医学的な価値があります。例えば、動脈波は心機能や動脈系の形態学・力学特性などの変動に応じて変化することが知られています。実際、昔からも動脈波がしばしば重要な指標として循環器病の診断に使われています。

循環器系が複雑であることで計算モデルの構築上、問題になることがもう一つあります。例えば、ひとりの人間の体内にある動脈(静脈)、細動脈(小静脈)、毛細血管の数はそれぞれ数百、数千万、数十億に達します。これらの血管は繋がりが合、極めて複雑なネットワーク構造を築いています。しかし、血管の幾何学形状や材料力学特性など計算モデルの構築に必要なデータは全ての血管に対して計測することは困難であります。従って、全身の血管を対象とした実形状の計算モデルの構築は極めて困難であります。

循環器系の動的特性と複雑さを対応できるように計算モデルを構築するために、我々は血流のマルチスケールモデリング手法を開発しました。まず、毛細血管や細動脈、小静脈などの微小血管系、動脈系、局部の動脈をそれぞれ集中係数(0次元)モデル、1次元モデル、3次元モデルを用いて表します。更に各モデルを互いに連成させ、閉じた系の複合モデルを構成します(図1)。このモデルは異なるスケールのサブ・モデルを含んでおり、マルチスケールモデルとも言われます。このモデリング法のメリットは、最も複雑な微小循環を計算コストの最も低い0次元モデルで扱い、モデル全体の計算コストを大幅に削減できることです。また、局所の3次元血流モデルの境界条件を他のモデルと連成解析させることにより、様々な生理・病理条件下での血流解析が可能となることも血流解析上の大きなメリットです。

本モデルを用いて様々な血行力学現象を調べることで、様々なことが分かってきました。例えば、若いときに見られた心臓から大動脈の末梢に向けての収

縮圧の上昇と圧力波の形状の変化が、加齢につれなくなるという現象を計算で再現し、加齢に伴う動脈硬化がその原因であることを明らかにしました。また、本モデルを上腕カフのモデルと連成させ、自動血圧計の計測精度やカフを利用した動脈硬化度の評価の妥当性などを定量的に検討することにも成功しました。現在、臨床で取られた血管画像、血流量などの患者別情報を本モデルに導入し、術後の血流状況を評価し、手術の効果やリスクを予測する研究を進めています。将来は、スーパーコンピュータを利用し、より広い範囲の動脈を3次元モデルで表し、より多くの患者別情報、生理学メカニズムをモデルに導入することで、よりリアルな血流解析を行いたいと考えています。

【参考文献】

- [1] Liang F.Y., Takagi S., Himeno R., Liu H., Biomechanical characterization of ventricular-arterial coupling during aging: a multi-scale model study, J Biomech, 42:692-704, (2009)
- [2] Liang F.Y., Takagi S., Himeno R., Liu H., A computational model of the cardiovascular system coupled with an upper-arm oscillometric cuff and its application to studying the suprasystolic cuff oscillation wave concerning its value in assessing arterial stiffness. Comput Methods Biomech Biomed Engin. (in press)

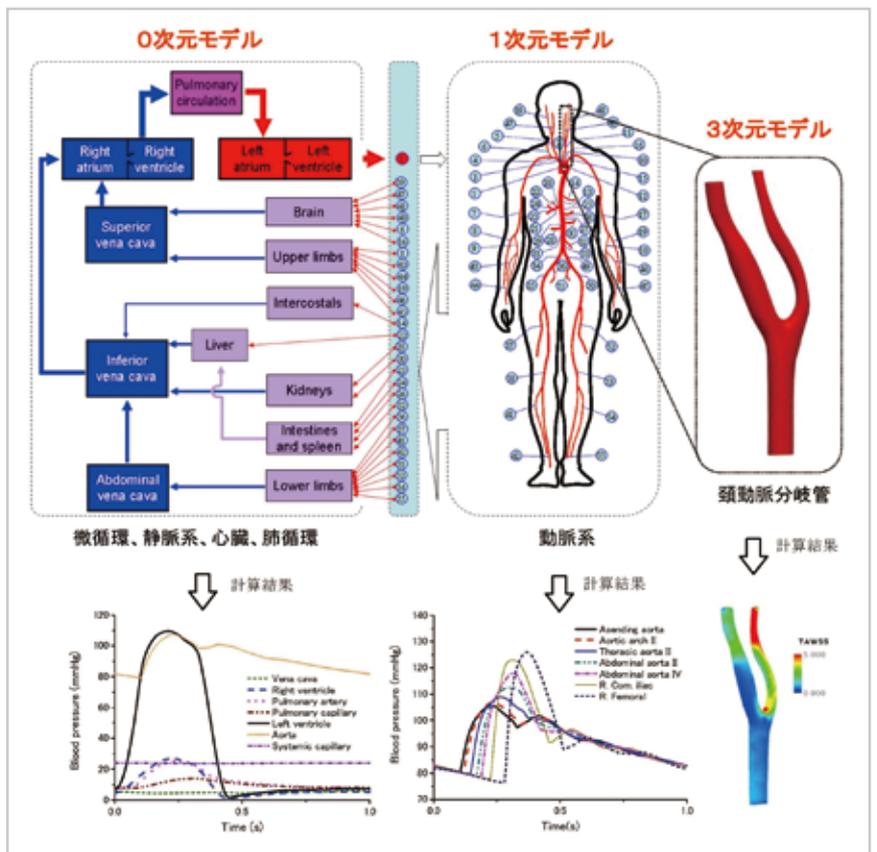


図1: ヒト循環器系のマルチスケールモデル。(上: モデルの仕組み、下: 各サブ・モデルで計算された結果の例)

初期サッカード視覚運動系のスパイクングニューロンレベルでのモデル化



ジャン・モーレン (写真) ①

柴田 智広 ②

銅谷 賢治 ③

(脳神経系WG)

- ① 京都大学
- ② 奈良先端科学技術大学院大学
- ③ 沖縄科学技術大学院大学

巨視的な視覚運動の挙動は、脳内におけるシステム規模の機構間の高いレベルの相互作用に依存しますが、そのような機構の機能的メカニズムは、低いレベルの神経生理に基づくものです。生物の挙動を十分に理解するために、これらの機構について両方の規模で見る必要があります。記述レベルが多角的になることで、構造的な複雑さや結果として生じるモデルの計算要求が急速に増えています。

我々の最終目的は、我々および他のグループと共同で、網膜から眼筋運動系までの完全な知覚行動ループを構築することであり、現在の目的は、特に上丘 (SC) に重点を置いた、衝動性眼球運動 (サッカード) を生成するニューロンレベルのモデル化です。サッカードは再現性および常同性が高いという点で良い標的となります。

我々は、現在の神経生理学の最高の知識に基づいた、上丘中間層 (SGI) における大規模なスパイクニューロンレベルのモデルを構築するために NESTシミュレーションツールを使用します (図1)。また、NMDAのシナプス入力が増強した、コンダクタンスに基づく順応性の指数関数的な laFニューロンモデルを使用します。このモデルのニューロンの数はおよそ100kで、入力発生および活動統合の支援ユニットのほか、バースト、ビルドアップ、半視野、強抑制、および抑制介在ニューロンを伴います。SGIは網膜位相で、眼球運動系の最初の部分を形成しています。網膜は、皮質の視覚系だけでなく上丘浅層にも投射します。上丘浅層は局所的に突出した領域を選択し、中間層部の対応する点を直接あるいは間接的経路を介して活性化します。浅層部は知覚領域ですが、中間層部は運動系の部分であり、これらがこの知覚行動ループのセンサーとモーターの集合部を構成しています。

サッカードは、SGIで、突出点の周りに拡散しているニューロンの活性化を増強して惹起されます。ここでは、外側のSGIを単一の抑制入力に単純化していますが、通常は、脳幹や他の領域のニューロンによって抑制されています。外部抑制を放出すると、バーストニューロンが脳橋と中脳内でバーストニューロン系を活性化するスパイクのNMDAシナプス介在性バーストを生成します。バーストニューロンは次々と眼球へ水平および垂直に移動し、小脳によって制御されます。

SGIはかなり解明されていますが、この領域での活性に関しては未だに

多くの問題が残っています。バーストは網膜部位の指標や移動ベクター、速度コマンドになり得るものの、ビルドアップニューロン間での活性化拡散の重要性についても明らかにはなっていません。我々は、この領域における既知の神経生理を踏まえつつ、スパイクングニューロンの活性化とビルドアップニューロン間の活性化拡散の最近の研究結果について説明を試しています (図2)。

バーストニューロン活性は、サッカードの距離依存的な眼球運動の特性を経時的に示します。バースト特性は、NMDAシナプス生成バーストと中脳網様体で蓄積したスパイクからの直線的抑制フィードバックにより構築されます。その活性により、バースト特性を実験的に測定することができます。ビルドアップニューロン活性の拡散は、局所的な結合内および抑制的なニューロン間で生じます。我々は、活性の位置ではなく量によって、サッカードの進行を経時的に追跡できることを提唱しています。これにより、上丘深層部でバーストニューロン活性を直接抑制する抑制ニューロンが誘発され、間接的に吻極に制御を回復させます。これがサッカードの終了時期の正確性を高めています。

我々は、上丘浅層の構築や、我々のモデルと、同様に詳細な網膜および脳幹系のモデルとを組み合わせることを目指しています。また、そのモデルを物理的入出力に結合することで具体化し、これによって計算的に重い行動モデルもより自然な設定にできる可能性について探索しているところです。

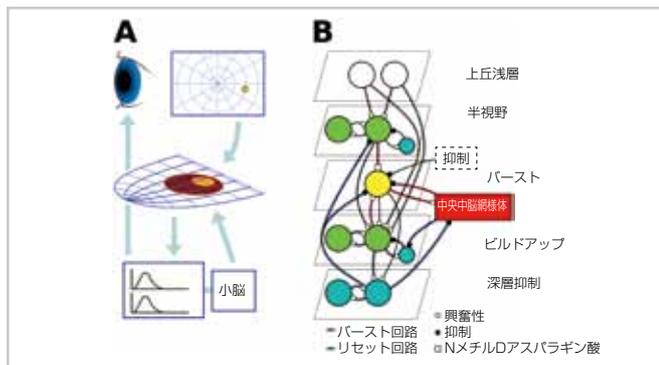


図1: A) 皮質下サッカード系の主成分と流れ。網膜からの出力は網膜位相の上丘を活性化し、網膜位相の上丘は脳橋および中脳内の水平方向と垂直方向の眼球運動系を活性化する。小脳は抑制因子として作用する。
B) 上丘中間層 (SGI) モデルの回路。赤の結線はバースト生成回路、青の結線は活性化の拡散とサッカード系のリセット、グレーの結線は両方に共通の回路を示す。

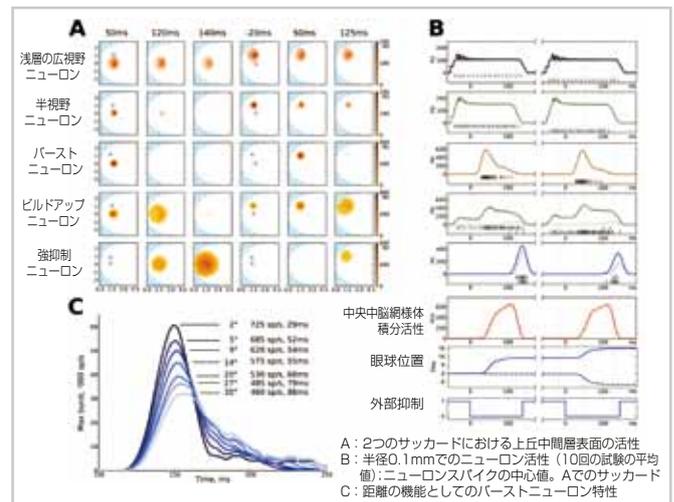


図2: A) 半径9°、角度0°の時、半径8°、角度45°のサッカードにおけるSGI表面の経時的活性変化。各点は、1モデルのニューロンで平均10msの活性。時間はサッカードの脱抑制に対して相対的。上から、浅層の広視野ニューロン、半視野ニューロン、バーストニューロン、ビルドアップニューロン、強抑制ニューロン。左から、バーストピーク、ビルドアップ活性ピーク、強抑制リセット、第2サッカードの準備活性、第2バーストピーク、第2ビルドアップピーク。
B) A) での各サッカードの中心点から半径0.1mmでの10回の試験の平均活性。同じ順序で示す。黒のバーは1回の試験でのニューロンスパイクの中心値で、上列は第1サッカード、下列は第2サッカード。各サッカードのビルドアップニューロンと強抑制ニューロンの点線は、もう一方のサッカードでの活性を示す。下に、全cMRF積分活性、水平方向 (実線) および垂直方向 (破線) の眼球位置の測定値、外部抑制を示す。
C) サッカード距離の機能としてのSGIのバーストニューロン活性の経時的活性変化。サッカード距離が長いほど活性ピークは低下し、持続時間は延びる。

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います。

分子動力学 (Molecular Dynamics, MD) シミュレーションは分子を構成する原子に働く力と運動を計算することで分子の運動・構造変化等を再現する手法の一つです。

生命科学においては、生命現象の基盤となるタンパク質のような生体分子の働きを解明するのに利用されています。タンパク質等の生体分子のシミュレーションでは、原子に働く力として、共有結合による力、ファンデルワールス力、クーロン力を扱います。共有結合は調和振動子のような単純な古典力場で代用し結合が連続したせいぜい4個の原子までしか扱わないので計算量は共有結合する原子の個数にほぼ比例しますが、任意の原子間に働くファンデルワールス力とクーロン力では全ての原子の組み合わせを扱うので計算量は原子の個数の二乗に比例し、原子の個数が多い場合はファンデルワールス力とクーロン力の計算が処理時間の大半を占めるようになります。高速化の手法としてもっと単純なのは遠距離の原子間の力を無視するカットオフ計算です。ファンデルワールス力は距離の7乗、13乗に反比例して小さくなるので、1.4nm以遠を無視してもほとんど計算精度に影響しません。クーロン力は距離の2乗に反比例して小さくなるだけなので、単純なカットオフ計算では計算精度に影響します。そのため、遠距離のクーロン力を高速に計算する手法としてFFTを利用するPME (Particle Mesh Ewald) 法が主流となっています。しかし、FFTは広い範囲の通信が必要で並列度が高くなると通信時間の増加が問題になるため、高速多重局展開法 (FMM) 等のFFTを使わないの計算手法を導入するといった対策を検討しています。

カットオフ計算による大規模テスト計算として、“分子の組み合わせ”のタンパク質動態への影響を調べるためのシミュレーションを行ないまし

た。細胞内環境では通常の実験より高濃度でタンパク質が存在しており、タンパク質の組み合わせが構造や相互作用に影響を与えていることが実験的にも示されています。タンパク質の組み合わせを再現するには多数のタンパク質を含んだシミュレーションを行なう必要となるため、大規模計算が可能な「京」の活躍が期待されます。テスト計算では、水中と卵白のタンパク質ovalbumin (原子数6000弱) が30%含まれる環境で、TTHA1718 という1000原子程度のタンパク質の構造を比較しました。図1はTTHAを構成するアミノ酸の位置の揺らぎの大きさを示しています。計算時間が短く実験を再現するには至っていませんが、水中の場合(赤線 vit1、青線 vit2)とovalbuminを含む場合(緑線viv)でTTHAの動きが異なることは確認できました。図2はTTHAがovalbuminに囲まれている様子を可視化したものです。

「京」の性能評価(注)での実行では、1億8000万原子、カットオフ2.8nmの計算で、効率30%以上、実効計算速度1.3PFlops という性能が得られました。カットオフ計算部分は効率40%以上を達成しています。

現時点でカットオフ計算とPME法・FMMが実装済みです。PME法・FMM等は詳細なアルゴリズムの変更も含めた最適化を行なっています。「京」のメッシュ/トラスネットワークの構造に合わせた通信方式を実装していますが、ノードあたりの原子数が少ない場合は計算時間より通信時間の方が長くなるため、通信時間を短縮する改良も継続しています。

本研究は高度化チーム小山洋、舩本現、長谷川亜樹、森本元太郎、沖本憲明、平野秀典、生命システム研究センター宮下尚之との共同研究です。

本研究の結果の一部は、RIKEN Integrated Cluster of Clusters (RICC) システムを利用して得られました。

(注) 理化学研究所次世代スーパーコンピュータ開発実施本部が実施している次世代スーパーコンピュータの開発・整備における京速コンピュータ「京」の性能評価。

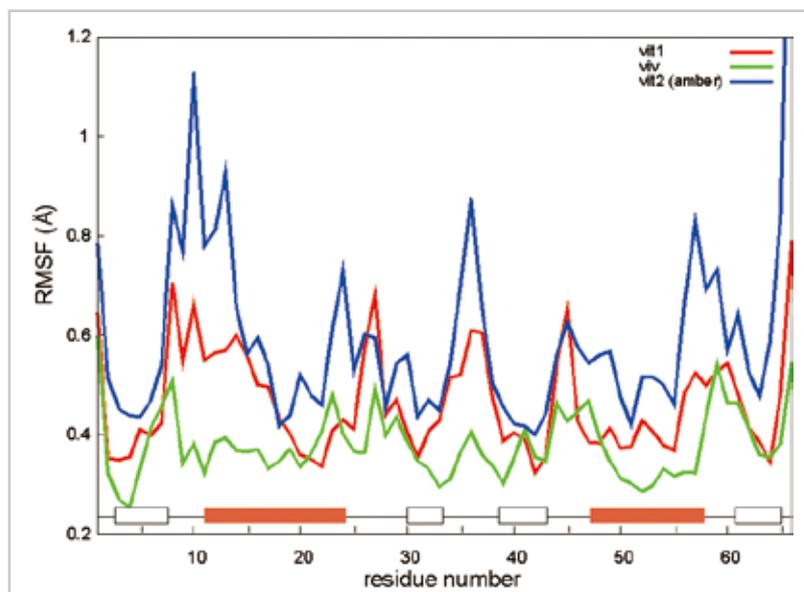


図1: TTHAの構造の揺らぎの大きさ 赤・青は水中、緑はovalbuminを含む場合。横軸はアミノ酸の番号、縦軸は揺らぎの大きさ。

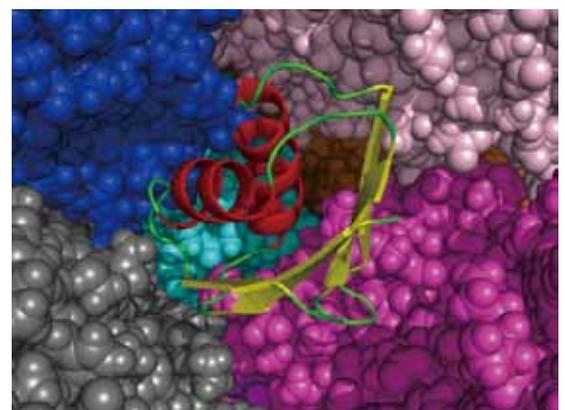


図2: ovalbuminに囲まれたTTHA中央のらせんや矢印で表現されているのがTTHA、周囲の球はovalbuminの原子で、一色が一つのovalbuminを構成。

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、柳田、江口

「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、4つの研究開発課題を設定している。それはどのような内容なのか、そして新たな研究開発はどのように進められていくのか。前号に続き、統括責任者・柳田敏雄氏、副統括・木寺詔紀氏、江口至洋氏にお話いただいた。

●4つの研究開発課題

柳田（敬称略） まずは、これから分野1がどのようなテーマに取り組んでいるのか、課題1～4の紹介から始めましょうか。

木寺 私たちのゴールは、「京」の性能を最大限に活かして生命科学に貢献する成果を出していくことです。そのためには全くゼロから走り出すのではなく、すでに



確立された“使える計算手法”を活用していくことが基本になるわけです。その意味では、課題1「細胞内分子ダイナミクスのシミュレーション」、課題2「創薬応用シミュレーション」では、分子シミュレーションの技術を活用した生体分子のシミュレーション、課題3「予測医療に向けた階層統合シミュレーション」では、構造解析に流体の計算を取り入れた流体構造連成解析、そして課題4「大規模生命データ解析」では、ゲノムを基軸とした生命データの解析と、大きく3種類に分けられます。さらに詳しく見ていくと、課題3には脳神経系シミュレーションが入るというように、それぞれ広がりがありますが、いずれにしても、これまでに開発されてきた手法が使われます。例えば課題1、2の生体分子シミュレーションでは、量子化学計算、分子動力学計算、粗視化モデル計算といった既存の手法で、「京」を使ってスケールアップしてやっていくわけです。

分子シミュレーションが、なぜ2つの課題に分かれているかということ、分野1の名称にも「医療および創薬基盤」と明示されており、このシミュレーション研究の成果として、医療と創薬への貢献を重視していることを明確にしたいからです。分子シミュレーション分野の応用として、集中的に創薬に向けた取り組みを推進していくために、課題2が設定されています。もちろん課題2だけでなく、すべての課題で、医療および創薬への貢献を意識しながら進めていくわけですが。

江口 課題2の方はある程度モデルが構築されているのに対して、課題1の方は、これまでの延長線上でスケールアップするだけではない、さらに新しいモデルを考えていく必要があるのではないですか。そうしたチャレンジングなところが課題1には課せられているよう

に思うのですが、いかがでしょうか。

木寺 まさにおっしゃる通りです。課題2においては、確立されたモデルが用意されています。そのモデルの効率的な利用によって、大きな成果が期待されています。しかし、それだけが唯一の手法ではなく、今後さらに優れたものへと発展していく可能性があります。その意味では、まだまだ課題2もチャレンジングであると思います。

江口 課題1についてはどうですか。

木寺 課題1でするべきことはむしろ、既存の分子動力学レベルのモデルで「京」を最大限に使うことで、既存のシミュレーションのレベルを拡大して、より長時間、より巨大なシステムの計算を行い、今までは見ることができなかった初めてそこで分かったといえる、分子レベルで起こる生物学的なイベントを再現することから始めなければなりません。そのなかから、次世代のシミュレーションのモデルをさがしていくという順番だろうと思います。

柳田 生命システムは、いってしまえば超自由度の世界ですね。今まで、いろいろな仮定などを入れることによって、コントロールできる程度の自由度に落とし込んで、生命システムを理解しようとしてきました。もちろん、「京」ができたからといって、超自由度のものをポンと放り込めば、ひょいと答えが出てくるわけではありません。それでも、試行錯誤しながらモデルをつくり、そのモデルを超自由度の方へ戻すときに計算機を使ってシミュレーションを行い、予測を導き出していくわけです。ですからシミュレーションは、これまで見てきた現象を説明するだけではなく、次の予測につながるシミュレーション、そのためのモデル化でないといけません。それには、実験・モデル・シミュレーションを組み合わせていくことも重要です。

木寺 モデル化と予測は、実は非常に難しい組み合わせです。予測するためには、モデルが予測可能なものになっていなければいけません。過去のデータを説明するだけでは、全くその先へ進めませんからね。そのためには、アブイニシオ（非経験的）な、できるだけ多くの自由度を放り込んだようなシミュレーションを行い、その結果の中に新たな低自由度で表現できる現象を発見していくことによって、モデルを改良していくことが必要です。もちろん、実験そのものをモデルの改良につかってもいいわけですが。

●最終目標を見据えた萌芽的なチャレンジも重要

江口 課題3も、非常にチャレンジングですね。

木寺 ここは、連成・統合ということに重きをおいたプロジェクトになっています。例えば、今までの血流のシミュレーションは、血管や

心臓などの臓器は構造として、血流は流体としてとらえていたが、構造と流体を連成したシミュレーションをやっています。さらに、これまで個別に行われていた心臓と全身血管のシミュレ



タを統合したり、そのうえに全身の筋骨格系、脳神経系といった全く別のシミュレータを持ち込んで、それらをすべて結び付けたヒトの全身シミュレーションをやっているというわけです。また、医療支援ということを強く打ち出し、病態の予測や治療法の検討、薬効の評価に貢献するシミュレーションを、はっきりと目的として掲げています。

江口 本当に統合できるのだろうかという思いを持つ研究者もいるようです。

柳田 ただ、臨床医の先生たちからすれば、いちばんやってほしい研究であることは間違いありません。実は、私は大阪大学第一生理学教室で教授をしていたんですが、今まで生理学ではこうした研究を、いわばクラシックにやってきたわけです。こういう研究が計算機科学と結合した形で発展していけば、医学にとって非常に大

きなインパクトがあると思います。確かにチャレンジングには違いありません。しかし、生物学はそこをめざしていかなければいけないはずです。「京」を使って成果を出すということはもちろん大事なことです。最終目標がどこにあるのかを見据えることも重要です。

木寺 達成できる明確な目標を掲げなければいけないというのは正論です。しかし、一方には、「京」を使って、スケールアップして見えてきたことだけを目的にしていけばいいのか、という考え方もあります。もっと発見的な要素があるチャレンジングな試みを「京」でやっていきたい、本当のゴールをめざした研究を「京」からスタートさせたい、そうした思いも確かに重要です。課題3のように、階層を超えて全身の統合的なシミュレーションをやっているというのは、まさにそうした今後の芽を育てていく、萌芽的なチャレンジという意味でも重要ではないかと考えています。

●生命データ解析によって医療に貢献

江口 課題1～3は、それぞれ他分野とオーバーラップする部分がありますが、課題4はデータ解析というやや独特な研究ですね。

柳田 ゲノムを基軸とした生命データの大規模な解析によって、生命プログラムとその多様性を理解するとともに、個別化医療やゲノム情報の産業利用などに役立てていくことを目標にしています。

木寺 ささまざまな最新の実験技術によって生命情報が大量に生産されており、それを解析することにより、その意味を理解し、さらに予測するところまで踏み込んでいかなければいけないわけです。そうした生命情報を「京」によって解析しているというプロジェクトです。その最も重要な部分であり、先行して注力してきたのが、次世代シーケンサーによって産出される大規模なゲノム情報です。大規模化・精緻化が急速に進んでいる、こうした生命システムデータの解析をやらなければいけないわけですが、現状として、これまでの計算機資源では全く間に合わなくなってきています。そのため、産出された生命システムデータを解析する基盤を「京」のなかに求めようというのです。まさに、直接的に実験データに結び付いたテーマですが、「京」を使うことによって、新たな生命科学におけるデータ解析的なサイエンスが実現すると考えているわけです。

江口 課題4の代表を務める宮野悟先生は、生命科学において第一原理的な方程式で説明されている部分は非常に限られていて、多くは第一原理がはっきりしていないということを強調しておられます。生命科学をサイエンスとしてとらえていくためには、データを持って語らしめる、つまり大量に出てきているデータのなかから法則

性を見出すための努力をしていく必要があるというお考えだと私は理解しています。そのために、最初の1、2年は、次世代シーケンサーで得られたデータをサクサクと解析できるようなシステムを、「京」上につくることに重点を置きたいと話しておられました。

木寺 少なくとも、そうした要請が極めて強いということがひとつ、さらにそうしたシステムが「京」で十全に動かないと、そこから先の大規模生命データ解析が全面展開できないということがあります。シーケンサーの発展によって、今後、ゲノム解析のデータがどんどん膨らんでいくことは間違いありませんから。

柳田 まずはデータをきちんとしなければいけない。その膨大なデータのなかから、例えばある人が病気になるという予測につながるような意味のある情報を抽出していくわけですね。そのためには、さらに大きな計算機資源が必要になりますね、言ってみればデータマイニングをしなければいけないわけですから。

江口 一方で、最近宮野先生が発表されたがんの転移に関する研究では、遺伝子ネットワーク解析によって転移に関与していると思われる遺伝子を見つけ出すことができたそうです。予測だけにとどまらない、こうした取り組みも重要になっていくでしょうね。



●発見的シミュレーションをめざす

木寺 4つの研究開発課題をそれぞれ見てきたわけですが、全体を通して「京」を使うことによって、今までとどこが変わるのかをしっかりと考えておきたいです。HPCI戦略プログラムにおける研究開発の正統的な考え方としては、「京」を使わなければ得られない、計算規模を拡大したことによる研究成果が求められるわけですが、それが単にすごいスケールの計算をやりましたというだけで、本来めざすべき生命科学の成果に結び付かないものであるなら、それは私たちの求める目的や成果ではないと思うのです。

柳田 確かに、「京」を維持していくための象徴的な研究をしていかなければいけないとしたら、それは本末転倒です。

木寺 生命科学の分野で、計算機科学はまだ広く浸透しているとはいえない、特に計算機資源のハイエンドにある「京」を使った生命科学の研究のプロトタイプは、残念ながらありません。そうしたなかで、とにかく使ってみなさい、結果を出しなさいといわれるままに使うのではいけないと思うのです。このプロジェクトの最終的な評価は、本当の意味でのサイエンティフィックな成果がどれだけ出せるか、そこが重要だと思うのです。

柳田 先ほども話ができましたが、今までは超自由度な生命システムのなかから、コントロール可能な数少ない自由度に落とし込んで理解しようとしてきたわけですが、「京」では、とりあえずより多くの自由度を入れてみるというのでもいいかもしれません。そこからキーとなる自由度が浮かび上がってくるかもしれません。

木寺 予定調和的なシミュレーションから、結果の中に予想だになかったことを発見するというような「発見的シミュレーション」という領域へジャンプする。「京」を使うことによって、それが可能になるということですね。

柳田 「発見的シミュレーション」というのは、非常にいい言葉ですね。超自由度の生命システムのなかでの発見的なシミュレーションをめざす、これが今日のまともになるんじゃないですか。

木寺 もうひとつ忘れてはいけないのは、それが「京」によってできるようになるということです。非常にポジティブなことが、これから生物学で始まるわけです。そこに立ち会えることの幸運を十分に認識しながら、ゴールをめざして頑張っていくことが必要だと思います。

全原子モデルにもとづく ヌクレオソームポジション変化の 自由エネルギープロファイル計算

日本原子力研究開発機構 量子ビーム応用研究部門
(上から) 河野 秀俊、石田 恒、米谷 佳晃、池部 仁善
(分野1-課題1)



ヒトの遺伝子は約30億塩基対からなるゲノムDNAにコードされ、その全長は約2mにもなります。ヒトを含めた真核生物のゲノムDNAは、ヒストンと呼ばれるタンパク質を芯にして、DNAが2回弱巻きついたヌクレオソーム構造を形成することで、直径約数 μ メートルの細胞核内にコンパクトに収納されています。ヌクレオソームはさらに、クロマチンと呼ばれる高次構造を形成しています。近年、それらの構造変化が遺伝子の発現制御において重要な役割を果たしていることがわかってきました。私たちの遺伝子はヌクレオソーム構造をとることによって、損傷から守られ安定に存在しています。一方、生命機能の根幹となる転写、複製、修復、組み換えといったDNAの代謝の際には、制御タンパク質やポリメラーゼと直接相互作用するために、ヌクレオソーム構造が壊されたり、その位置を変えたりすることが起こります(図1)。この一見すると相矛盾するような動きがヌクレオソーム構造にはあります。果たして、このようなヌクレオソームの構造破壊や位置変化(ポジション変化)はどのようなメカニズムで起こるのでしょうか。

さらに、ヌクレオソーム構造の変化は、細胞分化のメカニズムにも影響を与えることがわかってきました。私たちの体を構成する個々の細胞は、基本的に同じ遺伝情報(ゲノムDNA)を持っています。しかし、一度細胞が分化すると、一般には皮膚細胞からは皮膚細胞、肝細胞からは肝細胞しかできません。最近の研究により、そのからくりはヌクレオソーム構造の構成要素であるヒストンタンパク質の化学修飾と密接に関係することが明らかになってきました。このように、遺伝子の発現パターンがある種の記憶として維持、継承されていくことはエピジェネティクスと呼ばれ、盛んに研究されています。

本研究では、コンピュータシミュレーションを使っ

て、ヌクレオソームのポジション変化を自由エネルギープロファイルの観点から明らかにしていきます。また、DNAの塩基配列の違い、ヒストンの化学修飾、変異体の影響が自由エネルギープロファイルにどのような形で現れるのか調べることによって、これらのエピジェネティクス研究を進めていきます。私たちはこれまで、SCUBAという大規模分子動力学シミュレーションに適したプログラムを開発してきました。SCUBAでは、対象の系を空間分割して並列に計算できるアルゴリズムをいち早く導入し、DNA相同組換え時に見られるHolliday Junction 分岐点移動や、200万原子を超える系でのリボソーム分子のダイナミクス解析を実行してきました。図2に示すように、ヌクレオソームは、熱揺らぎによって自発的にゆっくりと呼吸をするかのように揺らいだり(breathing)、ヒストンタンパク質に巻きついたDNAがほぐれたり(opening)することが知られています。本課題では、そのようなヌクレオソームの運動メカニズムや構造安定性が、ヒストンの化学修飾、その変異体、さらに、巻きついたDNAの塩基配列の違いによってどのように変化するのかを、京コンピュータを使って明らかにしていきたいと考えています。このダイナミクスはサブ秒の時間オーダーであるため、世界最速である京コンピュータをもってしても、計算中に熱揺らぎの現象を待っているだけでは追いつけることはできません。そこで、適切な反応座標を設定し、その座標に沿って構造変化を誘導することによって、構造変化に対する自由エネルギープロファイルを計算します。このようなシミュレーション計算を通して、遺伝子発現がどのようにコントロールされているかという生命活動の基本的な分子メカニズムを、物理化学的に理解することを目指しています。

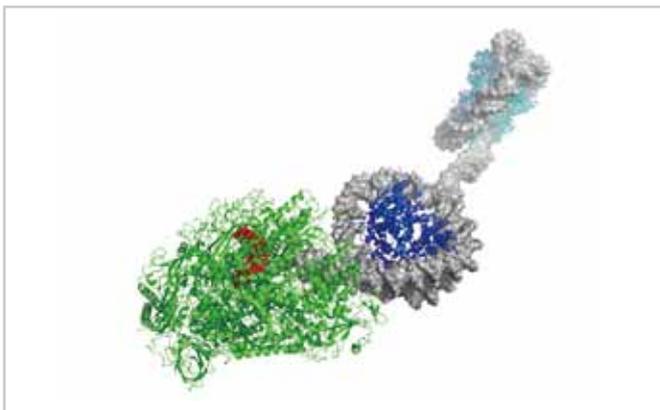


図1: RNAポリメラーゼによるmRNA合成。RNAポリメラーゼは、ヌクレオソーム構造をとっているゲノムDNAからどのように転写反応を進めているのでしょうか。そのような問題にシミュレーション計算を通して挑んでいます。RNA pol II: green; RNA: red; DNA: white; histone: blue and cyan.

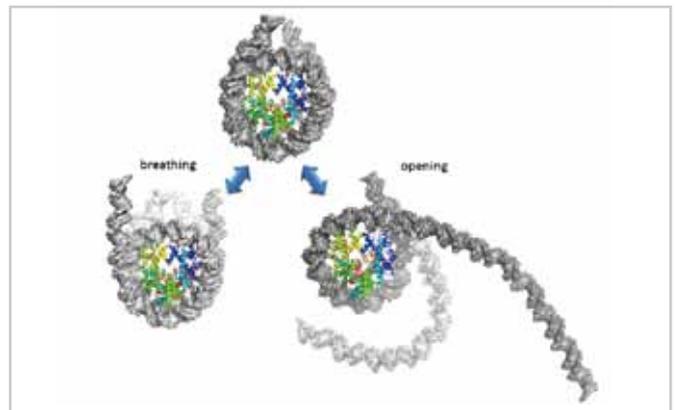


図2: シミュレーション計算によって、図に示したようなヌクレオソームの構造変化に対する自由エネルギープロファイル解析を行っています。ゲノムDNAは呼吸するかのようにヒストンタンパク質から離れたり巻きついたり、また、大きくほぐれたりしていることが実験的に示唆されています。

骨格筋の活動の推定と 脊髄反射の神経モデル



東京大学情報理工学系研究科
中村 仁彦
(分野1-課題3)

ロボティクスで開発されてきた効率的な剛体系の動力学計算を基礎にして「逆システム」^[1]として、人間の身体の中の情報を力学とフィードバック構造から推定する技術を開発してきた^{[2][3]}。身体の外からの計測と計算で身体の中の情報を覗き見ようというわけである。

CRESTの領域「脳を創る」(研究統括：甘利俊一)の一つのプロジェクトを1998～2003年まで実施し、ヒューマノイドロボットの身体の運動パターンを抽象的な非線形ダイナミクスの引き込みを用いて生成・制御する研究をスタートさせた。一方で、ロボットの順動力学計算法として $O(\log N)$ で計算できるAssembly and Disassembly Algorithmと呼ばれる並列計算法を開発した^[4]のがきっかけで、人間の全身の筋骨格系を物理モデルとして計算することを始めた。全身の155自由度の骨格系を997本のワイヤーで引っ張るモデルである^[5]。その後の2つの科研費 基盤研究 (S)「知能の力学的情報処理モデルの展開」(2003-2007年度)「身体運動と言語を統一した人間・機械コミュニケーションの成立」(2008-2012年度)を通じての柱の一つとなっている。

推定はモーションキャプチャーで計測した運動の幾何情報から、質量をもった骨格(人間の全質量を骨格に配分している)に加わる力を逆動力学計算法によって計算することから始まる。質量をもたないワイヤーモデルで表した全身の筋の張力を逆算することが次の計算である。ワイヤー張力は正または零とする不等式拘束条件を受ける。155<997の大規模な冗長性をもつ問題を2次計画問題として解を探索する。拮抗筋の活動を考慮できるように、モーションキャプチャーの際に無線筋電計を身体に取り付け最大32チャンネルの筋活動度を計測する。最適化問題で

はこの活動度を参考にしながら2次計画問題を解くのである(図1)。

好奇心はさらに内へと進み、筋活動から神経活動を推定する逆モデルを構成する方法に向かい、脊髄から末梢へ向かう運動神経系を扱おう。脊髄神経とは、脊髄から出る合計31対の末梢神経(頸神経8対、胸神経12対、腰神経5対、仙骨神経5対、尾骨神経1対)であり、これらはそれぞれ前枝、後枝となり124本の神経束に分かれる。各神経束が投射する骨格筋は明らかになっている。124本の神経束を同数の信号路と見立てて、その活動を推定する。脊髄の中には全結合型の数学的ニューラルネットワークを置き、神経束の活動度と、脳から下りてくる随意信号(ここに様々な仮説を用いる)によって学習させる。こうして得られた脊髄のニューラルネットワークは、神経系が解剖的な身体モデルと人間に特有の運動パターンに特化して最適化されたものとして、脊髄による反射弓を表している(図2)。

戦略プログラム課題3:「予測医療に向けた階層統合シミュレーション」(代表:高木周)では、筋や内臓などに質量を分散させたモデルを作り、弾性変形に伴う収縮・接触の応力分布を計算することで全身神経筋骨格モデルの高次元モデルとその並列計算アルゴリズムの開発に取り組んでいる。高木Gの骨格筋の筋繊維からのマルチスケールモデル、銅谷Gの脳基底核モデルと統合することで、例えば、パーキンソン病において脳基底核のドーパミン代謝系の変化が歩行やリーチングなどの動作の変化として出現するメカニズムを再現することを目指している。これによって、創薬、診断、治療に役立つ予測医療の基盤計算技術を確立する。



図1:筋活動の可視化

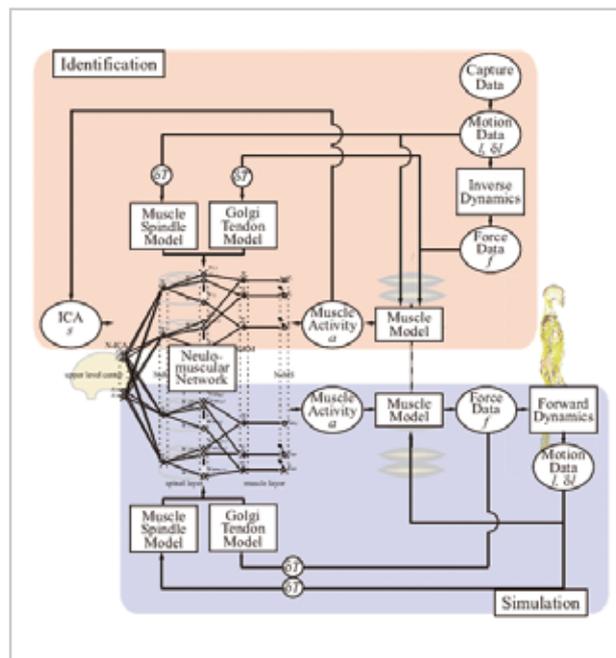


図2:脊髄反射モデルの構成

【参考文献】

- [1] 金子勝, 児玉龍彦. 逆システム—市場と生命の仕組みを解き明かす. 岩波新書(新赤版875), 2004.
- [2] A. Murai, K Yamane, and Y. Nakamura. Modeling and identification of human neuromusculoskeletal network based on biomechanical property of muscle. Proceedings of the 30th IEEE EMBS Annual International Conference, Vancouver, Canada, 2008.
- [3] A. Murai, K Yamane, and Y. Nakamura. Effects of nerve signal transmission delay in somatosensory reflex modeling based on inverse dynamics and optimization. Proceedings of the 32nd IEEE International Conference on Robotics and Automation, pp. 5076–5083, 2010.
- [4] K. Yamane and Y. Nakamura. “Efficient Parallel Dynamics Computation of Human Figures”. Proceedings of IEEE International Conference on Robotics and Automation (ICRA2002), 1:530–537, Washington D.C., May, 2002.
- [5] Y. Nakamura, K. Yamane, Y. Fujita, and I. Suzuki. Somatosensory computation for man-machine interface from motion capture data and musculoskeletal human model. IEEE Transactions on Robotics, 21:58–66, 2005.

ISLiM 成果報告会2011

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

2011年も押し迫った12月21日(水)~22日(木)の両日、東京大学本郷キャンパスにある武田ホールにおいて「ISLiM成果報告会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ソフトウェアの研究開発 (ISLiM) プログラムは、スーパーコンピューター「京」の開発に併せて、生命現象に対する理解を深め、特に創薬・医療に役立てることを目的に2006年10月に発足しましたが、終了まで一年を残すところとなりました。14研究機関の研究者の努力により、大規模かつ高精度のシミュレーションを効率良く実行するためのISLiM開発ソフトウェアの並列化チューニングも、33本のソフトウェアを対象に、京を使用して着々とすすんでいる状況です。そうした中、ISLiM成果報告会2011には、産業界、大学、公的研究機関等から141名の創薬・医療等の専門家にご参加をいただき、ISLiM ソフトウェア開発者が二日間にわたってその最新成果を発表しました。

はじめに次世代計算科学研究開発プログラムの茅幸二プログラム・ディレクターがオープニング・スピーチを行ないました。続いて、姫野龍太郎副プログラム・ディレクターが、2011年4月以降(一部は3月)から京を実際に使用して開発を進めているアプリケーションソフトウェアの最新の開発状況と性能データ(暫定値)を紹介しました。2012年11月の京の本格稼働に向けてISLiM ソフトウェアの最適化を鋭意進めている状況にあり、すでに一部のソフトウェア(cppmd やZZ-EFSI)は、京のピーク性能の20%~30%の実効性能値(暫定値)に達しています。今後

さらに性能向上に努めると共に、これらのソフトウェアを普及させる方向に力をシフトするということでした。このあとISLiM 全6チーム(分子スケール研究開発チーム、細胞スケール研究開発チーム、臓器全身スケール研究開発チーム、データ解析融合研究開発チーム、脳神経系研究開発チームおよび生命基盤ソフトウェア開発・高度化チーム)のチームリーダーとISLiMソフトウェア開発責任者が、その最新成果を発表しました。これらの発表資料はISLiMのウェブサイト(URL: http://www.csrp.riken.jp/2011/islim-houkokukai2011_j.html)からダウンロードできます。

二日間という限られた発表時間では、全開発ソフトウェアの詳細についてまでは報告できないため、武田ホール・ホワイエで全開発ソフトウェアについて38枚のポスター展示と説明も行われ、ソフトウェア開発者と参加者による議論がポスターの前で活発に行われていました。

今回、企業からの参加者から、ISLiMがいかに広くかつ深いレベルで実験も含めたソフトウェア研究開発をしてきたか強く実感できた、という成果報告会での発表を評価するコメントも聞くことができました。しかし、そのことは、ISLiM開発ソフトウェアの価値がプロジェクトの外にあまり届けられていなかったということも意味します。2012年秋にISLiM成果報告会2012を開催すると共に、ISLiMの開発ソフトウェアの価値と利用を普及させる取り組みをさらに進めて行く必要があると考えています。

謝辞 京での計算に関しては京速コンピュータ京の試験利用、2011年3月の特別運用での結果です。



高等学校で行った計算生命科学の授業

理化学研究所 HPCI計算生命科学推進プログラム
鎌田 知佐、藤原 康広、江口 至洋

昨年12月、兵庫県の神戸高校と西脇高校で「スパコン京とは何？生物学は変わるの？」と題する授業を行いました。5年という期限の付いたHPCI計算生命科学プログラムの終了後も、多くの若い高校生に計算生命科学に合流して頂き、計算生命科学をさらに発展させて頂くことを願って取り組みました。授業の初めに、今日話したいこととして、

- 1) 「京」が世界最速のスパコンになりました。最速の鍵は並列計算にあります。
- 2) 生物学を学ぶにも数理（数学、物理、化学）が必要になります。
- 3) 「京」は生物学と数理を融合する触媒です。

の3つを示しました。5月に神戸高校の先生と授業について相談した際に「スライドだけの講義ではダメですよ」と言われ、びっくり、ない頭を絞り授業に2つの実習を組み込みました。生徒から「実習が多くて楽しかったです。」といわれると、その苦労も吹っ飛びます。

神戸高校は60名ほど、西脇高校は90名ほどの生徒が参加してくれました。「京」の説明、戦略分野が取り組んでいる5つの課題の紹介に続き、戦略分野1「予測する生命科学・医療および創薬基盤」の紹介をかね、塩基配列の結合編集の実習と、ペプチド結合の作成実習を行いました。結合編集は15塩基対ほどの11組の断片を全生徒に配布し、実際に結合編集して頂きました。5分ほどで完成する生徒もいれば10分経ってもできない生徒もいて、教室中わいわい、がやがやでした。生徒からは「実習で塩基配列を二人でつなげてみたが、あまり数が多くないにもかかわらず、なかなか正解を見つけることができなかった。コンピュータではこれよりずっと長い塩基配列を一瞬でつなぎ、しかもミスがないということに驚いた。」といった意見が寄せられています。ペプチド結合の作成ではアラニンとフェニルアラニンの分子模型を生徒全員に配り、脱水縮合でペプチド結合を作ってもらいました。こちらはみんな静かに真剣に取り組んでいましたが、かなり悩んでいたようです。ペプチド結合の作成実習の後、タンパク質の分子動力学計算のシミュレーション結果を映像で示したところ「タンパク質が動くのを見て気持ち悪いと思いました。実際に自分の体の中でこのような現象が起こっていると知ってすごいと思いました。」と新鮮な感動の意見を寄せてくれました。携帯電話などのゲームに関心があると思っていましたが、このようなアナログ的な



実習に興味をもっていただけて驚きました。

HPCI戦略プログラムは、研究開発チームとともに計算科学技術推進体制構築チームを併せ持ったプログラムとなっています。推進体制構築チームのミッションは「研究開発チームによる研究開発目的の達成を支援する」ことにあります。その支援には超並列計算プログラムの高度化推進とともに、研究開発チームを社会全体で支援していただく体制の構築があります。高校生への授業は後者の活動の一環になります。授業に参加された生徒から「京はただ性能のいいコンピュータぐらいにしか思っていないでした。医療や防災、次世代の物作りや新物質・エネルギーの創生など様々な分野で活躍していることをこの講義で知ることができました。僕も将来こういう仕事に関係していける人間になりたいと思います。」といった声や「今まで京とは計算を行うだけのコンピュータだと思っていた。生物学や物理学とつながっていると知り、驚いた。同時に、興味が大きくなった。」との声を聞くと、授業に携わった者としてうれしくなれます。今後もご要望があれば、このような授業を実施していきたいと思えます。

最後になりますが、神戸高校の授業で代打を快く引き受けていただいた大阪大学の鷹野先生に感謝します。この授業用に分子模型の特注品作成を快く引き受けていただいた日ノ本合成樹脂製作所さん、生徒の参加数の塩基配列パズルを作っていただいた難波さんに感謝いたします。また、この機会をお作りいただいた神戸高校の長坂先生、西脇高校の藤原先生に深謝いたします。



イベント情報

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 (ナノ) 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 (ライフ) 公開シンポジウム

- 日 時：2012年 3月5日 (月) 10:00～18:20 及び 懇親会【合同セッション日】
2012年 3月6日 (火) 9:00～17:55 【パラレルセッション日】
- 場 所：ニチイ学館 神戸ポートアイランドセンター (神戸市) 〒650-0047 兵庫県神戸市中央区港島南町7丁目1番5
- 参加費：無 料 (懇親会は会費制)

プログラム (予定)

第1日目 3月5日 (月) 10:00～18:20 ナノ・ライフ合同セッション日 (18:30～懇親会)

ライフ分子チーム+ナノ生体合同セッション

特別講演 「スーパーコンピュータ「京」誕生までの軌跡～世界一への挑戦の経緯と技術～」 伊東 広樹 (富士通)
合同パネルディスカッション 「京とその先に向けて (仮)」

第2日目 3月6日 (火) 9:00～17:55 パラレルセッション、ポスターセッション日

ナノパラレルセッション

9:00～9:55 ナノ統合拠点報告
10:10～12:10 次世代ナノ情報機能・材料
13:40～15:40 合同ポスターセッション
15:50～17:50 次世代エネルギー

ライフパラレルセッション

10:00～12:00 パネルディスカッション
「HPCは生命にどこまで迫れるか」
13:40～15:40 合同ポスターセッション
15:50～17:50 パネルディスカッション
「ISLiMは医療・創薬にどう貢献できるか」

詳細はこちらをご覧ください <http://nanogc.ims.ac.jp/nanogc/sympo2012>

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、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。

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、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。