

3.3 大規模実験施設との連携

3.3.1 X 線自由電子レーザー施設 SACLA 等の大型研究施設との連携

(1) 課題概要

X 線自由電子レーザー (XFEL : X-ray Free Electron Laser) は、波の位相がきれいにそろったレーザーの性質を持つ超高輝度の X 線を発生することのできる光源であり、電子銃、線形加速器、アンジュレータで構成される。XFEL 施設 SACLA (SPring-8 Angstrom Compact free electron Laser) は、第 3 期科学技術基本計画における国家基幹技術として SPring-8 サイトに建設が進められ、2011 年 3 月に完成した。全長 700m という施設規模は、欧米の類似施設と比べて 3 分の 1 以下であり、コスト面での利点に加えて小型化による信頼性の向上と安定運転が期待できる。2011 年 6 月に初の X 線レーザー発振を経て、2012 年 3 月より内外に開かれた共用施設として利用運転を開始した[1]。

XFEL では、以下に記すようにこれまで構造を解くのが難しかった非結晶粒子や微細結晶の構造解析に威力を発揮すると期待されている。その際には、大量のデータ処理が発生するため、HPC との連携の重要性が指摘されている。そこで、この節では、SACLA の生命科学分野への利用研究として期待されている CXDI (Coherent X-ray Diffraction Imaging) [2]について、そのデータ解析における計算科学の重要性と今後の展望について紹介する。また、最近アメリカの XFEL 施設 LCLS (Linac Coherent Light Source) で盛んに行われ始めた SFX (Serial Femto-second Crystallography) [3]について、HPCI における分子レベルシミュレーションとの連携について今後の展望を紹介する。

2012 年に供用が開始された「京」と SACLA は、その連携による新規ナノサイエンスの開拓を現在模索している段階である。その成果は、次世代の HPCI への引き継がれていくと期待される。

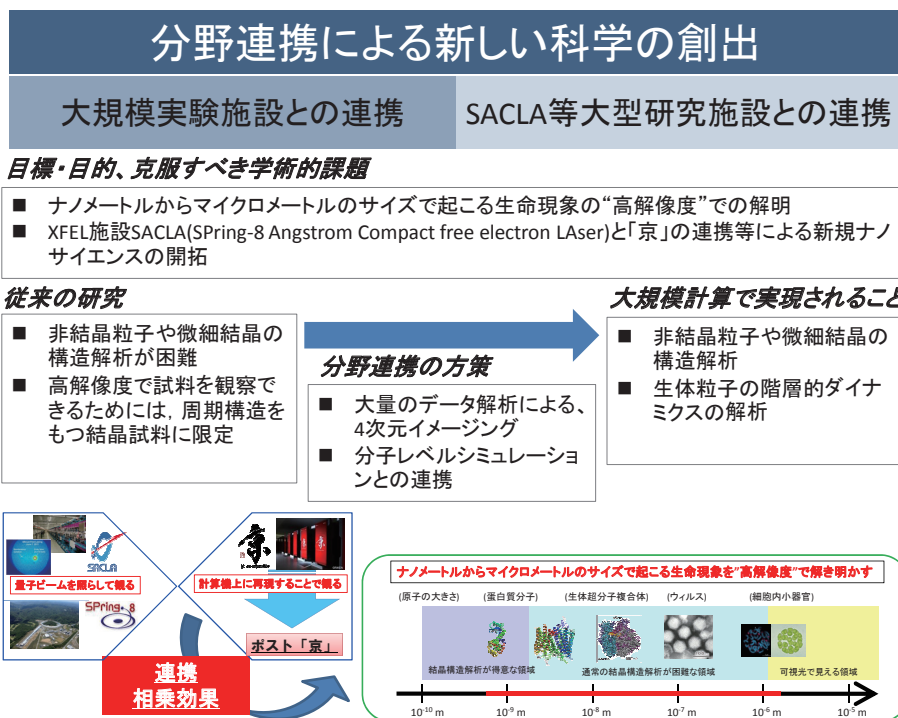


図 3.3.1-1 SACL A 等の大型研究施設との連携

(2) サイエンスの質的変化と長期的目標

(i) 大量データ解析による 4 次元イメージング

電子線に比べて透過力の高い X 線を利用した CXDI は、非結晶粒子の構造研究で期待される手法である。放射光を含む従来の X 線光源では、高解像度で試料を観察できるのは周期構造を持つ結晶試料に限定されていた。CXDI では、試料に位相のそろった X 線を照射し、スペックルと呼ばれる斑点模様状の 2 次元回折強度パターンを高い空間分解能で取得する。この回折強度パターンは、試料の電子密度（構造）のフーリエ変換である構造因子の 2 乗に電子散乱断面積等の因子を乗じたものである。構造因子は干渉波の様態を記述するが、波としての重要な情報である位相が検出時に欠落し、振幅情報のみが得られ、そのままでは元の構造を求めることはできない。この問題に対し CXDI では、高い空間分解能で取得した回折振幅データに位相回復法と呼ばれるアルゴリズム[4]を用いることで、試料の電子密度を決定することが可能である。

2012 年度は、SACL A 利用実験の立ち上げ時期であり、2 次元イメージングが CXDI 実験の主流となっているのが現状であるが、今後は 3 次元構造解析へと発展していくと期待される。

XFEL は、非常に高輝度であるため、一発の照射で試料を破壊してしまうという問題がある。XFEL の照射で得られる回折パターンは、XFEL のパルス幅である数十フェムト秒間隔で、試料をある方向から見た 2 次元のスナップショットである。このため、試料の 3 次元構造を得るためには試料を破壊しながら多数の回折パターンを取得することになるが、試料の向きを実験的に決めることはできないため、計算科学的に試料の向きを解析しなければならない。最大で毎秒 60 回の X 線パルスが生成される SACL A では、1 日に最大で約 500 万枚の回折パターンが取得される。イメージングの解像度を上げるためには、あらゆる向きの粒子からの回折パター

ンが必要である。このような大量実験データを効率よく解析し、試料の高解像度 3 次元イメージングを実現するためには、HPCI との効率的な連携が不可欠となる。

ここまでは生体粒子の静的な構造を仮定した議論を進めてきたが、生体粒子は遺伝子に規定された固有の立体構造をとり、その周りを揺らぐことで機能を発揮する。1 日で最大 500 万枚取得される大量実験データには、当然のことながらこの粒子の動きの情報が含まれる場合が多い。具体的には、生体粒子にはエネルギー的に準安定な多数の構造状態があり、構造間を熱揺らぎによって遷移しているが、そのさまざまな準安定構造からの回折パターンが含まれている。この大量データを生体粒子の“かたち”に基づき分類することができれば、3 次元構造の動態を記述するべく時間軸を含めた 4 次元イメージングが実現できる可能性がある。そのためには、多変量解析など統計科学的計算手法を駆使する必要があるだろう。例えば、生体粒子のダイナミクスを多次元空間内の超平面である「manifold」に対応させて解析する手法が近年提案された[5]が、実際の実験データを用いた応用・実証研究が期待される。3 次元イメージングと比べて次元が一つ増えることから、解像度を同じに設定した場合、3 次元イメージングによる静的構造解析と比べて、データ解析に必要な回折パターンの数も飛躍的に増大する。良質の回折パターンを効率よく大量に取得するための実験システムの構築と併せて、次世代の HPCI とのスムーズな連携を見据えた解析システムの構築が必要となるだろう。

(ii) 分子レベルシミュレーションとの連携

SACLA の XFEL 光は非常に高輝度であるが、その強度（現状: $\sim 10^{11}$ 光子/パルス）をもってしてもナノメートルサイズのタンパク質一分子からの回折パターンを取得するには不十分である。今後期待される加速器およびビームラインの高度化を考慮しても、SACLA 利用による CXDI のターゲットは、数十ナノメートルから 1 マイクロメートルのサイズの生体粒子となるだろう。到達可能な解像度はナノメートル程度と予想される。一方、化学反応によって媒介される生命現象を真に理解するためには、0.1 ナノメートル程度の解像度が必要となる。SACLA による CXDI の不足分に補うためには、電子顕微鏡解析、核磁気共鳴法、通常の結晶構造解析などさまざまな解析手法を相補的に利用することが重要となってくるだろう。

ここでは分子レベルシミュレーションと CXDI の連携について考えてみたい。粗視化 MD の今後約 10 年間のターゲットは、10 ナノメートルから 1 マイクロメートル規模となり、SACLA の CXDI のそれとほぼ一致する。今後、SACLA のイメージング結果に基づく粗視化 MD の解析手法が標準化できれば、ナノスケール解像度での染色体動態観察などの新規サイエンスの開拓が期待できる。また、粗視化 MD を援用して CXDI 実験データを解析することで生体粒子モデルを精密化し、更には結晶構造解析などの実験結果を援用しながら全原子 MD にスムーズに接続することで、原子モデルの構造精密化を実現できる可能性もある。このような連携により、10 ナノメートルから 1 マイクロメートルサイズの生体粒子の階層的ダイナミクスの解析が進み、この空間スケールで起こる生命現象の理解につながると期待される。

XFEL による生命科学研究には、CXDI とは別のアプローチとしてマイクロメートルスケールの微小タンパク質結晶を用いる SFX がある。生命現象を理解するうえで重要であり、また、創薬における重要な解析ターゲットである膜タンパク質などの試料では、その結晶化が困難であり、数マイクロメートル以下の微小結晶しか得られない場合がある。超高輝度の XFEL を利

用すると、このような結晶からの回折光を検出することが可能となる。また、数十フェムト秒程度のパルス幅を持つ XFEL の照射ごとに「新鮮な」状態の結晶を用いることにより、放射光で問題となっていたラジカル等に起因する X 線照射損傷がない状態の解析が可能となる。SFX で得られる膜タンパク質の結晶構造は、創薬 MD の重要なターゲットとなるだろう。また、SFX に可視光レーザーに状態を励起するポンププローブ実験を組み合わせると、光化学反応を担う生体超分子の数十フェムト秒からピコ秒の時間スケールの時分割構造解析が可能となる。このような SFX 実験結果に基づき HPCI 利用による QM/MM シミュレーションを実行することで、生命現象における光化学反応のメカニズム解明につながると期待される。

(3) コミュニティからの意見

SACLA の利用実験が順調に立ち上がりつつあり、取得実験データ量が加速的に増大してきている。今後は HPCI を利用して大量実験データを解析するためのコミュニティが形成されるものと期待される。その試みとして、XFEL 重点戦略研究課題についての委託事業「SACLA における低温 X 線回折イメージング実験の展開と標準化」第二回会合（2013 年 8 月 8～9 日）の中で議論された SACLA と HPCI の今後の連携に関する意見を以下にまとめる。

- SACLA により細胞内の物質分布情報が得られれば、リアリスティックな細胞モデリングが可能となり、大規模分子シミュレーションによる細胞内環境の研究が加速される。連携のターゲットとして最適であり、将来ぜひとも実現させたい。
- 生命科学研究では創薬と結びつけた出口論が多くなされる傾向にある。しかし、これは構造解析の 2 次的産物でしかない。SACLA の利用実験の立ち上げフェーズである現状では、SACLA を用いた「生命の理解」にどのようにアプローチするかを確実に積み上げていくことが不可欠である。更に SACLA に限らず、SPring-8 における CXDI・結晶構造解析や電子顕微鏡など細胞内の生命活動をさまざまな空間階層で見渡すことのできる実験と HPCI 利用による大規模計算とを組み合わせた生命科学研究を推進していきたい。
- SACLA や SPring-8 などのさまざまな実験とマルチスケールシミュレーションの連携は、分子力場の改良など計算科学技術の向上にもつながる。「京」のような汎用スパコンでは、D.E. Shaw Research の anton のような MD 専用計算機にはかなわない面があるのは事実である。スパコンの計算性能に頼るだけでなく、計算手法を工夫することにより独創的な研究を進めていきたい。例えば、HPCI 利用による蛋白質と低分子の結合エネルギー計算技術の向上なしには in Silico 創薬の実現は難しいのではなかろうか？
- 日本の計算科学分野では高度なコーディングができる研究者・技術者が減少している。同様に、構造解析分野でも、回折・散乱の物理学を理解し、装置を組み立てたり解析ソフトウェア作ることのできる研究者・技術者が減少傾向にある。一方で、利用するだけの“ユーザ”は多い。今後の堅実な連携に向けては、ソフトウェア開発や技術開発を担う人材を中期的・長期的視野に立って育成することがきわめて重要である。

(4) 必要な計算機資源

SACLA による CXDI 実験を成功させるためには、解析に必要な回折パターンが十分に取得できていることを実験中に確認しながら実験を進める必要がある。例えば 3 次元イメージングでは、目的解像度を実現するのに十分な良質回折パターンが必要な向きを網羅するのに十分な枚数だけ取得できているか実験中に確認できなければ、その実験のスロットを無駄にしかねない。SACLA の利用実験の立ち上げ時期で、実験・解析手法が確立していない現状ではかなり粗い見積りになるが、 10^6 枚の回折パターンから 3 次元イメージングを目的とした取得セットの質を判断するために、1PFLOPS の実行性能を用いて約 3 時間を要する。4 次元イメージングでは、更に数桁高い計算能力が必要となるであろう。また、4 次元イメージングのための統計科学的計算では、回折パターン間の距離行列を扱う場合があり、その行列サイズは解析する枚数の 2 乗である。このような大規模行列を効率よく計算するためのシステム（ハード・ソフト）が望まれる。

課題	要求性能 (PFLOPS)	要求メモリ (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
大量実験データ解析による 4 次元イメージング	2	0	0.000001	0.000001	2.8E-11	1E+12		構造分類、3次元構造構築、200時間軸推定のための統計処理	10^7 イメージ $\times 10^5$ イメージの相関計算で、合計 10^{12} 回の相関計算。1回の相関計算を 1 ケースとした。1 イメージあたりのサイズは数百 MB	通信、ファイル I/O は引き続き調査必要。特にファイル I/O がボトルネックとなる可能性あり。京ではローカリティを考慮した I/O により最適化している。
実験解析結果に基づく動的構造モデリング	490	49	0.2	1.2	48	10	850000	全原子/粗視化分子力学シミュレーション	~ 1 億粒子	B/F=0.1

参考文献

- [1] T. Ishikawa, et al., Nature Photon. 6, 540 (2012)
- [2] J. Miao, et al., Anu. Rev. Phys. Chem. 59, 387 (2008)
- [3] S. Boutet, et al., Science 337, 362 (2011)
- [4] J. R. Fienup, Appl. Opt. 21, 2758 (1982)
- [5] P. Schwander, et al., New J. Phys. 12, 035007 (2010)