

小角散乱データからのモデリングに関する出版ガイドラインを読み解く

藤澤哲郎

岐阜大学工学部化学・生命工学科 〒501-1193 岐阜市柳戸 1-1

要旨

近年、ハード及びソフトの発展に伴い、蛋白質溶液からの溶液散乱、小角散乱データから三次元蛋白質構造モデルを作成することが可能となり、非常に多くの論文が小角散乱データに基づいた分子構造モデルを載せるに至った。並行して、蛋白質小角散乱に関する新たな測定法、解析法も活発に報告されている。こうした流れの中で、*ab initio* モデリングや Rigid body モデリングといった三次元座標を持つ構造モデルを生成するモデリング技法を念頭におき、構造モデルの妥当性を判断するためには、どのような小角散乱実験、解析手順が必要で、出版に際して何を記述すべきか、意見を集約する必要が生じて来た。2011年、国際結晶学会の小角散乱部会 (IUCr Commission on Small-Angle Scattering) が、小角散乱データからの蛋白質構造モデリングに関する出版ガイドラインを発表した。また、2013年には小角散乱データの正確さに関する Nature 誌の論文がアメリカのグループから発表され、小角散乱データからのモデリングの標準化に関する議論が高まっている。世界的に見ると生物利用の小角ユーザーは、放射光に特に集中しており、研究成果を維持するためにも、ビームライン運営者側、ユーザー共に出版ガイドラインの概要と理論的背景、方向性を周知させる必要があろう。本稿では、ガイドラインに盛り込まれているポイントを概説し、さらに、最近発表された小角散乱解析に関する標準化の話題にもふれてみたい。

1. 出版ガイドライン策定の背景

過去20年、小角散乱、特に、蛋白質溶液からの小角散乱法を用いた蛋白質複合体のモデリングが放射光を中心に急速に利用されるようになった。筆者自身も SPring-8 で世界初の生体高分子専用の小角散乱ビームライン BL45XU-SAXS を建設し、その恩恵に少なからずあずかった人間である¹⁾。現在、測定精度の向上と EMBL の Svergun 博士らが開発した汎用のデータ解析ソフト (Table 1²⁾) の普及に伴い、小角散乱は構造生物学のメジャーな手法としての地位を確立するに至っている。それに伴い世界的に数多くの生体高分子専用の小角散乱ビームラインが建設され、実験だけでなく理論解析も含むおびただしい数の論文が出版されるに至った³⁾。

まずは出版ガイドラインについて述べる前に、このガイドラインの対象となるモデリングとはどのようなものかを簡単に説明する。詳しくは本学会誌での総説等を参考にさせていただきたい⁴⁾。小角散乱のモデリングの基本は、実験散乱曲線と理論散乱曲線が一致するように、非常にたくさんのモデル群の中から最適モデルを選別することにある。現在、小角散乱のモデリング技法は大別して2つに分類できる。*ab initio* モデリング (ビーズモデリング) においては、最初に探索球の中にビーズ(球)を格子状に配置し、実験散乱曲線に理論が一致するように、スタートモデル無しに (*ab initio*)、ビーズを削り変形する。この過程は非

Table 1 Frequently used programs of AT SAS program package in SAS analyses²⁾

プログラム名	機能
PRIMUS PRIMUS/QT AUTORG	ブランク溶媒散乱の差分 ギニエプロット、Kratky プロットの表示、 $P(r)$ の計算
GNOM	間接フーリエ変換法による $P(r)$ の計算
AUTOPOROD	Porod 体積及び $P(r)$ の自動計算
DAMMIF DAMMIN	<i>Ab initio</i> モデリング (ビーズモデル) の計算
DAMAVER SUBCOMB DAMCLUST	3次元構造モデルの形状に基づく平均化、重ね合わせ、統計評価
CRYSOL	原子座標からの $I(q)$ の計算
SASREF	Rigid body モデリング (single fit)
EOM	Rigid body モデリング (ensemble fit)

常に確率的で、一つの実験散乱曲線に対して複数のモデルが算出される。通常、10ないしは12回計算を繰り返し、得られたモデル形状の統計的チェックと共に平均化を行い、一つの平均構造を得る。その後、得られた平均構造と高分解能モデルとの比較を行い生物学的な解釈を行う (Fig. 1)。一方、Fig. 2 で示される Rigid body モデリングとは、蛋白質の一部を既知の高分解能構造 (鋳型 PDB) やホモロジーモデルを使い、その部分の構造は変えずに剛

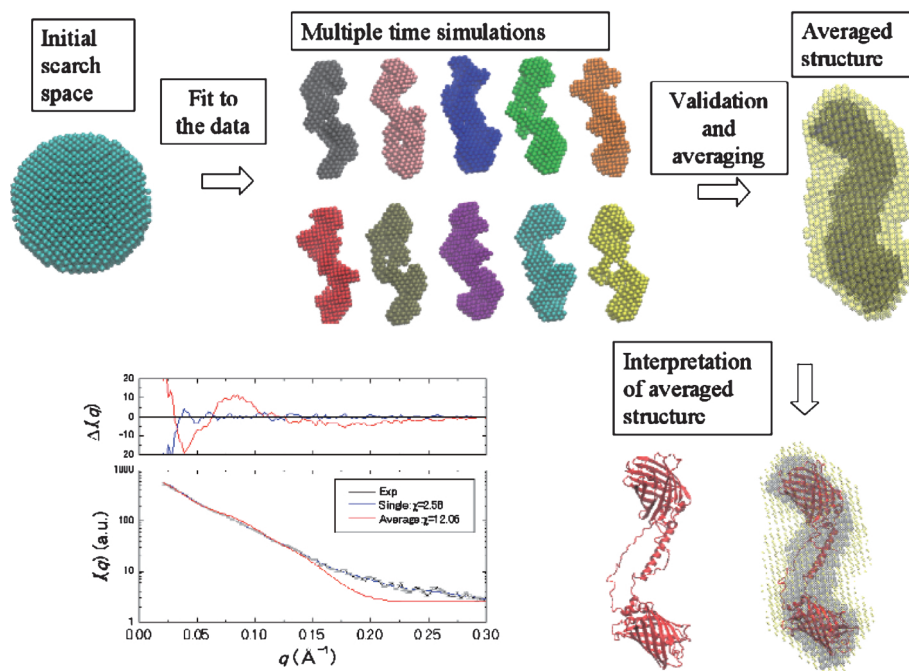


Fig. 1 (Color online) Analysis flow of bead modeling method. Densely packed spheres in the search space with diameter of D_{\max} is modified so as to fit the experimental data ($\chi=2.58$). The process of simulation is so stochastic that the models of each run deviate with each other. After statistical inspection, the ten models are averaged into an averaged structure (grey beads). Yellow beads represent the spatial spread of ten molecules. Interpretation of the averaged map is usually done by manual matching of high resolution model (red). Here, the high resolution model is highly consistent with the averaged structure, however, its fit to the data is rather poor ($\chi=12.06$). Experimental data were taken from Arai *et al.*, 2004⁵). *Ab initio* modeling was done by DAMMIN and averaged by DAMAVER²). The models are drawn by VMD. No distortion corrections has been made for the detector images.

体として扱い、それらの配置や方向を動かすことにより実験値にフィットさせる。一つの平均モデル構造を使って実験値をフィットさせる場合もあれば (single fit), 複数のモデル集団 (アンサンブル) として実験値を表すこともある (ensemble fit)。single fit の場合、ビーズモデルの平均構造と本質的に一緒に、分子の平均像を表現するが、ensemble fit の場合、溶液中における複数の構造種 (Fig. 2 では上側のドメインの配置が違う3つのモデル分子の集団) として実験値にフィットする。Fig. 2 のような分子の場合、溶液中では2つのドメインの相対配置はリンカー部の自由度によって揺らいでいると考えられ、一つの固定した平均構造を想定するよりは実態に近いはずである。このように、ensemble fit は分子構造のダイナミクスに関して情報が得られることになる。モデル選別の基本となるデータとの一致度は χ^2 値が用いられ、Table 1 のソフトウェアの χ^2 は、

$$\chi^2 = \frac{1}{N-1} \sum_i \left[\frac{I(q_i)_{\text{exp}} - I(q_i)_{\text{calc}}}{\sigma(q_i)} \right]^2 \quad (1)$$

で定義される²⁾。ここで N はデータ点数、 σ は実験誤差である。ensemble fit で抽出された複数の構造は単に計算機

上のバーチャルなものではなく、NMR (核磁気共鳴法) や単粒子 FRET (蛍光エネルギー移動) 測定と組み合わせることによって生物学的に意味があることが数多くの例で実証され、生体機能時の新しい構造体などが続々と発表されている^{3,6)}。

これら小角散乱モデリング解析は大変魅力的であるが、Figs. 1, 2 から推察できるようにデータに対する質 (サンプルの純度, 単分散性, 計測処理の妥当性など) に特に影響されるので、従来からの判断基準は適用できない。また、大量に小角散乱データを取得するハイスループト型のビームラインも登場し、解析の自動処理を進める上でも測定、解析に関する共通の判断基準が必要となっている⁷⁾。残念ながら、生体高分子モデリングに関する明確なガイドライン、標準化は現在のところ確立されておらず、本稿で紹介するガイドラインは発展途上であることを了承していただきたい。

小角散乱の最大の利点は、結晶化や急速凍結保存などのサンプル状態に関する制約がないことにある。反面、小角散乱はいかなる質のデータに対しても Figs. 1, 2 のような解析を機械的に行うことができ、また仮に質の高いデータであっても、その解析結果は手順によって少なからず依存する⁴⁾。小角散乱の構造モデルが提示される場合、必要な

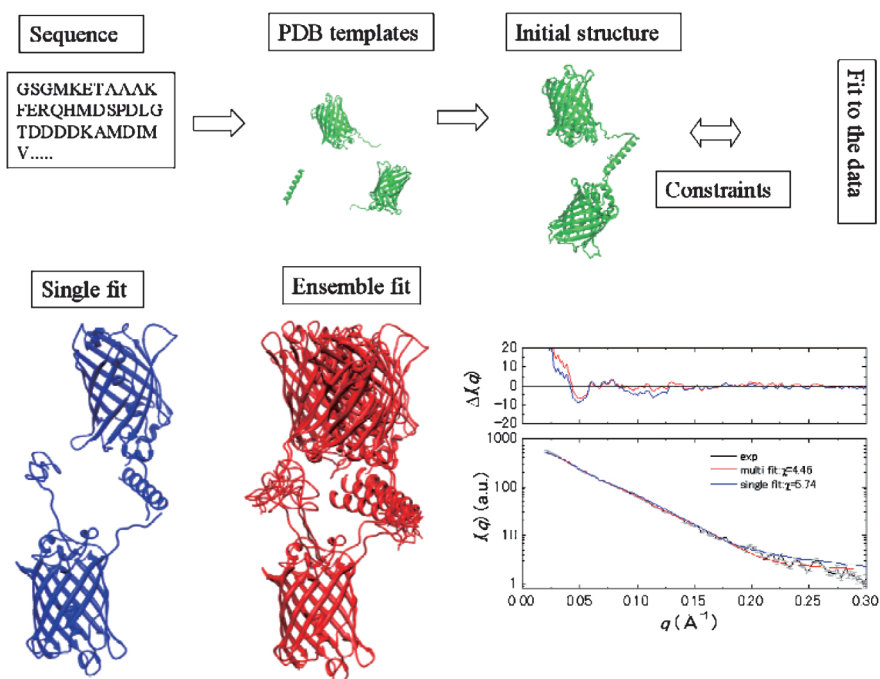


Fig. 2 (Color online) Analysis flow of rigid body modeling method. A rigid body model consists of pieces of high resolution structures (PDB templates) and linkers or loops that connect between them. First, based on the amino acid sequence together with PDB templates, an initial structure is created. Then, the structure of the model is modified to fit the data with some constraints. In single fit (blue), a single molecule tries to fit the data, which usually does not give sufficient fit to the data ($\chi = 5.74$). In ensemble fit (red), multiple structures (three conformations in this case) with different fractions, i.e., an ensemble of structures, are used to represent the data ($\chi = 4.46$). Single fit and multiple fit calculations are done by BUNCH and EOM², respectively. Experimental data were taken from Arai *et al.*, 2004⁵). No distortion corrections has been made for the detector images, which seems one of the reasons for the high χ values.

情報が記載されないとそのモデルの妥当性を判断することができない。このような論文における実験、解析手順の必要十分な記述の欠如の問題は小角散乱法においてのみではない。急速に技術が発達している構造生物学では、他の分野、結晶構造解析法⁸、電子顕微鏡法⁹においても問題となっており、2011年に相次いでガイドラインを策定している。小角散乱の分野においても、2011年、放射光小角散乱で小角散乱解析の第一人者、EMBLのSvergun博士、中性子小角散乱のシドニー大Trehwella教授を中心とする国際結晶学会小角散乱部会 (IUCr Commission on Small-Angle Scattering) が、IUCr関連の学会誌における生体高分子小角散乱データの出版ガイドラインを発表した¹⁰。翌2012年には、本ガイドラインの解説論文がJacquesらにより発表された⁵。論文の冒頭で、「本ガイドライン策定の目的は、あくまで出版基準ではなく、読者が著者の主張の是非を判断できるような方式を決めることである⁷ (著者訳)」としているが、文字通り、出版基準に直結することは間違いないであろう。

ガイドラインの構成は、出版に際して記述すべき必須事項、推奨事項が、サンプルの品質、データ収集、データ検証及び表示、モデリングの実験、解析手順の各項目にわたって記載されている。小角散乱全般を念頭におき策定され

ているので、放射光ユーザーには重要でない部分も含むが、本稿ではその部分は触れない。ガイドラインの内容には、Rigid bodyモデリングを念頭におき、従来の掲載されている論文とは違ったスタイル記載方法を要求している部分もある。上記に述べたように小角散乱における測定と解析の標準化は、この1、2年時々刻々と変化している。特に最近の研究の中で重要なのは、Nature誌に載ったRamboとTainerによる小角散乱データとモデリングに関する新たな評価方法¹¹で、まだ本稿執筆時点(2013年6月時点)ではガイドラインには反映されていない。

本稿は、放射光の利用者、特に、分野外の生物研究者、院生を念頭に、日本語で手軽に読めるよう、最初に、小角散乱出版ガイドラインを、次に、RamboとTainerにより提案された評価方法の概略と解説を試みるものである。

2. 出版ガイドラインの概略

2.1 サンプルの品質：単分散性を何で評価すればよいのか？

小角散乱による三次元モデル構造は、蛋白質濃度が無限に希釈された、単分散蛋白質溶液からの散乱を前提としている。純度の高い蛋白質溶液であることは当然として、そ

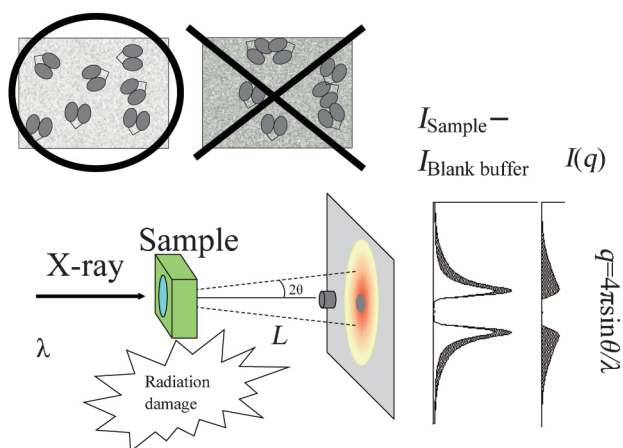


Fig. 3 (Color online) Sample quality and data collection. The incident x-ray is irradiated on the sample (protein solution and blank buffer, 10–30 μL) under identical condition. The isotropically scattered images (I_{Sample} and $I_{\text{Blank buffer}}$) are collected by two-dimensional detector and I_{Sample} is subtracted by $I_{\text{Blank buffer}}$. The subtracted image is azimuthally averaged into one-dimensional data, $I(q)$. In the figure, λ ; wave length, L ; camera length, and 2θ ; scattering angle. Sample should be highly purified and mono-dispersed; aggregation state should be single. Blank buffer is also very important for data quality. In synchrotron experiment, radiation damage often occurs, which results in aggregation.

れが単分散、すなわち、会合状態が一定であることが必要である。また、会合してなくても、蛋白質間に強い粒子間相互作用が存在すると、散乱曲線の小角側が歪んでくる¹²⁾。サンプルの品質とは、蛋白質の純度だけでなく、粒子間相互作用の特性、さらには、ブランク溶媒の調整に及ぶ。

サンプルの品質に関する必須記載事項の大半は、精製手順、溶媒条件、最終純度の推定と推定方法、試料濃度と決定方法(吸光係数を含む)など一般の生化学雑誌のフォーマットと特に変わらない。最終純度は SDS ポリアクリルアミド電気泳動で99%以上となっているが、蛋白質の物性研究では一般的な純度である。小角散乱測定において試料濃度は散乱曲線の補正や分子量校正の際、直接的に影響を与えるので可能な限り正確に測定しなければならない。通常は280 nm における吸光度を用いるが、吸光係数のみならず、核酸の混在やジチオトレイトール (DTT) などの紫外光吸収物質の存在により誤差が生じるので溶媒条件の記載も必要となってくる。見落としがちであるが、ブランク溶媒の調整はサンプル調整と同じくらい重要である。蛋白質溶液の小角散乱では、蛋白質溶液からの散乱とブランク溶媒からの散乱との差分を行う (Fig. 3)。高純度のサンプルが単分散であっても、蛋白質溶液の溶媒成分とブランク溶媒の成分が透析などで化学的に平衡に達しており、同一でなければ、小角部では言うに及ばず、広角部を使う分析でも影響があり、使い物にならない。平衡に達していない多くの場合は、ブランク溶媒の調整法にある。透析ま

たはゲルろ過液が推奨され、濃縮ろ過液などは避ける。意外な点は、蛋白質の配列も補酵素、蛋白質修飾も含めて必須事項として要求されている点である。これは散乱に大きく影響する、蛋白質の分子量と偏比容(平均電子密度の逆数)や静電相互作用などを推測する材料となるからである。

ガイドラインでは単分散性を保証する評価の記載が推奨されている。具体的には、動的光散乱法 (DLS)、角度分散光散乱法 (MALLS)、サイズ排除ゲルろ過—角度分散光散乱法 (SEC-MALLS) による評価である¹³⁾。光散乱は小角散乱と比較して凝集、会合などにより敏感なので小角散乱実験前の予備実験としては最適である。しかし、これらは高価な機械であり、筆者の経験から単に会合、凝集の予備チェックであるならば、サイズ排除ゲルろ過法や Native ゲル電気泳動法も十分活用できると考えている¹⁴⁾。単分散性に関しては、測定の直前のチェックでないとダメだという考えが急速に広がっており、後述するが、多くの小角散乱のビームラインでは SEC-MALLS と屈折計がサンプルセルに直結した装置を備えている³⁾。

2.2 データ収集：放射線損傷防止処置

放射光における小角散乱実験の詳細は、基本的にビームラインの論文を参照して波長、カメラ長、サンプルセル環境(セル光路長、温度を含む)、測定時間を記載する。 $I(0)$ などは絶対強度測定(単位は cm^{-1})か標準サンプルに対する相対値(任意単位)でスケールするよう要求されているので、 $I(0)$ の評価のための標準サンプルの測定も必要である。

必須事項として、放射線損傷を監視し、確実に低減する方法 (DTT, トリス (2-カルボキシエチル) ホスフィン塩酸 (TCEP), アスコルビン酸などのラジカルスカベンジャーの添加, サンプル流動, 時分割フレーム測定の解析結果)を取ったかどうかなどの記述が挙げられている。時分割フレーム測定とは、短時間の測定を連続して行うことである。放射線損傷があれば分子の会合が起こり強度が上昇するので、放射線損傷が起っていないデータを選別することができる。PILATUS¹⁵⁾のような時分割測定可能な検出器が無いビームラインは、減衰板を挿入し光量を減らして測定することになるので非常に不利になるであろう。ユーザーとしては検出器に必要な投資をしていただくよう施設側の努力をお願いするしかない。

推奨事項として、二次元の蛋白質溶液とブランク溶媒の散乱像から、実際にモデリングに使用する一次元の散乱強度 $I(q)$ ($q = 4\pi \sin \theta / \lambda$, 2θ ; 散乱角, λ ; 波長)に変換する初期データ処理ソフト名も記載することが求められている。個人で算出ソフトを書いている場合は、処理プロセスを別途記載する必要がある。この初期データ処理の際にサンプルの吸収、透過率、検出器の感受性、一様性の補正、絶対尺度への変換方法などなされたのなら記載すべきである。

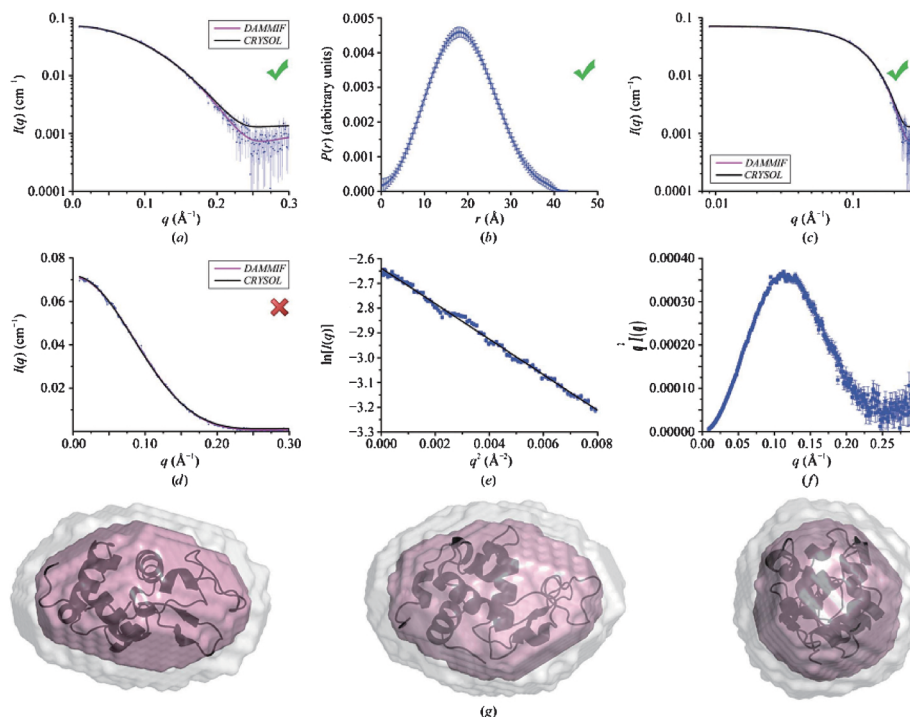


Fig. 4 (Color online) Example of representation of data and model (Figure modified from Jacques *et al.*, 2012⁷⁾). Representation of experimental data (a–f) and the final model (g). The vertical axis of (f) in the original paper was q^2 , which was an erratum.

2.3 散乱データの表示と検証

小角散乱データは、構造解析に適さない、凝集体が混在するようなサンプルや不適切に収集されたデータに対してもデータ解析操作を行うことができる。取得されたデータが意味があるかどうか、データ解析を行う際、各ステップ毎に数値とグラフで厳密に検証しなくてはならない。小角散乱解析の基本は、異なる特定の散乱領域 q に焦点をあて構造情報を導出する。解析の各ステップで小角パラメータに基づき定量的に判断するだけでなく、様々な形で散乱データ $I(q)$ を表示して判断するといった定性的な手法も用いられる。後者の場合、どのような形で $I(q)$ を表示するかによってデータの質に対する判断が左右されることもあるであろう。したがって、ガイドラインでは $I(q)$ の表示形式に関してもリゾチームを例として制約を設けている (Fig. 4(a)–(f))。

必須事項として挙げられているのは、散乱曲線 $I(q)$ 対 q だけでなく動径分布関数曲線 $P(r)$ 対 r (Fig. 4(b)) も本原稿に記載されるべきとされている。さらに球状蛋白質の場合、ギニエプロット $\ln(I(q))$ 対 q^2 (Fig. 4(e)) が必要である。

2.3.1 散乱曲線 $I(q)$ の表示

ガイドラインでは、散乱曲線 $I(q)$ は、線形 x 軸—対数 y 軸 (Fig. 4(a)) ないしは対数 x 軸—対数 y 軸 (Fig. 4(c)) のいずれかとして表示するよう記載されている。前者は、散乱曲線全体を後者はサンプルの多分散性の指標となる。

粒子間干渉関数や会合による小角部の立ち上がりをマスクすることは厳禁である。多くの場合、線形 x 軸—対数 y 軸 (Fig. 4(a)) で $I(q)$ は表示される。モデリングは $I(q)$ 全体を使うので線形 x 軸—対数 y 軸 (Fig. 4(a)) は、 $I(q)$ の特徴がよく分かってよい。反面、Fig. 4(a) では小角部が視覚的に圧縮されるので q の小さな領域の評価基準が必要となる。散乱曲線 $I(q)$ の q の小さい領域においては、近似式、

$$\ln(I(q)) = \ln(I(0)) - \frac{Rg^2}{3} q^2$$

$$Rg^2 = \frac{\int_{V_r} \Delta\rho(\vec{r}) r^2 dv}{\int_{V_r} \Delta\rho(\vec{r}) dv} \quad (2)$$

が成立する。データに対しギニエプロット、 $\ln(I(q))$ 対 q^2 を行えば、(2)式よりデータ点が直線で近似ができ、y-切片から $I(0)$ が、傾きから回転半径 Rg が得られる (Fig. 4(e))。ここで回転半径 Rg は、電子密度の2次モーメントであり、構造の違いを簡便に判別できる最も重要なパラメータである。ギニエプロットでは、単分散性がデータの直線性として反映され、粒子間相互作用や会合の有無が判別しやすい。最近の雑誌ではスペースに余裕がないのでギニエプロットを割愛している論文が多い。モデリングをしていると小角部の占めるデータ点数が少ないので特に必要

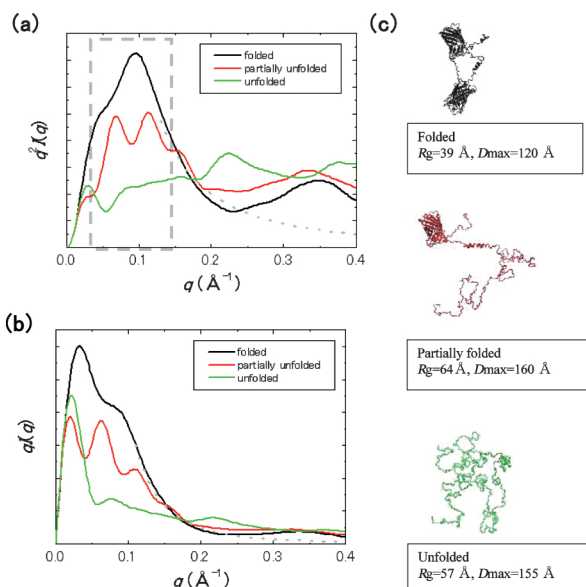


Fig. 5 (Color online) Detection of conformational flexibility shown in different plots. Theoretical calculations of a two-domain molecule with different degree of structural order are shown in $q^2 I(q)$ vs. q plot, Kratky plot (a) and $q I(q)$ vs. q plot (b). Unfolded structures are generated by RANCH²⁾ and their $I(q)$'s are calculated by CRYSOLO²⁾. The q region in the boxed area (a) reflects the structural order: more disordered in the structure, more flattened boxed area in the plot. The color of plots corresponds to each molecule shape in (c): folded (black), partially folded (red), and completely unfolded (green), respectively. Both R_g and D_{max} of each molecule are provided as a reference. Dotted lines are functions of q^{-2} (a) and q^{-3} (b) according to Porod's law¹⁸⁾, which are used for extrapolation to $q = \infty$.

が無いと思われているのであろう。

蛋白質が折りたたんでいるか、柔軟な領域を持つかを示す際には、Kratkyプロット $q^2 I(q)$ 対 q (Fig. 5(a)) が非常に便利である¹⁵⁾。Fig. 5では、同じアミノ酸配列の蛋白質で構造状態が変わると散乱曲線 $I(q)$ にどのように反映されるかを示した。かなり伸びた形状の分子であるが、折りたたんでいる天然状態と一部分構造が崩れている状態、完全に構造が崩れた状態の3つの状態で、Kratkyプロット、特に、四角で囲んだ部分で大きく変わっている (Fig. 5(a))。また、完全に変性するとKratkyプロットでは平らになり、 R_g だけではわからない蛋白質の折りたたみ状態がよく分かる。後述するPorod体積 V_p を計算する際には、 $q^2 I(q)$ を q が0から ∞ まで積分しなくてはならない。現実的には、ある範囲の q から $q^2 I(q)$ を q^{-2} の関数 (Fig. 5(a)における点線) で近似し、外そうを行い積分する。変性した蛋白質 (Fig. 5(a)) では明らかに外そうの誤差が大きくなってしまふ。したがって、Porod体積計算が妥当であるためには蛋白質が折りたたんだ状態でなくてはならず、そのことを示すためにKratkyプロットが必要であるとされている。Kratkyプロットは、散乱強度 $I(q)$ の変化を示す際にも有用である。しかし、今回のガイドライ

ンはあくまで「モデリング」の検証が中心なので必須事項とはされていない。

2.3.2 動径分布関数 $P(r)$ 関数の表示

動径分布関数 $P(r)$ 関数 (Fig. 4(b)) は散乱曲線のフーリエ変換によって得られ、散乱体内の2点間距離分布を示す関数である。動径分布関数 $P(r)$ は、実空間での $I(q)$ の表示となり、データの質を検証する際に大きな役割を担う。なぜなら、動径分布関数 $P(r)$ は単調な減少関数である $I(q)$ とは違い、理論的に以下の特徴を持つからである。

- 動径分布関数 $P(r)$ は、 r がある値 (最大長 D_{max}) 以上で0になる。
- 動径分布関数 $P(r)$ の r が原点近傍 $r \approx 0$ あるいは最大長付近 $r \approx D_{max}$ で平滑的に $P(r) = 0$ に漸近する。
- 動径分布関数 $P(r)$ は一般的に平滑な関数である。

動径分布関数 $P(r)$ が上の特徴を満たさないときはどのようなことが考えられるのだろうか。 $r \approx 0$ で $P(r)$ が0でない場合は、ブランク溶媒の差分がうまくいっておらず、 $r \approx D_{max}$ で $P(r)$ が0でない場合は、会合、凝集体が混在している可能性がある。もし $P(r)$ 関数が平滑でなければ著者は何らかの説明を入れる必要があろう⁵⁾。

動径分布関数 $P(r)$ は、Fig. 1のようなDAMMINやDAMMIFなどの *ab initio* モデリング^{2,4)} において利用される。特に解析に影響するのが D_{max} である。モデリングにおいて、探索空間の大きさがモデルの形に影響するからである。動径分布関数 $P(r)$ の質は(2)式からの小角散乱パラメータ、 R_g 、 $I(0)$ と動径分布関数 $P(r)$ から導出される量とが一致しているかで定量的に判断できるので両者とも記載する^{4,7)} (Table 2)。データの質、 $I(q)$ の範囲、散乱体の形によっては、 $I(q)$ では見かけよくフィットしているが、小角散乱パラメータがずれていることがあるからである。動径分布関数 $P(r)$ の提示は不可欠なのである。

2.3.3 蛋白質濃度の無限希釈外そう

表示以外の必須事項として、濃度依存性会合、粒子間相互作用の可能性を排除するため複数の蛋白質濃度でデータを測定し、散乱曲線 $I(q)$ の蛋白質ゼロ濃度への外そうを行わなくてはならない。蛋白質濃度を変えても R_g 、 $I(0)$ が変わらなければその事実を記載する。複数点の蛋白質濃度でデータを測定することは粒子間相互作用を排除するだけでなく、データの信頼性にもつながる。

最後に、小角パラメータなどの実験データ全てに (計測統計に基づいた) エラーバーを表示する必要がある。GNOM²⁾ などの $P(r)$ 計算ソフトは D_{max} の誤差を表示しないが、入力 D_{max} を変えてみて R_g の値が(2)式からの値と誤差内で一致する範囲に基づき D_{max} 値の広がりを出すことは有用である⁴⁾。

2.4 モデリング

基本的に、情報量の少ない次元散乱曲線から三次元構造を計算することはオーバーパラメータ問題である。三次

Table 2 Recommended representation for data-collection and scattering-derived parameters (Table was modified from Jacques *et al.*, 2012⁷⁾)

Data-collection parameters	
Instrument	BL45-SAXS (SPring-8, Harima)
Beam geometry	KB mirrors + guard slits (0.45 mm × 0.25 mm)
Wave length (Å)	1.5418
q range (Å ⁻¹)	0.009–0.30
Exposure time (sec)	7 × 18 frame or 10 × 18 frame
Concentration range (mg ml ⁻¹)	2–10
Temperature (K)	283
Structural parameters	
$I(0)$ (cm ⁻¹) [from $P(r)$]	0.114 ± 0.001
R_g (Å) [from $P(r)$]	14.27 ± 0.03
$I(0)$ (cm ⁻¹) (from Guinier)	0.112 ± 0.001
R_g (Å) (from Guinier)	14.5 ± 0.1
D_{max} (Å)	45 ± 3
Porod volume estimate (Å ³)	16500 ± 1000
Dry volume calculated from sequence (Å ³)	17570
Molecular-mass determination	
Partial specific volume (cm ³ g ⁻¹)	0.724
Contrast ($\Delta\rho \times eA^{-3}$)	0.098
Molecular mass M_r from sequence	14300 (calculated from CRY SOL)
Software employed	
Primary data reduction	PILATUS2INT
Data processing	PRIMUSQT
Ab initio analysis	DAMMIN
Validation and averaging	DAMAVR
Rigid-body modeling	N/A
Computation of model intensities	CRY SOL
Three-dimensional graphics representation	PyMOL

The table was modified from the original paper so as to meet synchrotron experiments.

元構造モデルを提案する際は、いかなるモデリングであろうと正当化され徹底的に説明されなくてはならない。例えば、Rigid body モデリングで single fit から ensemble fit へとパラメータを追加しモデルを詳細化する際は、著者自身の責任で説明がなされなければならない。ガイドラインで必須事項とされている事項を順次紹介する。

2.4.1 ソフトウェアの記載

動径分布関数 $P(r)$ 計算も含めモデリングのために使用されている全てのソフトウェアを記載する。Table 1 のような実績のあるソフトであれば、どうふるまうかの蓄積があるので、より第三者が他人のデータを判断しやすくなる。特に、(1)式で示す χ^2 値の定義などはまだソフトウェア間で標準化されていない部分もあるので注意が必要である。

2.4.2 ベストモデルのデータとの整合性

ガイドラインでは、少なくともベストモデルに対しては、 χ^2 値と実験 $I(q)$ 対 q プロット (Fig. 1(a)), を示せとされている。特に指摘は無いが、Fig. 4(a) では Figs. 1, 2 のようにモデルと実験値の残差プロット $\Delta I(q) = I_{exp}(q) - I_{cal}(q)$ もプロットし、系統的ずれがないかどうか確認しておくべきであろう。CRY SOL などは χ^2 値ではなく χ 値を

出力しているのに要注意である。(1)式の χ^2 の定義から明らかのように、実験誤差 (計測誤差) σ が大きいと小さな χ^2 値を出すことも可能である。理想的にフィットすれば χ^2 値は 1 前後になる。モデルが明らかに悪い場合は除き、プロットするとよくフィットしていて χ^2 値が 1.0 より大きい場合は、統計誤差の過小評価によることが多い。放射光の場合、計測誤差が非常に小さいので 1.0 より大きくなりがちである⁷⁾。Figs. 1, 2 で用いた実験データは放射光で取得されたデータであるが、明らかに実験誤差を過小評価しているのに χ 値が大きい。検出器の補正を行っていない影響もでているかもしれない。こういった点もデータの質にかかわってくる。いずれにせよ、 χ^2 値が 1.0 より大幅に小さい、ないしは、大きい場合は構造解析の正当性を著者が記述しなくてはならない。また、複数の散乱曲線に対するモデリングが行われたのなら、全ての χ^2 値とデータに対するフィッティングが示されなければならない。

2.4.3 モデルの統計値

いかなる小角散乱のモデリングであれ複数の解が存在するであろうから、モデル間の構造ばらつきを記述しなくてはならない。多くの場合、異なる分解能モデルでも適用できる規格化空間変位 NSD (Normalized Spatial Discrepan-

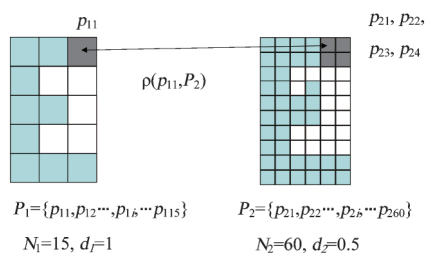


Fig. 6 (Color online) Schematic representation of normalized spatial discrepancy (NSD)¹⁷⁾. Suppose that two data point sets, P_1 and P_2 with different resolution d_1 and d_2 , inside which N_1 and N_2 data points exist. $\rho(p_{1i}, P_2)$ designates the minimum distance between p_{1i} and all the points in the set P_2 . In this case, the distance between p_{11} and p_{21} is indicated.

cy) という量を用いる¹⁷⁾。Fig. 6 のような分解能の違う 2 つのモデル 1, 2 において、それぞれが 2 つの点列 $P_1 = \{p_{1i}, i=1, \dots, N_1\}$, $P_2 = \{p_{2i}, i=1, \dots, N_2\}$ で表現されるとき、 P_1 と P_2 の間の NSD は

$$\rho(P_1, P_2) = \left(\frac{1}{2} \left[\frac{1}{N_1 d_2^2} \sum_{i=1}^{N_1} \rho^2(p_{1i}, P_2) + \frac{1}{N_2 d_1^2} \sum_{i=1}^{N_2} \rho^2(p_{2i}, P_1) \right] \right)^{1/2} \quad (3)$$

で定義されている。ここで N は P の要素数で d は各 P 内での隣接する要素間の平均距離 (分解能), $\rho(p_{1i}, P_2)$ はモデル 1 中の p_{1i} 点と最も隣接するモデル 2 中の点との距離を示す。

規格化空間変位 NSD は異なる分解能に適用でき、完全に一致する場合は 0 である。モデルの統計に関しては、得られたモデル群の平均 NSD 値とその標準偏差, 及び、NSD 平均の計算に除外しなくてはならないほどの高い NSD 値を持つモデルが存在したかどうかを記載する。通常 DAMAVER を用いて簡便に計算できる。

2.4.4 モデリングの仮定, 初期条件, 拘束条件

Rigid body モデルではいかに初期構造が得られたかを記述する。たとえば, あるドメインの結晶構造とかホモロジーモデルかなどである。電子顕微鏡法では, ホモロジーモデルを初期モデルにする時には, その鋳型構造の組み合わせも記述するが⁹⁾, 小角散乱においてはまたその必要性はない。また, モデリングの際, ドメインの接続序列とか距離制限をどのように設定したかを説明しなくてはならない。対称性などモデリングに際して用いた制約条件も全て記載しなくてはならない。

最後に, モデリングの際, 導入した仮定を支持する付加的な実験事実は報告されるべきである。

2.5 論文掲載形式

2.5.1 必須記載事項の表

ガイドラインでは, 上記の必須記載事項を表にしてまと

めることを勧めている。Table 2 は, データ収集と散乱パラメータの表である。Jacques らの論文の表を, 放射光測定用にデータ収集と解析ソフトの部分のみ変えてみたもので, あくまで例として考えていただきたい。Table 2 では分子量に関する重要な量 Porod 体積 V_P が記載されている。Porod 体積 V_P は散乱 $I(q)$ から以下の式に基づき求めることができる¹⁸⁾。

$$V_P = 2\pi^2 \frac{I(0)}{\int_0^\infty q^2 I(q) dq} \quad (4)$$

この Porod 体積 V_P は水和を含んだ散乱体の体積である。(4)式における $q^2 I(q)$ の積分範囲は 0 から ∞ であるが, 上記で説明したように蛋白質が揺らいだ構造を含むときには使えない。AUTOPOROD や PRIMUSQT などは動径分布関数 $P(r)$ 計算と同時に Porod 体積 V_P 計算を行ってくれるので便利である。Porod 体積 V_P (\AA^3) はおおよそ 20% 程度の誤差を含み, その値に 0.625 倍すれば分子量 MW (Da) になる²⁾。

$$MW = \begin{cases} 0.625 V_P \\ 0.5 V_{\text{exclud}} \end{cases} \quad (5)$$

Porod 体積 V_P と似た量で, ビーズモデルの平均構造を計算した際の排除体積 V_{exclud} (Excluded Volume, Filtered Volume ではないことに注意) (\AA^3) は, その値に 0.5 倍すれば分子量 MW (Da) になることが知られている²⁾。散乱から得られた V_P , V_{exclud} 値から得られる分子量 MW は標準試料の $I(0)$ との比較から得られる分子量 MW と全く独立しているので, 3 者間で整合性をチェックするとよい。通常では, これらの分子量はアミノ酸配列からの分子量と 10% 以内に収まる⁷⁾。Table 2 の例では若干 Porod 体積 V_P が小さめに出ている。Dry Volume は水和していないアミノ酸配列からの分子体積である。通常は Porod 体積 V_P の方が Dry Volume よりも 30% ほど大きい。偏比容は, 多くの場合, アミノ酸配列から計算される。コントラストは, 溶媒と蛋白質平均電子密度との差である。純水の電子密度が $0.334 \text{ e}/\text{\AA}^3$ なので溶媒の電子密度を計算すればコントラストは算出できる。これらの量は, 標準蛋白質の $I(0)$ との比較によって分子量を推定する際に重要となるので必要となる。

2.5.2 ベストモデルの表示

Ab initio モデリングの結果の表示例が Fig. 4 (g) である。ビーズモデルの平均構造と結晶構造を重ね合わせている。この図でのポイントは, 平均する前の 12 個のビーズモデルのばらつきを表示していることである。濃い色の部分は DAMMIF を使って 12 回 *ab initio* モデリング計算を行い, DAMAVER で平均化し得られたビーズモデルの平

均構造、薄い色の部分は全てのモデルの電子密度の広がりである。この *ab initio* モデリングの場合、モデル12個の規格化空間変位 *NSD* の平均は0.507、標準偏差は0.009で、ビーズモデルの平均化の際、形が特異で平均化に使われなかったモデルは存在しない。ビーズモデルのばらつきが少ないことは結果の信頼性につながる。

Fig. 4(a), (c)では結晶構造との比較を行っている。青線は *ab initio* モデリングである DAMMIF モデルの散乱曲線 ($\chi^2=1.27$)、赤線は結晶構造から CRY SOL で計算された散乱曲線 ($\chi^2=1.56$) である。「**Fig. 4(a)**における広角側での CRY SOL フィットのずれは、論文ならデータの解釈の際コメントが必要 (著者訳)」と Jacques らは記述している⁷⁾。**Fig. 4(a)**でリゾチームを用いているので、コンフォメーションの違いが検出されることは期待されない。Jacques らはさらに、「*P(r)* 関数が $r \approx 0$ でゼロでないのでブランク溶媒の差分がうまくいっていない可能性があるので、査読者なら再実験を要求するかもしれない (著者訳)⁷⁾」と付け加えている。

Ab initio モデリングは、単純に散乱体をビーズモデルで表し、生物学的制約も入れずに *I(q)* に対してフィットを行うので、計算は簡単で常に χ^2 値は小さい。反面、ビーズモデルは球の塊なので解釈の自由度があり、結果の是非に関しても客観的な判断が難しい。一方、原子座標に基づいた Rigid body モデリングは、初期構造を準備し、顕著なコンフォメーションを持つアンサンブルを選別するので、 χ^2 値を探すのも手間である。しかし、他の実験手法の結果を参考にして生物学的構造制約を入れているので、**Figs. 1, 2**から明らかなように、その分データの質に関してより詳細に判断できる。結果がより生物学的に厳密に判断される分、Rigid body モデリング技法の判断基準が、解析、実験両面で今後の標準化に大きく影響するだろう。

3. ガイドラインに反映されていない最近の小角散乱データとモデリングに関する評価方法

2013年、Nature 誌に小角散乱データとモデリングに関する新たな評価方法が Rambo と Tainer により公表された¹¹⁾。その論文においては

- 構造評価のための新しい指標となる相関体積 V_c 。
- 散乱曲線からのより正確な分子量決定方法
- モデルとデータとの一致度 χ の新しい定義 χ_{free}
- モデルの実効分解能の定量的な指標 R_{SAS}

の4点について新たな提案がされている。豊富な実験例と PDB データベースからのシミュレーションに基づいて、これらの指標は、小角散乱における蛋白質モデリングにとってバイアスのない指標となりうるかと主張されている。全てが非常に重要な知見であり、今後のガイドラインに大きく影響するであろう。詳しくは原著論文を参考にし

てもらおうとして、以下に簡単に解説を試みる。

3.1 構造評価のための新しい指標となる相関体積 V_c 。

近年、多くの蛋白質で生化学的に活性はあるが決まった構造を持たない蛋白質、いわゆる、天然変性蛋白質に関する関心が高まり、小角散乱へも広く応用されている¹⁹⁾。従来、揺らいた部分がどのように変わるか、構造状態の違いを評価するには回転半径 R_g の大小により定量化してきた。しかし、 R_g は、ギニエプロットという小角部の散乱領域を使用して決定されるため、同じく小角部に反映される粒子間相互作用に影響されやすく、また、(2)式の近似範囲によっても誤差が生じやすい。さらに、**Fig. 5**の構造と R_g を比較してみると、必ずしも構造状態を判断するパラメータとしては的確でない面もある。Rambo と Tainer はより定量性の高い構造状態の指標として、以下の式で定義される相関体積 V_c を提案した¹¹⁾。

$$V_c = \frac{I(0)}{\int_0^\infty qI(q) dq} = \frac{V_p}{2\pi l_c} \quad (6)$$

ここで l_c は相関長と呼ばれる量で、散乱体内で r の距離だけ離れたときの相関を示す。この相関体積 V_c は、“体積”とは言っても Porod 体積 V_p と違い構造状態に影響される。相関体積 V_c を用いる利点は、(4)式の $q^2I(q)$ の積分ではなく(6)式の $qI(q)$ の積分を計算する点にある。**Fig. 5(a)**で明らかなように、 $q^2I(q)$ では蛋白質の構造状態により大きく異なり、それ故、積分範囲に(4)式の積分結果が依存してしまう。一方、**Fig. 5(b)**の $qI(q)$ では $q > 0.15 \text{ \AA}^{-1}$ の領域で構造状態に大きく依存しない。それゆえ、積分範囲に対しても(6)式の積分結果が鈍感で安定している。また、相関体積 V_c は蛋白質濃度の項を含まず、蛋白質濃度の決定誤差をひきずらない。Rambo と Tainer は、核酸結合蛋白質の構造状態を変え、その構造状態が回転半径 R_g だけでなく相関体積 V_c とも高い相関があることを示した¹¹⁾。

3.2 散乱曲線からのより正確な分子量決定方法

散乱曲線 $I(q)$ から正確に分子量 MW を評価することは小角散乱において最も困難なことである。通常は、 $I(0)$ の絶対強度測定や標準サンプルとの $I(0)$ 値の比較が必要であった。標準サンプルによる分子量 MW は、蛋白質濃度、偏比容に依存するのでそれらの値もガイドラインで要求されている。2.5.1で述べたように、Porod 体積 V_p からの分子量 MW 決定は $\pm 10\%$ 前後の誤差を含み、(4)式の q の計測範囲に依存し、フレキシブルな構造を持つ蛋白質には適用できない。Rambo と Tainer は V_p の代りに、

$$Q_R = \frac{V_c^2}{R_g} \quad (7)$$

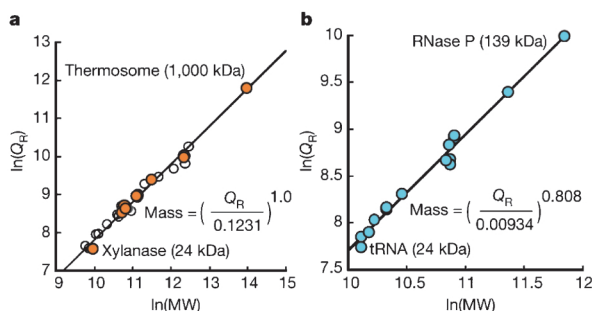


Fig. 7 (Color online) Power-law relationship between Q_R and particle mass (Figure taken from Rambo & Tainer, 2013¹¹). a. Q_R calculated from experimental SAXS data for protein-only samples. b. Q_R calculated from experimental data for RNA-only samples. See more details in the original paper¹¹.

で定義される指標 Q_R を使う、

$$\ln(MW) = \ln Q_R - \ln(0.1231) \quad (8)$$

という近似式が構造状態によらず最も正確に蛋白質の分子量 MW (Da) を推定できることを示した¹¹。Fig. 7a は横軸に $\ln(MW)$ 、縦軸に $\ln Q_R$ をプロットしたものであるが、非常によく直線に沿っていることが確認できる。図では、分子量を SEC-MALLS で厳密に精製したデータ（色つき）と通常のサンプルの両方が示されているが違いは見受けられない。この近似曲線を用いた10から1000 kDaの範囲での蛋白質の分子量 MW の誤差は、 $\pm 4\%$ 以下である。同様の近似は、RNA のような核酸でもよく成立する (Fig. 7b)。 V_c を計算する際の(6)式における q の最大積分（計測）範囲を $q_{\max} = 0.5 \text{ \AA}^{-1}$ から $q_{\max} = 0.3 \text{ \AA}^{-1}$ に狭めても、分子量が0.6%しか変わらず、積分（計測）範囲の違いに対しても大変安定である。つまり、(8)式による分子量の算出は、サンプルの構造、状態、計測方法に影響されず信頼性が高いことがわかる。今後、 Q_R による分子量が必須事項となるだろう。また、 Q_R による分子量決定は、偏比容や濃度、測定直前のゲルろ過精製などに影響されないので評価指標として受け入れられれば、Table 2 における一部の表記が緩和されることが期待される。

3.3 モデルとデータとの一致度 χ の新しい定義 χ_{free}

Fig. 2 でも明らかなように、Rigid body モデリングにおいては、single fit から ensemble fit への移行など、モデルと実験値とのずれの指標、 χ^2 値が決定的役割を担う。Fig. 2 での $\chi = 5.74$ と $\chi = 4.46$ の実験値 $I(q)$ に対するフィットのプロットを見てみると、 χ^2 値の違いというのは広角側でのずれが大きく影響していることがわかる。多くの実験データにおいては、S/N 比が低い広角側のデータ点数が全データ点数に対してかなりの割合を占める。しかも、(1)式の定義だと S/N 比が低いデータは χ^2 値を小さくするように働く。その結果、過度に広角側のデータが重

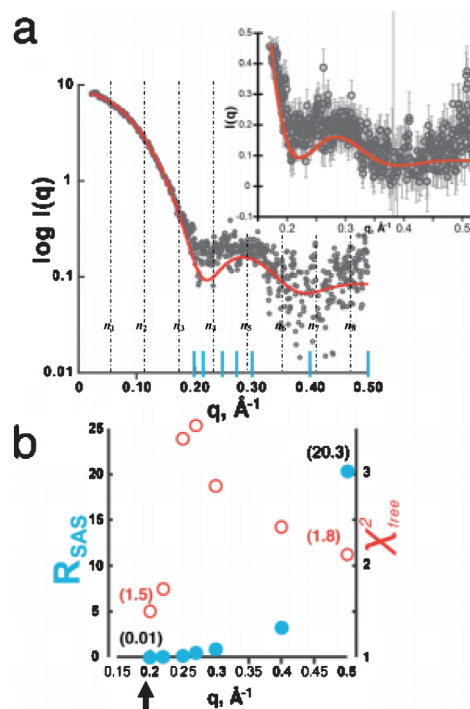


Fig. 8 (Color online) Determining resolution limit of model-data agreement by R_{SAS} . (Figure modified from supplementary information of Rambo & Tainer, 2013¹¹). (a) Scattering data of xylanase and its fit to PDB file (1REF). Dotted lines show Nyquist-Shannon sampling points. (b) χ_{free}^2 and R_{SAS} as a function of q used for calculation. An arrow suggests the practical resolution of this modeling.

視され、モデリングに際して予期せぬバイアスがかかる恐れがある。Fig. 8a は xylanase の実験 $I(q)$ と結晶構造から計算された理論 $I(q)$ を示す。 $q > 0.2 \text{ \AA}^{-1}$ の領域で両者にずれがあり ($\chi^2 = 1.39$)、xylanase は結晶構造と溶液中の構造が違うことが期待される。Rambo と Tainer はこのサンプル系で Rigid body モデリングを行い、 χ の定義を変えその影響を調べた。彼らは χ_{free} を以下の式で定義した¹¹。

$$\chi_{\text{free}}^2 = \frac{1}{n_s + 1 - p} \sum_i \left[\frac{I(q_i)_{\text{exp}} - I(q_i)_{\text{calc}}}{\sigma(q_i)} \right]^2 \quad (9)$$

ここで、 n_s は $n_s = q_{\max} D_{\max} / \pi$ (q_{\max} は最大計測散乱角ベクトル、 D_{\max} は最大長) で与えられるシャノン数で、 p はパラメータ数である。Fig. 8 の蛋白質では、 $D_{\max} = 53 \text{ \AA}$ 、 $q_{\max} = 0.5 \text{ \AA}^{-1}$ なのでシャノン数 $n_s = 8.4$ 、つまり、独立した情報量としては 8 前後、すなわち、8 個のサンプリング関数で $I(q)$ が近似できる。 χ_{free}^2 の考え方は、観測された $I(q)$ を Fig. 8 の点線のようにシャノン数で分割し、 n_1, n_2, \dots, n_8 で表されたナイキスト-シャノンサンプリング点間のそれぞれの q 領域でランダムにデータを抽出し χ_{free}^2 値を計算する。8 個のサンプリング関数の係数は分割されたそれぞれの q 領域で主に決まるので、 q 領域を分割して

χ^2_{free} 値を計算することにより、8つのサンプリング関数の寄与を均等に評価できるのである。また、分割して計算するので、結晶学で用いる R_{free} のように、最適化の際にデータの一部分を意図的に外して最適化のチェックを行うことも可能となる。Rambo と Tainer は、モデル選別の基準として統計学的にデータの自由度をより正確に見積もった χ^2_{free} 値を採用した方が、選別された構造モデル間の形のばらつきが小さく真のモデルに近いこと、ノイズにも強いことを示した¹¹⁾。

3.4 モデルの実効分解能の定量的な指標 R_{SAS}

次に重要な点はモデルの分解能である。小角散乱のモデルを解釈したり、クライオ電子顕微鏡像など他の構造手法と比較するためには定量的なモデルの分解能の議論は不可欠である。しかし、Rigid body モデリングではモデルの実効上の分解能が今まで不明確であった。Fig. 2 でもわかるように、モデリングが最大計測散乱角 $q_{\text{max}} = 0.3 \text{ \AA}^{-1}$ にフィットさせているからといって、 q_{max} に対応する分解能 $2\pi/q_{\text{max}} = 20.9 \text{ \AA}$ がモデルにあらわれているかというところではない。実際、single fit では $q_{\text{max}} > 0.15 \text{ \AA}^{-1}$ で実験値とずれており、分解能が 40 \AA 程度ではないかと考えられる。Rigid body モデリングの ensemble fit の場合、分解能の定義はさらに複雑になる。Ensemble fit では、Fig. 2 の例のように、溶液中の蛋白質のゆらぎに則した顕著な複数のモデル構造のアンサンブルとして小角散乱に対してフィットを行う。その際、抽出された構造アンサンブル中の個々のモデル自体は、必ずしも広角部の観測 $I(q)$ に対して十分フィットできない。アンサンブルとして始めて実験値にフィットするのである。その場合、実質上の分解能はどうなるのだろうか。このような分解能を判断する際には、 χ^2 値だけでは判断は無理で q に依存しない指標が必要である。Rambo と Tainer は、回転半径 R_g と相関体積 V_c を組み合わせた指標 R_{SAS} を以下の式のように定義した。

$$R_{\text{SAS}} = \frac{(R_g^{\text{exp}} - R_g^{\text{model}})^2}{(R_g^{\text{exp}})^2} + \frac{(V_c^{\text{exp}} - V_c^{\text{model}})^2}{(V_c^{\text{exp}})^2} \quad (10)$$

(10)式は、 R_{SAS} が 0 に近ければ近いほどモデルがデータと一致することを示す。 R_{SAS} は、 R_g など小角部のデータ一致度に大きく影響される小角パラメータの関数なので、小角部のデータ一致度を反映する指標であると考えてもよからう。Fig. 8 において $q_{\text{max}} = 0.5 \text{ \AA}^{-1}$ までの範囲でモデリングを行うと、 χ^2 値は 1.39 でありデータとのずれが明白ではないが、 R_{SAS} と χ^2_{free} 値はそれぞれ 20.3 と 1.8 となり、 R_{SAS} が 0 より大幅に大きく明らかにデータとフィットしていない (Fig. 8b)。単純に考えると、より正確なモデルを得るにはできるだけ広い範囲の q でフィットすれば良いように思われ、実際、Fig. 8a の $I(q)$ 上はモデルが実験値とあっているように見える。しかし、モデルの小角パラ

メータ、 R_g や V_c が実験値とずれているのである。すなわち、上記に述べたバイアスがかかってしまっているのである。バイアスの傾向は、小角部の相対的な寄与が少なくなる大きな q_{max} でより顕著に現れる。モデリングを行う領域を $q_{\text{max}} = 0.4 \text{ \AA}^{-1}$ 、 0.3 \AA^{-1} と小さくしていくと R_{SAS} が急速に小さくなり、 χ^2_{free} 値も $q_{\text{max}} = 0.25 \text{ \AA}^{-1}$ あたりで 2.5 といった大きくするがしだいに小さくなる。 $q_{\text{max}} = 0.2 \text{ \AA}^{-1}$ で R_{SAS} と χ^2_{free} 値はそれぞれ 0.01 と 1.5 となり、小角パラメータも含めてデータとモデルの一致度が高くなる。このようにモデリングの分解能を評価するには、フィッティングを行う q_{max} 領域を変えて、 R_{SAS} がゼロに収束し低い χ^2_{free} 値 ($\chi^2_{\text{free}} < 1.5$) を持つ q_{max} 領域から分解能が決定できる。Fig. 1 の *ab initio* モデリングの平均化によるずれも、ビーズモデリングの分解能が実験者が思うほどには良くないと考えると納得がいく。ビーズモデルの解釈はあくまで分解能に則したものでなければならない。

4. 今後の方向性

最近、小角散乱から非常にインパクトのある構造生物学の仕事が出ているが、そのほとんどが Rigid body モデリング技法を用いている^{3,6)}。小角散乱データによる Rigid body モデリング (ensemble fit) と NMR や単粒子測定 of データに基づけば、単に平均構造だけでなく溶液中のダイナミクスを知る手がかりが得られると多くの人が思い始めている⁶⁾。しかし、Fig. 8 のような例で結晶構造との違いを断定するためには、質の高いデータとそれに基づいた経験が必要である。現在、多くの小角ビームラインで、三次元構造モデルを得るためのデータ収集、解析実験の完全な自動化を目指して開発と投資が行われている。この状況は、過去自動化を大幅に進めた結晶構造ビームラインにおける開発と全く様相が違う。結晶構造では、データ収集、解析に関してかなり標準化が進んだ上で自動化を推し進めたが、小角散乱の場合、自動化はデータ収集・解析の標準化と平行して手探りで行われる。実際、技術的にどのようなメリットがあるのだろうか。

第一に、自動化はサンプルの品質とデータ収集技術の向上に貢献する。データの質を上げるにはサンプルの質を上げる必要がある。そのためには、実験室でできないことが多い。私の長年の経験から、実験室での光散乱を使ったスクリーニングはあくまで必要条件であり、十分条件とはなりえない。光散乱の精度や放射線損傷など放射光特有の問題も存在するからである。あるいは、ブランク溶媒がきちんと平衡になっていない場合もある。サンプルが解析に耐えるかは、ビームラインで得られた実験データからしか判断できない。非常に小さい χ^2 値で解析できるようなデータを出す、最適な溶媒条件、蛋白質変異体などを探索する必要があり、そのためには人間の手では限界がある。また、自動化は解析の自動化も含む。小角散乱ではデータ

の質というのは解析の最終段階でないと判別できないことが多い。スクリーニングのためにも解析の自動化は不可欠である。自動化と関連して、測定直前にサンプルの質をコントロールする手段を講じるのは有効であろう。SEC-MALLSのような、サンプルセルにHPLC、光散乱計、屈折計を直接 on-line で接続する方式は、サンプルの分散を著しく狭め、海外では今や標準的なセットアップとなっておりつつある。現在の日本の小角散乱研究では Fig. 1 のような *ab initio* モデリングが主に使われている。*ab initio* モデリングは、1回のシミュレーションでは常に χ^2 値が小さく、よほど $P(r)$ 関数がおかしい場合を除いて、導出されたモデルの形と χ^2 値でサンプル、データの質を判断するのは難しい。筆者の経験でも、検出器の補正などにも*ab initio* モデリングは比較的鈍感で、II-CCDのような縁の受光面で歪みと感受性が劣る検出器でも計算されたビーズモデルの不具合は判別できなかった²⁰⁾。ビーズモデルと結晶構造が一致する場合は問題ないが、仮に違っていた時にどこまでの議論が許されるのか、上記の分解能の議論とも関連して、今後のデータの蓄積によりビーズモデルに対するガイドラインも変わってくるかもしれない。

第二に、自動化により三次元構造と一次元散乱曲線との相関情報が著しく蓄積することによって、新たな小角散乱解析手法を生み出していることである。典型的な例は、Rambo や Tainer らによる小角散乱の新しい評価を明らかにした例である。彼らのグループでは、早くから小角散乱測定の標準化、自動化を目指して質の高いデータを蓄積してきた。それに基づきアイデアを出し、検証するために必要なサンプル系(キメラ蛋白質や変異体蛋白質)の実験をした結果、従来の小角散乱理論では取り扱えなかった蛋白質の構造変化が定量化できるようになっている。小角散乱の対象は非常に多岐にわたり同じ基礎理論を共有していることを忘れてはならない。小角散乱で扱うサンプルの中でも蛋白質溶液は特異である。単分散で、結晶構造や電子顕微鏡像など対応する実空間のデータベースが豊富である。今後蛋白質の小角散乱での経験、解析が合成高分子やミセル・コロイド系といった他の学問領域での小角散乱利用を促進するであろう。

ガイドラインは、単に解析手順だけでなく実験についても標準化が必要であることを要求している。標準化の議論の中で日本はほとんど貢献できておらず、技術的にかなり遅れをとっているような印象がある。これは、ハード及びソフト共に現在世界の主流となっている Rigid body モデリング技法に対応できていないからである。溶液状態というデリケートなサンプルを扱う性格上、単に外国の実績のある施設に行き測定をすればよいということにはならない。ある程度ウェット作業ができる、信頼ができる「ご当地放射光施設」が不可欠なのである。今、一番必要なのは、装置を洗練化するのに足るだけの質の高いサンプルが日本の小角散乱ビームラインに集まることであると筆者は

考える。

謝辞

本稿をまとめるのを奨励してくださった SPring-8, BL45-SAXS の担当者である理化学研究所専任研究員引間孝明博士に感謝いたします。また、私が理研 SPring-8 在籍中、BL45-SAXS の運営にお世話になった様々な方々にも深く感謝いたします。

参考文献

- 1) T. Fujisawa *et al.*: J. Appl. Cryst. **33**, 797 (2000).
- 2) M. V. Petoukhov *et al.*: J. Appl. Cryst. **45**, 342 (2012).
- 3) M. V. Petoukhov and D. I. Svergun: Int J Biochem Cell Biol. **45**, 429 (2013).
- 4) 藤澤：放射光 **19**, 429 (2006).
- 5) R. Arai *et al.*: Proteins **57**, 829 (2004).
- 6) R. P. Rambo and J. A. Tainer: Annu Rev Biophys. **42**, 415 (2013).
- 7) D. Jacques, J. M. Guss, D. I. Svergun and J. Trewhella: Acta Crystallogr D Biol Crystallogr **68**, 620 (2012).
- 8) R. J. Read *et al.*: Structure **19**, 1395 (2011).
- 9) R. Henderson *et al.*: Structure **20**, 205 (2012).
- 10) <http://journals.iucr.org/services/sas/>
- 11) R. P. Rambo and J. A. Tainer: Nature **496**, 477 (2013).
- 12) 藤澤：“生体系、ソフトマターの X 線分光”，X 線・放射光の分光，東京，講談社サイエンティフィック，85-107 (2009).
- 13) D. A. Jacques *et al.*: Acta Cryst. **D65**, 574 (2009).
- 14) T. Fujisawa *et al.*: High Pressure Res. in press (2013).
- 15) J. Aishima *et al.*: Acta Cryst. Section D, Biol. Cryst., **66**, 1032 (2010).
- 16) J. M. Flanagan, M. Kataoka, T. Fujisawa and D. M. Engelman: Biochemistry **32**, 10359 (1993).
- 17) M. B. Kozin and D. I. Svergun: J. Appl. Cryst. **34**, 33 (2001).
- 18) O. Glatter and O. Kratky: Small angle x-ray scattering (Academic Press, London, 1982).
- 19) M. Hammel: Eur. Biophys. J. **41**, 789 (2012).
- 20) Y. Takahashi, Y. Nishikawa and T. Fujisawa: J. Appl. Cryst. **36**, 549 (2003).

著者紹介



藤澤哲郎

岐阜大学工学部化学・生命工学科教授，名古屋大学客員教授，理化学研究所播磨研究所放射光科学研究センター客員研究員

E-mail: fujisawa@gifu-u.ac.jp

専門：生物物理学，タンパク質溶液散乱，高圧生物学

【略歴】

1989年大阪大学基礎工学研究科物理系生物工学専攻博士課程終了。工学博士。1989-1990年米国 Yale 大学博士研究員，1990年理化学研究所研究員，1998年同前任研究員。1996-2006年理研ビームライン BL45XU 小角散乱ステーションの建設運営責任者。2007年岐阜大学工学部生命工学科教授。2013年改組により現職。

How to read publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution

Tetsuro FUJISAWA Department of chemistry and biomolecular science, Faculty of Engineering, Gifu university, 1-1 Yanagido, Gifu City 501-1193, JAPAN

Abstract Small-angle scattering (SAS) technique is now widely used for many structural biological studies. However, the technical development is so fast and the user community is so expanding that the community requires to establish the minimum reporting for assessing the quality of data and models. In 2011, the IUCr Commission on SAS published a paper on the preliminary guidelines for the representation of biomolecular structures from small-angle scattering data. In 2013, new assessment methods for the data and modeling have been proposed by Rambo & Tainer. This manuscript overviews these recent important advances for the standardization of SAS modeling, and introduces them to synchrotron x-ray small-angle scattering users.