

「生命科学的側面」と「物質科学的側面」は相補的であり、この二つを融合することによって、創薬や生体分子を活用したものづくりなど社会的に重要な課題において飛躍的な発展が可能になる。例えば、薬剤のターゲットであるタンパク質の生命活動における働きは生命科学的な研究対象であるが、薬剤とタンパク質の結合は物理化学的相互作用であり、そこでは物質科学的方法論が有効である。今後のスーパーコンピュータの高度化と合わせ、物質科学分野で培われた精度の高い計算方法をタンパク質-薬剤結合解析に応用していくことで、創薬分野でのブレークスルーが期待できる。更には、タンパク質やDNAなどの生体分子のみならず、ウイルスや細胞環境などの超巨大システムの動力学解析が実現し、タンパク質の電子状態や原子レベルの機能構造を解明することが想定される。

生命科学分野、物質科学分野、ものづくり分野の分野横断連携において今後、必要となる計算機性能を下表に示す。

課題	要求性能 (PFLOPS)	要求メモ リバンド 幅 (PB/s)	要求ファ イルI/O 性能 (TB/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・計算手法	問題規模	備考
創薬などMD・ 自由エネルギー計算	1	0.5	0	0.001	0.001	1.2	100,000	430,000	全原子分子動力 学シミュレーション	~10万原子	>16GB/node
高精度創薬	10	1	0.001	1	1	1	100	3,600	薬品とタンパク質 間相互作用の量 子化学計算	フラグメント分子軌 道法で~500残基 程度までの計算を 統計的ゆらぎを含 めた複数サンプル で行う	計算要求は「物質科 学」のフラグメント分子 軌道法のところを参照
バイオデバイ ス設計	100	1	0.001	1	1	1	100	36,000	200-500残基程 度のタンパク質の分 光計算		計算要求は「物質科 学」のフラグメント分子 軌道法のところを参照
細胞環境・ウ ィルス	1000	500	2	4	1.2	48	10	1,700,000	全原子/粗視化分 子力学シミュ レーション~1億粒 子		

※ 本見積もりは、9月末日での見積もりである。未だ精査の余地があり、最終版では、より精度の高い数値を記載する予定である。

3.2 ビッグデータの有効利用

社会的課題の解決を促進する研究や、従来の計算技術の限界を超えた知の探求を行うためには次世代の計算科学技術基盤が必要である。特に近年、大規模数値シミュレーション、衛星・観測データ、個人のゲノム情報、大規模実験施設で得られる実験データなどの量が飛躍的に増加しており、いわゆるビッグデータの効率的な利用技術などの計算科学技術基盤が求められている。ここでは、冒頭に必要な計算科学技術基盤のあり方と高度化について述べ、その後、2つの代表的な連携事例を示す。

(1) 計算科学基盤技術の創出と高度化

実世界で起こる現象は現象自体が複雑であり、その因果関係を正確に把握することは難しい。多くの場合、現実世界のありのままを模擬することはできず、シミュレーションは対象とする実現象（問題）を理想化して計算することになる。したがって、現実世界で起こる現象とシミュレーションする現象との間には、少なからずギャップが存在する。現実

分野連携による新しい科学の創出

ビッグデータの有効利用

計算科学基盤技術の創出と高度化

目標・目的、克服すべき学術的課題

- データ同化、可視化、ビッグデータ、知識処理といった計算科学基盤技術の横断的な統合による進化・高度化
- 実世界とシミュレーション世界との間のギャップ(シミュレーションモデル・境界条件・初期条件と実現象の違い)の克服
- 大量、多様、リアルタイム、そしてファイルI/O負荷の高いビッグデータへの対応、システム化

従来の研究

- 各アプリケーション領域で個別に進化
- 観測ありき、決定論的手法、結果からメタデータを作成・評価するといった課題

分野連携の方策

- データ同化と可視化・データ処理をシミュレーションと融合
- 個別に進化してきたデータ同化技術の領域間での技術の統合化と洗練化
- 大量データの効率的な管理方法の開発やデータの再利用性を高める取組み

大規模計算で実現されること

- 双方向的な現実世界とシミュレーションの融合による予測精度向上
- 確率論的手法の導入による、低確率でも重大な事象のリスク評価を可能に

世界の問題の特徴をうまく再現できるような方程式群の形式を選定し、関連するパラメータを決めるプロセスをモデリングと呼ぶ。モデリングは観測や実験結果に対する洞察から得られた知見を基にしているが、シミュレーションの正しさや精度に影響する非常に重要なプロセスである。

従来、モデリングおよびそれに関わる各種技術は各アプリケーションの分野ごとに個別の技術的進化を遂げてきたが、共通部分も多い。例えば、現実世界の影響をモデリングに適用する技術であるデータ同化は、気象をはじめ、石油掘削、制御、ものづくり、創薬や分子シミュレーションなどの分野でも研究・活用されている。また、結果の解釈の一助となる可視化技術もさまざまな研究領域で利用され、個別の現象に適した表現方法が研究されている。

計算科学基盤技術の創出と高度化において今後、必要となる計算機性能を下表に示す。

課題	要求性能 (PFLOPS)	要求メモリ/バンド幅 (PB/s)	要求ファイルI/O性能 (TB/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・計算手法	問題規模	備考
並列レンダリング	200	60	8	0.8	10	0.5	1	360	ボリュウムレンダリング(レイキャスト、ファイルベース)		
並列レンダリング	200	60	8	2	1	0.5	1	360	ボリュウムレンダリング(In situ)		
データ圧縮	500	25	24	8	10	0.5	1	900	POD圧縮(ファイルベース)		

※ 本見積もりは、9月末日での見積もりである。未だ精査の余地があり、最終版では、より精度の高い数値を記載する予定である。