

# 先進的かつ独創的な研究プロジェクトを、 22.8T<sup>テラ</sup>flop/sのスーパーコンピューターが支援

独立行政法人産業技術総合研究所(以下、産総研)は、2005年春、臨海副都心センターにバイオ・IT融合研究棟を開設。バイオテクノロジーの研究に、最新のIT(情報技術)を融合させた研究施設として活動をスタートさせました。

このビルの7~10階にラボを持つ生命情報科学研究センターは、日本を代表するバイオインフォマティクス(生命情報科学)の研究拠点として先進的かつ独創的な研究プロジェクトを推進するために、IBMのスーパーコンピューターであるIBM @server® Blue Gene Solution(以下、Blue Gene Solution)を導入。22.8<sup>テラ</sup>flop/sのスーパーコンピューターを活用して、さまざまなシミュレーションに取り組んでいます。

ここでは、生命情報科学研究センターの研究活動と、それを支援するBlue Gene Solutionの役割について、研究センター長の秋山 泰氏と、研究員の関嶋 政和氏に伺いました。

## Interview ②

### 22.8T flop/s Supercomputer Supports Innovative and Creative Research Projects

In spring 2005, the National Institute of Advanced Industrial Science and Technology ("AIST") established a Bio-IT Fusion research facility at the new Tokyo waterfront sub-center, and started research activities that merge the latest IT with biotechnology research.

The Computational Biology Research Center ("CBRC"), which holds research labs in the seventh to tenth floors of this building, has installed a supercomputer of IBM called IBM @server® Blue Gene Solution ("Blue Gene Solution") in order to promote innovative and creative research projects as a representative research center of bioinformatics in Japan. In this research lab, various simulations are being conducted by using 22.8T flop/s supercomputers.

In this edition, we have interviewed Dr. Yasushi Akiyama, the Director of CBRC, and Dr. Masakazu Sekijima, a research scientist at CBRC, on the research activities at CBRC and the roles of Blue Gene Solutions that support such activities.

## 産総研の最新の研究拠点「臨海副都心センター」

休日ともなれば家族連れやカップルだけでなく、観光客が多数訪れる「お台場」。今や、東京を代表する観光名所の一つとなっています。

2001年4月、産総研はこのお台場に九つ目の研究拠点として臨海副都心センター(図1)を開所しました。

そもそも産総研は、工業技術院(1948年に工業技術庁として設立)を前身とする公的研究機関であり、「日本の経済社会の明日に向けて、常に新たな可能性を切り開き、新しい技術を提案していく」ことを使命に、研究活動を展開してきました。その研究範囲は「ライフサイエンス」「情報通信・エレクトロニクス」「環境・エネルギー」「ナノテクノロジー・材料・製造」「地質」「標準・計測」の各分野に及び、産業技術の広い分野におけるさまざまな技術開発を総合的に行っています。

つくば学園都市にある「つくばセンター」をはじめとする九つの研究拠点がネットワーク化され、約2,500名の研究員に加え、企業や海外からも約1,500名の研究者を受け入れている国内最大規模の公的研究機関です。産業界や大学との連携・協力も盛んであり、新しい産業の創造や技術移転にも力を注いでいます。

産総研に期待されている役割は次の三つです。

- ・日本の産業技術競争力の強化。先端的な研究によ

り、国際社会で勝負できる産業技術を育成していくこと。具体的にはナノテクノロジーやバイオ・ITの研究がこれに当たります。

- ・日本の長期的な政策を推し進める上で必要な研究により、日本全体の産業活性化につなげていくこと。エネルギーや環境関連の研究がこれに当たります。

- ・基盤的な研究を押し進めることで高い技術的裏付けを獲得し、自立した産業立国を目指すこと。

こうした研究活動を通じて、産総研は社会への貢献を果たそうとしています。

そして新しく開設された臨海副都心センターでは、生命情報科学研究センター・生物情報解析研究センターをはじめ、七つの研究ユニットが入居し、新規産業の創出や市場拡大につながるライフサイエンス、ナノテクノロジー・材料・製造、情報通信、環境エネルギーの各分野の研究分野において、独創的かつ最先端の研究を推進しています。

## CBRCの果たすべき役割

産総研の最新の研究拠点として、特にお台場が選ばれたのは、その地の利を生かすことで、外部と融合してのオープンな研究をより進めることができると考えられたからです。実際、産総研研究者が約70名

であるのに対し、外部から参加している研究者は270名を超え、ほかの研究拠点には見られない構成比となっています。

また、臨海副都心センターは、近隣の東京国際交流館・日本科学未来館とともに国際研究交流大学村(以下、国際大学村)を構成しています。国際大学村は、この三つの機関がそれぞれに国際交流・情報発信・産学官連携の機能を果たしつつ、有機的に結び付けることを目的としてい

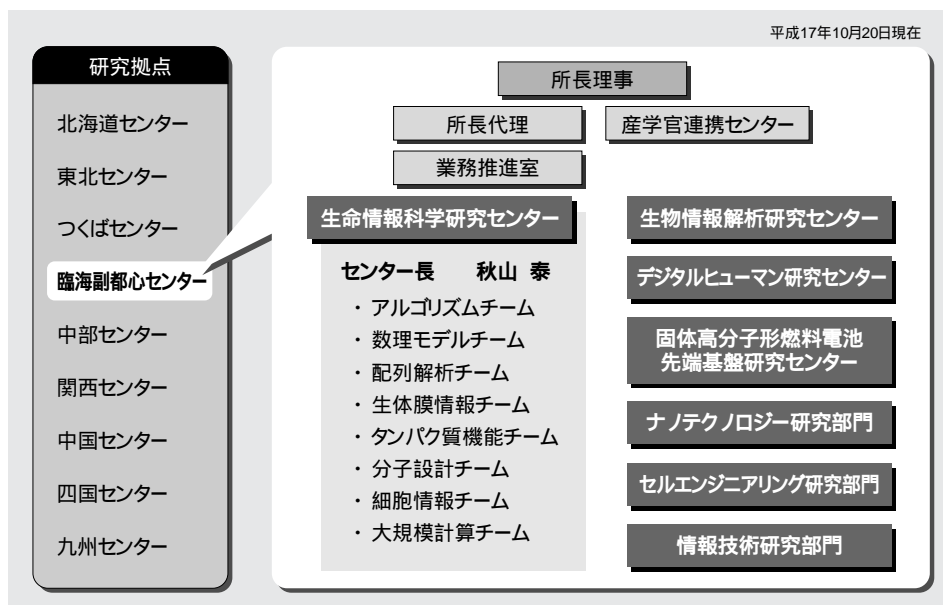


図1. 産総研組織図



独立行政法人 産業技術総合  
研究所  
生命情報科学研究センター  
研究センター長  
秋山 泰氏

**Dr. Yutaka Akiyama**  
Director  
Computational Biology Research  
Center (CBRC)  
National Institute of  
Advanced Industrial Science and  
Technology (AIST)

ます。これにより、世代や分野を超えた新しい思想や、科学技術の創造、地球規模の知的ネットワークの形成、国際的な指導者の養成などを図り、国際社会と人類の発展に貢献するという役割を担います。

このように臨海副都心センターは、産学官連携の研究拠点として、国内外の研究者が集い、交流や情報交換の場として機能することが期待されているのです。

中でも臨海副都心センターの別館として2005年春に竣工したバイオ・IT融合研究棟は、文字通り、バイオロジーの研究に最新のITを融合することを目指した研究施設であり、前述した七つの研究ユニットのうち、主に生命情報科学研究センター(以下、CBRC: Computational Biology Research Center)と生物情報解析研究センター(以下、BIRC: Biological Information Research Center)が連携して研究を進めています。

バイオ関連の研究活動は、産総研の臨海副都心センターのみならず、全国の公的研究所・大学・企業の研究施設などで行われていますが、今回お訪ねしたCBRCは、その規模・内容から見て、国内を代表するバイオインフォマティクス分野の研究拠点といえるでしょう。

バイオインフォマティクスは、生命科学と情報科学が融合した科学であり、生命実験により収集・データベース化された膨大な遺伝子情報やタンパク質の配列・構造に関する情報を、最新の情報数理理論と大規模高速計算システムを活用して解明しようとする学問です。

今日、バイオインフォマティクスは、二つの顔を持っています。一つは、ゲノム情報を解析するためのツールとしての側面であり、世界的な開発競争や特許取得競争に打ち勝つための手段としての顔です。もう一つは、生命現象の機構を真に分子レベルから深く理解するための新しい学問体系としての側面です。将来的には、計算機内に構築したモデルを使って計算機実験を行うことが、分子生物学の主要分野の一つになることは間違いなく、それにより高コストな実験に頼らずに研究開発を進めることが可能になるばかりでなく、倫理・安全の面からも社会に貢献することが期待されています。

CBRCが実際に取り組んでいる研究テーマは次の通りです。

#### ・バイオインフォマティクスの基盤となる数理モデルの整備

確率モデルなどのバイオインフォマティクス研究の共通基盤となる数理理論、アルゴリズム、および生命情報の高度な格納・検索・演えき方式に関する研究を行っています。

#### ・大規模ゲノム配列から遺伝子領域および機能予測技術の研究開発

ゲノム配列から精度良く遺伝子領域を決定し、遺伝子がコードするタンパク質の構造や機能を含めた総合的な解析を高速に行う手法を開発しています。

#### ・統計情報と物理計算の融合によるタンパク質構造予測技術の研究開発

タンパク質立体構造予測のための物理エネルギー計算と統計的知識を融合した計算手法や、化合物とタンパク質のドッキング予測の手法を開発しています。

#### ・細胞内の生命現象のシミュレーション技術の研究開発

細胞内の代謝ネットワークや遺伝子制御ネットワークなどを解明するための情報解析手法、およびそれらの統合的シミュレーションシステムを開発しています。

研究センター長である秋山 泰氏は、CBRCが設立されたいきさつについて次のように語ります。

「実はこのセンターができるまでは、国内にバイオインフォマティクスを単独で扱う大規模な研究拠点がなく、この分野の研究者や研究チームは、生命実験プロジェクトで収集したデータを計算機で解析する一部門として位置付けられることがほとんどでした。

そのために、バイオインフォマティクスの研究者はさまざまな研究所・大学・企業に散逸せざるを得ず、1990年前後にバイオインフォマティクスの研究拠点が設立された欧米に比べると、研究成果の共有や人材の蓄積が遅々として進みませんでした。また、実験現場では汎用的な技術は生まれにくいだけでなく、開発も近視眼的になりがちでした。

その一方で、実験プロジェクトの現場で、実験者の声を聞きながらバイオインフォマティクスの研究に取り組むことにも極めて多くの利点があります。

そこで、生物情報解析の研究者とも緊密に連携してシナジー効果を生み出すことを目的に、バイオ・IT研究棟が建設され、BIRCと一緒に研究を進めることにしました。こうしてCBRCは、日本初の本格的なバイオインフォマティクスの研究拠点として活動を始めたのです。

とはいえ、この分野における優れた研究が全国各地の研究グループにより進められていることも見逃すことはできません。そこで、既存の関連研究組織とのアライアンスや、バイオインフォマティクスの研究拠点として人材の結集・養成を行うことも大きな目的となっています。企業や大学などの外部研究員に門戸を開いたオープンな組織として、国内の研究機関・大学などと役割を分担しつつ、国内の研究コミュニティの人的交流の中心的な場所として機能することが、当センターの使命なのです」

## バイオインフォマティクスと超高速コンピューター

CBRCでは、バイオインフォマティクスの基盤となるアルゴリズムや情報表現手法の研究をはじめ、多くの実験データを統合して生物学的な事実を推論・計算するための情報処理システムの構築、およびこれらを駆使したさまざまな生物ゲノムの網羅的解析など、多数の研究チームが幅広い活動を行っています。

そして、バイオインフォマティクスに不可欠なのが、強力なコンピューティングパワーです。

この分野の計算機活用を歴史的に見ていくと、当初は、配列や立体構造間を比較する処理が中心でした。「相同性検索」や「モチーフ発見」などの計算では、

独立行政法人 産業技術総合  
研究所  
生命情報科学研究センター  
タンパク質機能チーム 研究員  
関嶋 政和氏

**Dr. Masakazu Sekijima**  
Research Scientist,  
Protein Function Team  
Computational Biology Research  
Center (CBRC)  
National Institute of  
Advanced Industrial Science and  
Technology (AIST)



既知の幾つかの典型的なデータと比べて、入力情報が似ていれば帰結も同一だと仮定するのがこの方法の特徴です。こうしたアプローチにもコンピューティングパワーが必要でしたが、今日では、コンピューターは「Aならばa」「Bならばb」という膨大な既知データから、「Cならばc」という新しい結果を類推する推定処理に用いられることが多く、こうしたアプローチに求められるコンピューティングパワーは前者とは比較になりません。さらに今後は、初期値と一般的法則から結果が得られる分子動力学法やMO法(分子軌道法)によるシミュレーションへの期待が高まりつつあり、さらなる高速計算が必要となっています。こうした状況の中で、世界をリードする先進的な研究を進めていくには、計算資源の効率的運用と拡充を図るなど、積極的に大規模計算環境を整えていかなければなりません。

「バイオインフォマティクスにおけるコンピューターの役割は、当初は、膨大な実験データを記録したり、あるいはDNA(デオキシリボ核酸)配列間の相同性を比較するという単純なものでした。

ただ、それだけでも数年ごとにデータベース容量が倍になるほどデータの蓄積が進んでいるため、処理すべきデータはますます膨大となり、それだけに高速なコンピューターが必要となっていました。

しかも最近では、そういった『情報の整理』にコンピューターを利用するだけでなく、ある情報から質的に異なる別の情報を生み出す『情報の変換』に使われることが増えてきています。膨大なデータを処理

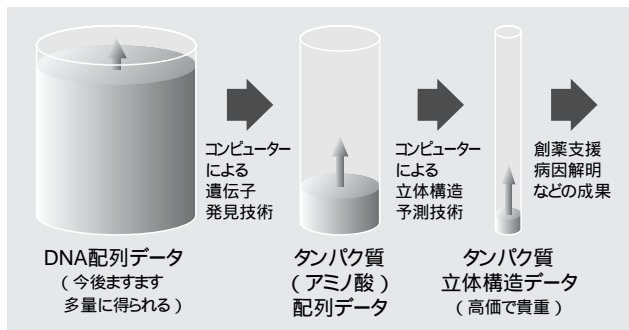


図2. 情報整理から情報変換へ

するために高速なスーパーコンピューターが必要なだけでなく、情報変換のためにさらに強力なコンピューティングパワーが求められているのです。

例えば、実験装置が進歩した結果、さまざまな生物種のゲノム(ある生物種が個体を完全な状態に保つために必要な遺伝的情報)が実験により次々と読み取れるようになってきています。ゲノムの実体であるDNA分子の配列データは今後ますます多量に蓄積されていくことでしょう。その一方で、DNAの一部にとびとびにコードされているアミノ酸(タンパク質)の配列を得るには、高コストな実験手法が必要であり、データの蓄積がなかなか進んでいません。実際、アミノ酸配列データはDNA配列データの25分の1程度しか蓄積されていないのです。さらに、アミノ酸配列データからタンパク質の立体構造データを得るには膨大な実験費用が掛かるため、アミノ酸配列データの80分の1程度のデータしか得られていません。つまり、着々と蓄積されつつあるDNA配列データと比べると、そこには2,000分の1という情報格差が存在するのです(図2)。

このギャップを、コンピューターによる情報変換で埋めることができれば、この分野の研究は大きく前進します。仮に、コンピューターによる計算で完全な情報変換ができなくても、計算により興味深い結果が得られたときに実験を集中することで、実験コストを抑えるだけでなく、手間を掛けずにゲノム解析を進めることができることとなります。つまり、超高速コンピューターを導入することで、実験に比べて低コストかつスピーディーに分析結果を得ることができるのです(秋山氏)。

## Blue Proteinの導入へ

バイオインフォマティクスの研究には超高速コンピューターが欠かせないことから、CBRCは以前からIBM @server BladeCenter®をはじめとするさまざまなPCクラスターを活用してきましたが、他国の研究に負けないためには、さらに強力なコンピューティングパワーを持つコンピューターを導入することが必要でした。

そこで、バイオ・IT融合研究棟の建設に際して、新しくスーパーコンピューターを導入することになりました。仕様調査に当たってCBRCが目にしたのが、IBMのBlue Gene Solutionです。

Blue Gene Solutionは、「スーパーコンピューターTOP500リスト」において、2004年11月、2005年6月および11月の3期連続で1位にランクされている世界最速のスーパーコンピューターIBM Blue Gene/Lをベースにしたスーパーコンピューターであり、高速演算能力のみならずコストパフォーマンス、低消費電力、省スペース性で、世界中の研究機関から高い評価を得ています。

入札に当たって、日本アイ・ビー・エム株式会社(以下、日本IBM)はバイオインフォマティクスに特化したBlue Gene Solutionを開発して提供することを提案し、その結果、CBRCへの導入が決定されました。本来、IBM Blue Gene/Lは、512ノード(1,024プロセッサ)を最小単位に超並列計算を行う設計になっていましたが、それを32ノード(64プロセッサ)ごとに並列計算を行えるように設計変更したのです。このコンピューターは、4ラック・8,192CPUで構成され、22.8Tflop/sという世界でもトップクラスの演算性能を発揮します。

「当センターでは、以前から遺伝子発見やデータマイニングなどにPCクラスターを使いこなしてきました。

ただ、タンパク質や低分子の分子動力学シミュレーションや、分子のドッキングの計算では、さらに膨大なコンピューティングパワーが必要であり、



Blue Protein

PCクラスターはコストパフォーマンスの点で必ずしもこうした計算に適していませんでした。

バイオ・IT融合研究棟の建設に当たっては、こうした計算が得意なバランスの良いマシンを探していたのですが、日本IBMから、強力なコンピューティングパワーを持つ上に、消費電力が少なく、省スペースで、かつメンテナンスしやすいシステムをご提案いただきました。それがBlue Gene Solutionだったのです。そして導入後、わたしたちはこのコンピューターにBlue Proteinという名前を付けました（秋山氏）。

## 日本のHPCを盛り上げたい

バイオインフォマティクスの分野に限らずスーパーコンピューターの利用には、最適化チューニングなどの技術支援が不可欠です。そこで日本IBMは、2005年6月にディープコンピューティング開発研究所を新設して、ハードウェアの技術者だけでなく、ソフトウェア、先進技術、基礎研究、さらには産業別の専門知識を持つ技術者のスキルとノウハウを結集し、高度な計算能力のニーズに対応する体制を整えました。

「やはりスーパーコンピューターは特殊なコンピューターであり、出来合いのサーバーを買ってきて動かすようにはいきません。十分なパフォーマンスを発揮するには、きめ細かなチューニングが必要ですから、その意味では日本IBMのディープコンピューティング開発研究所には期待しています。

実を言えば、HPC（High Performance Computing）の分野は、ベンダーにとってはビジネスとして見たときにさほど魅力的とはいえないのかもしれませんが、しかし、科学の中核ともいえる分野ですから、研究者だけではなく国や産業界の協力が必須だとわたしたちは考えています。

わたしたちがユーザーとして蓄積したノウハウと、ディープコンピューティング開発研究所の技術を組み合わせることで、大きな進展が期待できるのではないのでしょうか。その意味では、ベンダーのエンジニアの方たちと一緒に、日本のHPCを盛り上げていきたいと思っています（秋山氏）。

## 分子動力学シミュレーションに威力を発揮

CBRCでBlue Proteinを最も酷使している研究者の一人が、タンパク質機能チームの研究者 関嶋 政和氏です。

「わたしの主な研究テーマの一つが、プリオンタンパク質の分子動力学シミュレーションです。

生体を構成しているアミノ酸は約20種類ですが、それが複数接続されることによってペプチドやタンパク質になります。BSE（Bovine Spongiform Encephalopathy：牛海綿状脳症）の原因物質になると考えられているプリオンタンパク質もタンパク質の一種ですが、生体内でのダイナミクス（揺らぎ）や働きはほとんど知られていません。

しかし、プリオンタンパク質がほかのタンパク質と異なる法則に従っているわけではないので、一般的な生体分子の働きについての知見を基にして、物理や数学、コンピューターサイエンスを活用することで、解析できるのではないかというのがわたしの研究テーマの一つです。

生体内に存在しているタンパク質は、ある一定の構造を取っています。しかし、ある条件のときに正常な構造が異常な構造に転移して、疾病の原因となると考えられています。これは、普段の自分たちの姿を考えていただくと分かりやすいのですが、わたしたちはなるべく楽な格好でいるはずですが、その格好でいると楽で疲れにくいということです。タンパク質もそれと同じで、ある構造を取ることが多いのは、その構造が最も安定しているからです。ただ、何かストレスが掛かったり、普段とは異なる状況になると、楽な格好も微妙に変わってきます。病気になるというのは、まさにそういうときなのです。つまり『正常』から『異常』に転移するわけです。『異常』な状態というのは、ある特殊な環境においては最適な形なのです。わたしたちの研究では、簡単に言えば、どんな条件のときに『異常』な状態に転移するかを分子動力学というシミュレーションによって求めようとしているのです。

例えばヒトのプリオンタンパク質では、102番目のアミノ酸残基はプロリン（野生型）と呼ばれるアミノ酸です。それがロイシン（変異型）に変わると遺伝病にな

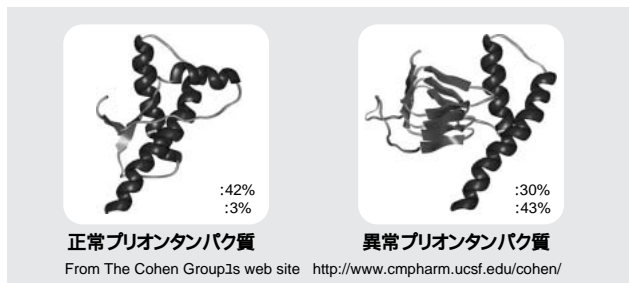


図3. 正常プリオンと異常プリオンの違い

ることが分かっています。アミノ酸は『残基』という単位で数えますが、たった1残基が異なることで病気になってしまうということです。そこで102番目のアミノ酸残基がプロリンか、ロイシンかという点を除いて、初期構造がすべて同じ条件でシミュレーションを実施し、遺伝病を引き起こす原因を探ったのです(図3)。従来、こういったシミュレーションを行おうとするとPCクラスターを用いても、それぞれ1カ月近くかかっていましたが、Blue Proteinを用いることで大幅に時間を短縮することができました。

もちろん最近のPCクラスターは、性能向上によりかなり速い計算が可能になっていますが、初期値の異なる物理シミュレーションを同時に行うのに限界があります。一方、Blue Proteinでは全8,192CPUを64CPU単位に分割し、同時に128通りの計算を効果的に行うことができるので、プリオンタンパク質の分子動力学シミュレーションには最適なのです。Blue Proteinをしのぐコンピューティングパワーを持つスーパーコンピューターも存在しないわけではありませんが、世界でもトップクラスの高速コンピューターを、バイオインフォマティクスの分野で占有して利用できるという点では、わたしたちは非常に恵まれていると思っています(関嶋氏)。

### 「網羅的分子シミュレーション」にも威力を発揮

関嶋氏が取り組んでいるもう一つ大きなテーマが、ウシ・ネコ・イヌなどの多くの動物種のプリオンタンパク質を調べる「網羅的分子シミュレーション」です。「例えば、ネコとイヌのプリオンタンパク質は、アミノ酸配列の中で4残基しか異なっていません。ところが、ネコとは異なり、イヌのプリオンタンパク質はプリオン

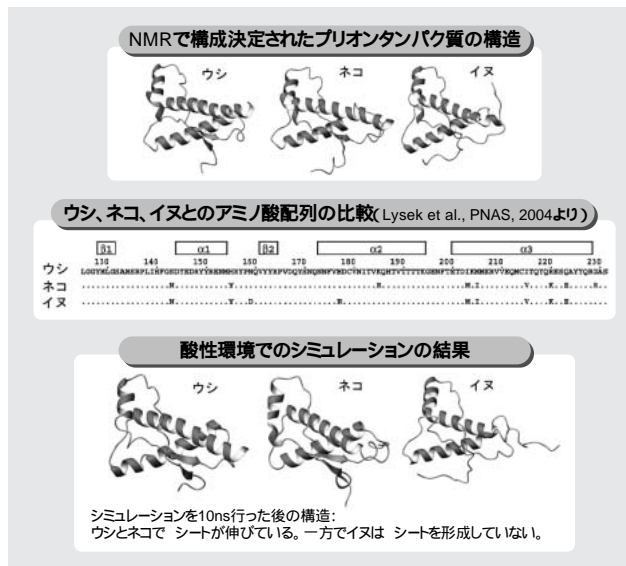


図4. 種にわたるプリオンタンパク質の網羅的分子シミュレーション

病を引き起こしません。ですから、この4残基の違いがプリオンタンパク質のダイナミクスや機能にどのような違いを明らかにできれば、プリオン病の原因に迫ることができるのです。

そこで、ウシ・ネコ・イヌのアミノ酸残基の違いがどのようにプリオンタンパク質のダイナミクスに影響を与えているのか、網羅的なシミュレーションを実施しています(図4)。

実はこういったシミュレーションは、従来のPCクラスターでは、コンピューティングパワーと計算時間の制約から実行が困難でした。複数ジョブを同時に並行処理する際に爆発的なコンピューティングパワーを発揮するBlue Proteinだからこそ(図5)、初めて可能になったシミュレーションなのです。

先ほど研究センター長から、日本IBMのディープコンピューティング開発研究所との今後のコラボレーションに期待したいという話がありましたが、その通りだと思います。Blue Proteinは、PCクラスターなどに比べればはるかに安定していますが、やはり一般のビジネス向けコンピューターと同じというわけにはいきません。使っている中で、さまざまな問題が生じるのは仕方のないことだと思います。ただそういう中で、わたしたちと一緒に問題を考えて、解決していただける体制が存在するというのは心強いですね。

コンピューターのいいところは、研究者がほかの作業をしていたり、あるいは寝ていても(笑)休みなく働いてくれることです。『ジョブが問題なく動いている

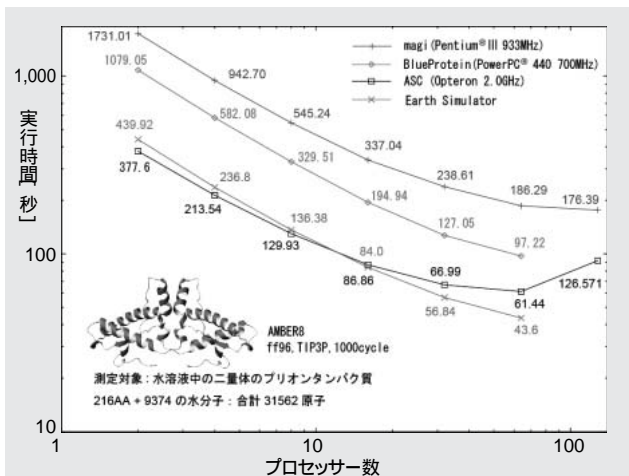


図5. コンピューティングパワーの比較

だろうか』とか『異常がないか』などとチェックするのは煩わしいだけでなく、何よりも研究がはかどりません。ですからアベイラビリティの問題を含め、例えばメモリーサイズやバッチサイズ、ネットワークなどについても、ディープコンピューティング開発研究所のエンジニアの方々と、お互いに知恵を出していければいいかなと思っています（関嶋氏）。

## 「お台場ドックプロジェクト」がスタート

秋山氏・関嶋氏をはじめとするCBRCの6名の研究者は、2004年10月、Blue Proteinの導入に先立って、Blue Gene/Lのテスト機で性能を測定し、研究の事前準備をするために、米国IBMのワトソン研究所を訪れました。

「ワトソン研究所では、研究者の方々にはずいぶんインスパイアされましたが、研究所の施設についても印象的なことが二つありました。それは食堂が立派なことと、トイレ近くのスペースにホワイトボードが用意されていたことです。人が集まるスペースを快適に



バイオ・IT融合研究棟のリフレッシュコーナー

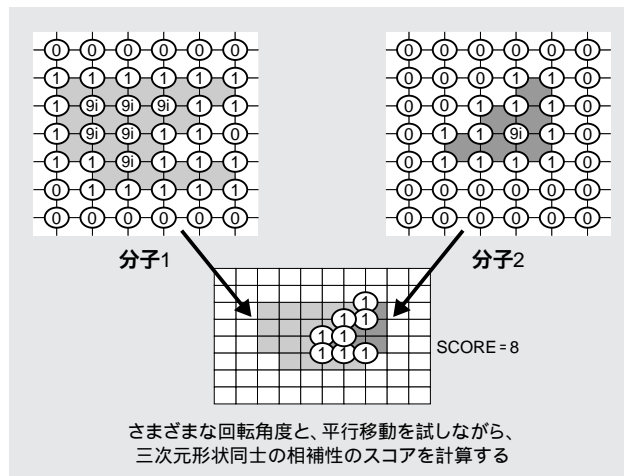


図6. 形状相補性による粗いドッキングの過程

したり、いつでもディスカッションできるように工夫されていることに感心しました。そこで、バイオ・IT融合研究棟にも各階にリフレッシュコーナーと呼ばれるスペースを設け、ホワイトボードを置きました。快適な場所、人が集まりやすい場所をつくって、そこで気楽に議論できるようにしようということです。

冒頭で説明したように、臨海副都心センターは開かれた研究拠点を目指していますし、外部から参加している研究者もたくさんいます。しかもコンピューターサイエンスや数学、さらには薬学、物理学といった各フィールドの専門家が集まっているのですから、そういった異分野とのコラボレーションや交流の中から新しい成果や、日本発の知見や技術を発信していけるのではないのでしょうか（関嶋氏）。

こうしたシナジー効果という点で、CBRCはBIRCと組んで大きなプロジェクトに取り組もうとしています。それは、BIRCが行う巨大なタンパク質同士のドッキングの実験であり、それをCBRCがシミュレーションにより支援しようというものです。

CBRCはこのプロジェクトで、二つのタンパク質をX軸、Y軸、Z軸方向に6度刻みで組み合わせたドッキング計算をスピーディーに実行するために（図6参照）、東京大学と共同でフーリエ変換を用いたアルゴリズムを開発している最中です。このアルゴリズムを使うことで世界最速のドッキング計算がBlue Protein上で実行されることとなります。この「お台場ドックプロジェクト」により、日本のバイオインフォマティクスは世界中の研究者から注目を浴びることとなるでしょう。