

## 4. 各計算科学分野の社会的・科学的課題

---

### 4.1 生命科学

#### (1) 現在行われている課題

生命科学分野でのシミュレーションの学術的意義が、今後ますます重要になっていくことは疑う余地がない。分子レベルにおいては、実験によって原子座標を解き明かすことの重要性は変わらないが、細胞内でタンパク質や酵素などの分子機械が働く姿を知ることが本来の目的であり、そのためにはどうしても計算科学により実験科学で得られない情報を補完する必要があるためである。

「京」を用いて行ってほしい研究を一般の方々にアンケートすると、「創薬や医療に役立つ計算」という意見は非常に多い。実際、ゲノム配列の次世代シーケンサーの発達によって、個人ゲノムの同定が日常的になされるようになると、病気などの予防のために、人によるゲノムの違いが与える影響を分子レベルで解明する必要があるが出てくる。従来は、実験による特徴づけが中心であったが、今後は、タンパク質の立体構造に基づいたタンパク質の動きや揺らぎ解析が重要になってくるであろう。そのためには、精度のよい分子モデリングや結合過程のシミュレーションなどの分子動力学シミュレーションが社会的要請として重要になると思われる。また、医薬品開発では、熾烈な国家規模の競争が行われている。計算機を用いて医薬品開発を加速することは、国内医薬品産業の国際競争力を高めることにつながる。複数細胞ネットワークの解明、疾患の解明、詳細な血中薬物動態、血管形状と血液内の物質移動など、実際に医療現場に役立つものまで、数値シミュレーションに対する期待は大きい。血栓症や糖尿病など、血液内のさまざまなイベントを通じて引き起こされる疾患は数多く、その予測に対する社会的要請は大きい。疫学的に知識が構築されてきた研究分野において、数値シミュレーションがもたらす学術的な影響は大きく、例えば、患者個別の血管幾何形状を利用した直接シミュレーションと、疫学的データを照らし合わせることで、これまでの事実を裏づけ、更には、新たな因果法則を与えることも可能となる。また、患者個別の情報を用いた血流中での薬効評価が可能となれば、有効な医療を創出することにつながる。超音波治療シミュレーションについて、低侵襲治療によって患者の Quality of Life (QOL : 生活の質) が向上すれば、患者の社会復帰が容易になり社会が活性化する効果も期待できる。

生命科学研究分野におけるシミュレーション方法の特殊性は、時間スケールおよび空間解像度の異なるさまざまな現象を解析する必要があることである。そのため、研究の必要性や研究者の興味・関心となる現象に対してさまざまなモデルや方法論が利用される。方法論的には、おおまかに、生体分子を対象とする生体分子シミュレーション、細胞や臓器を対象とする細胞・臓器シミュレーション、脳・神経を対象とする脳神経系シミュレーション、ゲノム情報などの大量に蓄積されたデータを対象とするゲノム解析・バイオインフォマティクスに分けられるので、その分類に沿って紹介する。

### (i) 生体分子シミュレーション

生体分子シミュレーションには、全原子モデルを用いた分子動力学（全原子 MD）、全原子 MD を創薬に応用した創薬 MD、酵素反応などを解析する量子化学／分子力学混合モデル（QM/MM）、粗視化モデルを用いた分子動力学（粗視化 MD）がある。

全原子モデルを用いた分子動力学である全原子 MD では、小タンパク質の折れたたみ過程から、膜タンパク質、DNA タンパク質相互作用、リボゾームやウイルス全体などが研究対象としてよく用いられる[1]。原子数としては数十万から数百万原子、計算時間としては、数百ナノ秒から数十マイクロ秒程度の研究がよくなされている。海外では、最近アメリカの D.E. Shaw Research が開発した分子動力学専用スパコン Anton[2]が特筆すべき性能を示している。現時点では 10 万原子程度までしか扱えないが、1 ミリ秒を超える計算を実現し、80 残基程度までのタンパク質の折れたたみや薬剤の自発的結合を再現できている。しかし、専用スパコンは計算速度に関しては秀でているが、計算法の柔軟性には欠けており、ある一定の計算しかできないのが問題である。一方汎用スパコンは、さまざまな計算法の工夫、例えば、上記のマルチスケール・マルチレゾリューション法の導入や、複数のレプリカを用いてレプリカ間を疎結合で結ぶ手法で分子の統計量（自由エネルギーなど）を計算する手法には向いている。

従来、創薬プロセスにおいては、発見的側面が大きく試行錯誤にかかる時間的経済的コストが高いものであった。近年、病気の原因となるタンパク質の構造が原子レベルで解かれるようになり、分子間相互作用を見ながら分子を設計する SBDD（structure based drug design）という開発戦略が取られるようになってきた。（図 4.1-1）しかしながら、数多くの化合物を調査する必要性から精密な計算モデルを採用できず、計算科学で創薬プロセスをリードしていくほどの精度は得られていない。典型的なドッキングプログラムでは、タンパク質の動きは固定、水分子の影響は無視されており、こうした近似が結合親和性の予測を困難にしていたと考えられる。これに対して、ペタスケールの京速計算が可能になった現在、タンパク質や水分子の自然な運動を考慮した全原子 MD を活用して化合物探索が可能になりつつある。

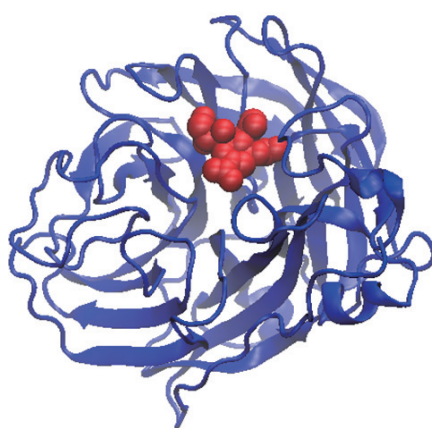


図 4.1-1 医薬品（リレンザ、赤）とインフルエンザウィルスのタンパク質（青）が結合している様子

実際、創薬にターゲットを絞ったシミュレーションである創薬 MD においても全原子 MD を用いて数十種から数百種の化合物を調べることができるようになってきた。HPCI 戦略分野においても、定量的に化合物と標的タンパク質の結合強度を評価し、薬としての活性を予測するという試みが始まっている。D.E. Shaw Research も製薬会社と協力して医薬品開発を始めており、今後は計算機を用いた創薬支援は実践的なものに発展して行くと期待されている。こうした流れを受けて、世界的にもスパコンによる合理的かつ効率的な医薬品開発を目指す風潮ができてきた。

分子レベルのシミュレーション研究における QM/MM 法は、酵素活性中心などの分子の電子状態を求める必要がある部分を量子化学 (QM) 計算で行い、その活性中心に与える触媒活性に重要なまわりのタンパク質環境を全原子 MD シミュレーションで用いている分子力学 (MM) に基づく分子力場で記述するハイブリッド法である。非常に効率的に反応性解析が行えるため、現在では、酵素反応解析においては標準的な手法である。この方法では、近年の実験によって重要性が認識されてきているタンパク質の遅い揺らぎの効果も取り入れることができる。

QM/MM 法では、依然 QM 計算が圧倒的に遅く (1 万倍の差)、かつ QM 計算は MM 計算に比べて非常に複雑であり、効率の高い並列化が非常に困難なため、並列化度を上げると、より計算時間のミスマッチが大きくなる。この問題を解決する方針として二つの方向性の研究が進んでいる。一つ目は、近似的に DFTB (Density Functional Tight Binding) 法や EVB (Empirical Valence Bond) 法、Car-Parrinello 法などの QM 手法を用いて直接的な QM/MM-MD を行う方法

(QM/MM-MD 法) である。もう一つは、QM 構造の統計サンプリングを諦める代わりに高精度な QM 計算に基づく自由エネルギー面上での構造最適化を行う方法 (QM/MM 自由エネルギー構造最適化法) である。QM/MM-MD 法が、現時点では数十 psec から nsec までに限定されている一方で、QM/MM 自由エネルギー構造最適化法では、QM/MM 計算と MM 領域に対する MD サンプリングを分離し、MM 領域の分布で定義される自由エネルギー曲面上で QM 分子の最安定構造を決定する手法であるため、現時点でも酵素反応と関連するサブマイクロ秒のタンパク質の構造変化を見いだすことができる。

一方、巨大な生体超分子に対しては、全原子 MD の適用範囲が拡大したとはいえ、時空間スケールに限界があり、それを突破する試みとして、粗視化モデルを用いた分子動力学である粗視化 MD が盛んに行われている[3]。タンパク質折れたたみ研究で開発された郷ポテンシャル

(天然構造の相互作用のみを導入する) が、分子モーターや膜タンパク質などの構造解析にも適用可能なことが示され、世界的にもユニークな研究が数多く発表されている。一方で、過去 5~10 年の間に、欧州を中心とした生体膜関連の粗視化モデル構築と、米国を中心とした粗視化モデル導出の方法論の開発が進み、広く応用されている。これらのモデルは、上記の郷ポテンシャルの流れとは異なり、生体膜や生体分子間の相互作用の記述に力点を置いている。欧州の中心は、Marrink らが開発する MARTINI 力場であり、最初の生体膜モデルから拡張されてタンパク質系も取り込んで発展している。方法論では米国の Voth らが精力的に理論展開をしており、それに基づいて巨大系の計算が行われている。核酸への適用は遅れていたが、ここ数年で徐々に整備され始めている。

## (ii) 細胞・臓器シミュレーション

細胞シミュレーションについては、E-Cell [4]、SBML[5]、セルイラストレーター[6]、米国産の Neuron[7]など、連続体的なアプローチに基づいた細胞内の反応シミュレータの開発が国内外で活発に進められてきた。細胞レベルの中核的な生化学機能は代謝、信号伝達、そして遺伝子発現の三つである。下記では、連続体細胞シミュレーションによる代謝系の計算および1分子粒度シミュレーションによる信号伝達経路のシミュレーションを中心に述べる。これら HPC への展開が比較的に見通されている2方面に加え、今後の展望を考えるうえで取り上げるべき最近の顕著な動向として、スタンフォード大学が最近発表したマイコプラズマ菌の全ゲノムシミュレーションがある[8]。この全ゲノムシミュレーションでは、反応ネットワークモデルを採用しており、技術的にはわが国の E-Cell プロジェクトが 1999 年に発表した E-Cell 1 全細胞モデル[9]から大幅な飛躍は見られないが、ゲノムアノテーション（ゲノム中の遺伝子の機能などの注釈）やマルチオミクスデータ（遺伝子発現量などの網羅的データ）の蓄積が進んだことで E-Cell 1 全細胞モデルのように一部の遺伝子セットを選択するのではなく、特定のゲノムの全遺伝子を組み込んだモデルの構築に成功している点が特徴である。現状ではモデルにアドホックな部分が散見され、多数の経験的パラメータをどう決定していくかといった未解決の問題も多いが、今後継続的に改良を続ければ、多くの技術的な蓄積が進むと考えられる。統合的な細胞シミュレーションの実現は実用面でのインパクトが大きく、システム生物学の発祥の地の一つであるわが国としても長期的な取り組みを一層深く考えるべき時期に来ていると思われる。

連続体細胞シミュレーションでは、物質量をモル濃度で細胞内反応を Michaelis-Menten 式や質量保存式などで記述し、細胞内の物質の濃度変化が計算される。これまでに、連続体細胞シミュレーションはシステム生物学の分野で細胞内の反応や物質相互作用のネットワークの検討に利用されてきた。しかしながら、創薬や疾患を検討するうえで重要なのは、個体レベルでの挙動である。個体を理解するためには血液による物質移動や物質循環、相互作用を無視することができない。それを捉えるには、反応ネットワーク的な視点を超えて、細胞・臓器の持つ空間的な広がりや、血流との連成を考慮した解析が必要になる。米国の The Virtual Cell プロジェクト[10] や、EU での Virtual Physiome Project [11]では、細胞モデルの多次元化が進められ、特に心筋細胞、血管内皮細胞の刺激に対する反応について現在も検証が行われている。国内では、RICS プロジェクト[12] において、超大規模並列計算機の利用を見据え、細胞内構造の不均一性をボクセルデータで表現し、複数細胞内の数百の代謝物に対する数百の酵素反応を捉えようとする新たな取り組みがなされている。

一方、レーザー顕微鏡などを用いた高解像度でのライブイメージング技術が急速に発展したことにより、更に解像度を高めた1分子粒度での細胞シミュレーションへの取り組みが要請されている。1分子細胞シミュレータの代表的なものに ChemCell（米サンディア国立研）[13]や Mcell（米 Salk Institute、コーネル大学、ピッツバーグスーパーコンピューティングセンター）[14]、Smoldyn（米国 Fred Hutchinson Cancer Research Center および The Molecular Sciences Institute）[15]、GFRD（理研および蘭 AMOLF）[16]、Spatocyte（理研および慶応大学）[17]などがあるが、計算対象は大腸菌などの単純な原核細胞の信号伝達経路が多い。HPC への展開はこれまで比較的ゆるやかであったが、1分子粒度の真核生物への展開には高並列化は不可避である。HPCI 戦略プロジェクトにより、Spatocyte の京への実装が進められているほか、理研が GFRD



の並列版の開発を開始、また Smoldyn の GPU への実装が進んでおり今後数年での急速な進展が期待される情勢である。

臓器レベルのシミュレーション研究においては、血流解析の例を取り上げる。血流解析シミュレーション技術の開発には、高齢化社会において深刻な問題となっている循環器系疾患の治療法の開発、腫瘍治療などの高度化を進めるうえで重要な意味がある (例えば[18]参照)。先に挙げた細胞シミュレーションとの連成を進めることにより、創薬を支援することに直結する。そのために必要不可欠なのは、方法論の確立である。血流中における単一の赤血球あるいは血小板など生体膜の力学挙動に関するシミュレーションはここ十数年、数多く研究されてきた。特に米国では、応用数学分野との連携により、ストークス力学、境界埋込法による解析に適した膜表面要素やバネ要素を用いる膜のモデル化が進められてきた [19]。計算機性能の向上にともない、国内外で、複数の赤血球と血流を連成した 3 次元シミュレーションが報告されるようになってきた。このような適用可能範囲の急速な広がりにより、現在、血流シミュレーションに対する計算力学 (とりわけ超大規模計算) の分野における研究者の注目が集まっている。赤血球を含む流体シミュレーションが 2010 年、2011 年の Gordon Bell 賞を受賞したことは記憶に新しい[20]。国内では、「京」での超大規模並列計算を意識した (あるいは親和性のよい) 血流解析手法が開発され、血管径 0.2 mm 程度の血流解析が進められている[21]。こうした動向を受け、国内では血流中における血栓形成シミュレータの開発が、医療応用上の重要課題として積極的に進められている[18]。血栓形成の初期過程である血管壁への血小板接着のモデル化は、実験と平行して十数年前から行われており、現在単一の血小板接着シミュレーションに関して方法論が確立されつつある状況にある。「京」の計算資源は、数百  $\mu\text{m}$  径の血管内に含まれる血球の「群」としての振る舞いを捉えるのに適している。「京」を利用する研究課題には、血小板血栓の生成と、それに起因する生化学反応の数値シミュレーションが挙げられる。また、戦略分野では、物性値の変化、凝固・溶解過程のモデル化が進められている。これらの至近の課題に続き、疾患の再現と薬効の評価、治療法の検討などが長期的な研究対象として発展していくものと考えられる。

医療へのシミュレーションの応用では、低侵襲治療の一つである超音波治療の実現に向けて、現在、機器開発の段階で厳しい競争が繰り広げられている。その数値的研究は現在進行中の「ものづくり」的課題や、市場の動向と密接な関係がある。近年、世界的な流れとして、患者の Quality of Life (QOL) 向上への要求とともに、低侵襲治療への関心が高まっている。なかでもエネルギー治療の一つである超音波治療が注目を集めている[22]。超音波治療の機器開発は、欧米と中国で盛んに進められており、日本はその後塵を拝している。欧米では、前立腺肥大や子宮筋腫などに対する超音波治療装置がすでに実用化され、認可もされている。更に最近では、MRI を併用した MR ガイドの超音波治療装置の開発が進められ、より深部の腫瘍に対する超音波治療の臨床試験も進められている[23]。それに対し日本では、超音波診断装置や MRI などの診断機器の技術開発のポテンシャルが高く、国内需要も高いため、積極的に新手法や新装置が研究開発されてきた。しかしながら肝心の治療器については、景気後退、健保財政の悪化、医療機器審査の高いハードルなどの複雑な国内事情により、製品化が困難であり消極的であった。このような消極性が、1980 年代以降、医療用具全体の貿易収支における赤字の膨大、更には、国民医療負担の増大をもたらした原因の一つとして指摘されており、わが国は、戦略的な治療

器開発の必要に迫られている。現在、低侵襲治療を実現する HIFU (High Intensity Focused Ultrasound) 装置の開発がオールジャパン体制で進んでいる[24]。特に装置の制御や生体へのリスク評価において、数値シミュレーションの果たす役割が大きく位置づけられている。海外ベンダーの状況に関して、精力的に開発が進められていることは伝わってくるものの、正確な計算機利用情報は、企業秘密の部分が多く不明である。だが、制御法に関しては 1990 年代に欧米で提案され、Finite-Difference Time-Domain 法との親和性の高い Time Shift 法[25]や Time Reversal 法[26]などが方法論として確立しており、これらの開発競争においては計算機性能が鍵を握る。わが国では、これまで、HIFU シミュレータの開発が進み、脳内腫瘍を対象とした実寸スケールの超大規模計算が実現可能となっており、世界的に高い競争力がある[27]。「京」の計算資源は、ある条件の計算を一つ行うことは能力的に余裕があるので、パラメータの探索を効率よく進めることが可能であり、開発競争における戦略上、非常に魅力的である。そのため、「京」のハードウェア構成に合わせたソフトウェア開発が積極的に進められてきた。

### (iii) 脳神経系シミュレーション

コンピュータにおける情報処理とはそもそも人間の脳で行われる情報処理を手本として作られているが、ここでは、その中で特に、神経細胞レベルからその大規模なネットワークを構築してその情報処理を再現するものを脳神経系シミュレーションと呼ぶ。神経回路の情報処理のモデリングとしては、80 年代頃からアーティフィシャルニューラルネットワークモデル、すなわち同種のニューロンの膜電位の一種の平均を「活動度」として扱い、関数を介して学習能力を持つシナプスに出力する McCulloch & Pitts 型モデル[28]が流行した[29]。この活動度というアナログ量がシナプスを通じて伝搬するアーティフィシャルニューラルネットワークモデルは情報処理の方法として一定の成功を見る。しかし、このようなモデルではシナプス後電位がスパイク列に変換されて運ばれる実際の脳の生理的データと詳細に対応させることが困難であり、脳という超大規模・並列なネットワーク構造で起こっている行動と関連した記憶・学習やさまざまな可塑性、感覚統合や、環境に適応する行動指令信号の生成を知るためには、より精密な生物物理的詳細モデルが必要とされることがわかってきた。このような要請に従い精密化、すなわち、神経形態をマルチコンパートメントなほしご状回路上で表現し、コンダクタンスの変化を決定論的にシミュレーションする Hodgkin-Huxley formalism[30]を用いたものが行われるようになり、細胞モデルをスクリプトベースで記述してシミュレーションさせる C/C++ベースのプラットフォームシミュレータである NEURON や GENESIS、NEST 等が使われるようになった。2000 年代になって、ある程度神経回路のシミュレーションが現れ始めた。その多くのは Hodgkin-Huxley formalism を採用しているが、計算量の問題で、多くのはシングルコンパートメントモデルを採用しており、また、活動電位の記述を簡略化することでモデルを簡単にする Izhikevich モデルや、積分発火モデルのような簡約モデル[31]も使われていた。2005 年に、人の脳のシミュレーションを Hodgkin-Huxley formalism のマルチコンパートメントモデルを用いて行うことを最終的な目標にする BLUE BRAIN プロジェクトが始まったあたりから、大規模並列計算をとまなう脳のシミュレーションプロジェクトが増え始めた。その例として、100 万個ニューロンに簡約化したヒト脳シミュレーション[32]や、ネコの脳と同数のニューロンで行った Synapse プロジェクト[33]などが挙げられる。しかしながら、これら

は大域的な信号の広がりはある程度シミュレーションできても、脳の **function** の再現とはほど遠いものであり、**functional** な全脳のシミュレーションは現状においては全く達成されていない。日本においては、2008 年の理研の次世代計算科学研究開発プログラムにおいて、脳神経系研究開発チームが発足し、日本初の大型の脳神経系シミュレーションの開発プロジェクトが始まった。ここでの昆虫の嗅覚＝運動系を対象としたマルチコンパートメントのシミュレーション（図 4.1-2）は、「京」上での 15% 程度の実効効率と細胞分割テクニックにより、神経回路シミュレーションのほぼリアルタイム実行が可能となっている。また、ほ乳類の視覚情報処理系を対象に、シングルコンパートメント積分発火モデルを用いた  $10^9$  ニューロンのシミュレーションも達成されている。（図 4.1-3）このように神経シミュレーションにおいて、そのモデリングは Hodgkin-Huxley formalism を最も精密なモデルとしつつも、さまざまな簡約モデルが試行されているのが現状である。大規模計算もその歴史は浅く、どちらかといえば HPC を用いた大規模計算よりも PC ベースのものが主流である。プログラミング言語（プラットフォームシミュレータを用いるものを含め）は C/C++ ベースなものが多数であるが、MATLAB ベースや python ベースのものも多い。神経系はもともと高度に階層化・並列化した情報処理システムで、細胞の詳細な生物物理的な性質から新回路の機能までをカバーする大規模なシミュレーションが広く望まれており、HPC による脳神経系のシミュレーションがこれから大きく発展する素地はそろっている。最新の国際的動向としては EU の「未来および発展期にある技術（FET）プログラム」として、人間の脳の働きを解明するヒューマンブレインプロジェクトが 2013～2023 年の期間で動きだしている。

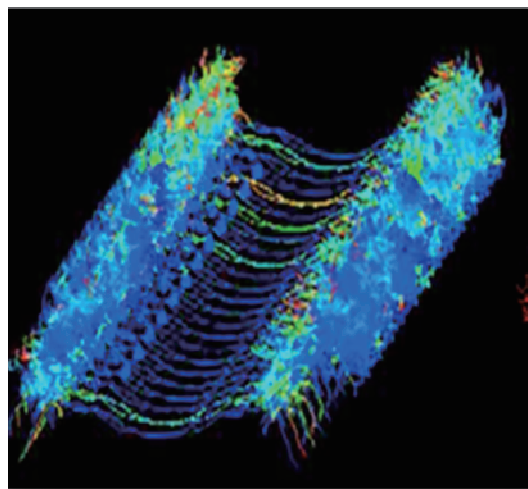


図 4.1-2 昆虫前運動中枢シミュレーション（NEURON）

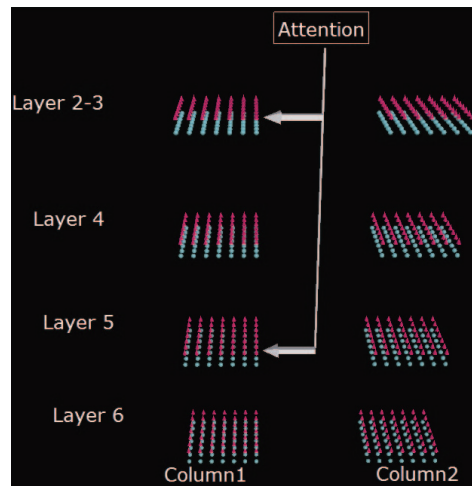


図 4.1-3 ほ乳類皮質シミュレーション (NEST)

#### (iv) ゲノム解析・バイオインフォマティクス

ゲノム解析・バイオインフォマティクス分野では、近年観測技術が大幅に進展し、DNA マイクロアレイによる遺伝子発現データや、次世代 DNA シークエンサーによる DNA 配列データに代表されるような、これまで得ることのできなかった短時間で大量に出力されるデータ、いわゆるハイスループットデータが急速に蓄積されるようになってきている。米国 NCBI が運営するデータベースには 2012 年 12 月の時点で 84 万サンプルを超えるマイクロアレイによる遺伝子発現データが蓄積されており、これらの膨大なデータからの大規模な計算リソースを活用した生命科学、とりわけ医学や医療の発展が期待されている。特に、ヒトにはおよそ 3 万の遺伝子があるとされ、そこから産生されるタンパクは 10 万種とも言われている。バイオインフォマティクスやシステム生物学と呼ばれる分野では、このような膨大な因子間の関係を大規模な観測データから統計科学や情報科学に基づき推定・予測する研究が盛んに行われている。遺伝子発現パターンの比較解析や、発現データに基づくクラスタリング、共発現解析や仮説検定などの複雑な計算を必要とする遺伝子ネットワーク解析などがその代表的な例である。「京」を用いるグランドチャレンジアプリケーション研究開発プロジェクトや HPCI 戦略プログラムにおいても、遺伝子発現データからの遺伝子ネットワーク推定・解析ソフトウェアの研究開発が進められており、ペタフロップス級の大規模計算による観測データ解析が当たり前になりつつある。膨大な計算リソースを前提とした新しいアルゴリズムも登場しており、「京」時代ではこれまで難しかった全遺伝子を含む遺伝子ネットワークの推定も現実となった。

また一方で、DNA 配列に代表されるゲノム配列解析では、ショットガン法による DNA 配列決定法の登場により大規模な計算機の利用が一般的になった。最近になって登場した次世代 DNA シークエンサーではこれまで数年かかった全ゲノム解読を数週間のレベルにまで短縮することが可能になり、個人一人一人が自分自身のゲノム配列を知りそれに基づく医療を受けることができる個人ゲノム時代が目前に迫っている。次世代 DNA シークエンサーでは、1 回の計測で 1TB ほどのデータが発生し、そこから意味のあるデータに処理するために膨大な計算が必要となる。次世代 DNA シークエンサーは、世界各地に 1700 台以上導入されており、例えば



中国では、2010年にすでに100台以上の次世代シーケンサーと1 PFLOPS程度の計算リソースを投入しゲノム解読に当たっている。日本においても、30台程度のシーケンサーが全国で稼働しており、国際がんゲノムコンソーシアムのプロジェクトでは肝がん500サンプルの解読がヒトゲノム解析センターの計算リソース（225TFLOPS）を用いて進められている。HPCI戦略プログラムにおいても次世代DNAシーケンサーを用いたメタゲノムおよびがんゲノム解析の研究を行う予定となっている。今後、解析サンプル数は急激に延びることが予想されるため、これらのデータ解析に必要な計算機の導入が、今後の研究の進展に必要不可欠である。また個人の配列データが蓄積されるようになれば、全ゲノム関連解析（GWAS）もより大規模化・複雑化し、またゲノム配列データと発現データを組み合わせた解析もより複雑度を増していくことが容易に予想され、必要な計算リソースもそれに応じて膨大になると予想される。このように取り扱うデータが膨大であるため、生命科学におけるデータ解析での計算における一番のボトルネックはディスクI/O性能となる。また一般に、多種多様なソフトウェアが組み合わせて利用され、また観測技術の進展・変化が非常に速いため、常に新しい計算・解析手法が提案されており一つのソフトウェアの寿命（利用期間）が短いのも特徴である。統計科学やシステム生物学における数値計算から、生物ネットワーク解析に必要なグラフ探索、DNA配列解析における離散的な計算、あるいはそれぞれを複雑に組み合わせた計算など多種多様であり、前述の次世代DNAシーケンサー解析ソフトウェアもそのうちの一つに過ぎない。観測データから計算・解析手法を開発し、実際に計算し、計算結果を生物学的な実験にフィードバックし、最終的に生物学的な発見を論文としてまとめ上げるまでの時間の短縮が当該分野で求められていることである。逆に、すでに確立した一つの方程式をいかに高速に計算するか、という点は当分野ではあまり求められてはいない。また並列計算という点からは、一般にバイオインフォマティクスやシステム生物学ではデータ並列性があるため、計算の独立性が高く並列性能を出しやすい。したがって、複雑な並列アルゴリズムの構築を経なくても、数万コアから数十万コアに達するような超並列計算機の恩恵を即座に受けやすいと言える。

## (2) 長期的目標

すでに述べてきたように、生命現象の複雑さや多様さなどにより、これまでの生命科学の分野では異なる計算手法やモデルが用いられてきた。また、計算機の演算性能の限界により十分な時間およびサイズの計算が困難だった。今後、長期的研究ではその制約の多くが取り払われ、研究したい生命現象をまるごとシミュレーションできる日が来ると期待される。そして、理論・シミュレーションの専門家だけでなく、実験・医療関係者にもシミュレーションの有効性が明らかになっていくと思われる。そこで必要となる理論や手法の開発を行い、有効な実験とシミュレーションの共同研究を行っていく必要がある。更に、今後の生命系のシミュレーションでは、予測可能性という点が重視されると思われる。社会応用という観点からも、創薬や医療機器開発においてシミュレーションが盛んに利用されるようになると予想される。下記では、各スケール・方法における今後の発展について記述する。

### (i) 生体分子シミュレーション

全原子 MD では、計算可能な時間スケールおよび空間スケールが更に増大していくが、研究対象はあくまで個々の研究の必要性によって多様性を持つと思われる。例えば、溶液中でのタンパク質の折れたたみなどを解析するためには、10 万原子程度の系で十分であるが、マイクロ秒から数ミリ秒に至る長時間の計算かレプリカを用いたアンサンブル計算が必要である。分子モーターや膜輸送体などの構造変化を通して機能を発現する過程を見るためには、やはり数マイクロ秒から数ミリ秒のシミュレーションが必要であるが、脂質膜も計算することを考慮すると、数十万から百万原子程度のサイズになる。更に大きなリボゾームやウイルス、ヌクレオソーム複合体などの生体超分子複合体、細胞環境を考慮したシミュレーションなどでは、数百万原子から 1 億原子に至る巨大な系のシミュレーションとなり、計算可能な時間はやはり、その時代の計算機の限界に依存する。全原子 MD をオーソドックスな力場を用いて計算する場合には、すでに手法が確立しているため、その手法自身をハードウェアで組んだ MD 計算専用計算機が今後とも計算速度はリードしていくと思われる。しかし QM 計算とのハイブリッドなど、今後とも開発される新たな計算法では、汎用計算機の利用が重要な役割を果たす。また、高並列な計算機で全原子 MD を行う場合には、多数のレプリカを疎結合して系の統計的性質を解析するアンサンブル計算の重要性が増していくと思われる。

創薬 MD においても、計算機パワーの向上は、より大規模な化合物探索（「京」を利用した場合、数百種から数千種）を可能にすると予想される。高精度な全原子モデルでの大規模な探索が可能になることで、予測の精度は格段に向上し、計算による創薬の有効性はますます高まるだろう。また、膜タンパク質のような複雑な系に対しても計算機で薬を設計するという挑戦が始まると想定される。現在は、標的タンパク質と化合物の結合強度に焦点が当たっているが、副作用を抑えるという観点から、標的タンパク質以外のタンパク質と化合物との相互作用を調べることは重要である。計算機の進歩により、こうした医薬品候補の特異性に関わる予測もシミュレーションに基づき行われると予想される。

今後、SACLA（3.3.1 項を参照）をはじめとする実験技術も大きく進展し、指数関数的に創薬応用シミュレーションの対象となるタンパク質構造が明らかになるであろう。こうした標的タンパク質の情報をすぐに計算に持ち込み、医薬品候補を探し出すというスピード感も重要な側面となる。そのためには、実験との連携を十分に高めていくことは当然ながら、数多くの標的タンパク質に同時並行で対応できる計算機環境が重要である。

QM/MM 法は、酵素反応活性解析のみならず、新規酵素活性設計にも適用されていくと考えられる。特に、緑色蛍光タンパク質（GFP）やチャネルロドプシンのような、機能性タンパク質ツールの改変設計に大きな力を発揮していくと考えられる。また、生物学と化学のシームレスな記述を生かし、バイオミメティックな化学的触媒活性化合物の設計にも応用されていくと期待される。更に、創薬研究においては、これまで問題であった高精度な相互作用の記述とタンパク質構造変化の考慮の両立が可能であるため、QM/MM 法が重要な役割を果たしていくと考えられる。方法論的には、QM/MM 自由エネルギー構造最適化法による酵素反応活性における、マイクロ～ミリ秒のタンパク質構造変化の解析が今後主流になっていく。これまで、実験的にも直接的に追うことができていない酵素に特徴的な可塑的触媒場の新たな概念が明らかになる可能性がある。これは、上記の酵素活性設計や薬剤設計に、概念的にも、実際の設計ス

キームを構築するうえでも、決定的に重要となるであろう。一方、直接的な QM/MM-MD 計算では、計算時間のミスマッチが増えてしまい、並列化効率の高い MD の有効性を生かせない。その際、レプリカを用いた高い並列度の全原子 MD との組み合わせが発展し、タンパク質構造変化をとまなう酵素反応経路解析などが精力的に行われると予想される。QM 計算自身は、その複雑なアルゴリズムにより、問題固定での並列化効率向上の問題解決はかなり困難であるだろう。計算精度とパラメータ化の困難さに目をつぶれば、DFTB 法が可能性として挙げられる。一方、MD 計算に必要な力の計算をエネルギー差分で行うような単純な疎結合アルゴリズムにより、比較的小規模な QM 分子系に対する非常に高精度な反応性解析を、超並列のメリットを生かして行う方向に発展していくことも想定される。並列数の増加に従い、問題を大規模化・高精度化するような QM 計算の場合には、QM 計算においてよくスケールする手法がいくつか開発されており、QM/MM 法の適用が進むと考えられる。サブナノ～ナノ秒程度のシミュレーションであっても重要な系は多くあり、特に、大規模分子全体にわたる電子波動関数の振る舞いが重要となる、電子移動系の解析には大きな力を発揮すると思われる。以上の分野は、物質科学的側面もあるので、物質科学分野、ものづくり分野との連携も重要である。生命分子に関わる分野間連携については、3.1.3 項ですでに詳細を述べた。

粗視化 MD は、一つには全原子モデルからの高精度の導出法が課題であり、それによりタンパク質・核酸・膜を統一的に扱える粗視化モデルの整備が進むであろう。方法論の進展に応じて、より高次の細胞生物学的な課題に適用されていくであろう。10nm から 1 $\mu$ m 規模で、ミリ秒から秒規模の現象が対象となる。具体的には、シグナル伝達系などのタンパク質の相互作用ネットワーク、転写や翻訳過程の構造機能計算、核内の遺伝子動態解析、膜・細胞骨格系、ウイルス動態などが重要な課題となる。

## (ii) 細胞・臓器シミュレーション

一分子粒度細胞シミュレーションにおいては、まず原核細胞よりも大きな真核細胞 (10 $\mu$ m 以上) 内の複雑な信号伝達経路モデルについて巨大分子が密に存在し混雑している、より現実的な細胞環境のもとで、転写因子が応答する時間スケールである 1000 秒程度の計算を実現することが大きな課題となる。シミュレーション技術面では、分子の構造多型や重合脱重合化など、分子レベルの状態変化の表現が可能な計算手法の開発が期待される。更に、1 細胞または少数の細胞のモデルから、多数の細胞から構成される集団および組織の計算が可能になると期待される。加えて、細胞の形、大きさの決定、また分化や発生過程のモデル構築のためには、化学反応と力学構造とのカップリングを取り込んだ計算が必要になり、この方面での研究開発が期待される。中長期的には遺伝子発現系や代謝系との連成により 10,000 秒以上の細胞周期を基準とする時間スケールでの予測性獲得が重要課題である。このためには、前述の E-Cell 1 全細胞モデルやスタンフォード大学のマイコプラズマモデルのようなゲノムスケールモデルとの連携、連成、また次に述べる連続体的アプローチによる代謝系モデルとの連携、連成が必要となる。

連続体細胞シミュレーションの適用可能な系が、計算機性能の向上にとまなない、均質な場から不均質な場へと広がってきた経緯を振り返ると、今後は、細胞の構造・形態のダイナミクスの捕捉を目指す方向へと発展するであろう。5 年以内には、一番小さい組織である腓ランゲル

ハンス氏島（5000～1 万細胞）や血小板による血栓形成のシミュレーションが可能になると考えられる。より長期的視野から展望すると、細胞スケールの現象は、分子スケールから臓器スケールまでの異なるスケールの現象を渡す役割を果たすため、今後は、粒子系、流体・構造力学系の研究と連携、統合して研究が進むだろう。細胞の形態変化の仕組みは、実験観測による定性的な知見はあるものの、計測に限界があるため、それをもたらす力学的機構については定量的に不明な点が多い。この状況の打開には、粒子系の数値シミュレーションにより、ボトムアップ的なモデルの構築が期待される。また、細胞内反応を基とした組織・臓器・器官の生化学血流連成シミュレーションにより、疾患解明や創薬につなげる試みも進んでいくと思われる。

臓器レベルのシミュレーションにおける血流解析シミュレーションについて、これまでは流体・構造・膜の力学的連成解析法の開発が大きな目的であったが、現在は方法論が確立しつつある。したがって今後の発展は、計算機性能の向上にともなう計算対象の大規模化に向かうものと考えられる。数年～十数年先においては、数百  $\mu\text{m}$ ～数 mm 径の血管内の血流解析が、計算機で捕捉可能になると考えられる。ただし、生命研究として意義のある結果を得るには、依然問題が山積みである。血栓生成は数秒～数分の時間スケールの現象であり、数値シミュレーションで有意な結果を得るには、生体膜の力学挙動を捉えるための離散時間幅 1 マイクロ秒に対し、 $O(10^6)$ ～ $O(10^8)$ 回もの時間積分を行う必要がある。そのため、数値的安定性の向上および数値誤差蓄積の低減が不可欠である。また、超大規模並列化に耐えうる並列化アルゴリズム、あるいは方法論そのものの見直しも必要となる。上記の問題を解決する方策として、粗視化法の構築も課題となる。そのためには、現象の本質と関わる血流の多重スケール性を踏まえたモデル化が重要となる。更に、血栓シミュレーションにおいては、上述のように、細胞の生化学反応をモデリングしシミュレーションに取り入れる必要があるが、個別要素の積み重ねから得られる知見が重要であり、モデル構築においても詳細な解析が必要不可欠である。

医療への応用における超音波治療シミュレーションの今後は、現在と同様、製品化に至るまでの治療器の開発状況や、市場の動向と密接な関係を持つはずである。長期的な予測は不確定要素が多く困難であるが、これまでの超音波治療を取り巻く状況を踏まえて、今後の発展について具体例を挙げていく。MR ガイドの超音波治療装置は低侵襲治療を実現することが実証されており、すでに、実際の治療に用いられつつある。しかし、装置導入に掛かる高コストが、普及の妨げとなっている。そこで、治療部位を特定する際に、MR ガイドの代わりに低コストの超音波ガイドを利用した、超音波診断治療器の開発が検討されている（例えば[34]）。超音波ガイドの実現には、軟らかい生体組織の温度変化で生じる微妙な剛性の変化をいかに精度よく捉えられるかが鍵となる。そしてその開発においては、剛性の温度依存性を捉える大規模数値シミュレーションによる支援が大きく期待されている。また近年では、薬物の体内動態を制御し、目標とする患部に必要な量の薬物を投与するドラッグデリバリーシステム（DDS : Drug Delivery System）において、超音波照射により治療効果を向上させる試みもなされている[22]。海外の超音波治療器メーカーは、薬剤メーカーと提携してその開発に注力し始めている。DDS で用いられる数ミクロンのマイクロカプセルが、血管中に 3 次元的に分布した生体中では、干渉によって超音波音場が大きく影響を受けることが知られている。マイクロカプセルの動的挙動には強い非線形性があるため、そのような場の予測は、数値シミュレーションによってのみ



可能となる[35]。よって、超音波診断治療装置を用いた DDS の実用化に向けた効率化や安全性の評価において、数値シミュレーションが果たすべき役割は大きい。

### (iii) 脳神経系シミュレーション

神経科学における最大の目標は、脳で用いられている情報処理機構、究極的には人間の高度な知的情報処理を解明することである。それがまだ不可能であるのは、脳を構成する神経細胞の数が莫大で計算能力が不足していることと、脳内は非常に複雑であるため、実験による観測で神経細胞の活動を観測しきれないという原因がある。例えば、人間の脳には  $10^{11}$  とも言われる莫大な数の神経細胞が存在し、それらが  $10^{15}$  のシナプスで複雑に結合していると推定されている。後述するようにこれを詳細にシミュレーションすることは、エクサフロップスの規模においてもやや不十分である。また現在、生理実験において同時に神経活動を観測できる細胞数は電気生理のホールセルブランプにおいてはたかだか 10 個程度であり、イメージング技術を用いても 100 個程度のものにすぎない。しかし近年、顕微鏡技術を（多くの場合は計算機的解析や制御と融合させて）精密に用いてカルシウムイメージングによる神経細胞を同時に高解像度で観察する技術や[36-38]、そこから電気生理で得られる情報に近いものを解析により復元する技術[39]、また遺伝子技術をミックスさせて光により細胞を制御する技術[40]などが進歩しており、2020 年頃には、1000 個オーダーの神経細胞の活動状態の観察と制御が技術的には可能になっていると推測される。静的な形態解析では、脳のスライスの電子顕微鏡写真から行う再構成には多くの努力が傾けられており、エクサスケールにも達する 3 次元のイメージデータを統合することにより、小さなものであればほ乳類の脳のシナプス結合の全パターンは明らかにできる可能性がある[41]。そのような、“回路構成はわかるがどのように動作するかわからない”という状況では、シミュレーションを行い、動作メカニズムを調べることが、より強力なツールとしての役割を担うだろう。リアルタイム以上で神経回路のシミュレーションができれば、次世代のエクサスケール級計算機では昆虫脳の全神経細胞を用いたリアルタイムシミュレーションが可能となり、生理実験（主に光生理学的なことになるだろう）と相補的に、その神経回路の機能の細部の解明は急速に進むと思われる。また、エクサスケール級計算機においては、単純モデルに限れば人間の全脳レベルのシミュレーションも可能な範囲に達する。その数学モデルが、脳のモデルとして妥当であるかどうかについては非常に議論があるところではあるが、人間の全脳のシミュレーションにおけるマイルストーンになると思われる。形態学や電気生理学的な性質を再現した詳細な全脳のモデルについては、2020 年代においてゼブラフィッシュやマウス等のモデル脊椎動物で行うことが一つのチャレンジとなるだろう。プロセスルールの微細化が終焉したのちも、何らかの技術革新によって HPC の FLOPS 向上に指数関数で性能向上するムーアの法則が適用されつづけるならば、2030 年前に、その速度が 2/3 に落ちたとしても 2030 年代前半に十分詳細な人間の脳のシミュレーションがリアルタイム以上で可能な計算速度が得られる。そして 2040 年頃には、「人間の知的情報処理を解明する」というレースは一応のゴールを迎えると予想されるが、その前後には、個人の脳をシミュレーションして医療に用いたり、BMI（Brain Machine Interface）によって人間の脳とコンピュータの計算能力を組み合わせるなど、全脳のシミュレーションの社会への普及も実現されていくのではないだろうか。ただし、人間の脳の機能を再現した詳細なシミュレーションを人間より速く実行でき

るようになることは、人間以上の存在が生まれることを意味し、その実現が可能になる少し前には、倫理的議論が必要になるだろう。

#### (iv) ゲノム解析・バイオインフォマティクス

次世代 DNA シークエンサーについては 3.2.3 項を参照いただくとして、ここではバイオインフォマティクスにおけるもう一つの代表的な研究として、遺伝子発現データの解析について述べる。3.2.3 項で述べたとおり、バイオインフォマティクスにおいて今後は、長期的にどのような観測技術がどの程度のコストで利用可能になるかは非常に流動的であり不明確な部分が多い。したがってここでは非常におおざっぱな議論をする。遺伝子発現データの解析、なかでも遺伝子ネットワーク解析は、遺伝子発現データから複雑な計算によって遺伝子間の関係を予測・推定するもので、必要とする計算資源は他の発現データ解析手法と比べて圧倒的に多い。遺伝子発現データは、かつて計測に用いるマイクロアレイが非常に高価だったこともあり、数サンプルを用いて特定の疾患に特異的な遺伝子を発見するなどの研究が行われていた。しかし近年急激にコストが下がり、現在では数百サンプルを一度に取得・計測し解析することが当たり前になっている。研究で計測された遺伝子発現データは、公共データベース上に登録し公開することが一般的である。図 4.1-4 は、代表的な公共データベースである NCBI Gene Expression Omnibus に登録されている遺伝子発現データのサンプル数と今後の予想である。現在およそ 84 万サンプルが登録されており、2015 年頃には 150 万サンプル、2020 年頃には 280 万サンプルほどになると予想されている。遺伝子ネットワーク解析は、多くの計算資源を必要とすることから、まだごく一部の先駆的な研究でしか用いられていないが、今後大型計算機の利用が一般的になるに従い、一般的に行われるようになると思われる。また、遺伝子ネットワーク解析自体も、これまでは研究者が興味のある疾患データや薬剤応答データなど特定の条件下で観測された比較的小規模なデータから遺伝子ネットワークを推定し解析をするということが行われていた。しかし「京」などの計算リソースの登場により、解析対象をそれほど絞らずに数百程度のデータセット（数万サンプル）を抽出し、それに対して網羅的に推定ソフトウェアを適用するというアプローチが現実的になった。具体的には、現状で 256 データセット・30,517 アレイ（サンプル）での網羅的計算を国内の研究グループで行っているが、これは公共データベースに登録されている全約 3 万データセット・84 万アレイのごく一部であり、すべてを計算対象とすることはできていない。2020 年に想定されるサンプル数を仮定すると現在の 100 倍の規模の計算が必要になる。また、サンプルの量が増えるだけでなく解析手法も多様化することが予想されるため、それらを見越した計算機資源の整備が研究推進に絶対不可欠であると言える。

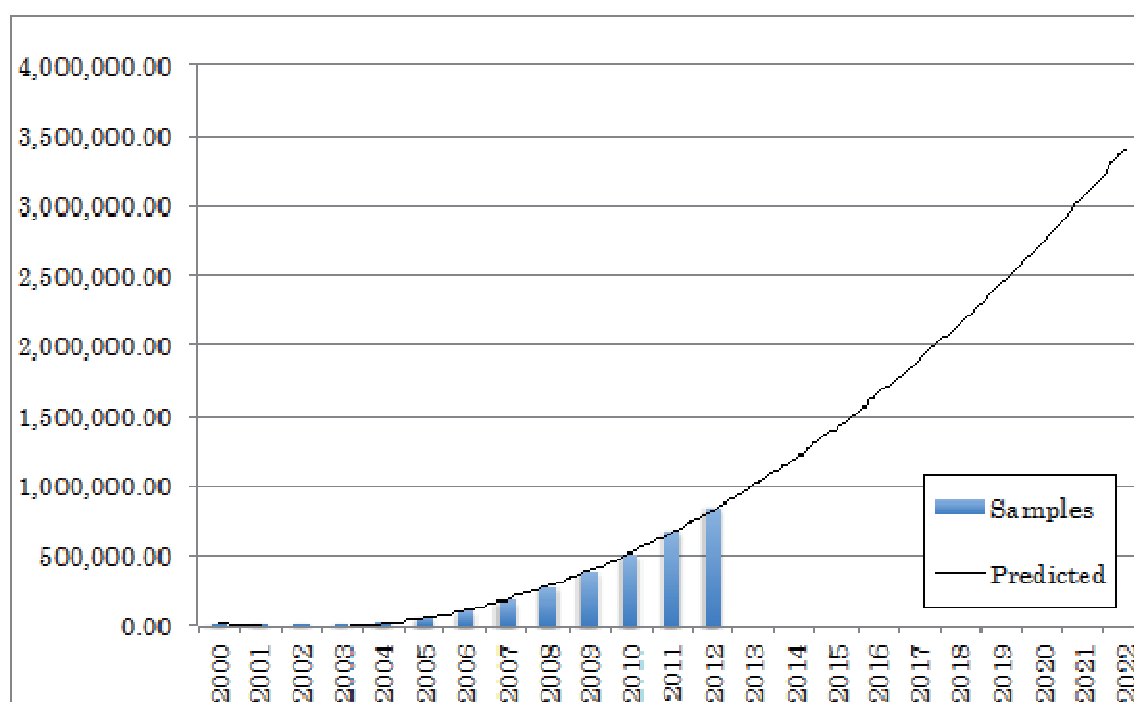


図 4.1-4 NCBI GEO に登録されているアレイ数と今後の予測

### (3) 次世代に解決すべき課題

#### (i) 生体分子シミュレーション

全原子 MD では、構造生物学的に重要なタンパク質が次々と計算対象になるだろう。X線結晶解析などで構造が決定されたタンパク質の機能を解析するために、全原子 MD を実行することがルーチンになっていくと思われる。研究対象に従って、数多くの課題に対応する計算が行われることになるが、数万から数百万原子を含む小規模～中規模系が、多くのユーザが選択する研究対象となると思われる。そのなかでもやや大きめの数十万から百万原子規模は、分子モーターや膜輸送体などの分子機械において生体膜環境などを露わに取り込んだモデルに対応する。一方、数百万原子系のような大規模系では、単一のタンパク質というよりはリボゾームに見られるようなタンパク質・核酸などを含む生体超分子複合体が主要な研究対象となるであろう。更に、細胞環境における生体膜だけでなく、細胞質やオルガネラなどの細胞内小器官も研究対象になるだろうが、そのような巨大系（数百万原子から1億原子系）のシミュレーションを行うためには、粗視化 MD などとの連成が必要となる。

創薬 MD では、可溶性タンパク質に関してより実践的な応用段階に入り、数千種から数万種の化合物から医薬品候補を選び出すため、化合物力場構築の自動化を行う必要がある。更に、類似タンパク質との結合強度を評価し、医薬品候補のタンパク質への結合が特異的なものか否かが判断される。また、膜タンパク質を標的とした医薬品設計も始められる。計算する原子数が10万を超え、計算の高速化が必要となる。また、膜を構成する脂質についても高精度なモデリングを行うことで、膜タンパク質を標的とする計算精度の向上を目指す。今後は実験技術が進歩し、解かれる標的タンパク質の3次元構造は指数関数的に増加していくと予想される。

これに応じて、計算分野でも、即時に数多くの標的タンパク質に対応する体制が求められると予想される。

QM/MM 法では、まずは分子モータータンパク質や膜輸送体・G タンパク質共役受容体などの分子機能解析がターゲットとなるであろう。これらは、タンパク質機能において化学反応とタンパク質構造変化が強く相関しており、生体分子機能のよいモデル系となる。また、実験的研究も進んでおり、計算科学的アプローチの有効性の検証に必須である緊密な共同研究も期待できる。また、酵素反応性解析で得られた知見に基づく酵素設計も行われる。酵素のアミノ酸配列の大幅な改変に基づく新規機能設計は、非常に挑戦的で、バイオインフォマティクスや全原子 MD 計算などと組み合わせた高度なスキームの洗練が必要となるであろう。一方、小規模改変による設計は、実験的な要求に応え得るものが十分期待される時期に来ている。実験研究との連携を通して実績・経験を積みながら、高い仕様要求を満たす新規機能性タンパク質設計へのスキーム構築を目指すとともに、バイオミメティック触媒設計への応用を行う。

粗視化 MD では、その高速性を生かして、細胞スケールの生命現象に挑むべきである。遺伝子動態、シグナル伝達系、膜・細胞骨格系、ウイルス動態などが挑戦的課題である。例えば、真核細胞における遺伝子の本体であるクロマチンの構造解析や、さまざまな転写因子の作用機序、DNA 複製機構など、構造情報に基づいて研究すべき課題が多い。そのうえで、老化やがんなどの疾患とかかわるエピジェネティクスの理論計算は大きな挑戦である。シグナル伝達系は、分子構造に基づく MD と 1 分子粒度シミュレーションやシステム生物学的な研究の接点になる可能性が高い。シグナル伝達を支えている足場タンパク質複合体などの構造基盤とその可塑性は、システム生物学に直接つながる粗視化 MD の課題となる。

## (ii) 細胞・臓器シミュレーション

1 分子粒度細胞シミュレーションでは、1 細胞での 1 分子粒度での計算を実現、精密化することで、細胞の分化、癌化、増殖などの機序に信号伝達経路のダイナミクスからアプローチすることを目指す。計算を更に大規模化することにより、取り組む事が期待される問題として応答不均一性がある。ヒト細胞を用いた 1 分子観察実験および免疫染色実験では、同一の遺伝情報を持つ均一な細胞集団に対して上皮成長因子や神経成長因子などの同一の因子を同一の条件で与えても細胞の応答には一定の不均一性が観察されることが確認されており、これが発生などの高次の生命現象に持つ意味が議論されている。このばらつきが生じる要因には、分子の構造多型によるとする説、信号伝達ネットワークにおける内因的ノイズの影響によるとする説、転写制御ネットワークなどの遺伝子発現系におけるノイズ、エピジェネティクスによるとする説など複数あり決着していないが、実際にはこれらの要因が複合的に働いている可能性が強い。このうち、構造多型や信号伝達系の内因的ノイズに関しては、現在の 1 分子粒度計算の大規模化でアプローチできるが、遺伝子発現系やエピジェネティクスなどの要因に迫るには、10,000 秒以上の計算でも妥当性を失わない計算モデルが必要である。そのためには、長期的目標の項で述べたようにゲノムシミュレーションと代謝系シミュレーションとの連成が必要であるが、この方面では計算量の問題以前に大規模モデル構築における総合的、長期的な取り組みと技術の蓄積が必要とされ、長期目標の達成に至る筋道として微生物などの原核細胞などの比較的単



純なモデル生物を用いた先行的なプロトタイプアプリケーションの構築をポストペタ世代のうちに進めることが求められる。

細胞・血流シミュレーションについて、対象とする問題は多数考えられるが、例として、血栓性血小板減少性紫斑病[42]に関して説明する。これは、細動脈（血管径  $100\sim 200\mu\text{m}$ ）において過剰に血小板が積み重なることで、血流全体の血小板数が減少し、それに伴い出血（皮下出血による紫斑）が見られる疾患である。脳の細動脈で形成されると神経障害、腎の細動脈では腎障害を引き起こし、放置した際の致死率は90%以上と知られている。この疾患の原因は、血漿中に存在する ADAMTS 13 と呼ばれるフォンビルブランド因子の切断酵素の欠損であり、通常、この ADAMTS 13 は血中の超高分子量フォンビルブランド因子多量体を適度な大きさに断片化することにより、過度な血小板凝集を抑制している。近年では、ADAMTS 13 の立体構造と血流のずり応力との関係が指摘されており、徐々にその原因因子が明らかになってきているが、創薬の観点ではまだまだ研究の初期段階である。今後、分子動力学法あるいは実験的アプローチにより、単一の分子に対する創薬開発が期待できるが、実際の人体においては、血流と細胞の代謝反応が連成して起こり、通常、そのような代謝反応は正のフィードバック機構が働くため、血流も踏まえたシミュレーションによる薬効評価が重要となると考えられる。このため、対象はある程度限定しつつ、解析アプローチを確立することが研究の対象となる。

超音波伝播シミュレーションについて、これまでは生体内の縦波の伝播のみが扱われてきたが、エクサスケール級計算機では、縦波に比べて2桁程度遅い伝播速度を持つ横波の影響[43]を考慮した解析を行う。これにより、治療において報告されている皮膚の火傷や骨の痛み等の生体組織界面における横波による加熱の影響を加味することができ、超音波治療のリスク評価の高精度化や、温度上昇による剛性の変化を利用した超音波による治療部位特定手法の開発が可能になる。更に、ドラッグデリバリーにおけるマイクロカプセルと超音波が干渉した音場の解析や、超音波を援用した場合のドラッグの拡散に関するシミュレーションも可能になると考えられる。

### (iii) 脳神経系シミュレーション

人間の脳の高度な知的活動は、どの部分とその複雑な神経メカニズムによるのか、あるいはその大規模性によるのかも判明しているとは言いがたい。したがって、現段階での HPC において詳細化を追い求めたものと規模の拡大を求めたものの双方の戦略を用いつつ、使用可能な計算資源を活用してシミュレーションを行っていくことになるだろう。計算技術的には、詳細シミュレーションでは膜電位伝搬とさまざまなイオンチャネルのダイナミクスの計算が多く、計算量を要する部位であり、そしてシナプス開口放出から後シナプスまでの一連の過程とその可塑性がネットワーク通信の大枠を規定する。そして、個々の神経の違いを階層的に記述する複雑な構造の上で、ローカルメモリ/キャッシュの効率的な利用が行えるかが一つの鍵になるだろう。また、大規模シミュレーションにおいては、メモリの効率的な利用によるメモリ使用量の削減が課題となるだろう。シナプス間通信などのネットワーク利用を効率的に行うために、通信の階層化や最適化アルゴリズムを介することになることが必要になる可能性も高い。このようななかで、 $10^{11}$  のニューロンとその 10,000 倍のシナプスからなる人間の全脳規模の神経回路を用いた比較的単純なモデルを用いた神経回路シミュレーションと、入力から出力までの信

号処理を単一ニューロンレベルから詳細に調べることができる系として詳細なモデルを用いた昆虫脳の全脳規模の神経回路シミュレーションの二つが、マイルストーンとして重要であり、この二つを進めることで神経回路の機能に対する理解が深まると考えられる。すなわち、神経活動の大規模な同期や感覚情報処理・運動指令信号の形成での神経回路の情報処理における雑音の性質、形態や信号伝達・可塑性などの統計性や、同等なニューロンが集合する意味などがクローズアップされ、その内のいくつかははっきりと進展すると考えられる。また、HPC 計算能力をより小規模な神経回路や単一細胞の複雑性の評価・最適化に用いることもできる。このような系においては、進化的アルゴリズムのような並列シミュレーションや信号情報処理のマルコフ過程(代表的にはイオンチャネル)のダイナミクスの確率的なシミュレーション、更に、リアルタイム神経シミュレーションを神経インターフェースと結合することで神経活動のなんらかの制御や神経メカニズムの同定が行われることが期待される。細胞成長や筋肉などの運動系を含めたマルチフィジックスシミュレーションも一つの大きな課題となるだろうが、そのような時間・空間スケールの異なる現象の結合は、すでに脳神経シミュレーションそのものとは別の問題なのでここでは踏み込まない。

#### (iv) ゲノム解析・バイオインフォマティクス

遺伝子ネットワーク推定ソフトウェアは、マイクロアレイなどの遺伝子発現データから統計科学的・情報科学的方法を用いて遺伝子間の制御の予測・推定を行うソフトウェアである。遺伝子ネットワーク推定の方法としてはさまざまなモデル・推定法が提案されているが、それぞれ一長一短があり、今後も解析対象・目的に合わせたさまざまな遺伝子ネットワーク推定法が現れると思われる。また前述のとおり、現在は「京」などの計算リソースの登場により、解析対象をそれほど絞らずに数百程度のデータセットを抽出し、それに対して網羅的に遺伝子ネットワーク推定ソフトウェアを適用するというアプローチが現実的になっている。現状で 256 データセット・30,517 アレイ(サンプル)での網羅的計算を行っているが、公共データベースに登録されているサンプル数は全約 3 万データセット・84 万サンプルと一部であり、これまでのデータの増加ペースを考えると 2015 年には 150 万サンプル、2020 年頃には 280 万サンプルほどになると予想される。つまり 2020 年には全体のサンプル数では現在の 3 倍ほど、全データを解析しようとする 100 倍程度の計算資源が必要となる。また解析手法が多様化することが予想されるため実際に必要な計算資源は更に大きくなる。公共データベースにある発現データをまるごと計算対象にすることが可能になれば、膨大な遺伝子ネットワークが蓄積可能となり、そこから更に解析・知識発見を行う、というもう一段上の研究ステージに上がることができる(次世代 DNA シークエンサーについては 3.2.3 項を参照されたい)。

ゲノム解析・バイオインフォマティクスではこれまでの HPC での計算にない特徴的な部分がいくつかある。ゲノム解析・バイオインフォマティクス分野で特異的に必要となる定性的スペックを挙げておきたい。当該分野ではこれまで主に PC クラスタを用いて解析が行われてきたが、現世代機のフラッグシップである「京」では十分な対応がなされていないのが現状である。これらの大部分は PC クラスタでは当たり前を実現されていることであり、将来のフラッグシップシステムでも同等以上の利用環境になることが当該分野では必須事項である。

- 脳神経系シミュレーションにおいても共通な必要事項であるが、さまざまな既存アプリを利用するために、GCC を前提にしている一般的なオープンソースアプリがコンパイル・動作する必要がある。Java/R/Python/Ruby などの高級言語、あるいはミドルウェアが高速で動作することが生命科学のデータ解析では求められている。あるいは、こういったミドルウェアレベルでの並列プログラミングサポートが必要である。また、一步進んでスクリプト言語やミドルウェアレベルでの並列化のサポートが望ましい。
- 複数のアプリを組み合わせて順番に利用することがあり、ジョブ管理システムの API がスクリプト言語や C/C++ などから利用できることが望ましい。現状、計算ジョブを監視・制御する別の計算ジョブを走らせる、ということで実現しているがプログラミング言語レベルで対応されれば解析に必要な計算手順の構築が容易になり研究全体がより効率的になる。
- 前項のとおりさまざまな既存アプリが使われるため、よく使われるものについては、あらかじめシステムにインストールされていることが望ましい。例えばヒトゲノム解析センター（HGC）に導入されているスパコンには、バイオ系アプリに限っても 41 種のアプリがプリインストールされユーザに使われている。
- 計算の背景に膨大なデータを利用しており、高速なデータとのアクセスが可能であることが望ましい。特にバイオインフォマティクスでは、システム内に公開データベースのミラーデータがあり高速にアクセスできることが必要である。具体的には、NCBI のアレイデータおよび配列データなどである。HGC のスパコンでは、実際に各種データベースのミラーリングを提供している。
- データ解析では、ジョブ実行時間が長期間に及ぶこともある。伝統的な HPC の世界では長くても数日が最大のジョブ実行時間として運用されていると思われるが、HGC では数週間あるいは数ヶ月継続実行可能なキューが用意され、実際に使われている。
- ゲノム解析やバイオインフォマティクスでは、データ並列性が高いためアプリ自体は高並列化せずパラメータスweep的な大量（具体的には数万から数十万）のアレイジョブによる並列実行が多い。
- ゲノム解析では DNA 配列という究極の個人情報扱うためセキュリティにも配慮が必要である。

#### (4) 課題を解決するために必要なアプリケーション群（要求性能）

##### (i) 生体分子シミュレーション

分子動力学計算は、運動方程式を時間について数値積分することで、分子における粒子の動きのシミュレーションを行う計算手法である。計算する相互作用については、近距離の粒子間に働く相互作用と、遠距離の粒子間に働く相互作用があり、それぞれ計算コストのあり方がかなり異なる。近距離間相互作用は、空間的に系を分割し粒子を計算ノードに割り当てれば、隣接ノード間通信だけで計算に必要な情報はカバーできる。一方の遠距離間相互作用では、情報を大域的に共有する必要があるため、遠距離ノード間通信が発生する。現時点では、遠距離間相互作用計算では、FFT を利用する計算アルゴリズムが普及しているが、通信量を削減するため FFT を利用しないアルゴリズムの開発も求められる。

一方、実際の生物中では、分子は多数存在しアンサンブルとしてその機能を発揮しているので、一分子だけでなく、多数分子の統計量（自由エネルギー等）を計算すべきという考えがある。それに基づいた手法としては、対象系の複数のレプリカを計算機中に置き、それぞれのレプリカでシミュレーションを行いつつ、そのレプリカ間で情報交換を行いながら計算を進めていく方法がある。幅広いサンプリングを実現するレプリカ交換や、タンパク質の機能発現にともなう構造変化を扱うレプリカパス法が代表例だが、これらの方法では、レプリカ間は疎結合ですむので、大規模並列計算に適しているという特徴がある。

全原子 MD では、演算性能が重視され、高いネットワーク性能（特に低レイテンシ）が必要となる。計算したい対象（タンパク質など）があらかじめ決まっていることが多いため、ある特定の対象に対して並列性能を上げる **Strong Scaling** が必要となる。分子レベルでの重要な生命現象の時間スケールは、マイクロ秒から秒ぐらいであるため、1日に100マイクロ秒のシミュレーションを行い、10日で1ミリ秒の計算を行う場合について試算を行う。通常、時間刻みは2 fs/step なので、1ステップ当たり1.7マイクロ秒で計算を終える必要がある。演算量のおよその見積もりでは、 $\sim 50k$  演算/particle/step である。 $10^6$  原子では、50G 演算/step となる。ここから計算すると、30PFLOPS の実効性能があれば、上記の計算が可能となる。実効性能が10%出るコードであれば、300PFLOPS の計算機ということになる。1ステップ当たり、通信は複数回存在するので、サブマイクロ秒以下のネットワークレイテンシが必要になってくる。ただし、この見積もりは対象のサイズに依存する数字である。対象が決まっても、複数のレプリカを連携させながら計算するアンサンブルシミュレーションでは、計算機中で扱う原子数がレプリカの数だけ増えるため、問題サイズがかなり増大する。一方、レプリカ間の通信は疎であることが多いので、ネットワークへの要求は緩和される。上記の1ステップ当たり1.7マイクロ秒という速度は、使用する計算機の実績によっては、単独のシミュレーションで実現するのは厳しい目標値だが（例えば単に並列度が高いだけの計算機では実現困難）、複数のレプリカでの MD 計算を行い、それをマルコフモデルなどのアンサンブルモデルを用いて統合すれば、それぞれの単体シミュレーションでの要求速度は緩和されるため現実的となる。ストレージは、10 ピコ秒ごとに単精度の座標履歴をとったとして、 $10^6$  原子の1ミリ秒の計算では1.2 PB 必要である。計算時間のうち10分で出力するとすると、ストレージ速度は2 TB/s 必要である。長距離力の計算に PME 法(3D-FFT を利用)を用いた実測によると、メモリ速度は  $B/F=0.4$  程度であり、各ノードに分割されたデータだけ持つようになれば必要メモリ量はそれほど多くない。そこで効果的にプログラミングをすればキャッシュで計算が完結しローカルメモリが有効となるため、適合する HPC の形態としては、アクセラレータやメモリオンチップ型の計算機を利活用できる可能性がある。ただし、高速化のために相互作用を行うペアリストを持てば、 $\sim 1000$  ペア/粒子程度が必要となり、各ノードに1,000 原子を割り当てられれば（ $10^6$  原子、100 レプリカ、10 万ノードを想定）、4 MB 程度が必要となる。さらに、プログラミングの自由度を考えると、各ノードに全原子のデータがある程度持っておいたほうがよく、実装にもよるが、各原子に対し力場パラメータなどを80 バイト持つとすると、各ノードのメモリ量はペアリスト分も含めて  $10^6 \times 80 + 4MB = 84MB$  となる。10 万ノードを利用するとすれば、全体で8.4 TB 必要になる。単体演算性能では、SIMD などのデータ並列機構も有効である。ネットワークに関しては、上記のように、低レイテンシのネットワークが必要であり、キャッシュ・ローカル



メモリからネットワークへ直接通信できることが望ましい。また、高速な同期、とくに reduction が必要である。現在、広く普及している長距離力の計算法は、3D-FFT を用いるものであるが、典型的な問題サイズに対しては FFT のグリッド数が $\sim 100 \times 100 \times 100$  程度であり、超大規模な並列計算にて効率を出すのは困難かもしれない。この場合、FFT に依存しない相互作用計算法を導入する必要がある。一方、既に分子力場が確立したものであれば分子動力学計算専用のアクセラレータ（近距離および遠距離相互作用計算を含む）が利用できると高速計算が可能となる。

複数のタンパク質を含む多成分系の超大規模系に対しては、次世代機の全ノードを活用した計算が有効であると思われるため、ここでは、その計算量の見積もりを行う。例えば、理論性能 1 EFLOPS の計算機を実効性能 50% のコードで使用すると、500 PFLOPS の実効演算能力があれば、細胞環境やウイルス全体、細胞膜を模した脂質膜環境やドラッグデリバリシステムについて、全原子レベルで表現した 1 億原子系に対する 1 マイクロ秒の分子動力学シミュレーションを 2 日程度で完了可能である。メモリ速度はソフトウェア MODYLAS の実測に基づき  $B/F=0.1$  程度とし、ネットワーク速度は「京」程度の性能を想定した。必要メモリ量については、最低限、各ノードに分割された原子位置、速度、力、力場パラメータ等があれば、原理的には計算可能であり、MODYLAS の実測に基づく試算からも 0.2PB 程度あればよいという結果を得ている。ストレージについては、1 ps ごとに、単精度の 3 次元座標を出力するとすると、1 億原子系に対し、1 マイクロ秒では、 $4 \times 3 \times 10^8 \times 10^6 = 1.2 \text{ PB}$  必要になる。これを計 10 分で出力するとすると、ストレージ速度は 2 TB/s 必要である。大規模な対象に対する MD については、第 4.2 節の長距離古典分子動力学シミュレーションの計算機見積もりも参照されたい。

創薬への応用において、標的タンパク質と化合物との結合強度（結合自由エネルギー）を定量的に求めることが、薬としての活性を予測するうえで重要である。結合自由エネルギーを精確に求める方法（例えば MP-CAFEE 法）は、全原子分子動力学計算に基づいているため、全原子分子動力学計算の高速化と直接的に連動する。そのため、全原子分子動力学の計算手法の改善を行うことが重要な課題となる。また、大規模な化合物探索に向けて、化合物力場の作成の自動化・高精度化を行っていく必要がある。また、実験チームとの連携体制を整え、シミュレーションに基づく予測を検証し、経験を蓄積して行くことが重要である。

創薬 MD では、標的タンパク質に対して 1000 種類の化合物を MP-CAFEE 法で結合強度を調べる場合、10PF マシンを使えば 5 日くらいで計算が終了する。1EF の計算パワーがあれば、同じ時間内に 10 万種類の化合物について標的タンパク質との結合強度が評価できる。ここでは、各々の化合物に対する結合強度は密に通信する並列化ジョブで計算されるが、その並列化ジョブは各化合物別のアレイジョブとして実行されたとした。MP-CAFEE 法のような高精度計算で探索できる化合物が増加することは、医薬品開発を大きく加速できる可能性がある。また同時に、多数の標的タンパク質に対応できることは、産業応用にとっても大きな魅力である。

QM/MM 法は、QM 計算と MM 計算を用いたハイブリッド法である。MM 部分の計算については全原子 MD と同等の計算である。QM 計算は用いている近似法に基づきさまざまな手法がある。一般的に用いられているのが HF 法や DFT 法などの SCF に基づく手法である。DFT 法では、原子基底と平面波基底を用いたアプローチに大別される。誤解を恐れずに大雑把な表現をするのであれば、前者は HF 交換相互作用を用いたハイブリッド汎関数に基づく高精度な計算が可能であり、後者は高い計算効率を志向している。さらに、SCF 法で得られた波動関数を

出発点とし、電子相関の効果を考慮する Møller-Plesset (MP) 法や coupled cluster 法、また電子励起状態を計算する TDDFT や CASPT2 法などの post-SCF 法がある。このアプローチは、計算コストが高いが、非常に高精度な計算が可能である。SCF 法では、大きく分けて、行列要素の計算と、その対角化という二つの複雑な計算プロセスの繰り返しとなる。したがって、strong scaling するアルゴリズムの開発が非常に困難である。近年開発されている DFTB 法は、前者の行列要素の計算を経験的パラメータを用いて非常に簡略化することにより、高速化を図っている（精度とパラメータ決定の困難さがトレードオフとなる）。また、後者に関しては、Car-Parrinello 法により、対角化を避けて時間発展を計算する手法がよく用いられている（短いステップ長がトレードオフとなる）。これらを組み合わせることにより、並列化による大幅な計算速度の向上は見込めるが、DFTB 法の汎用性や精度に関しては更なる検討を要する段階である。一方、weak scaling を目指す計算（大規模系や高精度）に関しては、分子科学や物性物理などの分野で SCF 法および post-SCF 法の両者に対して、領域分割や実空間積分などに基づくさまざまな方法及びアルゴリズムの開発が行われている。したがって、それらを開発している分子・物性などの研究チームとの密接な共同研究が重要となるであろう。QM/MM 計算によるエネルギーおよび力の計算においては、ほぼ QM 計算が律速になり、QM-MM 相互作用も QM 計算のアルゴリズムに基づくことになる。したがって、QM/MM-MD 計算は、QM 計算の効率如何となる。一方、QM/MM 自由エネルギー構造最適化法では、QM/MM 計算と MD 計算が完全に分離される。QM/MM 計算では、従来の QM/MM 計算に加えて、MD 計算により得られた大量のトラジェクトリスナップショット構造における QM-MM 間の平均場 MM 相互作用の計算が SCF の繰り返し毎に必要となる。後者に関しては、スナップショット構造ごとに分割した計算が行えるので、超並列化に対して非常に高いスケーリングが期待され、長時間および大規模 MM 系に対する拡張は容易であると思われる（実効性能に関しては検討の余地あり）。前者に関しては、Strong Scaling が困難であるため、後者との同時実行を少数のプロセッサを用いて行うなどのアルゴリズム開発が必要であろう。QM/MM 計算は、QM および全原子 MD 計算のハイブリッド法のため、要求計算性能は、QM 計算（物質科学の 4.2 節参照）および全原子 MD に準じる。

粗視化 MD は、分子動力学計算の一種であり、運動方程式を時間について数値積分することで、分子中の粒子の動きをシミュレーションする計算手法である。その意味で全原子 MD と類似点は多いが、通常用いられる相互作用の差異などにより、アプリケーション開発においていくつか異なる状況がある。全原子 MD に比して、粗視化 MD のポテンシャル関数は多様であり一概に論じるのは難しいが、多くの場合、静電相互作用に起因する遠距離相互作用は、カットオフを用いるか Yukawa ポテンシャルなどダンピングがある形で記述されることが多い。そのために、全原子 MD の場合に比して、計算量に占める遠距離相互作用の割合は幾分少なくなる。これは、計算を速くするのに寄与しているが、並列計算における Strong Scaling を大規模まで実現するには困難がある。また、溶媒分子を陽に含まない粗視化モデルの場合には、粒子密度に大きな偏りがあり、全原子 MD とは異なる並列化アルゴリズムを工夫する必要がある。計算量は計算するシステムサイズに依存するので見極めは難しいが、通常、全体で  $10^{10}$  程度のステップ数が要求されるため、1 ステップあたり 0.1 ミリ秒、遅くとも 1 ミリ秒程度が望まれる。

## (ii) 細胞・臓器シミュレーション

一分子粒度細胞シミュレーションでは、格子法と粒子法の2種類の計算手法がよく用いられる。どちらも基礎方程式として Langevin 方程式の過減衰極限である Smoluchowski の反応拡散方程式を用いる。ペタ世代では、大規模並列化で格子法が先行しており、主なアプリケーションとなっている。格子法は格子気体法を拡張したアルゴリズム（例えば Spatiocyte 法）が用いられ、0.1 から 1PF の能力があれば、1 細胞でのレセプターへのリガンド結合から転写制御因子活性化までの一連の応答反応が計算できる。1EF であれば、1,000 から 10,000 細胞で構成される細胞集団、あるいは組織における応答不均一性の計算が行える。粒子法では、通常のブラウン動力学ではステップ幅がナノ秒を切り、生理学的な時スケールである 1,000 秒に到達するためには最低  $10^{12}$  のステップ数が必要であり、実用的ではない。このため、グリーン関数を用いたステップ幅の伸長（GFRD : Green's Function Reaction Dynamics 法）や経験的パラメータによる近似解法（Smoldyn 法）などの工夫が用いられる。近似を用いない正確な解法として現在最も性能がよい改良 GFRD 法（eGFRD 法）はイベント駆動型であるため高並列化が困難であり、1,000 分子程度の系までが適用限界であったが、階層的時ステップを用いて高並列化を試みる新世代の手法である pGFRD 法の開発が理研で進んでいる。これが順調に進めばポストペタ世代には百万分子程度の系の計算が可能になると見込まれ、粒子法が実用段階に入ると思われる。この分野のハードウェアへの要求の特徴として、まず第一に、格子法、粒子法の双方において反応ネットワークや境界条件などが複雑であり、イベント処理や例外処理が多いことがある。このため、SIMD も含めたベクトル系の命令の利用が十分に性能に反映されない場合が多く、整数系の性能が高いほうが有利である。特に、離散事象法においては Intel Haswell アーキテクチャ以降一般的になると考えられるハードウェアトランザクショナルメモリの有効性およびその Simultaneous Multi-Threading との親和性が指摘されている。第二に、長時間のダイナミクスを対象とするため、必要なステップ数が多く、ネットワークのレイテンシが容易に律速となることがある。以下に、具体的な例を挙げての計算機要求の見積もりを示す。まず、格子法で 10,000 細胞のシミュレーションを 10 ケース程度行う場合を考える。細胞の体積が 10pL ( $10^{-12}$  L)、含まれるタンパク質すべての濃度が 10 $\mu$ M とすると、1 細胞当たりおよそ  $10^9$  個のタンパク質分子が存在する。格子法における典型的な時ステップサイズは 1 $\mu$  秒であり、3,600 秒のシミュレーションを行う場合、総ステップ数は  $3.6 \times 10^9$  となる。1 粒子当たりの演算量はおおよそ 10 FLOP であり、その時の総演算量は、 $3.6 \times 10^5$  EFLOP となり、これを 10 ケース行う。10 日間でシミュレーションする場合、必要な演算性能は、0.42 EFLOPS と見積もられる。同様に、粒子法で、一細胞の信号伝達経路のシミュレーションを行う場合を考える。粒子法での典型的な時ステップサイズは、1 ナノ秒であり、3600 秒のシミュレーションを行う場合、総時ステップ数は、 $3.6 \times 10^{11}$  となる。1 粒子当たりの演算量はおおよそ 1000 FLOP であり、総演算量は、 $3.6 \times 10^5$  EFLOP となり、これを 100 ケース行う。格子法と同様に、10 日間でシミュレーションする場合、必要な演算性能は 0.42 EFLOPS と見積もられる。

これまでに連続体細胞、臓器レベルで取り上げてきた応用例では、いずれも有限差分法に基づき、カーテシアン座標系で立方体格子を用いて基礎方程式を離散化する。このような方法は、並列計算において均等に領域分割することが容易であり、台数効果が高く、計算プロセッサ数の増加にともなう計算機性能の向上の恩恵を直接受けてきた。細胞・血流シミュレーションで

は、連続体に対する移流・拡散・反応式を扱い、時間積分によって生化学反応、流体・構造・膜の動力学挙動を再現する。圧力の更新には、擬似圧縮性の解法を用いることで、圧力ポアソン式の求解の際に不可欠な反復をとまなう隣接通信を避ける。時間刻みを、クーラン条件を満たすように移流速度と格子解像度から決めると、粘性項と拡散項の一部には陰解法を適用する必要があるものの、計算コストのかかる反応項、移流項、構造・膜の運動学構成式には陽的な時間積分の適用が可能である。カーチアン格子ベースで、（ほぼ）陽解法を採用する細胞・血流シミュレーションでは、格子点当たりの演算量が多いため、今後、B/F が下がるとしても、それを捉える程度にキャッシュサイズが増えるならば、実行性能を出しやすい。一方、コードの書き方の選択肢が少なく、アルゴリズムの工夫で格段に性能向上することは期待できない。長さ 100mm、径 100~200 $\mu$ m の血管を対象とし、流速を  $10^{-2}$  m/s と考える。血球の変形挙動までを取り扱うため、0.1 $\mu$ m 程度の格子幅を用いるとすると、格子点数は  $10^6 \times 10^3 \times 10^3 = 10^{12}$  である。クーラン数を 0.1 とすると、時間解像度は 1 $\mu$ s と見積られ、これは膜の動的挙動と捉えるための時間スケールと同程度である。血栓生成の時間スケールを考え、10 秒の現象を扱おうとすると、全計算ステップ数は  $10^7$  程度である。格子点上で定義される変数は速度ベクトル、圧力、VOF 関数、バルクひずみテンソル、膜ひずみテンソル、参照曲率、反応種であり、合わせて  $10^2$  個とするならば、必要な主記憶容量は、 $8 \times 10^2$  (変数)  $\times 10^{12}$  (自由度)  $\sim 10^{15}$  Byte = 1PByte である。例えば、ノード当たりの性能、計算規模が、それぞれ、100TFLOPS、 $10^8$  格子 $\times 10^2$  変数、実効性能が 40%だとすると、1 ステップ当たりの計算時間は  $10^{-2}$  s 程度と見積られる。 $10^2$  個の変数を持つ 1 格子点、1 ステップ当たりの演算量は約 25,000 であり、 $10^{12}$  の格子点に対する  $10^7$  ステップ数の計算の場合、全演算数は  $2.5 \times 10^{23}$  程度となる。「京」の全ノードでの実効性能が 40%の性能を有しているため、1 EFLOPS の計算機を用いた場合、実効性能が 400 PFLOPS となり、約 174 時間の計算時間と見積もられる。理想的なキャッシュであれば B/F は  $2 \times 8 \times 10^2$  (変数) / 25,000 = 0.064 と見積もられる。したがって、必要なメモリバンド幅は  $100$  (TF/s)  $\times 0.064$  (B/F) = 6.4 TB/s と見積もられる。

超音波治療に用いられるコードの「京」での実効性能は 20%程度である。したがって、超音波治療で用いられる周波数 1~2MHz の超音波の軟組織中での波長が、0.75~1.5mm であり、これを解像するのに 0.05mm 程度の格子幅を用いるとすると、京では、 $10,000^3$  格子点 ( $5.12 \times 10^{11}$  自由度) の計算が可能であり、400mm の立方体領域中の超音波の伝播が 1 日程度で計算できる。しかし、軟らかい生体組織の温度変化で生じる微妙な剛性の変化を捉えるためには、音波である縦波に比べて 2 桁程度遅い伝播速度を持つ横波を再現する必要がある、より高い解像度と長時間の時間積分が必要になり、最低でも京での計算の 1,000 倍の自由度の計算が求められ、「京」の 100 倍 (1EFLOPS) の計算機を利用した場合、10 日程度で計算ができる見積もりとなる。また、マイクロカプセルとの干渉によって生じる超音波音場の解析を行うためには、時間スケールが小さいマイクロカプセルの時間積分を行うためには計算コストが通常の超音波伝播計算に比べて数十倍程度必要となるため、現状の 100 倍程度の計算処理速度が将来的に必要となると考えられる。



### (iii) 脳神経系シミュレーション

脳神経系シミュレーションにおいては、理想としての詳細なモデルを用いた、ヒトの全脳規模の神経回路シミュレーション、2020年頃の実現可能な二つのモデルとしての単純なモデルを用いた、ヒトの全脳規模の神経回路シミュレーション、詳細なモデルを用いた昆虫(カイコガ)の全脳規模の神経回路シミュレーション、の三つのシナリオについてシミュレーションのプロファイルを検討する。詳細モデルにおいて神経細胞の膜電位変化のダイナミクスを十分に再現するには、Hodgkin-HuxleyモデルのNaとKチャンネルの他に、Na<sub>p</sub>、K(AHP)、Ca(high)、Ca(low)、Hなどのチャンネルを追加することが必要であり、各コンパートメントは、Hodgkin-Huxleyモデルの5倍程度の複雑度を持たせるとして、20変数、600FLOP/ステップとする。これはHodgkin-Huxleyモデル1ステップの計算には、120FLOPかかるというIzhikevichの見積もり[44]に基づく。コンパートメント数を5,000とする。シナプスには、コンダクタンスベースのモデルを用いる。シナプス数は、Matrix exponential法[45]を用いると、1ステップに5FLOPかかる。1ステップは0.1msとする。シナプス可塑性は、時間スケールの異なるものがいくつかあり、短期可塑性を2種類、長期可塑性としてSTDP (Spike timing dependent plasticity)、短期と長期をつなぐ時間スケールの可塑性を1種類想定する。神経細胞の発火頻度は平均10Hzとして、発火1回に対する指数関数の積和演算13FLOP (STDPは26FLOP)を可塑性の演算量として見積もる。これに、人間の脳の規模のパラメータ、すなわち、 $10^{11}$ の細胞数、1個の神経細胞当たり10000個のシナプスという量を当てはめれば計算規模が見積もれる。1神経細胞の1ステップあたりの計算に必要な、主記憶メモリバンド幅は、キャッシュメモリを使用しない場合、約3.4MB (神経細胞変数: 800KB、シナプスコンダクタンス: 160KB、シナプス可塑性: 320KBのload/storeと神経細胞定数: 800KBのload)のメモリ転送と、約3MFLOP ( $600 \times 5000 \text{FLOP} + 5 \times 10000 \text{FLOPS} + 650 \text{FLOPS}$ )の演算が必要と考えるならば、B/F値は1.1と見積もれる。ネットワークバンド幅は、発火情報(神経細胞のIDと発火時刻)のデータ量で決まる。平均発火頻度を10Hzとし、シナプス遅延である1msごとにデータ転送を行うとすると、発火情報:  $10^{11} \text{neurons} \times 10 / 10000 \times 10 \text{Hz} \times 16 \text{Byte} = 16 \text{GByte}$ が発生し、その間の計算量は、 $600 \text{FLOP} \times 10 \text{step} \times 5000 \text{compartments} \times 10^{11} \text{neurons} = 3 \times 10^{18} \text{FLOP}$ となり、ネットワークバンド幅のBF値は $5.3 \times 10^{-9}$ となる。

単純なモデルを用いた、ヒトの全脳規模の神経回路シミュレーションでは、神経細胞数やシナプス数などの基本的な設定は、詳細なモデルを用いたヒトの全脳規模モデルと同様であるため、重複する説明は割愛する。単純モデルでは、神経細胞モデルにLeaky Integrate-and-Fire neuronモデルを用いることとし、1ステップで5FLOP[38]かかる。シナプスには指数減衰型のモデルを用いて、1ステップで、10FLOP、減衰の計算に更に1FLOPかかる。シナプス可塑性は詳細なモデルと同様とする。

カイコガ全脳の神経回路のリアルタイムシミュレーションにおいては、計算の基本は詳細なモデルを用いたヒトの全脳規模の神経回路シミュレーションと同じであるが、神経数は $10^6$ 程度であり、1個のシナプス当たりのシナプス数も数百個オーダーと規模は小さい。また、計算ステップを0.05msとする。三つのシナリオについてシミュレーションのプロファイルを概算した結果を表4.1-1に示す。

表 4.1-1 脳神経系シミュレーションの三つのケースの計算規模見積もり

項目	詳細なモデルを用いた人の全脳規模の神経回路シミュレーション	単純なモデルを用いた人の全脳規模の神経回路シミュレーション	カイコガ全脳の神経回路のリアルタイムシミュレーション
細胞数	$10^{11}$	$10^{11}$	$10^6$
コンパートメント数	5000	1	10000
シナプス数/細胞	10000	10000	500
リアルタイムシミュレーション必要 FLOP			
細胞	$3 \times 10^{21}$	$5 \times 10^{15}$	$7.2 \times 10^{16}$
シナプス	$5 \times 10^{19}$	$1.1 \times 10^{16}$	$5 \times 10^{13}$
可塑性	$6.5 \times 10^{17}$	$6.5 \times 10^{17}$	$6.5 \times 10^{10}$
データ量 (byte)			
細胞	$1.6 \times 10^{17}$	$2.4 \times 10^{12}$	$2 \times 10^{12}$
シナプス	$4.8 \times 10^{16}$	$2.4 \times 10^{16}$	$2.4 \times 10^{10}$
可塑性	$3.2 \times 10^{16}$	$3.2 \times 10^{16}$	$1.6 \times 10^{10}$
BAND 幅 (byte / step * CELL)			
双方向	$1.28 \times 10^6$	349	$1.02 \times 10^6$
ロード	$8 \times 10^5$	32	$9.6 \times 10^5$
FLOP/STEP	$3 \times 10^6$	666	$3.6 \times 10^6$
B/F	1.1	1.1	0.83
キャッシュ活用時			
キャッシュ必要量(byte)/チップ	$10^6$		$10^6$
B/F (主記憶)	0.11	-	0.041
ネットワーク B/F (全系通信として評価)	$5.3 \times 10^{-5}$ ( $5.3 \times 10^{-9}$ )	$2.5 \times 10^{-1}$ ( $2.5 \times 10^{-5}$ )	$1.1 \times 10^{-6}$ ( $2.2 \times 10^{-9}$ )

脳神経での詳細なモデルを用いた、ヒトの全脳規模の神経回路シミュレーションの見積もりから、ヒトの全脳規模の神経回路モデルの計算には、神経細胞の変数計算の割合が圧倒的に大きく、物理時間の1秒のシミュレーションを行うのに、 $3 \times 10^{21}$ FLOPが必要である。B/Fは1程度あることを考慮するとメモリバランス型が適しており、更に0.23EByteという大容量のメモリ容量が必要であると言える。ただし、B/F要求値については、必要最低限のメモリアクセスのためにシナプス遅延である1msの間はメインメモリにアクセスせずに1チップにおいて100~200MBのキャッシュ利用による計算が可能なシナリオを考えると、1桁低くなる。

単純なモデルを用いたヒトの全脳規模の神経回路シミュレーションの計算には、リファレンスモデルとメモリモデルの中間的性能が適していると考えられる。主記憶のメモリバンド幅の要求が高く、ネットワークバンド幅のB/F値も詳細なモデルに比べ高い。また、シナプス可塑性に関するメモリ転送量が大きく、キャッシュの効果はほとんど期待できない。1秒のシミュレーションに0.7EFLOPs程度の実行性能が必要であり、10%程度の実効効率ならば、メモリバランス型で、現実時間で1秒のシミュレーションが、数十秒程度で実行可能になると見積もられる。

カイコガ全脳の神経回路のリアルタイムシミュレーションの計算には、B/F重視型が適していると考えられる。ただし、カイコガの全脳モデル開発は、非常に複雑であり、開発を容易にするためGCC等一般的なコンパイラと互換性のある環境が望まれる。現在シミュレーションが実行できている「京」と同様のアーキテクチャと予測されるリファレンス、メモリバランス

型で、0.7EFLOPs の最大性能であれば、実効効率 10%程度で、リアルタイムシミュレーションを行える可能性がある。

昆虫神経回路シミュレーションを用いて精密なシミュレーションの状態推定・パラメータ推定を行うときの計算負荷はモデリング依存であり、また大規模神経回路になるとまだ見積もりが難しいところがある。1 細胞の 10,000 コンパートメントモデル程度の規模においては、物理時間 1 秒程度のシミュレーション ( $7.2 \times 10^{10}$  FLOP) の 1,000 遺伝子 $\times$ 100 世代 $=10^5$  倍ほど計算資源があれば可能である。1,000 細胞の神経回路を対象にするとすれば、 $10^6$  遺伝子 $\times$ 100 世代程度とすれば、 $7.2 \times 10^{21}$  FLOP 程度の計算実行がとりあえずの目安になり、この 10 倍程度の計算資源を用いれば実際の神経回路の動作についての有用な推定が行うことができると見込まれる。しかし、この状態推定のためにはおそらくまだ相当の試行が必要であり、2015~18 年程度までは、 $10^2$  程度の神経を用いた神経回路シミュレーションを用いた神経回路状態推定の方法を多くの人間がプロダクト計算実行が行える程度に定着させる必要があるだろう。この開発として、 $10^{19}$  FLOP $\times$ 100 回程度が見込まれる。

リアルタイムシミュレーションは、リアルタイムな実験とのデータ同化や生体コントロールの可能性を押し開く。したがってその段階においてはシミュレーションと観測装置とのリアルタイムに近いコミュニケーションが必要となる。2020 年頃に想定されるデータ量として、カルシウムイメージングにおいて、 $512 \times 512 \times 10 \times 16$  bit 程度のイメージデータが 100ms ごとに得られると仮定すると、これをシミュレーションを介した昆虫脳のリアルタイムのデータ同化や制御に用いるなら、シミュレータと外部環境と 100 MByte/s 程度の通信が可能である必要がある。カイコガ全脳の神経回路のリアルタイムシミュレーションにおいては、シミュレーション対象はたかだか 10~100 秒の物理時間であり、リアルタイムシミュレーションであるならば、同一の計算時間しか使用しないので、1 回の計算実行はきわめて短い。しかしながら、生物実験との同期をとるための特別なスケジューリングの運用が必要になる可能性は留意すべきである。それに見合った生物実験ができることが前提であるが、このような特別にスケジュールされたジョブを 10 回程度実行できれば、シミュレーションを用いた昆虫の神経回路の制御や同定の可能性を一通り評価することが可能であろう。

以上の個々の性能が見積もられた神経系シミュレーションに対して 2020 年頃に必要な総計算量は

①ヒトの脳シミュレーションにおいて

詳細モデルにおいては  $10,000 \text{ EFLOP} \times 1$

簡約モデルはにおいては  $7 \text{ EFLOP} \times 100$

②昆虫脳シミュレーションに対しては

全脳シミュレーション単体に対しては  $7 \text{ EFLOP} \times 100$

実験とのリアルタイム通信を含む神経回路の状態推定には  $7200 \text{ EFLOP} \times 20$

程度であり、合計  $1.5 \times 10^{23}$  FLOP が見込まれる。

脳神経系シミュレーションにおいては、実験神経生理学との連携は必要不可欠である。形態学や同時多点計測が可能なイメージング技術などと連携することも望ましい。しかしながら、現実的にはそのすべてと強く連携することは不可能であり、その一部と連携しつつさまざまな

情報を集積・再利用可能にするニューロインフォマティクスと連携することになると思われる。一つはニューロインフォマティクス分野のさまざまなデータベースとも連携し、自動的にオンタイムでビッグデータを使用してシミュレーションを構築する動きが進むと予想される。また、現在は PC もしくは小規模なクラスターマシンでしか行われていない神経形態抽出や標準脳座標系へのマッピングのような 3 次元画像処理も重要性が年々増していくと思われ、これらの計算を HPC に乗せていく動きも強まるであろう。後者は技術的には医療画像処理の分野の画像処理と非常に近い分野であり、連携しつつ進むと思われる。

#### (iv) ゲノム解析・バイオインフォマティクス

次世代 DNA シークエンサーについては 3 章で詳しく述べたのでここでは省略する。

遺伝子ネットワーク推定については、これまで「京」用ソフトウェア開発プロジェクトであるグランドチャレンジアプリケーションの研究開発プロジェクトにおいて「京」向けの並列遺伝子ネットワーク推定ソフトウェアを開発し「京」の計算リソースを活用して公開アレイデータ 256 セット・30517 サンプルのデータに対しベイジアンネットワークを用いた 2 種類の推定アルゴリズム (SiGN-BN NNSR/HC+Bootstrap) [46,47]を用いた網羅的遺伝子ネットワーク推定計算を行っている。これには 259 CPU Core 年を要し、FLOPS 換算するとおよそ 130 EFLOP の演算量であった (理論性能ベース)。またそれとは別に L1-正則化法を用いた個人ネットワークプロファイリングが可能な SiGN-L1 の開発も進めている。これらを用いて 2015 年に現在の 40 倍の規模、2020 年に 100 倍の規模の解析を想定すると、以下のような計算スペックが 1 ケースあたり必要となる。

##### 【CPU 速度】 2015 年：～1 PFLOPS、2020 年：～2.5 PFLOPS

2020 年には現在の 100 倍の規模である 26000 セットの計算を想定しており、1 セット分の計算を分割可能な 1 ケースとする。26000 ケースを 1 年間で処理すると仮定すると 1 ケースあたりの処理時間は 1200 秒である。プロファイル結果より 2020 年に想定しているパラメータの 200,000 分の 1 の規模の計算での総演算量は 15 TFLOP であった。単純にこれを 200,000 倍すると 1 ケースあたり 3,000 PFLOP である。これを 1200 秒で処理するので恒常的に必要な計算性能は 2.5 PFLOPS である。

##### 【メモリ量】 2015 年：～0.03PB、2020 年：～0.08 PB

出力が 1 ケースあたり 78 GB であるので、多めに見積もってその 100 倍のメモリを計算時に使用すると仮定する。従って  $78\text{GB} \times 100 = 78,000 \text{ GB} = 78\text{TB} = 0.08\text{PB}$  となる。

##### 【ネットワーク性能】 2015 年：～0.4 PB/s、2020 年：～1.1 PB/s

現行システムのプロファイル結果から求める。一番多い計算ノードでの通信量は 897,000,000 B でこの計算は 120 秒であった。従っておよそ 1 GB を 120 秒で通信しているため 1 ノードあたり 8 MB/s 程度である。実データではこの 100 倍程度を仮定する。またこれを 140,000 ノード程度での計算を仮定すると  $800 \text{ MB/s} \times 140,000 = 112,000,000 \text{ MB/s} = 112 \text{ TB/s}$  で十分と思われる。そこからさらに将来 10 倍の複雑度になると仮定しても  $112\text{TB/s} \times 10 = 1.1 \text{ PB/s}$  である。



【ストレージ容量】2015 年：～6.24 TB、2020 年：～ 15.6 TB

想定している 2000 分の 1 の計算規模で 7.8 GB の出力がある。したがって 1 ケースあたりに必要なストレージ容量は単純にその 2000 倍の 15.6 TB である。

課題	要求性能 (PFLOPS)	要求メモリ バンド幅 (PB/s)	メモリ量/ ケース (PB)	ストレージ量/ ケース (PB)	計算時間/ ケース (hour)	ケース数	総演算量 (EFLOP)	概要と計算手法	問題規模	備考
生体分子機能解析	29	12	0.0084	1.2	240	10	250000	分子動力学計算(全原子(代表)QM/MM,粗視化MDなど)	対象:100万原子,100レプリカ	サブマイクロ秒以下のネットワークレイテンシが必要。メモリ量/ケースは10万ノード実行を想定。
細胞環境・ウィルス 創薬などMD・自由エネルギー計算	490	49	0.2	1.2	48	10	850000	全原子/粗視化分子動力学シミュレーション	～1億粒子	B/F=0.1
細胞内信号伝達経路シミュレーション	1000	400	0.0001	0	0.0012	1000000	4300000	全原子分子動力学シミュレーション	ケース数:10万化合物×10標的蛋白質(10万原子程度)	B/F=0.4, 数百から数千ケース同時に実行することを想定している。実行時に必要な全メモリ量, 各ケースの実際の実計算時間は, 表の値の数百～数千倍となる。メモリ量/ケースは100ノード実行時を想定。
細胞内信号伝達経路シミュレーション	42	100	10	10	240	100	3600000	一分子粒度細胞シミュレーション(格子法)	1000 から 10,000 細胞で構成される細胞集団	格子法・整数系の演算性能を要求。ケース数は最低10回、100回程度が望ましいため100回とした。
細胞内信号伝達経路シミュレーション	420	0.010	0.001	0.001	240	100	36000000	一分子粒度細胞シミュレーション(粒子法)	グリーン関数反応動力学法・百万分子程度	共通・低ネットワークレイテンシを要求
血流シミュレーション	400	64	1	1	170	10	2500000	差分法、準陽解法(構造・流体・生化学連成シミュレーション)	100mm長×100um径, 0.1um格子, 流速10 <sup>-2</sup> m/s, 解像度1us, 10秒	
超音波シミュレーション	380	460	54	64	240	10	3300000	差分法、陽解法(音波・熱シミュレーション)	400mm <sup>3</sup> の計算領域を軟組織とマイクロバセル干渉音場を捉えるため、225兆点の格子と時間ステップ数として1459200ステップが必要である。また、1格子点あたり演算数1000程度となる。	
脳神経系シミュレーション・ヒト全脳簡約モデル	7	7.6	56	3600	0.28	100	700	単一コンパートメントIFモデル シナプス可塑性・通信	1000億ニューロン ニューロンあたり1万シナプス 10 <sup>5</sup> step	ネットワークのボトルネックはレイテンシー
脳神経系シミュレーション・ヒト全脳詳細モデル	71	78	250	25000	39	1	10000	マルチコンパートメントH-H(局所クランクニコルソン) シナプス通信	1000億ニューロン ニューロンあたり1万シナプス 10 <sup>5</sup> step	ストレージ量は最大想定 ネットワークはレイテンシーの影響も大きいと予測
脳神経系シミュレーション・昆虫全脳詳細モデルリアルタイム	71	60	0.002	0.2	0.028	100		マルチコンパートメントH-H(局所クランクニコルソン) シナプス通信	100万ニューロン ニューロン(10000コンパートメント)あたり500シナプス	通信パターン設計に工夫の余地がある
脳神経系シミュレーション・昆虫全脳詳細モデル神経回路パラメータ推定	71	60	0.2	20	28	10	72000	マルチコンパートメントH-H(局所クランクニコルソン) シナプス通信 進化的アルゴリズム	1000ニューロン 10 <sup>6</sup> 遺伝子 100世代	通信パターン設計に工夫の余地がある
脳神経系シミュレーション・昆虫全脳詳細モデル生理実験とシミュレーションの通信	71	60	0.2	20	28	10	72000	マルチコンパートメントH-H(局所クランクニコルソン) シナプス通信 進化的アルゴリズム	1000ニューロン 10 <sup>6</sup> 遺伝子 100世代	100MB/S程度の外部との通信も想定
遺伝子ネットワーク解析	2900	1500	0	0	0	26000	94000000	ベイジアンネットワークおよびL1正則化法	4万転写物×26,000データセット・280万アレイ	

※本見積もりは未だ精査中である。より精度の高い数値はWeb版(→「1.2. 本文書の構成」)を参照のこと。

## (5) 他分野との連携

タンパク質やDNAなどの生体分子およびそれらの複合体の立体構造に基づく解析は、「生命科学」分野と「物質科学」分野の境界に位置した横断的課題である。この点については、分野連携の章(3.1.3項参照)に詳述した。

脳神経系シミュレーションにおいては実験神経生理学との連携は必要不可欠である。形態学や同時多点計測が可能なイメージング技術などと連携することも望ましい。しかし、現実的にはそのすべてと強く連携することは不可能であり、その一部と連携しつつさまざまな情報を集積・再利用可能にするニューロインフォマティクス分野と協同して、ネット上のデータベースを構築し、自動的に、オンタイムにビッグデータを使用してシミュレーションを構築する方向性が考えられる。また、現在は1 PC もしくは小規模なクラスターマシンでしか行われていない神経形態抽出や標準脳座標系へのマッピングのような3次元画像処理も重要性が年々増していくと思われ、これらの計算を HPC に乗せていく動きも強まるであろう。後者は技術的には医療画像処理の分野の画像処理と非常に近い分野である。

#### (6) 公的な意見交換とそのサマリー

第4回アクセラレーション技術発表討論会（2012年9月6日～7日：福井大学）において、「人間の脳と同等のコンピュータがつくれないか？」をテーマに加沢・五十嵐がエクサフロップへの計算見積もりを紹介し、議論を行った。議論においては、どこまでの精密・巨大なシミュレーションが人間の脳の創造性を再現するために必要なのか、ということが一つの大きなテーマとなった。それは、規模の問題・粒度の問題ともに答えが出ない問題であり、当面は両面でやってみることが重要であるというところで落ち着いた。また、シミュレーションのメモリ使用量についての議論があり、現状ではそこがシミュレーションの性能向上の大きなファクターであるということで意見が一致した。

2012年9月23日の日本生物物理学会におけるシンポジウム「スーパーコンピューティング：分子ネットワークと細胞内ダイナミクス」や、2012年10月9日の計算物質科学イニシアティブ（CMSI）計算分子科学研究拠点第3回研究会においては、生命分野のサイエンスロードマップの紹介が行われた。

2012年10月19日のバイオスーパーコンピューティング研究会において、生命分野のサイエンスロードマップの紹介および意見交換が行われた。そこでは、放射光施設やX線自由電子レーザーなどの大型実験施設との連携を考慮して進めるようにとの意見が産業界から出された。また、医学分野との連携をより強く行うように医学界から意見が出された。

2013年6月12日～14日の第13回日本蛋白質科学会年会にて、「エクサフロップス時代の計算蛋白質科学」と題した生命科学・物質科学・ものづくりの各分野と連携したワークショップが開催され、意見交換が行われた。意見交換の詳細は、3.1.3項に記述している。

2013年6月21～22日 Neuro2013（第36回日本神経科学大会/第56回日本神経化学会大会/第23回日本神経回路学会大会合同大会）において本ロードマップの計算量見積もりの一部を紹介した。

2013年6月27日に、製薬企業関係者が多数集まる情報計算化学生物（CBI）学会の研究講演会にて、計算創薬の展望についての意見交換が行われた。放射光施設など大型実験施設との連携や、更なる並列数増加への対応、ものづくり分野との連携についての意見が出された。

2013年8月1日ニューロインフォマティクス国際統合機構日本ノード運営会議において次世代計算機に向けた計算量とロードマップの紹介を行った。将来の HPCI インフラに対しては高級言語の使用を含めた使いやすさの要望が多かった。

2014年1月23日～24日のバイオスーパーコンピューティング研究会「ウィンタースクール2014」において、バイオスーパーコンピューティングにおけるポスト「京」の展望について意見交換が行われた。

## (7) ロードマップ

年代 課題	2012 ～	2014 ～	2016 ～	2018 ～	2020 ～	2022 ～
分子創薬	タンパク質、生体膜、DNAなどのダイナミクス MD計算を用いた医薬品候補の探索(数百から数千)		生体超分子複合体や細胞環境など巨大系のダイナミクス 膜タンパク質を含めた複雑なターゲットタンパク質への適用			
細胞臓器医療	細胞内シグナル伝達反応経路のモデル構築 膵ラ氏島や血小板による血栓形成 血流解析シミュレーションの大規模化と並列化 超音波治療シミュレーション	分子混雑等の細胞環境の導入 細胞の分化、癌化、増殖等の機構の解明と細胞集団への拡大(1,000秒オーダーの予測性)	細胞反応を基とした組織・臓器・器官の生化学血流連成シミュレーション 生物の代謝反応のモデリングと、その血流解析シミュレーションへの導入	シグナル伝達に加え、代謝系、遺伝子発現系を導入へ(10,000秒オーダーの予測性へ) 代謝反応と血流解析連成による循環器シミュレーションへの適用 Drug Delivery System (DDS)において超音波制御を用いる		
脳神経	視覚野神経回路網モデルシミュレーション カイコガ詳細モデルでの嗅覚＝運動系リアルタイムシミュレーション	人間全脳単純モデルシミュレーション カイコガモデルでの全脳詳細モデルリアルタイムシミュレーション			人間の全脳の詳細モデルシミュレーション	
遺伝子ネットワーク解析	現状256サンプル512ネットワークの解析	40倍の規模の解析		100倍への規模の解析		
ゲノム解析	2,000人規模	第3世代シーケンサー		第4世代シーケンサー200,000人規模による解析		

図 4.1-5 生命科学ロードマップ

## 参考文献

- [1] T. Schlick, “Molecular Modeling and Simulation: An Interdisciplinary Guide”, Springer (2010)
- [2] R. Dror et al. “Biomolecular Simulation: A Computational Microscope for Molecular Biology”, Annu. Rev. Biophys. 41: 429-52 (2012)
- [3] G. Voth ed. “Coarse-Graining of Condensed Phase and Biomolecular Systems”, CRC Press (2009)
- [4] E-Cell: <http://www.e-cell.org/>
- [5] SIBL: <http://sbml.org/>
- [6] セルイラストレーター: <http://www.cellillustrator.com/>
- [7] Neuron: <http://neuron.duke.edu/>
- [8] Karr et. al. “A Whole-Cell Computational Model Predicts Phenotype from Genotype”, Cell, 2012, <http://dx.doi.org/10.1016/j.cell.2012.05.044>

- [9] Tomita et. al., E-CELL: software environment for whole-cell simulation. *Bioinformatics* (1999) doi: 10.1093/bioinformatics/15.1.72
- [10] The Virtual Cell プロジェクト: <http://www.nrcam.uchc.edu/>
- [11] Virtual Physiome Project: <http://www.vph-noe.eu/>
- [12] RICS プロジェクト: [http://cellsim.riken.jp/research\\_e.html](http://cellsim.riken.jp/research_e.html)
- [13] ChemCell: <http://chemcell.sandia.gov/>
- [14] Mcell: <http://www.mcell.cnl.salk.edu/>
- [15] Smoldyn: <http://www.smoldyn.org/>
- [16] GFRD: <http://gfrd.org/>
- [17] Spatiocyte: <http://spatiocyte.org/>
- [18] 高木周 (2012) 「生命と科学」(岩波書店) 第5章「人体シミュレーション」
- [19] Pozrikidis, C. (Ed.) (2003) ‘Modeling and Simulation of Capsules and Biological Cells,’ (Chapman & Hall/CRC, Boca Raton); Bungartz, H.-J. and Schafer, M. (Eds.) (2006) ‘Fluid-Structure Interaction - Modelling, Simulation, Optimisation,’ Lecture Notes in Comput. Sci. Engrg., 53; Pozrikidis, C. (Ed.), (2011) ‘Computational Hydrodynamics of Capsules and Biological Cells,’ (CRC Press, Boca Raton).
- [20] Rahimian, A. et al. (2010) “Petascale direct numerical simulation of blood flow on 200K cores and heterogeneous architectures,” Proc. of 2010 ACM/IEEE Int. Conf. High Performance Computing, Networking, Storage and Analysis, doi: 10.1109/SC.2010.42; Bernaschi, M. et al. (2011) “Petaflop biofluidics simulations on a two million-core system,” Proc. of 2011 ACM/IEEE Int. Conf. High Performance Computing, Networking, Storage and Analysis, doi: 10.1145/2063384.2063389; Grinberg, L. et al., “A new computational paradigm in multiscale simulations: Application to brain blood flow,” Proc. of 2011 ACM/IEEE Int. Conf. High Performance Computing, Networking, Storage and Analysis, doi: 10.1145/2063384.2063390.
- [21] Sugiyama, K. et al. (2011) “A full Eulerian finite difference approach for solving fluid-structure coupling problems,” *J. Comput. Phys.*, 230, 596-627; Ii, S. et al. (2012) “A full Eulerian fluid-membrane coupling method with a smoothed volume-of-fluid approach,” *Comm. Comput. Phys.*, 12, 544-576; Takagi, S. et al. (2012) “A review of full Eulerian methods for fluid structure interaction problems,” *J. Appl. Mech.*, 79, 010911; Sugiyama, K. et al. (2012) “Development of explicit Eulerian finite difference solver for large-scale fluid-structure interaction systems,” 2012 年ハイパフォーマンスコМПユーティングと計算科学シンポジウム, IPSJ-HPCS2012056.
- [22] Duck, F.A. et al. (1998) ‘Ultrasound in Medicine,’ (Medical Science Series, Institute of Physics Publishing, Bristol and Philadelphia); Kennedy, J.E. et al. (2003) “High intensity focused ultrasound: surgery of the future?,” *Br. J. Radiol.*, 76, 590-599; Wu, F. et al. (2004) “Circulation tumor cells in patients with solid malignancy treated by high-intensity focused ultrasound,” *Ultrasound in Med. Biol.*, 30, 511-517; Kennedy, J.E. (2005) “Innovation: High-intensity focused ultrasound in the treatment of solid tumours,” *Nat. Rev. Cancer*, 5, 321-327; Hesley,



- G.K. et al. (2008) “A clinical review of focused ultrasound ablation with magnetic resonance guidance: an option for treating uterine fibroids,” *Ultrasound Q.*, 24, 131-139.
- [23] <http://www.healthcare.philips.com/main/products/mri/systems/sonalleve/>  
[http://japan.gehealthcare.com/cwcjapan/static/rad/mri/MRgFUS\\_ExAblate2000/](http://japan.gehealthcare.com/cwcjapan/static/rad/mri/MRgFUS_ExAblate2000/)
- [24] 佐々木明 (2010) “バイオスーパーコンピューティングが拓くライフサイエンスの未来: 超音波治療の推進および治療機器開発に欠かせない生体の音響的シミュレーション研究,” *BioSupercomputing Newsletter*, 3, 4-5
- [25] Sun, J. and Hynynen, K. (1998) “Focusing of therapeutic ultrasound through a human skull,” *J. Acoust. Soc. Am.*, 104, 1705-1715.
- [26] Fink, M. (1992) “Time reversal of ultrasonic fields. Part I: Basic principles,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 39, 555-566; Fink, M. et al. (2003) “Time Reversal Acoustics in Biomedical and Engineering,” *Annu. Rev. Biomed. Engrg.*, 5, 465-497.
- [27] Okita, K. et al. (2010) “Numerical simulation of the tissue ablation in high-intensity focused ultrasound therapy with array transducer,” *Int. J. Numer. Meth. Fluids*, 64, 1395-1411; Okita, K. et al. (2011) “Development of high intensity focused ultrasound simulator for large-scale computing,” *Int. J. Numer. Meth. Fluids*, 65, 43-66.
- [28] W. S. McCulloch and W. Pitts (1943), A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- [29] 篠本滋 (2009) ニューラルネットワーク研究との遭遇日本神経回路学会会誌 Vol. 16, 3, 131-133.
- [30] L. Hodgkin and A. F. Huxley (1954), A quantitative description of membrane current and application to conduction and excitation in nerve, *J. Physiol.*, 117, 500–544.
- [31] A. V. M. Herz, T. Gollisch, C. K. Machens, Dieter Jaeger (2006), Modeling Single-Neuron Dynamics and Computations: A Balance of Detail and Abstraction, *Science*, 314, 80-85.
- [32] E. M. Izhikevich and G. M. Edelman (2008), Large-scale model of mammalian thalamocortical systems, *Proc. Nat. Acad. Sci.*, vol. 105, 3593 .
- [33] R. Ananthanarayanan, S. K. Esser, H. D. Simon, & D. S. Modha (2009), The cat is out of the bag: Cortical simulations with 109 neurons, 1013 synapses., In *Sc '09: proceedings of the conference on high performance computing networking, storage and analysis* (1–12). New York: ACM.
- [34] Sasaki et al. (2006) “Effect of echo-guided high-intensity focused ultrasound ablation on localized experimental tumors (theriogenology),” *J. Vater. Med. Sci.*, 68, 1069-1074; 中村弘文ら (2012) “超音波 CT を用いた超音波診断治療の基礎研究,” 第 11 回日本超音波治療研究会, 15.
- [35] Liu, Y. et al. (2011) “Numerical study on the shape oscillation of an encapsulated microbubble in ultrasound field,” *Phys. Fluids*, 23, 041904.; ・杉山和靖ら (2012) “超音波音場における膜被包性マイクロバブルの表面安定性の数値解析,” *超音波テクノ*, 24, 90-94.

- [36] A. J. Sadosky, P. B. Kruskal, J. M. Kimmel, J. Ostmeyer, F. B. Neubauer, and J. N. MacLean (2011), Heuristically optimal path scanning for high-speed multiphoton circuit imaging, *J Neurophysiol.*, 106, 1591–1598.
- [37] J. Rosen and G. Brooker (2008), Non-scanning motionless fluorescence three-dimensional holographic microscopy, *nature photonics*, 2, 190-195.
- [38] B. A. Wilt, L. D. Burns, E. T. Wei Ho, K. K. Ghosh, E. A. Mukamel, and M. J. Schnitzer (2009), Advances in Light Microscopy for Neuroscience *Annu Rev Neurosci.*, 32, 435.
- [39] J. T. Vogelstein, A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski (2010), Fast Nonnegative Deconvolution for Spike Train Inference From Population Calcium Imaging, *J. Neurophysiol.*, 104, 3691–3704.
- [40] S. Herlitze, & L. T. Landmesser (2007), New optical tools for controlling neuronal activity. *Curr. Opin. Neurobiol.* 17, 87–94.
- [41] D. Kleinfeld, A. Bharioke, P. Blinder, D.D. Bock, K.L. Briggman, D.B. Chklovskii, W. Denk, M. Helmstaedter, J.P. Kaufhold, W.C. Lee et al. (2011), Large-Scale Automated Histology in the Pursuit of Connectomes, *J. Neurosci.*, 31, 16125–16138.
- [42] Sadler, J.E. (2008) “Von Willebrand factor, ADAMTS13, and thrombotic thrombocytopenic purpura,” *Blood*, 112, 11-18.
- [43] Sugiyama, K. et al. (2012) “A priori modeling of the acoustic boundary layer effect on the heat source in ultrasound,” *J. Biomech. Engrg. Sci.*, 7, 84-101.
- [44] E. M. Izhikevich (2004), Which Model to Use for Cortical Spiking Neurons, *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 15, 5.
- [45] S. Rotter and M. Diesmann (1999), Exact digital simulation of time-invariant linear systems with applications to neuronal modeling. *Biol Cybern.*, 81, 381-402.
- [46] S. Imoto, T. Goto, S. Miyano (2002) Estimation of genetic networks and functional structures between genes by using Bayesian network and nonparametric regression, *Pacific Symposium on Biocomputing*, 7, 175-186.
- [47] Y. Tamada, S. Imoto, H. Araki, M. Nagasaki, C. Print, S. Charnock-Jones, S. Miyano (2011) Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8 (3), 683-697.