



## タンパク質溶液散乱法 —高分解能構造と低分解能構造の橋渡し—

藤澤哲郎 独立行政法人 理化学研究所播磨研究所放射光科学総合研究センター

**要旨** 近年、タンパク質結晶構造解析により非常に多くのタンパク質断片の結晶構造が解かれるにつれ、タンパク質の機能と関連した、より native な構造を研究したいという需要が高まっている。これを実現する最も有効な方法の一つがタンパク質溶液散乱である。これまで、タンパク質溶液散乱を解析する際に2つの大きな障害があった。第一は、高分解能結晶構造から水和されたタンパク質の溶液散乱を計算する計算法であり、第二は、実験データに基づき三次元のモデル構造を最適化する計算法である。1990年代半ば以降、タンパク質溶液散乱解析技術は、効率的な非線形最小自乗法の最適化アルゴリズム、計算機速度の高速化に伴い急速に発展した。我々が最近行った2つのドメインからなる人工タンパク質を例に取り、EMBL Svergun 博士らが開発した ATSAS プログラムパッケージとテキサス大 Wrigger 博士らが開発した SITUS プログラムパッケージを使ったタンパク質溶液散乱解析の応用例を紹介する。

### 1. はじめに

近年 X 線結晶構造解析の進歩により、大きなタンパク質複合体も含めて高分解能構造を得る事が容易になってきた。放射光は、タンパク質結晶ユーザーの裾野を広げ、より微小な、より不安定な結晶のデータ取得に重要な役割を果たしている。同時にゲノム科学的アプローチはおびただしい量のタンパク質の断片構造を産出している。タンパク質の機能を理解する上で“生”のタンパク質の構造はどのようになっているのかという疑問がますます着目されている。本稿ではマルチドメインタンパク質の溶液構造解析を例に挙げながら溶液散乱法を使ってタンパク質分子の形を見るモデリング技術の概略を解説する。より詳しい情報は最近出た総説<sup>1-3)</sup>あるいは教科書<sup>4,5)</sup>を参照されたい。なお、本稿ではタンパク質は塩ないし緩衝剤を含む水溶液に分散しているものとし、溶媒に脂質などを含む3成分系は取り扱わない。

### 2. 小角散乱パラメータの導出と動径分布関数の計算

#### 2.1 タンパク質溶液散乱の原理<sup>4,5)</sup>

タンパク質溶液散乱の原理は非常に単純である。タンパク質が溶液中で入射光に対し等方的に向いていると、タンパク質からの溶液散乱は散乱方位依存性を失くし散乱角のみに依存する一次元の測定量となる。通常散乱ベクトルは  $S = 2 \sin \theta / \lambda = 1/d$  で定義され、その時の  $2\theta$  は散乱角、 $\lambda$  は波長、 $d$  は Bragg Spacing に対応する。 $q = 2\pi S = 4\pi \sin \theta / \lambda$  もよく使用される。タンパク質の電子密度が溶媒の電

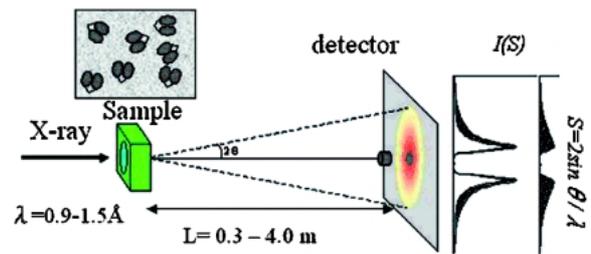


Fig. 1 Principle of protein solution scattering.

子密度よりも高いために、タンパク質溶液からの散乱と水からの散乱の差が、得られるタンパク質からの散乱  $I(S)$  となる (Fig. 1)。

観測される  $I(S)$  は

$$I(S) = N i(S) F(S)$$

で表すことができ、 $N$  は照射体積中の散乱体数、 $i(S)$  はタンパク質一個からの散乱、 $F(S)$  は構造因子と呼ばれる粒子間相互作用を表す関数である。 $F(S)$  は複数の濃度点で  $I(S)$  を測定しゼロ濃度外挿する事により  $F(S) = 1$  となる。また、 $i(S)$  は

$$i(S) = i_{\text{shape}}(S) + i_{\text{shape-fluctuation}}(S) + i_{\text{fluctuation}}(S)$$

タンパク質の形状項  $i_{\text{shape}}(S)$ 、内部微細構造の項  $i_{\text{fluctuation}}(S)$  及び両者の cross 項  $i_{\text{shape-fluctuation}}(S)$  から成る<sup>6)</sup>。 $d = 20 \text{ \AA}$  から  $30 \text{ \AA}$  ぐらいまでの角度領域では形状項の寄与がほとんどで  $S$  が大きくなると内部微細構造の項が大き

なる。本稿で取り扱うタンパク質溶液散乱の解析は、形状項の解析を中心にして、単分散で均一系の場合を想定して話を進める。

## 2.2 解析結果を歪める実験上の主な要因

解析の第一歩は、タンパク質溶液散乱を正しく計測する事である。それは経験の積んだ専門家でも生のデータからは分からない。筆者自身、多大な解析時間を費やした後、測定がどうもおかしいと気づいた経験が少なくない。ユーザーは正しい結果を導くのを歪める放射光実験上の要因として以下のような可能性がある事を知っておく必要がある。(1)散乱体のサイズに対して小角部の散乱データが不十分である。(2)タンパク質溶液からの散乱と溶媒からの散乱が同一条件で取得できていない。(3)タンパク質が単分散でない。(4)タンパク質の放射光による損傷。それぞれについて簡潔に説明すると(1)は、溶液散乱の原理から散乱体が大きくなればなるほど入射光近傍の散乱角の小さな部分に構造情報が局在する。理論上、散乱体最大長  $D_{\max}$  の2倍の大きさに対応する散乱角まで計測できていないとタンパク質溶液構造は決定できないとされている<sup>7)</sup>。(2)は、光学系の不安定性によるビーム位置・強度の変動、検出器測定回路の安定性など計測技術上の問題点と、タンパク質溶液と溶媒が化学的に平衡に達していないなどの化学的な問題点がある。後者の場合には溶媒散乱に透析外液を使用するのが最も適している。(3)に関しては最も問題が生じやすい。放射光の登場により従来の X 線発生装置よりも希薄なタンパク質溶液で計測できるようになったが<sup>8)</sup>、それでも分光学的なタンパク質解析方法に比べ1桁以上濃度が濃い。従って、サイズカラムクロマトグラフィーで精製した後、濃縮の過程で会合が発生したりする。また、散乱の性質上限定的に会合が起こっている場合、例えば細長いタンパク質で単量体分子と2量体分子が混在している場合は判断が非常に難しくなる。単分散性をチェックする事は解析の過程で最も大事で、以降の項においてより詳しく解説する。最後に(4)に関しては、タンパク質溶液に強力な放射光が照射されるとタンパク質が変性する前に水ラジカルによるタンパク質表面の性質が変化し、ある閾値以上では会合が起こる。閾値はタンパク質と溶媒条件によって異なり、今のところ残基配列、構造と溶液散乱における放射線損傷の相関は明らかでなく実際の実験で実証するしかない。ちなみにリゾチームなど比較的放射線損傷に弱いタンパク質では400 Gyである<sup>9)</sup>。閾値時間以降では照射時間に対して会合が指数関数的に増加するので最初に放射線損傷の閾値時間を決めておく必要がある。

## 2.3 小角散乱パラメーターの導出

小角散乱から得られる構造パラメーターの決定は解析の最初に必ず行うステップである。散乱関数  $I(S)$  は  $S$  が非常に小さい領域、小角領域では散乱体の形状によらずガウ

ス関数で近似できる。

$$I(S) = I(0) \exp\left(-\frac{4\pi^2}{3} Rg^2\right)$$

縦軸に  $\ln(I(S))$ 、横軸に  $S^2$  をプロット（ギニエプロット）し、内側の直線領域の傾きと  $y$  切片からそれぞれ散乱体の回転半径  $Rg$  と原点散乱強度  $I(0)$  とが求まる。この近似は厳密には  $2\pi SRg < 1$  の小角の範囲で成立する。大きな会合体が混在する時にはギニエプロットでの小角部の直線性が無くなる。 $Rg$  は分子量  $M$  に対しおおよその目安として

$$Rg = 0.762 \sqrt[3]{M}$$

が成立するので散乱体の分子量を予想する事も可能である<sup>10)</sup>。より厳密に分子量を知るには  $I(0)$  を重量濃度  $C$  (mg/ml) で割った量  $I(0)/C$  を使用する。 $I(0)/C$  は分子量に比例するので、分子量既知で濃度の分かっているタンパク質の  $I(0)/C$  と比較することにより散乱体の分子量が推定できる。 $I(0)/C$  を濃度に対してプロットすると、単分散の場合濃度が高くなるにつれて通常わずかに減少するが、単量体と会合体の平衡状態にある場合は濃度が高くなるに連れて急激に増加する。複数の濃度に対して  $I(S)$  を測定しゼロ濃度外挿を行う事は、データの信頼性のチェックのみならず、タンパク質溶液の単分散性を間接的に知る上で重要である。これ以降の文では  $I(S)$  を濃度外挿された散乱強度関数として扱う。

## 2.4 動径（二体間距離）分布関数

次に、散乱体電子密度の convolution (\*) を空間平均  $\langle \rangle$  した  $P(r) = \langle \Delta\rho(r) * \Delta\rho(r) \rangle$  で定義される  $P(r)$  関数を考える。 $P(r)$  関数は動径分布関数と呼ばれタンパク質の3次元形状情報を1次化したものであり、より具体的には、タンパク質内部の2点間距離の分布を表すので二体間距離分布関数とも呼ばれる。タンパク質の場合、 $P(r)$  関数はタンパク質の最大長  $D_{\max}$  より大きい  $r > D_{\max}$  で  $P(r) = 0$  となる。 $P(r)$  関数を使うと  $I(S)$  は

$$I(S) = 4\pi \int_0^{D_{\max}} P(r) \frac{\sin(2\pi rS)}{2\pi rS} \sin(2\pi rS) dr$$

と書ける。式の形から判断できるように  $P(r)$  関数は、散乱強度の逆フーリエ変換から直接計算できるが、非常に信頼性に乏しい。そこでより一般的な方法<sup>11,12)</sup>として、

$$P(r) = \sum_{i=1}^K c_i \varphi_i(r)$$

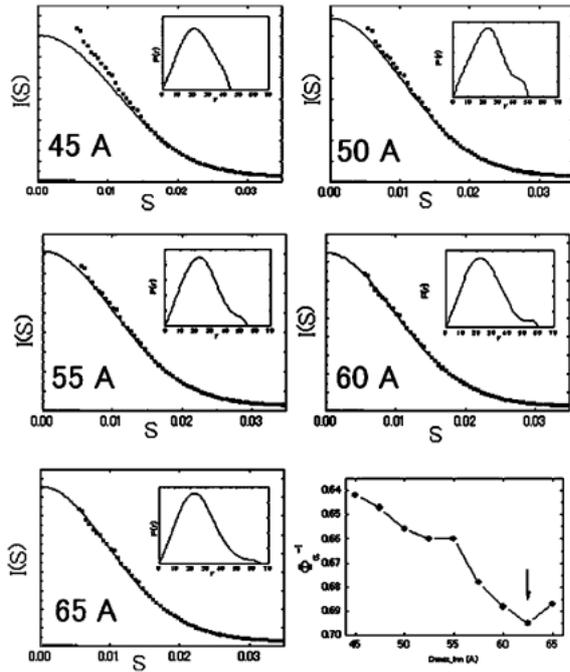


Fig. 2 Determination procedure of  $P(r)$  and  $D_{\max}$  of tropomodulin<sup>13)</sup>.  $D_{\max}$  is indicted by an arrow.

と  $0 < r < D_{\max}$  の範囲内で  $P(r)$  関数を多項式  $\phi_i(r)$  の線形和で表し、係数  $C_i$  は実験データに対して

$$\phi_{\alpha} = \sum_{k=1}^N \left[ \frac{I_{\text{exp}}(S_i) - \sum_{i=1}^K c_i \phi_i(S_i)}{\sigma(S_i)} \right]^2 + \alpha \int_0^{D_{\max}} [P'(r)]^2 dr$$

を最小にする形で決定する。パラメーター  $\alpha$  は第2項が示す  $P(r)$  関数の平滑性と第1項が示すフィットの残差をバランスさせるパラメーターである。

このような計算法を間接フーリエ変換法といい、今最も使われている GNOM<sup>12)</sup> というソフトでは  $D_{\max}$  を入力すると自動的に  $P(r)$  関数の平滑性や正値性などを満たす  $\alpha$  を計算する。従って  $P(r)$  関数を決定するには  $D_{\max}$  を決める事に他ならない。それでは未知の  $D_{\max}$  をどのようにして決めるのだろうか？ Fig. 2 は  $P(r)$  関数と散乱曲線へのフィットが入力  $D_{\max}$  に対しどのように変わるか示したものである。散乱曲線への近似が十分良くなる最小の  $D_{\max}$  の値がいわゆる  $D_{\max}$  の真の値、最大長になる。Fig. 2 の場合、 $D_{\max} = 62.5 \text{ \AA}$  となる<sup>13)</sup> (右下図矢印参照)。

### 3. Ab initio 法

#### 3.1 Ab initio ビーズモデリング

以下に扱うモデリングでは電子密度を一様と仮定する。この仮定は  $I(S) \equiv I_{\text{shape}}(S)$  が成立する小角領域を解析する上では十分成立する。従来、球、楕円体、円柱などの単

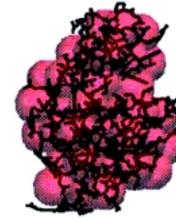


Fig. 3 Representation of lysozyme structure by space filling beads (red sphere). The beads were created by SITUS<sup>25)</sup> and drawn by VMD<sup>27)</sup>.

純な形でタンパク質を近似し散乱体の形を評価していた<sup>4,5)</sup>。しかし、より汎用性のある3次元モデルとして、タンパク質を Fig. 3 のように数百から千程度の小さな球からなる球充填モデル(ビーズモデル)で表現する事が広く使用されている<sup>15-17)</sup>。

一般的に3次元のビーズモデルを1次元データから再構成する事は不可能である。しかし、電子密度一様としてモデリングを簡略化し、モデル  $X$  のコンパクト性、充填球の連続性など、タンパク質構造特徴を取り入れたペナルティ関数  $\Pi(X)$  を導入すればタンパク質三次元モデルの信頼性、一意性は大幅に向上する。近年、モデルの散乱強度  $I_{\text{theor}}(S)$  を計算し、実験散乱強度  $I_{\text{exp}}(S)$  に対して下記式に従ったモデルの最適化が広く行われるようになった<sup>16)</sup>。

$$f(X) = \frac{1}{N-1} \sum_{j=1}^N \left[ \frac{I_{\text{exp}}(S_j) - I_{\text{theor}}(S_j)}{\sigma(S_j)} \right]^2 + \beta \Pi(X)$$

上式で  $\beta$  は重み係数である。最適化アルゴリズムとして遺伝アルゴリズム<sup>15)</sup>、焼きなまし法<sup>16)</sup>、モンテカルロ法<sup>17)</sup>、などを用いる3つのソフトが存在する。多くのタンパク質の場合、いずれのアルゴリズムも30-40万のモデルから20-30 Å 分解能で似たような形に収束し、結晶構造解析や高分解能 NMR などの原子構造とも整合性のある結果が得られる事が知られている<sup>18,19)</sup>。

#### 3.2 DAMMIN によるビーズモデルの計算

この節では最も一般的に使用されている、Svergun 博士らによって開発された、焼きなまし法の最適化アルゴリズムを用いた *ab initio* ビーズモデリングソフト DAMMIN を、2つの円筒状ドメインをリンカーで繋いだ人工タンパク質の解析に利用した例を取り紹介する<sup>20)</sup>。DAMMIN を実行させるには前述した GNOM による  $P(r)$  計算を行い、最大長  $D_{\max}$  を決定する事が必要である。Fig. 4 は DAMMIN によるプログラムの流れを示したものである。

プログラムは  $D_{\max}$  を直径とする探索空間を小球で充填する (Fig. 4A)。タンパク質の対称性があらかじめわかっている際にはモデルの対称性を縛る事も可能である。1次元の散乱曲線を満たす3次元モデルは多数存在するの

で、プログラム実行時毎に微妙に形が変化してくる (Fig. 4B)。従って、*ab initio* ビーズモデリングで低分解能構造を再構成する際には、10回程度ソフトを実行し得られたモデルの向きを揃え重ね合わせ (Fig. 4C 上図)<sup>21)</sup>、平均化して代表的なモデル構造として抽出する<sup>22)</sup>。この処理は DAMAVER というプログラムでバッチ処理により1段階ステップで計算する<sup>23)</sup>。特に、タンパク質が変性した状態のように揺らぎが大きい場合は、ビーズモデルはソフトの実行毎に大きく異なる場合がある。そこで、我々は2段階ステップによる平均構造の計算を行っている<sup>20)</sup>。最初のステップで平均化した後に、平均モデル構造の体積を通常より10%大きめにとり (Fig. 4C 下図)、さらにそれを探索空間として最終的に散乱曲線に対してモデルを計算す

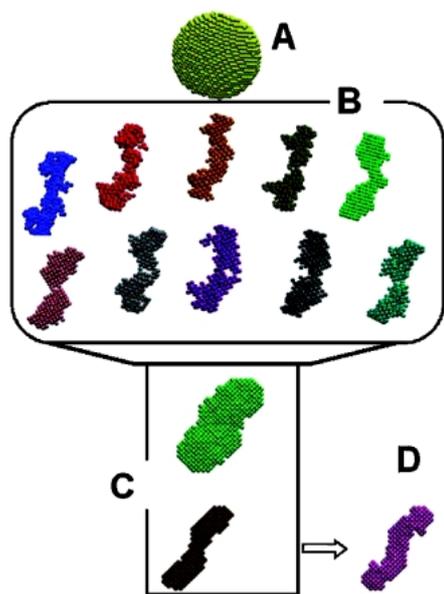


Fig. 4 Data flow of *ab initio* beads modeling by DAMMIN program. The results of multiple runs were averaged and taken as a low-resolution model. The protein has a linker sequence, HL5<sup>20)</sup>.

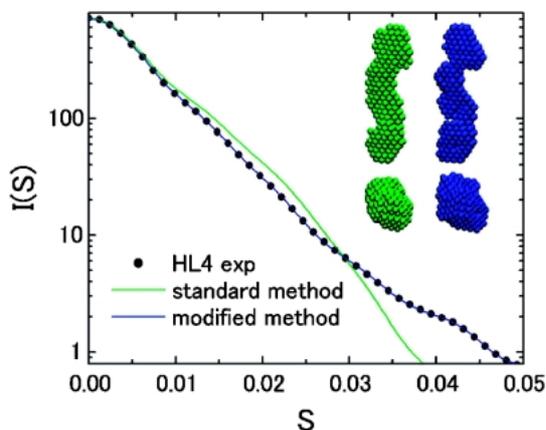


Fig. 5 Effect of averaging procedure of *ab initio* beads models. The protein has a HL4 linker.

る (Fig. 4D)。平均処理の方法により最終モデルがいかに変わるかを Fig. 5 で示した。図中の緑色で描かれたモデルは Volkov らの方法による1段階の標準手順による平均構造、青色が2段階ステップによる平均構造である。

Fig. 5 から *ab initio* モデリングで重要な2つの点が導き出せる。第一に、*ab initio* モデルは、タンパク質の形によっては解の線形性が成り立たないという事である。すなわち、個々に実験データによくフィットするモデルの平均構造からの  $I_{\text{theor}}(S)$  が実験散乱強度  $I_{\text{exp}}(S)$  と必ずしも一致しない。第二に、1次元の散乱曲線が微妙な3次元構造の違いに大きく反映される事である。Fig. 5 では、2つのモデルはいずれも細長く共通のドメイン構造の特徴を持つように見えるが、散乱曲線では意外と違っている。

### 3.3 SITUS<sup>24,25)</sup>による低分解能構造をベースとした高分解能構造による解釈

低分解能構造が Fig. 5 のように得られると次のステップは、低分解能構造の意味を知る事、すなわち、「低分解能構造の高分解能構造による解釈」が必要となる。例に挙げた人工タンパク質の場合、既知の2つのドメインを繋ぐリンカー部分を設計し、ドメインの配置と距離を変え、蛍光プローブとしての人工タンパク質の機能可変を目的としていた。ドメインの配置の自由度が高いため結晶構造解析は困難で、小角散乱による低分解能構造に基づき構造を予測しようという事になる。「低分解能構造の解釈」という点では電子顕微鏡における一粒子データ解析と共通している。一粒子データ解析を小角散乱の低分解能構造に応用す

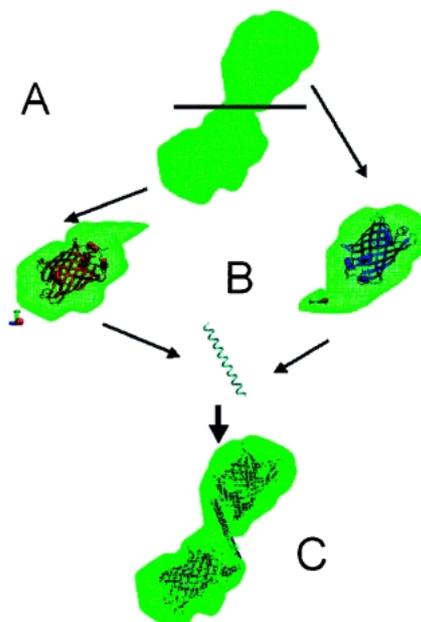


Fig. 6 Data flow of docking high-resolution and low-resolution structures. The protein has a HL5 linker. The contour map was generated by SITUS and drawn by VMD<sup>27)</sup>.

る事を Willy Wrigger 博士のグループが SITUS<sup>26)</sup> というプログラムパッケージで提唱している。

Fig. 6 は HL5 というリンカー配列を持つ人工タンパク質の小角散乱低分解能構造の例を示したものである。最初にビーズモデル構造を convolution 演算操作で電子顕微鏡のような電子密度マップに変換する。次に、マップを半分に切断し (Fig. 6A) 各部分に対してドメインの高分解能構造を当てはめる (Fig. 6B)。一般に低分解能マップは実験エラーを多く含む。SITUS は、スケルトンアニメーションの技術を応用して、低分解能、高分解能共にタンパク質の形状を code vectors という点列で抽象化し、点列間で高分解能構造とのドッキングを行う事で、低分解能データによるエラーを克服している<sup>24)</sup>。最後のステップでは、両ドメインを繋ぐヘリックスリンカー部分を当てはめ、分子動力学計算で形を整える。S-タグや His-タグと呼ばれる断片は無視して構築した最終的な高分解能モデル構造は Fig. 6C に示される<sup>20)</sup>。

### 3.4 Dummy residues modeling 法<sup>28)</sup>

*Ab initio* ビーズモデリングに対して、さらにタンパク質の構造性質を組み込んだ方法が Svergun らによって開発された dummy residues modeling 法<sup>28)</sup> である。タンパク質は C $\alpha$  原子が約 3.8 Å ずつ離れて存在し、5 Å の分解能では C $\alpha$  原子を中心とした模擬残基 (dummy residues) 球の集まりと考える事ができる。モデルの最適化はビーズモデリングの場合と同様であるが、大きな違いは散乱曲線をより広角 (5 Å) までフィットさせる事ができる事、模擬残基球と仮想水和水球を分けて計算する事、さらに単純にコンパクトさだけでなく C $\alpha$  原子の統計的幾何分布に基づき擬似折れ畳みを生成する点にある<sup>28)</sup>。計算結果は模擬残基球と水和水球のモデルとして与えられる。現在、アミノ酸残基のペプチド角や 2 次構造情報などを簡便に取り込んでいるが正しい折れ畳みを予測できるレベルではなく、むしろより広角情報を使用したビーズモデリング法の一つと考えた方が良からう。また、dummy residues modeling 法による低分解能構造は“充填された”球モデルではないため SITUS による解析には適さない事を付け加えておく。

## 4. Rigid body modeling 法

### 4.1 原子構造からの溶液散乱の計算法

*Ab initio* 法は汎用性が非常に高いが、溶液散乱から得られた低分解能構造を高分解能構造で解釈するのが容易ではない。断片的に高分解能原子構造が分かっている際には既知の断片構造を並進回転し、その  $I_{\text{theor}}(S)$  と  $I_{\text{exp}}(S)$  を直接比較する事によってモデルを最適化する、いわゆる散乱強度をベースとした rigid body modeling 法がより人気がある。このためには、原子構造からタンパク質溶液散乱

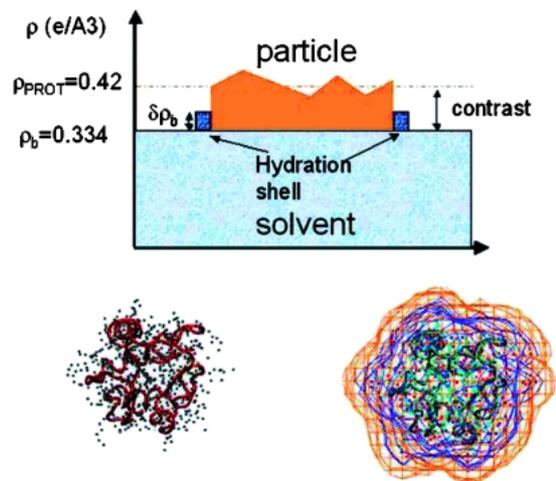


Fig. 7 Protein hydration in solution scattering. The lower left figure represents proteins in vacuum. Lower right figure shows protein surface (blue) and hydration layer (orange).

強度  $I(S)$  が正確に計算されなくてはならない。 $I(S)$  は以下のような式で定義できる<sup>29)</sup>。

$$I(S) = \langle |A(S)|^2 \rangle_{\Omega} = \langle |A(S) - \rho_{\text{PROT}}E(S) + \delta\rho_b B(S)|^2 \rangle_{\Omega}$$

$A(S)$ ,  $E(S)$ ,  $B(S)$  はそれぞれ真空中、排除体積中、水和層からの散乱因子で  $\rho_{\text{PROT}}$  がバルク水の電子密度、 $\delta\rho_b$  は水和層とバルク水の電子密度差すなわちコントラストである (Fig. 7)。特に、後ろの 2 項はタンパク質の体積、水和と大きく関わっているので計算は単純ではない。分子動力学法によりタンパク質の水和構造を計算し溶液散乱を計算する事も考えられるが<sup>30,31)</sup>、時間と手間がかかってしまい、モデルの最適化の際に使用するのはあまり現実的ではない。現在最も使用されており、多くの実験データの解析に使用されているソフトが Svergun 博士により開発された CRY SOL である<sup>29)</sup>。彼は、水和層の微細構造を無視して電子密度一様として単純化している。CRY SOL による理論曲線は実験データを 5 Å の分解能まで十分にフィットする事が示されているが<sup>32)</sup>、広角領域における散乱曲線の寄与は依然不明されていない。CRY SOL は実験データの存在の有無に関わらず使用できるが、実験データがあれば水和水層の厚さやコントラストを微調できるのでより正確な結晶構造と溶液構造の比較が可能となる。

### 4.2 MASHA による散乱強度をベースとした高分解能構造による解析

Rigid Body Modeling 法では、様々なサブユニット位置でのタンパク質複合体の散乱曲線を大量に計算する事が不可欠である。説明を簡単にするために、2 つのサブユニット A, B から成るタンパク質複合体を考える。A を固定し B を動かした時に、複合体の散乱強度  $I(S)$  は

$$I(S) = I_A(S) + I_B(S) + 2 \langle A(\vec{S}) C^*(\vec{S}) \rangle_{\Omega}$$

と書ける。 $I_A(S)$ ,  $I_B(S)$  はそれぞれサブユニット A, B の散乱強度,  $A(S)$  が原点位置におけるサブユニット A の散乱因子,  $C(S)$  が動かした後のサブユニット B の散乱因子である<sup>33)</sup>。1つのドメインを固定し, もう1つのドメインを動かすには3次元の回転(オイラー角,  $\alpha, \beta, \gamma$ ) と3次元の並進の計6次元パラメーターを必要とする。通常, 小角散乱で行う rigid body modeling 法では, 動かすドメインは一つだけで, それ以上の場合は対称性を仮定するのが実際的である。上式に基づき  $I(S)$  を計算するには速度が重要な因子である。Svergun 博士らは CRY SOL を用いて水和された各サブユニットの散乱因子を計算し, 球面調和関数を用いる事により  $I(S)$  計算速度を高速化している<sup>33)</sup>。本項では人工タンパク質全残基数の約15%を構成するリンカー部分とタグ部分を無視し, 2つのドメイン位置を rigid body modeling 法で決め, Fig. 6C で示した *ab initio* 法による高分解能モデルとの比較を行った<sup>20)</sup>。MASHA と呼ばれる対話的に rigid body modeling を行うソフトを使用して計算した結果が Fig. 8 である。

*Ab initio* 法と違い, MASHA による Rigid body model-

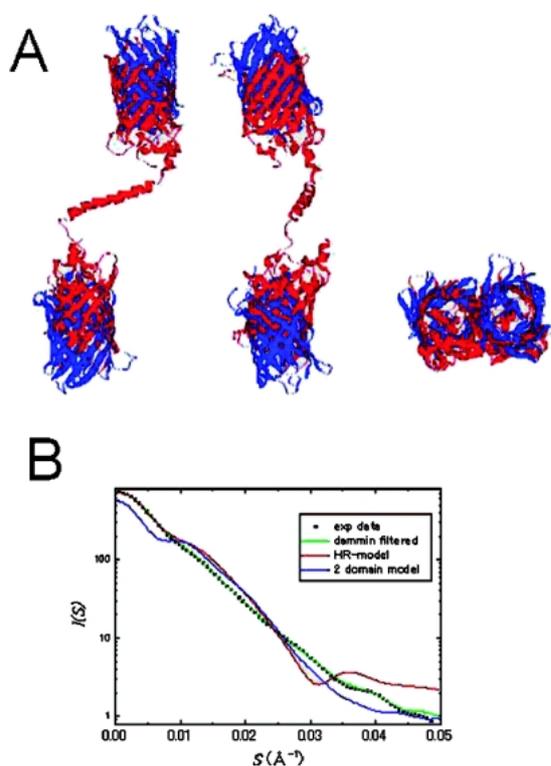


Fig. 8 Example of rigid body modeling method by MASHA. The model obtained by rigid body modeling method (blue) and high resolution model based on low resolution bead structure (red, same as Fig. 6C). The corresponding scattering curves were plotted in B. Some part of structure was ignored for analysis.

ing 法は初期構造の近傍でモデルを探索するので, local minimum に解が拘束されやすく解を最適化するのが簡易ではない。我々はできるだけ計算を効率化するために, *ab initio* 法でだいたいの配置の見当をつけてから初期位置を与えたり, 大きく解がずれないように一度に散乱曲線全体をフィットするのではなく, 小角部分から多段階でフィットする領域を広げたりした。Rigid body modeling 法のドメインの位置は高分解能モデルのドメインの位置と非常に似通っている事が分かる (Fig. 6C)。

#### 4.3 高分解能構造が部分的にしかわからないタンパク質の散乱強度をベースとした解析

前項では高分解能構造未知の部分の寄与は無視したがそれらをモデルに取り込むにはどうすればよいのだろうか? Svergun 博士らのグループでは BUNCH というプログラムを提供している<sup>34)</sup>。このプログラムは, 先に欠損している部分を dummy residues modeling で任意構造をシミュレーションし, お互いの部分が衝突しないようにしながら既知の構造の配置を探索する事を同時に行う。

このプログラムを実行させるためには既知 PDB 構造と dummy residues でシミュレーションする部分のアミノ酸

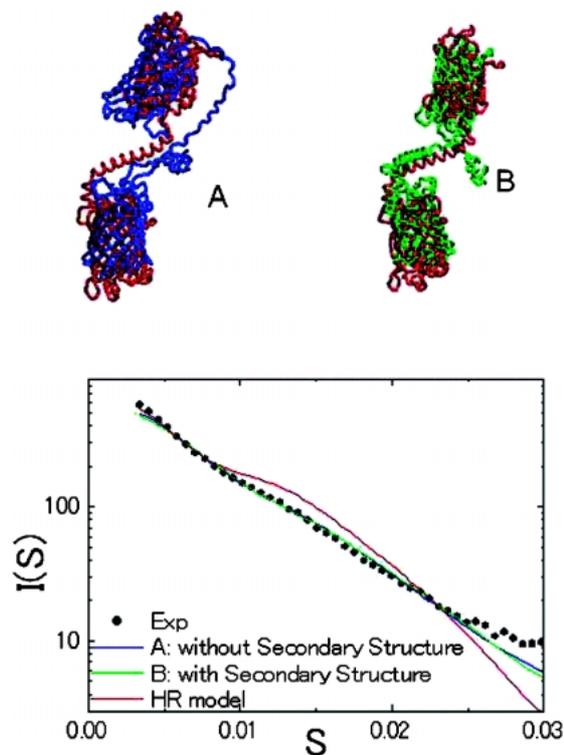


Fig. 9 Example of rigid body modeling method on the protein that incorporates the missing part of high-resolution structure. Experimental data was the same as Fig. 8. The blue model (A) ignored the secondary structure of missing part, while the green model (B) took into account of it. The red model is the high-resolution model same as Fig.6(C). The corresponding scattering curves were plotted in lower part.

配列を入力する。Fig. 9 は、His タグ、リンカー部を導入して rigid body modeling 法を行った結果である。欠損しているリンカー部分の 2 次構造を指定していない場合 (Fig. 9A) とヘリックス構造を指定した場合 (Fig. 9B) のいずれの場合も、散乱曲線に対するフィットが大幅に改善されている事が分かる。また、2 次構造の指定によらずモデルのドメインの位置は安定しており、*ab initio* 法に基づく高分解能モデル (Fig. 6C) の位置とも大きく変わらない。しかし、2 次構造の有無により予測されるドメインの向きが全く逆になってしまう。このように、タンパク質溶液散乱の解析において他の手法による情報は非常に重要な役割を果たす。

## 5. 考察

これまで紹介した解析例は現行のソフトの可能性を示すにはベストの系ではないかもしれないが、以下のような事が言えよう。第一に、タンパク質溶液散乱は解析アプローチに関わらず、タンパク質の 4 次構造と呼ばれるタンパク質及びサブユニット配置に関しては安定した結果が得られる。この点はタンパク質溶液散乱自体の情報量を示し、他の実験手法と比較して重要な特色となっている。第二に、単純に既存のソフトを実行したからといって一概に満足できる答えは得られないという事である。Fig. 9 の例のように、構造未知の部分をシミュレーションして散乱関数を最適化するなど、よりタンパク質のモデリング的作業が必要となっている。最後に、タンパク質溶液散乱のフィッティングに総じて言える事であるが、回転半径に代表される小角パラメーターを軽視してはならない。広い範囲の散乱曲線に対しフィットを行うと相対的に小角パラメーターを決定する小角部の寄与が小さくなってしまふ。球状のタンパク質複合体であるとそれほど問題にはならないかもしれないが、細長い形状や扁平な形状など異方性が増すにつれて小角部の情報が形状の決定に関して重要になってくる。

今後の方向としては、散乱曲線にいかにか構造情報を取り入れるかがポイントとなろう。散乱曲線の最適化のプログラムにより複雑なタンパク質の構造情報を取り込むのではなく、既存のタンパク質分子動力学のソフトを改良してタンパク質溶液散乱の情報を制限条件として構造モデルを作成するソフトが試みられているが<sup>35)</sup>、まだ完全とは言えない。また、溶液散乱自体の情報量を増やす試みが大事である。本特集号でも扱われているが、より散乱角の大きい中広角領域 (20, 30 Å<sup>-1</sup>~5 Å<sup>-1</sup> の領域) はタンパク質の折れ畳み構造に敏感に反映される事が知られている<sup>36,37)</sup>。実験的には今計測しているレベルの 2 桁, 3 桁小さいシグナルを計測しなくてはならず検出器等の開発も必要である。今後のタンパク質溶液散乱の解析に重要な役割を果たすであろう。

本稿は解析技術という事をテーマとしているが、放射光の重要性についても最後にぜひとも強調しておきたい。現在、PDB で登録されているタンパク質結晶構造で放射光を利用した例は最低でも 7 割は超すという。しかし、タンパク質溶液散乱では中性子を除けば実験室系の X 線小角散乱装置のデータで論文になっている例はほとんど無いといって過言でない。何故なら 2.1. の項で示したようにタンパク質溶液散乱の実験原理は非常に単純であるため、光源、光学系、検出器の性質が装置のバックグラウンド、寄生散乱を減らすのに決定的な役割を果たす。タンパク質溶液のようにシグナルが弱く、不安定で分子量も大きいサンプルから散乱を計測するには、放射光という施設においてタンパク質溶液散乱測定に特化したビームラインが不可欠であるというのが、長年ビームラインを建設・運営してきた筆者の信念である。昨今のように散乱データの情報をより細かく引き出す事が可能な時代においては、国際的に通用するデータを得る事、すなわち、散乱計測の正確さ、信頼性が大事である。欧米は言うに及ばず近隣アジア諸国でも小角散乱ビームラインの新規建設、更新が盛んに行われている。タンパク質溶液散乱の装置、解析はこの数年で確実に進んでおり、放射光を中心としたユーザー数も欧米では著しく増加している。タンパク質溶液散乱法は、長い歴史を持つが依然発展途上にあり、タンパク質構造のデータベースが発達する時代においては、その未来は明るいと筆者は考える。

## 謝辞

本稿の例となった人工タンパク質の溶液散乱は新井亮一博士との共同研究により計測されたものです。また本稿の執筆にあたり、私と共に理研構造生物学ビームライン BL45XU 小角散乱ステーションの建設、維持管理に携わった研究者、技術者、特に、前田雄一郎前主任研究員、西川幸宏博士、桑本滋氏、秋山修志博士、飯塚崇氏、芝田晃一氏、伊藤和輝博士、松尾博史博士の各位にこの場を借りて深く感謝いたします。

## 参考文献

- 1) M. H. J. Koch, P. Vachette and D. I. Svergun: Q Rev Biophys. **36**, 147-227 (2003).
- 2) D. I. Svergun and M. H. J. Koch: Rep. Prog. Phys. **66**, 1735-1782 (2003).
- 3) P. Vachette, M. H. Koch and D. I. Svergun: Methods Enzymol. **374**, 584-615 (2003).
- 4) L. A. Feigin and D. I. Svergun: Structure Analysis by Small-Angle X-Ray and Neutron Scattering (Plenum Press, NY, 1987).
- 5) A. Guinier and G. Fournet: Small-angle Scattering of X-rays (Wiley, New York, 1955).
- 6) H. B. Stuhmann: Zeit. fur Phys. Chem. Neu. Folg. **72**, 177-184 (1970).
- 7) P. B. Moore: J. Appl. Cryst. **13**, 168-175 (1980).
- 8) T. Fujisawa, K. Inoue, T. Oka, H. Iwamoto, T. Uruga, T.

- Kumasaka, Y. Inoko, N. Yagi, M. Yamamoto and T. Ueki : J. Appl. Cryst. **33**, 797–800 (2000).
- 9) S. Kuwamoto, S. Akiyama and T. Fujisawa: J Synchrotron Radiat **11**, 462–468 (2004).
  - 10) T. E. Creighton: Proteins (W. H. Freeman and Company, 1984) 242.
  - 11) O. Glatter: J. Appl. Cryst. **10**, 415–421 (1977).
  - 12) D. I. Svergun: J. Appl. Cryst. **25**, 495–503 (1992).
  - 13) T. Fujisawa, A. Kostyukova and Y. Maeda: FEBS Letters **498**, 67–71 (2001).
  - 14) O. Glatter and O. Kratky: Small Angle X-ray Scattering (Academic Press, London, 1982).
  - 15) P. Chacon, F. Moran, J. F. Diaz, E. Pantos and J. M. Andreu: Biophys J **74**, 2760–75 (1998).
  - 16) D. I. Svergun: Biophys J **76**, 2879–86 (1999).
  - 17) D. Walther, F. E. Cohen and S. Doniach: J. Appl. Cryst. **33**, 350–363 (2000).
  - 18) 藤澤哲郎：蛋白質・核酸・酵素増刊号「バイオ高性能機器・新技術利用マニュアル」(共立出版, 2004) 1687–1692.
  - 19) Y. Takahashi, Y. Nishikawa and T. Fujisawa: J. Applied Crystallography **36**, 549–552 (2003).
  - 20) R. Arai, W. Wriggers, Y. Nishikawa, T. Nagamune and T. Fujisawa: Proteins Structure, Function, and Bioinformatics **57**, 829–838 (2004).
  - 21) M. B. Kozin and D. I. Svergun: J. Appl. Cryst. **34**, 33–41 (2001).
  - 22) V. V. Volkov and D. I. Svergun: J. Appl. Cryst. **36**, 860–864 (2003).
  - 23) D. I. Svergun: <http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html>
  - 24) W. Wriggers, R. A. Milligan and J. A. McCammon: J. Structural Biology **125**, 185–195 (1999).
  - 25) W. Wriggers: <http://situs.biomachina.org/>
  - 26) W. Wriggers and P. Chacon: J. Appl. Cryst. **34**, 773–776 (2001).
  - 27) W. Humphrey, A. Dalke and K. Schulten: J. Molec. Graphics **14**, 33–38 (1996).
  - 28) D. I. Svergun, M. V. Petoukhov and M. H. J. Koch: Biophys J. **80**, 2946–2953 (2001).
  - 29) D. I. Svergun, C. Barberato and M. H. J. Koch: J. Appl. Cryst. **28**, 768–773 (1995).
  - 30) T. Fujisawa, T. Uruga, Z. Yamaizumi, Y. Inoko, S. Nishimura and T. Ueki: Journal of Biochemistry (Tokyo) **115**, 875–880 (1994).
  - 31) Y. Seki, T. Tomizawa, N. Khechinashvili and K. Soda: Biophysical Chemistry **95**, 235–252 (2002).
  - 32) M. Hirai, H. Iwase, T. Hayakawa, K. Miura and K. Inoue: J. Synchrotron Rad. **9**, 202–205 (2002).
  - 33) P. V. Konarev, M. V. Petoukhov and D. I. Svergun: J. Appl. Cryst. **34**, 527–532 (2001).
  - 34) M. V. Petoukhov and D. I. Svergun : Biophys J. **89**, 1237–1250 (2005).
  - 35) M. Kojima, A. A. Timchenko, J. Higo, K. Ito, H. Kihara and K. Takahashi: J. Appl. Cryst. **37**, 103–109 (2004).
  - 36) W. Zheng and S. Doniach: J. Mol. Biol. **316**, 173–187 (2002).
  - 37) A. V. Sokolova, V. V. Volkov and D. I. Svergun: J. Appl. Cryst., **36**, 865–868 (2003).

### ● 著者紹介 ●



#### 藤澤哲郎

独立行政法人理化学研究所 播磨研究所  
放射光科学総合研究センター 城生体金  
属科学研究室・先任研究員

E-mail: [fujisawa@spring8.or.jp](mailto:fujisawa@spring8.or.jp)

専門：生物物理学、タンパク質溶液散乱  
【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。工学博士取  
得，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年同先任研究員。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。2002年–兵庫  
県立大学連携大学院理学研究科客員助  
教授，2005年名古屋大学大学院理学研  
究科客員助教授。

## Solution Scattering from proteins—a bridge between low- and high-resolution structures—

**Tetsuro FUJISAWA** SPring-8 Center, RIKEN Harima Institute,  
1–1–1 Kouto, Sayo, Hyogo, 679–5148, Japan

**Abstract** As a number of fragmental protein structures are being solved, more interest is growing on the native structure of proteins in relation to their functions. There were two main obstacles to advance structural analyses from solution scattering data of proteins. The first was the calculation method of theoretical solution scattering for hydrated proteins from atomic coordinates. The second was the optimization of three-dimensional models from one-dimensional experimental data. Since the mid 1990's, the technology of scattering analyses has been greatly developed, owing to both modern efficient algorithm for non-linear optimization problem and speeding up the computational power. Exemplified our recent studies on designed proteins consisting of two domains with fluorescent probes, the usage of two program packages, ATSAS (Dr. Svergun, EMBL, Hamburg) and SITUS (Dr. Wriggers, Texas Univ. Houston), was outlined.