

表面分析技術選書

計算シミュレーションと分析データ解析

日本表面科学会 編

丸善株式会社

イントにつきまして解説いたしました。「理論計算はとっつきにくい」と思っていた方々が、本書をきっかけに「理論計算」をはじめていただけるようになりましたら幸いです。第2部は「実験解析のシミュレーション」の中から「電子線回折法」,「構造解析法」,「電子分光法」,及び「電子顕微鏡」を取り上げ、実験データ解析の計算シミュレーションを行う際の注意点につきまして解説致しました。実験データ解析におけるノウハウを本書で知っていただくことで、「専門外の実験」にもチャレンジしていただけることを期待しております。

最後に本書が複数の分野、複数の手法でのアプローチが必要不可欠な現代のナノテクノロジーの発展につながることを祈っております。

2007年11月

編集幹事

白石賢二

編集幹事・執筆者一覧

編集幹事

白石賢二 筑波大学
田沼繁夫 (独)物質・材料研究機構

執筆者 (五十音順)

荒井正浩	住友金属工業(株)	[7・2]
池谷鉄兵	首都大学東京	[6・2・2]
泉富士夫	(独)物質・材料研究機構	[6・1・1]
井田朋智	金沢大学	[7・1]
遠藤一央	金沢大学	[6・1・2]
大野隆央	(独)物質・材料研究機構	[3章]
甲斐荘正恒	名古屋大学・首都大学東京	[6・2・2]
川村隆明	山梨大学	[4・1]
熊坂崇	(財)高輝度光科学研究センター	[6・2・1]
重田諭吉	横浜市立大学	[5・2]
高羽洋充	東北大学	[1・1]
竹田精治	大阪大学	[8章]
竹中章郎	東京工業大学	[6・2・1]
舘野賢	筑波大学	[1・2]
田沼繁夫	(独)物質・材料研究機構	[7・2]
ベーター・ギンタート	首都大学東京/フランクフルト大学	[6・2・2]
水野清義	九州大学	[5・1]
村田顯二	大阪府立大学名誉教授	[4・2]
森健彦	東京工業大学	[2章]
安田雅昭	大阪府立大学	[4・2]

所属は2007年7月現在

[] 内は執筆担当箇所

引用・参考文献

- 1) C.Sauter *et al.*, *Acta Crystallogr.*, D57, 1119 (2001).
- 2) ソフトウェアの紹介として, 熊坂崇, 日本結晶学会誌, 42, 474 (2000) などがある.
- 3) タンパク質結晶解析法の成書として, 竹中章郎・勝部幸輝・笹田義夫共訳 (J. Drenth 著), “タンパク質の X 線結晶解析法”, シュプリンガー・フェアラーク東京(2000)がある.
- 4) T.Kaneko, N.Tanaka, T.Kumasaka, *Protein Sci.*, 14, 558 (2005).
- 5) T.Chatake, T.Hikima, A.Ono, Y.Ueno, A.Matsuda, A.Takenaka, *J. Mol. Biol.*, 294, 1223 (1999).
- 6) M.D.Winn, M.N.Isupov, G.N.Murshudov, *Acta Crystallogr.*, D57, 122 (2001).
- 7) D.E.Tronrud, *Methods Enzymol.*, 277, 306-319 (1997).
- 8) R.A.Engh, R.Huber, *Acta Crystallogr.*, A47, 392 (1991).
- 9) A.T.Brünger *et al.*, *Acta Crystallogr.*, D54, 905 (1998).
- 10) A.T.Brünger, L.M.Rice, *Methods Enzymol.*, 277, 243 (1997).
- 11) V.Schomaker, K.N.Trueblood, *Acta Crystallogr.*, B54, 507 (1998).
- 12) A.Kidera, N.Go, *J. Mol. Biol.*, 225, 457 (1992).
- 13) U.Stocker, W.F.van Gunsteren (M.G.Rossmann, E.Arnold, ed.), “International Tables for Crystallography”, Volume F, p.481, Kluwer (2001).

6・2・2 NMR による蛋白質構造決定の自動化

a. はじめに

蛋白質の立体構造決定の主要な技術には X 線結晶構造解析法と NMR 法があるが, 蛋白質構造データの国際的なデポジット機関である Protein Data Bank (PDB <http://www.rcsb.org/pdb/>) の年間登録総数の 85% は X 線によるものであり, NMR 法の寄与は現時点では 15% 程度に過ぎない. しかしながら, 蛋白質の結晶化が不可欠な X 線解析法と比べ, 蛋白質が機能を果たす状況により近い, 溶液内での立体構造解析が可能な NMR 法への期待は大きい. 今後, 蛋白質構造研究の中心が立体構造と生物機能の関連へとさらに大きくシフトするにつれ, NMR 法への期待はよりいっそう高まるに違いない. しかしながら, NMR による蛋白質構造解析技術は誕生以来 20 年足らずの, いわば発展途

上の手法である^{1,2)}. 従来の NMR 技術に頼る限り, 分子量が 20 kDa 程度の比較的小さな蛋白質を対象としてさえ, 高度な専門知識をもつ研究者が立体構造決定に数ヶ月の期間を要する. このような問題点は, 本稿で紹介する CYANA に代表される各種自動化アルゴリズムの開発により, 低分子量蛋白質 (20 kDa 以下) においては大幅に改善されつつあり, NMR 法による自動構造解析も実用的な段階になってきた.

一方で, 重要な生物学的機能をもつ蛋白質は 40-60 kDa 近傍の分子量をもつものが多く, 20 kDa 以下の低分子量蛋白質を対象にして発展してきた立体構造決定の迅速化技術をそのまま適用することは困難である. 30 kDa 程度の蛋白質においてさえ, 熟練した研究者の数ヶ月を超える長期の努力によってようやく蛋白質主鎖の折れ畳み構造 (global fold) が得られる程度である. X 線解析法においては対象となる分子量の限界は事実上存在しないこと, また多くの医薬開発においては側鎖立体配座を含む高精度な立体構造情報を必要とすることを考慮するならば, より高分子量の蛋白質の立体構造を迅速かつ高精度で解析できる実用的 NMR 技術を緊急に開発する必要がある. 甲斐荘らにより開発された SAIL (Stereo-Array Isotope Labeling; 立体整列同位体標識) 法^{3,4)} は, これまでの NMR 法のもつさまざまな問題点を一挙に緩和する革新的技術である. 従来の欧米を中心として展開されてきた技術開発は, NMR 測定・解析技術と構造決定アルゴリズムの改良にもっぱら向けられてきた. SAIL 法は, 全く視点を変え, 試料として用いる蛋白質を最適化させ, NMR 解析上の問題点を一挙に解決することを目指した新技術である. 近年の X 線解析法の急速な発展にとって蛋白質の発現・精製技術と結晶化技術の進歩が決定的な役割を演じたように, NMR 法においても蛋白質試料調製技術の高度化が不可欠であることは論を待たない. われわれは, 分子量範囲 40-60 kDa の高分子量蛋白質の自動 NMR 構造解析技術の開発においても, SAIL 蛋白質の利用を積極的に進めることが最重要課題であるとみなしている.

本章では, 主に既存の NMR 解析技術を基盤とした低分子量蛋白質の立体構造決定の自動化および構造計算の現状について解説するが, これらの自動化の取組みは今後 SAIL 法を基盤とした迅速, かつ高精度な全自動 NMR 構造決定解析法の開発へと繋がっていく. 章末において, SAIL 法についても若干ふれるのはこのためである.

b. 典型的な NMR 構造決定プロセス

NMR を用いた蛋白質の立体構造決定は、一般には次に示すような一連の流れに沿って行われる(図 6・19)^{1,2)}。

- (1) 試料蛋白質の調製
- (2) 各種多次元 NMR スペクトルの測定
- (3) NMR 測定データ処理
- (4) ノイズシグナルの選択的除去による試料シグナルの集積(ピークピッキング)
- (5) スペクトル解析による各原子由来のシグナルの帰属・同定
- (6) NOESY スペクトル解析による NOE, およびその他の構造制限データの収集
- (7) 集積された構造制限データを用いる蛋白質立体構造計算

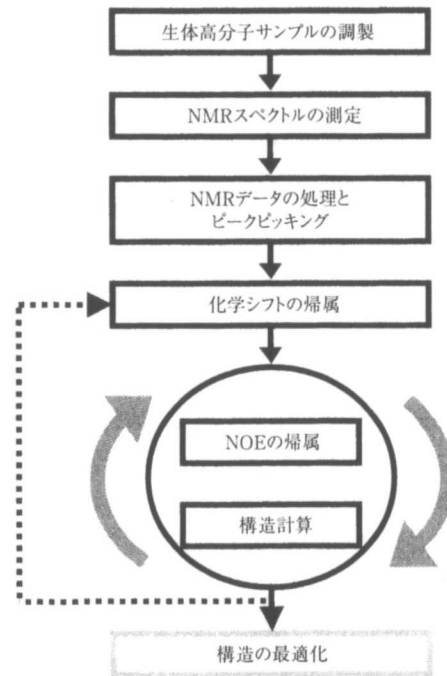


図 6・19 NMR 構造決定の一連の流れ。実線の矢印は、通常必須の流れを示し、破線の矢印は、必要に応じて追加される流れを示す

(8) 蛋白質立体構造の精密化

これら一連の流れのうち、(1)の蛋白質調製過程を除き、以下に簡潔にその内容を示す：

- (2) シグナル帰属のために必要な十数種類にのぼる多次元 NMR スペクトルの測定。NMR 測定装置を数週間にわたって占有する。
- (3) 複数の時間軸をもつ多次元 NMR 測定データを、シグナル感度の向上やノイズを抑制する数学的処理を加えた後に、フーリエ変換し多核種の共鳴周波数を軸とするスペクトルを得る。
- (4) 上記スペクトルにはノイズシグナルが含まれるため、試料由来のシグナルのみを選別し、それらのピーク位置(共鳴周波数)を確定し、ピークリストを作成する。
- (5)・(6) 確定したピークリストを実際の蛋白質の各原子に対応させる、いわゆる帰属を行う過程であり、一連の構造決定過程のなかで最も時間と経験を要する作業である。20 kDa 程度の蛋白質を対象とすれば、熟練したスペクトル解析の専門家が行ったとしても、手作業に頼る限り、数ヶ月程度を必要とする。
- (7)・(8) 帰属されたピークをもとにさまざまな構造制限データを集積し、それらを最もよく満足させる立体構造を計算し、さらにエネルギー最小化計算により立体構造の精密化を行う。

一連の流れの中で、(1)の試料調製以外は、アルゴリズムやコンピュータ処理の改良により、大幅な自動化と高速化が可能である。また(3)・(4)については、音声・画像認識等を含む異分野で発展しているさまざまな数学的手法を NMR スペクトル処理に応用する試みが報告されている。これらのうちの幾つかは将来大きく発展する可能性があるが、本稿では試料蛋白質の NMR ピークが得られた後の処理過程である(6)-(8)に焦点を絞り、コンピュータによるデータ解析の現状を解説する。

c. NMR 構造計算の自動化

(1) NMR 測定から得られる構造データ

NMR スペクトルから得られるデータのうち、生体高分子の構造決定に最も重要となるものは、上記の(6)で得られる ^1H - ^1H 間の NOE (核オーバーハウザー効果) データである。NOE ピークの強度 I と二つの水素原子 (^1H) 間距離

r との間には次式の関係式が成り立つ。

$$I \propto \frac{1}{r^6} \quad (6 \cdot 34)$$

よって、蛋白質を構成する NOE ピークから分子中の各水素原子間の距離情報をすべて収集できれば、それらの距離情報から蛋白質の立体構造を構築することが可能となる。式(6・34)からわかるように NOE は遠距離になると急速に減少する性質をもつため、利用できる NOE は、せいぜい 6 Å 程度以下の近距離の水素原子対由来のピークに限定される。このようにして得られた蛋白質構造の妥当性に関しては、1980 年代当時には大きな疑問をもたれたが、現在においてはそのような近距離情報であっても十分な数の NOE があれば、蛋白質の立体構造は精密に決定できることが実験的に立証されている。

いうまでもなく、NOE を測定するための NOESY スペクトル (NOE を測定するための手法) 中の各クロスピークは、分子中のどの水素原子対に由来して生じたピークであるかは測定直後の段階では不明である。したがって、構造計算に必要な距離制限に利用するためには、各 NOE ピークがどの水素原子対に由来しているかを同定する、いわゆる“帰属”が不可欠である。原理的には各ピークは対応する水素原子由来の二つの化学シフト座標により、それらの位置が決定されていることから、すべての水素原子の化学シフトが精密に決まっていれば NOE ピークの帰属は可能になるはずである。したがって、構造決定の前段階として、正確に、かついかに多くの ^1H -NMR シグナルの化学シフトを決定するかは、きわめて重要な段階となる。蛋白質は対称性を欠く不斉性をもつ分子であり、例えばメチレン (CH_2) を構成する二つの水素原子は異なった構造環境にあるため、化学シフトも異なる。したがって、蛋白質の分子量が増大するにつれシグナル数はきわめて増大し、スペクトル解析は急速に困難になる。現在はシグナルの分離を高めるためにすべての炭素を ^{13}C 、窒素を ^{15}N で均一標識し、 ^{13}C 、 ^{15}N の核スピンを利用し 10 種を超えるさまざまな多次元 NMR (三、四次元) スペクトルを 600-900 MHz (^1H -NMR の共鳴周波数) 級の高磁場 NMR 装置を用いて測定し、 ^1H -NMR シグナルの分離を高め、帰属を容易、かつ正確に行う手法が一般化されている。このような最新の手法を用いても、NMR スペクトルの解析には多大な労力を要し、いまだに 25 kDa 程度の“低分子量”蛋白質においても側鎖シグナルを含むスペクトル解析には多くの困難が存在する。

上でも述べたように、構造決定にとって最も重要な距離制限情報である NOE の帰属は、化学シフト表の完成度に大きく依存するために、その迅速化・精密化を目指した NMR スペクトル解析の自動化に関しては、これまでに多くの試みがなされてきた⁵⁾。残念ながら、現在までに開発されたプログラムは複雑な実際のスペクトル解析に幅広く適用できるものではなく、その多くは、かなり理想的なスペクトルデータを与えるモデルケースには有効であっても、構造解析に実用的レベルで利用するうえではさらなる改良が必要である。このため、化学シフトの帰属については、現在においても手作業、あるいは半自動解析が主流を占めている。手作業による化学シフトの帰属では、大部分のピークの帰属は比較的短時間で達成できたとしても、残りの少数のピークの帰属に難航し、多くの時間を費やしてしまうことがしばしばである。構造決定の過程をより円滑に進めていくためには、不完全な化学シフト表が入手できた段階で、NOE の帰属と構造計算に進み、必要とあれば、途中の粗い構造が得られた段階で、再度その構造情報を参照してより完成度の高い化学シフト表の作成に戻るほうがより効率的であろう。したがって、NOE スペクトルの自動帰属および構造計算のソフトウェアは、不完全な化学シフト表を用いても、適切な計算ができるようなアルゴリズムであることが求められる。なお、化学シフト表の不完全さと構造精度との関係については、J.G.Jee らが詳細に検討している⁶⁾。

(2) 自動構造計算のアルゴリズム

生体高分子のスペクトル解析では、NOESY スペクトルの帰属を行う作業はそれほど簡単なものではない。それらのスペクトルは、多くの偶然重なりあったシグナル、夾雑蛋白質・低分子物質に由来するノイズ、測定の際の人為的ピーク、蛋白質分子の動的状態を反映した多重ピークの出現やシグナルの部分的な欠失など、事前に予想しにくい数多くの可能性を含んでいるからである。このため、手作業による解析では、NOE の帰属と構造計算を何度も繰り返すことで、徐々に構造を組み立てていくという方法が一般的である。計算の途中経過の構造を参照しながら、ピークの帰属候補を限定していくことができるためである。それゆえ、構造情報をもたない解析の初期段階が、最も難度の高い段階といえる。初期構造に関する情報なしに自動構造解析を可能にするためには、以下の点を考慮したアルゴリズムの開発が必要である。

第一に、化学シフト表のみを参照して一義的に帰属できる NOE ピークのみ

では、粗い構造の出力さえも構造決定には不十分となる場合が多いことに留意すべきである。これは多くの場合、類似した化学シフト値をもつ水素原子が複数存在することに由来する。したがって、一義的に帰属が確定できない NOE ピークについても、構造計算の入力として用いることが必要となる。第二に、上述した NMR スペクトル中に含まれるノイズなどの“不正常”なピークの処理を自動的に行う過程が必要となる。最後に、現実的には化学シフト表は完全なものにはならず、一部欠失したものになってしまうため、それをもとに NOE の帰属を進めなければならない。このように、NOE の自動帰属アルゴリズムは、経験豊富な NMR 研究者がこうした不完全な NMR データを豊富な経験に基づいて処理する過程を模倣することが求められる。

現在、NOESY スペクトルの自動帰属と構造計算のプログラムとして幅広く用いられているものに、CYANA^{7,8)}と ARIA⁹⁾がある。ARIA は、構造計算プログラム CNS とともに用いられる。ARIA には、多くの新しいアルゴリズムが加えられているが、特に“曖昧さを含んだ距離制限”⁹⁾を構造計算に取り込んだ点が注目すべきところである。ARIA は、すでに初期構造が導かれている解析の中盤以降に用いると非常に効果的である。一方、CYANA は、化学シフトのリストを入力として二面角系の分子動力学計算と自動帰属を 7 回のサイクルに分けて交互に行う NMR 自動構造計算プログラムである。CYANA は、後述する Network-anchoring 法と Constraint combination 法⁸⁾を用いることで、効率的な自動帰属と構造計算を可能にしている。ここでは、CYANA を主に例にとりながら NOE の自動帰属と構造計算について解説する。

(3) 曖昧な NOE

すでに述べたように、最新の多次元 NMR 技術を用いたとしても、現実の NMR スペクトルの分解能は限界がある。したがって、解析の初期段階で作成した化学シフト表を参照して、NOE ピークのすべてを特定の水素原子対に一義的に帰属することは到底期待できない。よって、初期段階では一つの NOE ピークにいくつかの帰属候補を割り当てるが必要となる。われわれが利用できる距離制限としての NOE ピークはこのような一義的な帰属をもたないものを含むという意味で、“曖昧な NOE”である。NOESY シグナルの帰属に関して単純な数学的なモデルを立て、スペクトル中で一義的に帰属可能な NOE の割合を推定すると¹⁰⁾、概ね十数%程度となる。この程度の NOE のみを用いる限

り、正確な折畳み構造 (global fold) を得ることは困難である。一方で、精度は低くても基本的には正しい初期構造さえいったん得られれば、その構造情報を基に、さらに多数の“曖昧な NOE”を一義的に帰属することができる。単純化したモデル計算による見積りでは、正しい立体構造を参照できれば約 95% の“曖昧な NOE”の帰属を一義的に確定することが可能であると予測される。このことは逆にいえば、例え完全な立体構造を参照して NOE 帰属を行ったとしても、すべての NOE を一義的に帰属することは原理的にできないことも意味している。

以上の議論を踏まえて、“曖昧な NOE”を“曖昧な距離制限”として構造計算に効率よく用いるアルゴリズムの開発が自動 NOE 帰属、自動構造決定法の開発の重要な鍵となることは明らかである。一般に曖昧な距離制限は、“帰属の縮重”として処理される。そこでは、原子間距離の 6 乗分の 1 に比例した相対的な重みが使われる。一意に帰属できる曖昧さのない“一義的 NOE”の距離制限が、水素原子 α と β の距離 $d(\alpha, \beta)$ において NOE ピークの強度から計算される距離 b (距離上限) となるのに対し、 $n > 1$ 以上の帰属をもつ“曖昧な NOE”は、 n 個の縮重したシグナルの重なりとして表現され、 $\bar{d} \leq b$ の“曖昧な距離制限”として処理される。

$$\bar{d} = \left(\sum_{k=1}^n d_k^{-6} \right)^{-1/6} \quad (6 \cdot 35)$$

上式における d_k は、水素原子対 α_k と β_k の距離 $d_k = d(\alpha_k, \beta_k)$ を表している。式からわかるように、 r^{-6} で合計した距離 $\bar{d}$ は、つねにこの距離 d_k よりも短いので、曖昧な距離制限は、例えその中に誤った帰属候補を含んでいたとしても正しい帰属がある限り、その距離制限が算出される蛋白質の立体構造を大きく乱すことはない。

(4) Network-anchoring (ネットワークアンカリング) 法

Network-anchoring 法⁸⁾は、NOE から導かれた距離制限は、ある密度をもった複数の距離制限の“network”を形成するという考えに基づいて開発された (図 6・20)。一つの水素原子対に由来する NOE は、大部分は単独では存在することはなく、その原子対の周囲に同様な複数の NOE が存在するはずである。このような“network”の存在を想定することにより、Network-anchoring 法は、基本的に蛋白質の構造情報なしに、ピークの帰属候補を絞ることができる。

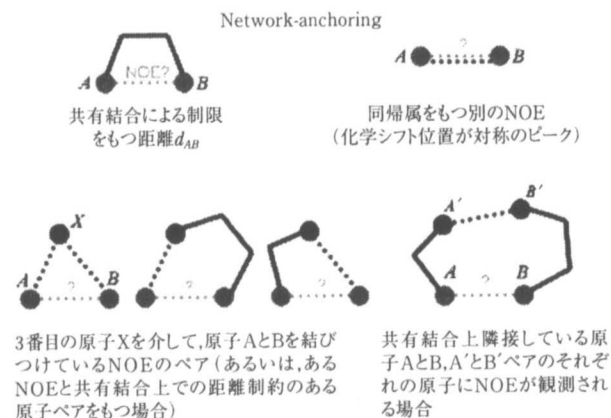


図 6・20 Network-anchoring 法の概略図

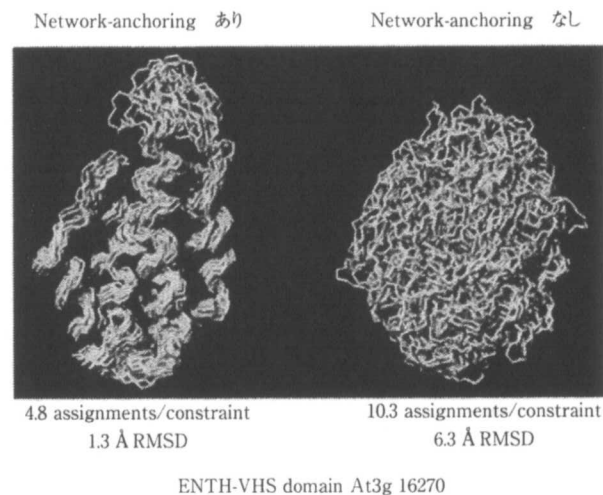


図 6・21 Network-anchoring 法を用いて構造計算を行った場合と、用いなかった場合の構造計算結果の比較 (Cycle 1 での構造)

Network-anchoring 法は、CYANA の最も重要なアルゴリズムになっており、構造計算の初期段階ですでにかなり精度よく折り畳まれた構造を得ることを可能にしている (図 6・21)。

CYANA では、Network-anchoring 法を以下のような形で計算に組み込んでいる (図 6・20)。各 NOE ピークの原子対候補 (水素原子 α , β) の Network-

anchoring のスコア P_{network} は、NOE ピークの強度から計算される距離 u (距離上限) よりも短い距離 $d_{\alpha\beta}$ をもつ原子対候補について計算される。一般には複数の候補をもつことから、 P_{network} は個々の確率の積として次式で表される。

$$P_{\text{network}} = 1 - (1 - P_1)(1 - P_2) \cdots \quad (6 \cdot 36)$$

P_{network} は、つねに個々の確率 $P_1, P_2, \dots$ より大きくなることから、確率 $P_1, P_2, \dots$ のいくつかの確率が高いと P_{network} 自体も高い値となる。確率 $P_1, P_2, \dots$ は、具体的には以下のように計算される。

(a) 距離上限 u よりも近い二つの原子対候補が蛋白質の慣性半径 R 内にある確率を計算する。

$$P_1(d_{\alpha\beta} \leq u) = (u/R)^3 \quad (6 \cdot 37)$$

(b) 共有結合構造では、2 原子間の距離 $d_{\alpha\beta}$ は、ある上限距離 c よりつねに短くなるはずである。

$$P_2(d_{\alpha\beta} \leq u) = \min((u/c)^3, 1) \quad (6 \cdot 38)$$

(c) NOE ピークは、対をなす水素原子の両方に観測できる場合が多い。しかしながら、対応する化学シフト位置に偶然ほかの NOE ピークが存在する可能性を完全に捨てることはできない。よって、帰属候補が同一 (α, β) である確率を P' とすると、

$$P_3(d_{\alpha\beta} \leq u) = P'(d_{\alpha\beta} \leq u') \cdot \min((u/u')^3, 1) \quad (6 \cdot 39)$$

と表すことができる。 $P'(d_{\alpha\beta} \leq u')$ は、距離上限 u' となる対称ピークが (α, β) に帰属される確率を示す。

(d) ある NOE ピークの帰属候補 (α, β) が、ほかの二つのピークの帰属候補 (α, γ), (β, γ) によって、帰属上で原子 α と β が 3 番目の原子 γ を介して network を形成している確率。

$$P_4(d_{\alpha\beta} \leq u) = P(d_{\alpha\gamma} \leq u_{\alpha\gamma}) \cdot P(d_{\beta\gamma} \leq u_{\beta\gamma}) \cdot f(u; u_{\alpha\gamma}, u_{\beta\gamma}) \quad (6 \cdot 40)$$

$P(d_{\alpha\gamma} \leq u_{\alpha\gamma})$ と $P(d_{\beta\gamma} \leq u_{\beta\gamma})$ は、それぞれ距離上限 $u_{\alpha\gamma}$ と $u_{\beta\gamma}$ をもつ帰属 (α, γ) と (β, γ) の二つの“間接的”NOE の確率を表す。関数 f は、距離 $d_{\alpha\beta}$ が距離上限 u より短い場合の存在確率を示す関数であり、二つの距離 $d_{\alpha\gamma}$ と $d_{\beta\gamma}$ が $u_{\alpha\gamma}$ と $u_{\beta\gamma}$ それぞれよりも短いときに与えられる。二つの NOE のうち一つは、共有結合の距離制限に置き換えることができる。その場合、NOE から計算される距離上限は、共有結合距離に置き換えられ、その確率は 1 になる。

(e) 共有結合上近接した原子 α と β が、同じく共有結合上の近接原子 α' と

β' と、それぞれの帰属によって結ばれるとすると、その確率は、

$$P_5(d_{\alpha\beta} \leq u) = P(d_{\alpha'\beta'} \leq u_{\alpha'\beta'}) \cdot g(u; c_{\alpha\alpha'}, c_{\beta\beta'}, u_{\alpha'\beta'}) \quad (6 \cdot 41)$$

となる。 $P(d_{\alpha'\beta'} \leq u_{\alpha'\beta'})$ と g は、式(6・40)とほぼ同様の関数になる。 $c_{\alpha\alpha'}$ と $c_{\beta\beta'}$ は、共有結合上の距離 $d_{\alpha\alpha'}$ と $d_{\beta\beta'}$ から導かれた距離上限である。

Network-anchoring 全体の確率は、間接的な“つながり”を反映させた上記(c)-(e)の関数値の積として、式(6・36)の形で計算される。Network-anchoring の確率は、ほかのピークの確率の情報を必要とするので、繰返し計算によって値が収束するまで実行される。こうして、NOE ピークリスト中のすべてのピークの帰属に関して同時に、Network-anchoring の寄与が計算される。

(5) Constraint combination 法

蛋白質の NMR 法による立体構造決定において、実測 NMR スペクトル中のノイズや人為的ピーク等の処理を、立体構造情報のない初期段階で不適切に実施することは、その後の解析過程に大きな影響を与え、ひいては致命的な立体構造の誤りを生み出しかねない。Constraint combination 法⁸⁾ は、一部の距離制限情報を一時的に犠牲にして、このような構造決定の初期段階における NOE データの解釈の誤りに起因する悪影響を最小限に留める目的で考えられた(図6・22)。この方法は、通常は無関係の異なるピークを結びつけて、新た

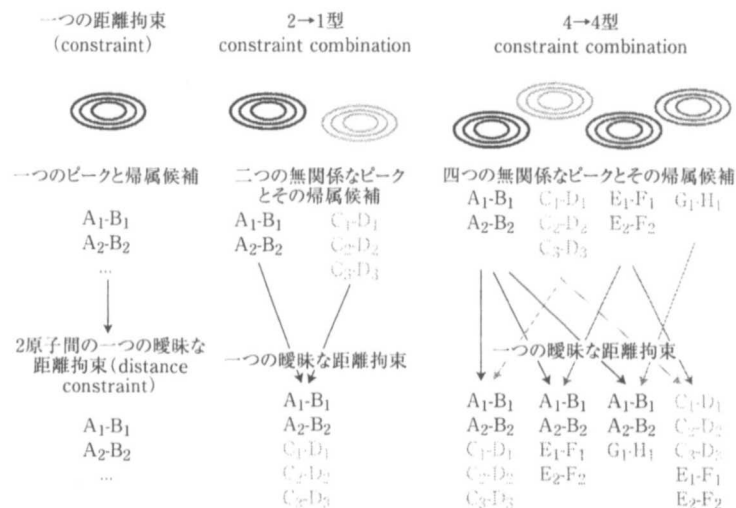


図 6・22 Constraint combination 法の概略図

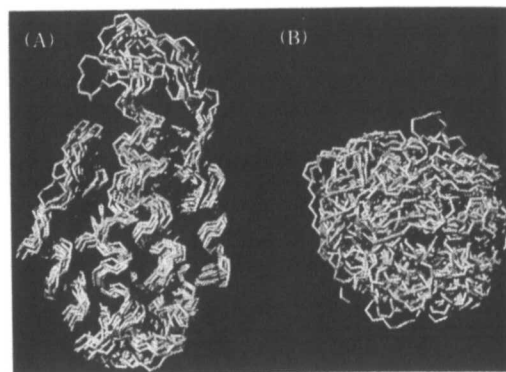


図 6・23 Constraint combination 法を用いて構造計算を行った場合 (A) と、用いなかった場合 (B) の構造計算結果の比較 (Cycle 1 での構造)

な距離制限を生み出す方法であり、CYANA プログラムでは、計算初期の2サイクル中で適用されている。すでに述べたように、“曖昧な距離制限”の基本的な考え方は、その帰属の候補の中に一つでも正しい帰属があるのであればたとえ誤った帰属が入っていたとしても、正しい構造を満たす距離制限を得ることができるというものである。すなわち、いくつかのピークを人為的に縮重させた距離制限を用いれば、元の距離制限をそのまま使うよりも誤りの確率を低くすることができる(図6・23)。

(6) 構造の最適化

NMR の構造計算では一般に非常に単純化した“柔らかい”エネルギー関数が利用される。これには主に、コンピュータによる計算効率と初期構造から最終構造への“折畳み経路”をスムーズに到達させる、という2点の理由によりとられる戦略である。構造計算の際に通過する“折畳み経路”として、標準的な分子動力学計算の力場で用いられる Lennard-Jones ポテンシャルのような急勾配の反発ポテンシャルを用いると、“折畳み経路”の途中でそれらが障壁となって、十分収束した構造を得ることができなくなり、いわゆる局所的安定構造“local minima”にトラップされる。そこで、Simulated annealing の初期ステージでは、原子どうしの“すり抜け”を許すことで、正しい立体構造とは異なった局所的安定構造にトラップされてしまうのを防ぐ。

しかしながら、このポテンシャルでは、物理的特性を表す分子間力や分子間反発力がかなり近似しているため、得られる構造は多くの場合、“低評価”のも

のになってしまう。ここでいう低評価とは、出力構造を、一般に使用されている蛋白質構造評価プログラムにかけると、比較的低い値を示してしまうということである。この理由は、これらの評価プログラムの多くは、Ramachandran plot やアミノ酸側鎖二面角周りの回転構造の妥当性、水素結合や静電相互作用を形成しやすい空間位置座標などを主な評価因子としているためである。したがって、このようなやや低評価の構造を修正するために、しばしば、水分子を陽(Explicit solvent model)もしくは陰(Implicit solvent model)に近似したエネルギー関数を用いて、短時間の分子動力学計算を実行するというような操作が行われる。ARIA プログラムでは、最終構造を最適化するために、水分子を蛋白質表面に薄く陽に加えた分子動力学計算を実行するようになっている¹¹⁾。

d. SAIL (Stereo-Array Isotope Labeling: 立体整列同位体標識) 法^{2,3)}

通常の NMR 構造解析用の蛋白質試料は、測定した NMR スペクトルには原理的には観測可能なすべての構造情報が含まれているものの、それらがすべて独立して観測できるわけではなく、またすべてが構造決定に必要というものでもない。むしろ、それらの構造情報には重複した、あるいは立体構造決定にはさして意味をもたない情報である場合も多い。現実的には蛋白質の分子量が増せば増すほど、NMR シグナルの重なるのためにスペクトルから、各残基あたりの距離制限として得られる情報はむしろ減少する傾向にある。最新の NMR 測定装置、複雑なパルス列を駆使した多次元 NMR 測定を用いても、分子量 25 kDa を実質的な限界として、それを超える分子量範囲にある蛋白質の NMR 構造決定がモデル構築的な色彩を帯びざるを得ないのはこのためである。先に述べたように、これまでの欧米を中心とした NMR 技術開発には、容易に調製できる試料を力任せに解析しようとするだけであり、NMR 法の特徴を見極め、蛋白質試料自体をその特徴に合わせて設計するという観点が完全に欠けていた。もし、われわれが重複した構造情報を与える水素を重水素化により徹底的に削減し、立体構造決定に有用な構造情報のみを残すような試料を調製できるならば、NMR スペクトルは必要な構造情報は一切失うことなく大幅に簡略化できるであろう。このような NMR 解析に最適化した蛋白質試料を利用すれば、本稿で述べてきた NMR 構造解析の自動化を困難としてきたさまざまな要因の多くを著しく軽減することができるであろう。また、そのような試料を用いれ

ば、構造決定速度の大幅な向上のみでなく、従来の限界を大きく超えた高分子量蛋白質の解析も可能になり、また得られる立体構造精度の大幅な改善も同時に達成される。

構造決定に必要な NMR 情報を削減する方法は実際には無数にある。しかしながら、ある一定の原則に従って蛋白質を構成する 20 種類のアミノ酸について最適な重水素置換体を設計することにより、それらから構成される蛋白質自体を NMR 構造情報の取得に最適化することが可能である。このためには、位置、および立体選択的な重水素化を徹底的に進めるとともに、多核種多次元 NMR の利用を前提として、NMR 測定技術上の必要性に応じてアミノ酸中の炭素や窒素原子を、適切に ^{13}C 、 ^{15}N 置換することが必要となる。これらの要件を満たすアミノ酸の設計指針は以下の 4 点にまとめられる：

- (1) アミノ酸残基中のすべてのメチレン基に存在する 2 個の“プロキラル”水素は必ずどちらか一方を立体選択的に重水素置換する(立体選択的重水素化)
- (2) メチル基の 3 個の水素のうち、2 個を重水素化する(プロトン密度の極小化)
- (3) 芳香環水素を位置選択的に部分水素化する(位置選択的重水素化+プロトン密度の極小化)
- (4) Leu, Val の 2 個の“プロキラル”メチル基のうち、一方のみを立体選択的に(2)のように残余水素が 1 個となるように標識し、他方のメチル基はすべて重水素化する(立体選択的重水素化+プロトン密度の極小化)

以上のような指針に従って設計するアミノ酸においては、メチレン(CH_2)とメチル(CH_3)は、実質的にはすべてメチン(^{13}C - ^1H 対)となり、立体構造決定のみならず、従来の標識蛋白質試料を利用した場合と比較し、あらゆる NMR 実験が遥かに高感度、高精度、かつ迅速に可能となる仕掛けが合成段階で組み込まれているのである。われわれは上記の設計方針に従って ^2H 、 ^{13}C 、 ^{15}N -三重標識したアミノ酸を用いて蛋白質の全アミノ酸残基を同時一斉に安定同位体標識する方法を“SAIL 法”(Stereo-Array Isotope Labeling: 立体整列同位体標識法)と名づけた³⁾。このような三重標識アミノ酸(SAIL アミノ酸)において、新たに生じる不斉中心をすべて高い立体選択性(95% d.e.以上)を保ちつ

つ合成することに成功している (図 6・24)。

図 6・24 のなかで, Leu を例にとって説明する. SAIL-Leu の 2 個のメチル基, および 2 個のメチレンプロトン, は, 立体化学の言葉で“ジアステレオトピック”な環境にあり, それぞれ異なった化学シフトをもつ. これらの NMR シグナルを, どちら側のプロトンに由来するのかを決定することを立体特異的帰属とよび, NMR により決定される蛋白質の立体構造精度に大きく影響する重要なステップである. しかしながら, 各二つ含まれるメチル基とメチレンプロトンのシグナルの各々は基本的には重複した構造情報をもたらすため, 立体特異

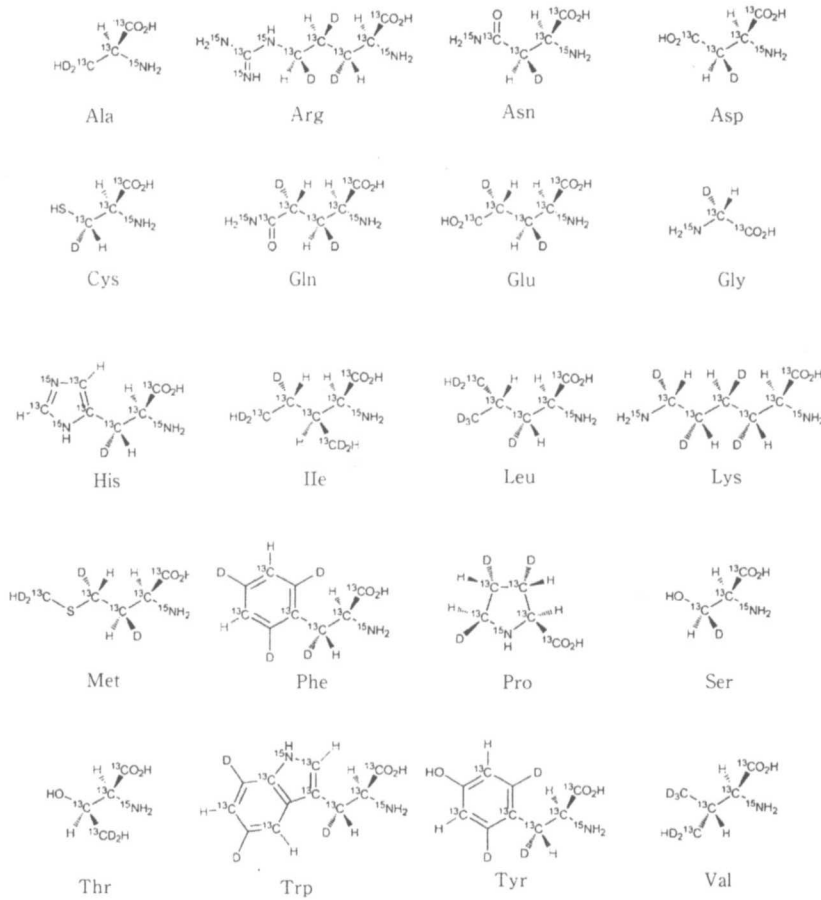


図 6・24 20 種類の SAIL アミノ酸の構造

的に帰属がなされてさえいれば, どちらか一方の NMR シグナルが観測できれば, 構造決定に必要かつ十分な構造情報となる. したがって, すべてのアミノ酸残基を SAIL-アミノ酸に置換した蛋白質を構造解析用の試料として用いれば, 必要な立体構造情報を失うことなく NMR スペクトル上から重複した構造情報量 (シグナル数) を削減することができる.

(2) SAIL 法による蛋白質の立体構造決定

従来の蛋白質立体構造決定技術において, 広く利用されてきた均一 [^{13}C , ^{15}N] 二重標識蛋白質と比べ, SAIL-蛋白質では観測されるシグナルの絶対数が選択的重水素化により大きく減少するために, より分離した NMR スペクトルを, 遥かに高感度 (短時間) で測定できることが明らかとなった (図 6・25). これらの特徴を利用することにより, 各残基の主鎖・側鎖の全炭素 (^{13}C), 窒素 (^{15}N), および水素 (^1H) 核の NMR シグナルの化学シフトの精密な決定が容易に達成できる. メチレンプロトンのようにプロキラル炭素上 ^1H -NMR シグナルの場

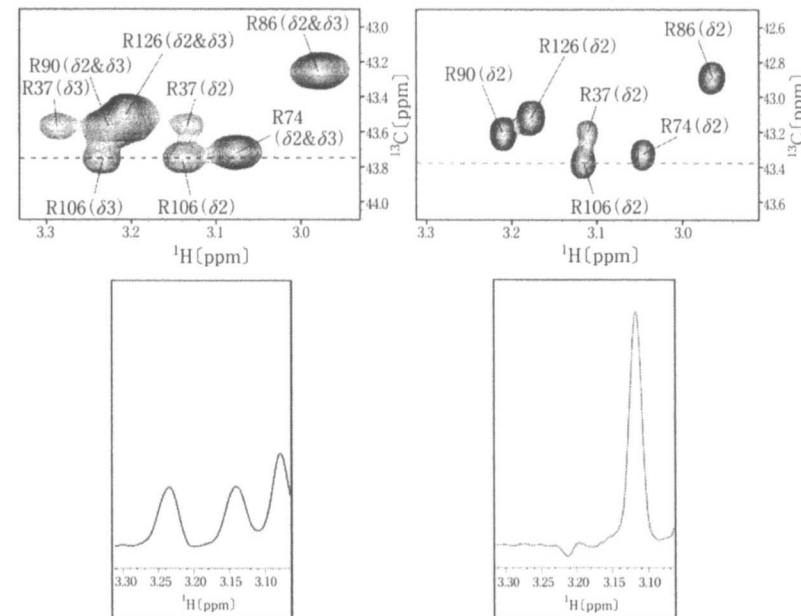


図 6・25 通常の [^{13}C , ^{15}N]-均二重標識カルモジュリンと SAIL カルモジュリンの ^1H - ^{13}C -HSQC (二次元相関スペクトル) の比較. アルギニンの H_α - C_α 部分のみを示してある. SAIL 試料のシグナルの線幅の減少と感度の上昇は破線部分のスライスから明瞭である. 青 (二重標識), および赤 (SAIL)



図 6・26 SAIL-CYANA 法による自動構造計算を行った際に得られたカルモジュリンの立体構造、リボンモデル(NMR 構造)、ワイヤーフレームモデル(X 線構造)

合には、蛋白質の精密構造解析にあたっては、それらの立体特異的帰属が必要である。こうした NMR 解析上の難点の一つとされてきた立体特異的帰属は、SAIL 法の場合、アミノ酸合成段階の立体選択的重水素化によりすでに情報として含まれている。すでに述べたように、NMR による蛋白質の立体構造決定は、核間距離が 5-6 Å 以下の水素原子対に現れる NOE 情報を集積し、それらの近距離にある水素原子対をすべて満足する立体構造を計算することにより達成される。NOE ピークの帰属は、分子量が大きくなるにつれその数が膨大となるために帰属は大変な作業となり、数ヶ月に及ぶことも普通である。しかしながら、SAIL-蛋白質を用いれば、立体特異的帰属を含めた“完全な”化学シフトデータ表の作成が可能となるため、CYANA などの NOE 自動帰属ソフトウェアの利用によって、精密な立体構造がきわめて短時間で決定できる。図 6・26 に示した構造は、SAIL-CaM の NOESY スペクトルのデータテーブルのみを用いて、人手による NOE 帰属を全く行わずに、CYANA で構造計算させた結果である。数十台の計算機クラスターを用いれば、わずか 30 分程度で構造が収束し、図に示したような最終構造を得ることができる。このようにして自動的に得られた構造は NMR 構造としてはきわめて高い精度をもっている。この方法は、分子量 41 kDa の MBP (maltose binding protein) など、従来の手法によっては精密構造が決定できない高分子量蛋白質にも適用され良好な結果が得られている¹²⁾。

(5) 終わりに

ポストゲノム配列時代の構造生物学・構造ゲノム学における NMR 法による

構造解析は、迅速、かつ高精度の解析と同時に、より高分子量の蛋白質に適用可能な技術の開発が求められている。本稿で述べた NMR 構造解析技術の自動化と構造計算への期待は、SAIL 技術との組合せによる生物学的により大きな興味もたれる分子量 40-60 kDa 領域にある高分子量蛋白質の完全自動構造解析、さらには膜蛋白質や超分子複合体への NMR 法に適用へと広がるであろう¹³⁾。これまで欧米で開発された技術を利用してきたわが国が発信する次世代の世界標準として、SAIL-CYANA 法が大きく成長することが筆者らの願いである。

引用・参考文献

- 1) Wüthrich K, "NMR of proteins and nucleic acids", Wiley, New York (1986).
- 2) Wüthrich K, "タンパク質と核酸の NMR-二次元 NMR による構造解析", 東京化学同人 (1991).
- 3) 大木進野, 寺内勉, 甲斐荘正恒, ゲノム蛋白質の高効率, 高精度 NMR 解析法の開発, ケミカルエンジニアリング, **47**, 7 (2002).
- 4) 鳥澤拓也, 寺内勉, 甲斐荘正恒, タンパク質の立体構造解析のための新しい安定同位体利用 NMR 技術—SAIL 法の開発, 生物物理, **44**(5), 200-205 (2004).
- 5) Moseley, HNB, Montelione, GT, Automated analysis of NMR assignments and structures for proteins. *Curr.Op.Struct.Biol.*, **9**, 635-642 (1999).
- 6) Jee JG, Güntert P, Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *J. Struct.Funct.Genom.*, **4**, 179-189 (2003).
- 7) Herrmann T, Güntert P, Wüthrich K, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J.Mol.Biol.*, **319**, 209-227 (2002).
- 8) Güntert P, Automated NMR protein structure calculation with CYANA. *Meth. J.Mol.Biol.*, **278**, 353-378 (2004).
- 9) Nilges M, Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J.Mol.Biol.*, **245**, 645-660 (1995).
- 10) Mumenthaler C, Braun W, Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting

- distance geometry. *J. Mol. Biol.*, **254**, 465-480 (1995).
- 11) Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M, Refinement of protein structures in explicit solvent. *Proteins.*, **50**, 496-506 (2004).
 - 12) Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P, Optimal isotope labelling for NMR protein structure determinations. *Nature*, **440**, 52-57 (2006).
 - 13) Ikeya T, Terauchi T, Güntert P, Kainosho M, Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA. *Magn. Reson. Chem.*, **44**, S152-S157 (2006).

電子分光法

7・1 化学シフト

X線光電子分光(X-ray Photoelectron Spectroscopy; XPS)の特徴は、原子の化学結合状態に関する情報を得ることができることである。電子は核に静電力で束縛され、内殻電子は価電子帯の電子より強く核と結合し化学結合に直接関与していないが、化学結合に寄与している価電子帯の電子の影響を受けて核との静電力に差を生じる。これは同一元素で異なる状態を有する原子が、異なる内殻電子結合エネルギーを最大 10 eV 程度生じさせるという XPS のピークシフトとして観測された。この内殻電子結合エネルギーのピークシフトは核磁気共鳴(NMR)のアナロジーから化学シフトとよばれることとなり、定性的に隣接原子の電気陰性度の大きさによって説明され、比較的簡単な電荷ポテンシャルを用いて説明されていた。しかし、このピークシフトは物質固有原子の内殻電子結合エネルギー差にほかならない。一方、最近のコンピュータのソフトおよびハードの進歩により、XPS で観測される物質原子の各電子の結合エネルギーは、量子化学計算法(分子軌道法や密度汎関数法)による物質モデルの計算からその電子の結合エネルギーを高い精度で対応させることができる。そこで、ここでは量子化学的な基礎理論から原子、分子および固体物質の順に内殻電子結合エネルギー(Core-Electron Binding Energy; CEBE)の計算方法とその計算値を XPS による実測値と比較して述べる。

7・1・1 原子、分子の CEBE

量子化学理論とは、原子や分子中の電子の運動をシュレディンガー方程式によって記述し、軌道概念や電子密度に対応するポテンシャルを導入して得られ