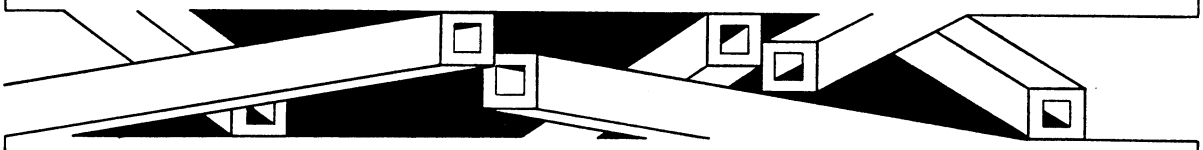
蛋白質原子座標データとその情報処理



郷　信広※　水野　裕重※　郷　通子※※
宮澤　三造※※　武富　敬※※※

X線解析により原子座標データの得られている球状蛋白質分子の数は増大しつつある。それに伴い原子座標データに基づく新しいタイプの蛋白質研究が可能となってきた。一般の蛋白質研究者にとっても，原子座標データに基づく情報を入手したいと思うことが多くなったと思われる。そのような一般の研究者を対象として，まず第1に原子座標データの入手法を述べる。これに関連してプロテイン・データ・バンク（PDB）の活動を紹介する。第2に，座標データの利用に関連し「蛋白質原子座標の情報処理プログラム・システム（PSPCS)」を紹介する。これは筆者らを含むグループで過去2年間開発を進めてきたもので，すでに多様なプログラムが作成されている。しかし，システムの理想から見ると現状には，まだ問題点が多い。これに関連し，わが国の蛋白質関係の情報収集・管理・処理体制の将来についての夢を述べる。

I. はじめに

X線解析がなされ，原子座標データの得られている球状蛋白質分子の数は，現在では 100 近くあるといわれている。X線解析のなされた蛋白質の数の増大に伴って，一般の蛋白質研究者にも，原子レベルにおける立体構造のイメージがより親しみを増してきた。その結果，どのような側面の蛋白質研究にも，そのイメージがなんらかの影響を及ぼしつつあるように思われる。今後は，ますますその傾向は強まるだろう。そして一般の蛋白質の研究者も，原子座標データを実際に手にして，そのデータを処理して自分に必要な情報を自ら作り出していく方向に動いていくのではないだろうか。

蛋白質原子座標データは，現在では世界的な組織であるプロテイン・データ・バンク（以下 PDB という）が収集・配布にあたっている。本文の目的の第1は，PDB を紹介し，わが国の研究者が PDB からデータの提供を受けるためにはどうしたらよいかを具体的に述べることにある。

データを入手する所までは，PDB の組織的活動のお陰で容易になった。しかし，データを目の前にしても，そこから自分の研究に必要な情報を引き出すことは必ずしも容易ではない。本文の目的の第2は，蛋白質原子座標を加工し，必要な情報を取り出すための計算機プログラムの開発には，どのような問題があるかを述べることにある。わが国では，いくつかの研究グループがそのような計算機プログラムの開発にとり組んでいる。第2の目的には，それらの活動を紹介することも含まれる。筆者らが属している PSPCS 作成グループ（後述）の活動の紹介が主となる。

そして最後に，わが国における蛋白質に関する情報収集・処理体制の将来についての夢を語りたい。

II. 原子座標を用いる蛋白質研究

蛋白質の数は大変に多いので，わずか 100 個たらずの蛋白質の原子座標が，新しく同定された蛋白質の立体構造の研究に直接に役立つことはない。それにもかかわらず，100 個たらずの蛋白質分子の原子座標データは，蛋白質研究一般に対して，大きな影響力を持っている。それは原子座標データが新しいタイプの研究を可能にしているからである。蛋白質原子座標データを用いる新しいタイプの研究は，おおよそ次の3つに大別できるように思われる。

第1は，X線解析のなされた蛋白質分子について，高度な物理的・化学的研究を，原子座標データから得られた立体構造の知見に基づいて，立案し実行するものである。このような蛋白質分子の数は多くはないが，それらについてなされてきた研究は，蛋白質研究における先駆

※ Nobuhiro Gō, Hiroshige Mizuno, 九州大学理学部物理
※※ Mitiko Gō, Sanzo Miyazawa, 九州大学理学部生物
※※※ Hiroshi Taketomi, 九州大学大型計算機センター
（〒812 福岡市東区箱崎 6-10-1）

Protein atomic coordinate data and their information processing

的役割を果たしてきたし，今後も果たしていくであろう。

第2は，蛋白質分子が生物進化の所産であることに由来している。多くの蛋白質分子は，個々ばらばらに存在するのではなく，進化の過程を通して，いろいろな類縁関係で結ばれている。そして，蛋白質分子の立体構造は，分子レベルにおける進化の過程を通じてかなり安定であることが認識されている。そのため，新しく同定された蛋白質分子についても，アミノ酸配列のホモロジーその他の理由により，既知の蛋白質分子との類縁関係が推定される場合には，その立体構造を論ずる可能性が開ける。また，類縁関係の推定される蛋白質分子の立体構造を比較することにより，進化の過程を解明していこうとする研究も盛んになりつつある。

第3は，原子座標が既知の蛋白質の立体構造を比較検討して，アミノ酸配列から立体構造が決定されていく仕組みを解明しようとする研究である。このタイプの研究の1つの典型として，次の例をあげることができる。まず，既知の立体構造中に，α-ヘリックス，β-シートなどの特徴ある構造を把握する。そして，それらの特徴ある構造を構成しているポリペプチド鎖中のアミノ酸配列になにか特徴がないかを探る。すなわち，アミノ酸配列と特徴ある構造との間の相関を求め，それを経験則として整理する。相関の度合が高ければ，逆にアミノ酸配列から立体構造を予言することにも応用できる。この種の経験則を積み上げていこうとする研究はかなりの成果を収めている。相関の経験則ばかりでなく，既知の立体構造の比較検討から，球状の特異的立体構造への折れたたみ過程における核形成の役割を議論する試みもなされている。

これらの研究においては，X線解析で得られた原子座標データを様々に処理する必要がある。処理プログラムには様々のものがある。今後の研究の進展に応じて，さらに新しいいろいろのプログラムの開発が必要となってくるであろう。ここでは，それらの中から任意に次の2つの例を考えてみよう。

第1の例として，原子座標データのある任意の蛋白質中の任意の1個または数個の残基の主鎖，および側鎖の原子および化学結合の立体配置の図を描くプログラムを考える。このような図は，たとえば活性部位のような特定の残基に着目し，それの置かれている分子内のミクロな環境を把握するのに大変に役に立つ。

第2の例として，蛋白質中の特徴のある局所構造（たとえば，α-ヘリックスやβ-シートやターンのような2次構造）を探索するプログラムを考える。このような構造を認識し探索することは，上で述べた蛋白質の原子座標を用いる研究のうち，第2と第3のタイプの研究ではしばしば必要となることである。

さてこのような処理プログラムが欲しいと思ったとして，簡単に書けるかというと決してそれほど簡単ではない。日頃計算機を使いなれていない人にとっては障害はかなり大きく，あれば良いとは思っても，結局なしですませる所におちついているのが現状ではなかろうか。本文は主としてこのような人々を読者として想定している。この小文を読んでいただくと問題が解決するかというと，残念ながらそうはいかないが，どんなことが問題かは少しわかっていただけると思う。プログラムの種類によっては，作成者に連絡を取っていただければ，問題解決となるかもしれない。まず順序として，蛋白質原子座標データの入手法から述べる。

III. 原子座標の入手
——プロテイン・データ・バンク（PDB）の紹介——

PDBは，蛋白質の結晶学的データを収集・標準化・配布する目的で，1971年にイギリスのケンブリッジの結晶学データ・センター（CDC）とアメリカのブルックヘブン国立研究所との共同で発足した[1,2]。現在では収集範囲を核酸にも広げ，さらにはデータ処理プログラムをも含むようになっている[3]。PDBからの最新のニュースレター，第6号（1978年5月）によれば，90個の蛋白質原子座標データと，3個のtRNA（すべて酵母のPheで，データの登録者が異なっている），および2個のデータ処理プログラムが含まれている。それ以外にも，非標準データとして，16個の蛋白質の構造因子と位相，および2個の蛋白質中の2面角が含まれている。原子座標データのある90個の蛋白質の中には，リファインメントの違いのみで区別しているものもあるので，本当に異なる蛋白質分子の数は，90個よりはいく分少ない。

蛋白質分子は，リゾチームのように比較的小さい分子でも水素原子以外の原子を約1000持っている。各原子がx-, y-, z-座標を持っているので，蛋白質分子1個分のデータには数千の数が含まれる。このように多量のデータは，一般には一次論文には発表されない。そのかわりに，PDBに預け入れられる。一般のデータ・バンクが，発表された一次文献から情報を抽出して構成されるのに対して，PDBはデータの預け入れ機関としての性格を持ち，データの生産者と利用者の仲介をしている。

PDBが発足した当時は，ケンブリッジのグループによる140字フォーマット（CDCフォーマット）が，標準フォーマットとして使用されていた。その後，データの入力，管理，保管がブルックヘブンのグループのもとで行なわれるシステムに移行したことに伴って，現在ではこのグループによる80字フォーマット（BNLフォーマット）を用いて標準化がなされている。構造因子と位

```
HEADER    ELECTRON TRANSPORT                              01-AUG-76   1CYC      1CYC   3
COMPND    FERROCYTOCHROME C                                                     1CYC   4
SOURCE    BONITO (TUNA) HEART                                                   1CYC   5
AUTHOR    N. TANAKA, T. YAMANE, T. TSUKIHARA, T. ASHIDA AND M. KAKUDO           1CYC   6
REMARK   1                                                                      1CYC   7
REMARK   1 REFERENCE 1. N. TANAKA, T. YAMANE, T. TSUKIHARA, T. ASHIDA           1CYC   8
REMARK   1 AND M. KAKUDO, THE CRYSTAL STRUCTURE OF BONITO (KATSUO)              1CYC   9
REMARK   1 FERROCYTOCHROME C AT 2.3 ANGSTOM RESOLUTION, II. STRUCTURE          1CYC  10
REMARK   1 AND FUNCTION, J. BIOCHEM., VOL. 77, P147 (1975)                     1CYC  11
REMARK   2                                                                      1CYC  12
REMARK   2 RESOLUTION. 2.3 ANGSTROMS                                           1CYC  13

SEQRES   1     103  GLY ASP VAL ALA LYS GLY LYS LYS THR PHE VAL GLN LYS        1CYC  22
SEQRES   2     103  CYS ALA GLN CYS HIS THR VAL GLU ASN GLY GLY LYS HIS        1CYC  23
SEQRES   3     103  LYS VAL GLY PRO ASN LEU TRP GLY LEU PHE GLY ARG LYS        1CYC  24

HET    HEM      1      43     PROTOPORPHYRIN IX PROSTHETIC GROUP.              1CYC  30
HELIX    1  H1 GLY      1   VAL      11  1                                     1CYC  31
HELIX    2  H4 CYS     14   CYS      17  5 14 AND 17 BOUND TO HEME GROUP       1CYC  32
HELIX    3  H5 THR     49   LYS      55  5 LOOSE FROM 49-53,3/10 53-55         1CYC  33
HELIX    4  H2 ASN     60   GLU      69  1                                     1CYC  34
HELIX    5  H3 GLU     90   SER     103  1                                     1CYC  35
TURN     1  T1 ILE    75   THR      78      TYPE II                            1CYC  36
TURN     2  T2 LYS    53   GLY      56      TYPE I                             1CYC  37
TURN     3  T3 CYS    14   CYS      17      TYPE I (NOTED AS H2 ABOVE)         1CYC  38
CRYST1   57.680   84.580   37.830  90.00  90.00  90.00 P 21 21 21    8         1CYC  39

ATOM      2  CA  GLY     1     -19.878  14.593  -7.963  1.00  0.00             1CYC  50
ATOM      3  C   GLY     1     -18.996  14.530  -6.703  1.00  0.00             1CYC  51
ATOM      4  O   GLY     1     -18.114  13.711  -6.640  1.00  0.00             1CYC  52
ATOM      5  N   ASP     2     -18.996  15.412  -5.821  1.00  0.00             1CYC  53
ATOM      6  CA  ASP     2     -17.610  15.097  -5.191  1.00  0.00             1CYC  54

HETATM  844  O1D HEM     1     -14.082   5.773  18.560  1.00  0.00             1CYC 886
HETATM  845  O2D HEM     1     -15.405   5.647  20.639  1.00  0.00             1CYC 887
HETATM  846  O   HOH     2     -16.496   9.304  17.402  1.00  0.00             1CYC 888
CONECT  104  103  825                                                          1CYC 889
CONECT  124  123  833                                                          1CYC 890
CONECT  134  132  133  803                                                     1CYC 891
```

図 1. プロテイン・データ・バンク (PDB) のチトクローム c 還元型のデータの抜粋
データ・フォーマットの概要については，本文に説明した（文献 4 より再録）

相とは，標準化がなされておらず，データ登録者によるフォーマットがそのまま使われている。BNL フォーマットは，PDB File Record Formats というタイプ印刷の説明書に記載されており，後述の方法で入手できる。ここではその概要を例について説明しよう。図 1 は，角戸・安岡による PDB などの紹介記事[4]から再録させていただいたもので，蛋白研で解析されたカツオの心筋のチトクローム c 還元型のデータの抜粋である。まず一見してわかるように 80 字のカード・イメージでできており，最初の 6 欄に record type identifier が記されている。最初の HEADER カードには，この蛋白質の機能，PDB への登録の日付，およびこのデータの identification code が記されている。この identification code は，3 文字で蛋白質を表わし，その前につけた 1 桁の数字で，精密化の程度・空間群の違いなどの異なるデータと区別する。SEQRES には，アミノ酸残基数とアミノ酸配列が与えられる。HET には，標準的アミノ酸残基以外の残基などの名称（3 文字からなる identifier）と，それに属する原子数などが記される。HELIX·SHEET·TURN のカードには，著者の報告した 2 次構造が記載される。CRYST 1 は，ユニット・セルの形状を記載し，図 1 には省略されているが，ORIGX と SCALE には，このデータ中でオングストローム単位で与えてある原子座標を，普通登録者がデータを預け入れるときに使う一般には直交系でない座標系に変換するマトリックス，および結晶学的な少数表示の座標に変換するマトリックスが与えられる。ATOM には原子の通し番号，原子名，属するアミノ酸残基名とその番号，x-, y-, z-座標値，占有率，温度因子がこの順で記載されており，これがデータの大半を占める。HETATM には，標準的アミノ酸残基以外の残基などに属する原子の x-, y-, z-座標値などが与えられる。1 つの残基中の原子を配列する順序は，定められており，前述の PDB File Record Formats に記載されている。CONECT には，アミノ酸残基名，原子名のみでは完全に決めることができない共有結合と，水素結合および塩結合に参加している原子の通し番号が与えてある。その他は，書誌的データが主で，その内容は容易に推察がつくと思われる。

データの入力・管理は，ブルックヘブンで行なわれ，配布は，4 ヵ所の窓口（ブルックヘブン，ケンブリッジ，東京，メルボルン）を通して行なわれる。わが国の利用者への窓口には，島内教授・田隅教授のご尽力により東

大がなっている。データは，東大計算センターのデータ・ライブラリーとして登録し公開されている。ブルックヘブンから送られてきた 2400 ft の磁気テープは，東大以外にも名大理・佐々木教祐博士，京大化研・大井竜夫教授，阪大蛋白研・角戸正夫教授，九大理・郷信広のところへ順々に送られてコピーが保存されている。そして，上記のどなたかに申しこめば，テープをコピーさせてもらえることになっている。テープ上のデータを利用するには，“PDB File Record Formats” を必要とするが，それも上記のどなたかに見せてもらえばよい。本格的に長期にわたってデータの配布を受けたいときには，田隅教授を通して PDB に利用者の登録をすれば，以後ニュースレターなどの送付を受けることができる。

なお，PDB からのニュース・レターでは，しばしばデータの訂正が報告される。九大では，これらの情報を管理するシステムが宮沢によって作られ利用されている。まず，その時点で使用可能な座標データのリストを，共用ファイル上に作成した。そのファイルには，蛋白質名，identification code，登録者名，登録の日付，データの収められているボリューム通番，ファイル名，訂正が必要か否か，必要な際には参照すべきニュースレターの号数などが記されている。各利用者は，利用の際必要となる情報をすべてこのファイルから知ることができる。

IV. 原子座標データの利用
——蛋白質原子座標の情報処理プログラム・システム（PSPCS）の紹介——

原子座標データが入手できたとして，次にそれを処理し研究上必要な情報を得ることを考えよう。そのような情報の例としては，前々節にあげた2つの例を一応思い浮べておくことができる。このような情報を得るための処理プログラムを書くことは，一般に片手間でできるほど簡単ではないけれども，個人レベルでできる程度のものである。従来わが国の蛋白質の立体構造の研究者は高いレベルの研究を維持してきたが，それと関連してかなりの数の研究者が，蛋白質の原子座標データの情報処理プログラムの開発にあたってきた。その経験を通していくつかのことが認識されてきたが，まず第1に，異なる研究者が異なるプログラムを書く際に，しばしば似かよった問題に悩まされることに気が付いた。第2に，異なったプログラムが，しばしばほとんど同じ機能を持つサブプログラムを含んでいることに気がついた。これは多様なプログラムが組織化しうること，またそのことによって多様なプログラム開発の努力が有効に蓄積しうることを示唆している。そこで，プログラム開発の経験者が集まって，蛋白質原子座標の情報処理プログラム・システム（Program Syetem for Protein Conformation Study，略して PSPCS）を作ろうということになり，科研費特定研究「情報システムの形成過程と学術情報の組織化」の中の1つのグループとして 1976 年度以来活動してきている。この小論は，このグループのうち，九州大学に所属するメンバーによる共同執筆である。本節では，PSPCS 作成の活動を通して浮び上ってきた原子座標データ利用上の問題点を議論する。PSPCS の中の個個のプログラムは，次節で紹介する。蛋白質の研究者が自分で原子座標データを処理しなんらかの情報を得たいと考えたときには，すでに開発されているプログラムを他人から借りるか，あるいは自分自身で必要なプログラムを開発しなければならない。本節における議論は，主にプログラムのシステムを作る立場からなされるけれども，1人の研究者が特定の目的のプログラムを書く上でも参考となる点を含んでいると思う。次節は，どのようなプログラムがすでに開発されているかを知っていただくのに役立つ。

第1の問題は，原子座標データのフォーマットと処理プログラムの接点にある。原子座標データの記録のフォーマットは，現時点では PDB の用いている BNL フォーマットが標準化されているが，これが整備されてきたのもごく最近のことである。PSPCS 作成グループが活動を開始した 1976 年度当初 PDB から入手できるデータは，140 字の CDC フォーマットで，その内容も充分に整備されているとはいえなかった。BNL フォーマットは，1976 年5月に最初の案が作られ，その後順次拡張・整備されて今日に至っている。それと同時に，PDB 中のデータも CDC フォーマットから BNL フォーマットに順次変換され，今日では，すべてのデータが BNL フォーマットで与えられている。BNL フォーマットは確かに良く整備されているが，これとても少し長い目でみると変化してゆくものと予想される。それはオンライン情報検索システムへの動きと関連している。この動きは，ブルックヘブンにもあるようだが[3]，ここでは角戸正夫教授を中心とする阪大グループの行なっている開発研究を紹介しよう。

BNL データ・フォーマットはカード・イメージで，データが順々に1列にはいっているいわば素朴な形態を取っている。そのため，大量のデータの中から必要な情報を探し出すために，その都度データを最初から見ていかなければならない。これはオンラインでデータを利用することを想定するとはなはだ非能率である。そこで，阪大グループでは，汎用データベース・マネージメント・システムを利用して，オンライン処理などに適した蛋白質データ・ベースの開発を進めている。データ・ベースには，格納・検索の技法から見て，大きく分けて2つの方法がある。第1は，チェーン型データ・ベースで，レコ

ード間の関係は，ポインターにより結びつけられ，格納および検索はポインターをたどって行なわれる。第2は，インバーテッド・ファイル型データ・ベースで，レコードに対してインバーテッド・インデックスを準備し，格納および検索を，これによって行なう。いわば，索引によって，本の中から必要事項を検索するのに似ている。将来は，計算センター間のネットワーク化や，公衆網 TSS の普及により，このようなデータ・ベース化された原子座標データの利用が，より容易になるものと思われる。BNL フォーマットにおけるデータ構造は，レコードが順々に1列にはいっており，一般に人間にとって読みやすいようになっている。BNL のデータは，今のところ磁気テープに収められ配布されているが，その内容は，一度紙に印刷され，人の目で読まれることを想定しているように思われる。阪大グループによるデータ・ベースも現在のところ，BNL フォーマットによるデータをもとにして作成されている。将来ともに，データが一度人間にとって読みやすい形態をとってからデータ・ベース化されるのか，あるいは，直接データ・ベース化されたデータのみが配布されるようになるのか興味ある点だ。いずれにしても，処理プログラムを書く立場からは，必要とする原子座標データの存在形式がいつまでも BNL フォーマットにとどまっていることはないことを示唆している。

1976 年度に PSPCS 作成グループが活動を開始したときには，CDC フォーマットしかなく，これはあまり充分整備されたものとはいえなかった。そこで，PSPCS 中のプログラムへの入力データの標準フォーマットを定めることにし，PSPCS フォーマットと呼ぶことにした。その後，BNL フォーマットが発表され，PDB で採用されたのに伴って，PSPCS フォーマットも改訂されてきた。現在使用中のフォーマットは，1977 年3月にまとめられた文献「タンパク質原子座標の情報処理プログラム・システム」[5] にまとめられている（この文献は希望者に配付している）。PDB がデータの預託・仲介機関であることを反映して，BNL フォーマットは，原子座標の生産者（X線解析の研究者）の側からみて有意義なデータを記述しやすくできているし，また利用者の側からみたとき，目で見て読みやすくできているという特徴を持っている。それに対して，PSPCS フォーマットは，一応ある処理プログラムの集まりを考え，それへの入力データのフォーマットとして考えられたために，処理プログラム・システムの構造を反映した形となっている。最も大きな特徴は，BNL フォーマットでは残基名・原子名などが英字名のみによって与えられているのに対して，PSPCS では，各原子に8桁の連続識別番号を与えて各種の検索が容易に行えるようになっている点である。また，BNL フォーマットでは，S-S 結合や，非標準原子の関与する結合についてのみ与えられている原子間結合の組を，PSPCS フォーマットでは，すべての共有結合について与えてある。この結合のデータは，残基名・原子名から計算できるものだが，複雑な処理プログラム中でしばしば必要となる情報なので，入力データ中にすでに計算して加えておくことにより，処理プログラムの構造を簡単化する上での効果が大きい。PSPCS フォーマットは，PSPCS 作成グループ内の水野により，BNL フォーマットに先立って立案され，グループの討議を経て決められてきたものである。その後 BNL フォーマットが整備され，PDB のデータもすべてこのフォーマットによって与えられるようになってきたので，宮澤は2つのフォーマットの長所を生かした1記録あたり 120 字のフォーマットを提案し，一部のプログラムで試用している。このフォーマットでは，BNL フォーマットのデータには一切手をつけず，ただ1記録 80 字を 120 字に拡張した余裕の 40 字に，PSPCS フォーマットの特徴をなす原子の連続識別番号などを与える。このフォーマットは，紙にリストしたときに1記録が1行に形よく納まる点も長所である。

結局，入力データ・フォーマットと処理プログラムの接点は，今後とも動き続けて行くものと思われる。純粋に技術的な問題であり大したことはないと一般には，思われるだろうが，実際に接点に立ってみると，そこが揺れ動くことによる消耗は少なくなく苦痛を感じる。これは大量のデータ処理に関連して常に起こることで，なにも蛋白質の原子座標データの処理にかぎった問題ではない。情報処理の専門家によれば，こういった問題は次のように考えるものらしい。入力データと処理プログラムの接点に，できるだけ一般的で柔軟な標準データ・フォーマットを設定しておく。入力データ形式がときとともに変化しても，それを標準データ・フォーマットに変換するプログラムを作ることにより，変化の影響が処理プログラムまでは及ばぬようにする。処理プログラムはもちろん標準データ・フォーマットを入力データ形式として書いておく。データが物理的に存在する状態は，その時々の入力データ形式によるものとし，実際に標準データ・フォーマットに変換しておくことはしない。処理プログラムの実行の際には，標準データ・フォーマットへの変換プログラムをコネクターとして用いる。

上のような本格的な考え方でいくのが有効になるのは，原子座標処理システムへの需要がある程度本格的になったときであろう。現状では，個人レベルで特定の機能を持つプログラムを書くのには，その場しのぎでいくほうが効率的であろう。しかし，システムとしてのプログラムの体系を考える上では，やはり上述の考えをとり入れる必要があると判断される。PSPCS フォーマットは，

標準データ・フォーマットとしての資格をそなえているものとはいえない。標準データ・フォーマットを考えていくことが，システム開発の立場からは，今後の1つの問題である。

以上では，プログラム・システムを作る際の第1の問題点，すなわち原子座標データのフォーマットと処理プログラムの接点の所を論じた。次に第2の問題として，入力データと処理プログラムの接点をどこに置くかについて考える。前述のように，BNL フォーマットでは，データとして，α-ヘリックス，β-シート，ターンなどに関する情報を含みうるようになっており，現に多くのPDB 中の蛋白質原子座標データが，これら2次構造に関する情報を含んでいる。しかし，これらの情報は，各原子の x-, y-, z-座標を1次情報とすれば，それから適当な判定条件を設定することにより計算することのできる2次情報である。2次情報は，それを1次情報から作り出す処理プログラムが完備していて，さらにそのプログラムの実行が容易であれば，1次情報と並列にデータに加えておく必要はない。必要に応じて作り出せばよい。PSPCS作成にあたっては，この立場に立って，2次的情報は，それを作り出す処理プログラムで置き換え，データそのものは，できるだけ1次情報のものにかぎった。ただし，前述のように，処理プログラムの構造を簡単化する効果の大きい結合に関する情報などは，2次情報であるけれども加えた。

PSPCS とは正反対の立場もありうる。たとえば，NIH の Feldmann によって出版されている AMSOM(Atlas of Macromolecular Structures on Microfiche)[6] がそれである。これは，蛋白質の原子座標のみでなく，それを用いて計算した種々の量（主鎖の内部回転角 ϕ と ψ；ベンダー模型組み立てに必要な数値；各残基内の結合距離，結合角，内部回転角，および 10Å 以内の非結合距離；等々），さらに蛋白質分子の各種の立体図（8通りの方向から見た骨格構造；所定の残基を加えた主鎖；所定の α-炭素から 10Å 以内の構造；等々）を数千の図表にまとめ，それをマイクロフィッシュ上に焼き付けて頒布しているものである。利用者は，これをマイクロフィッシュ・リーダーにかけて見る。これは，2次情報として必要となりそうなものを充分に多数生産しておき，それをデータとして貯わえておこうとする立場である。この場合，かなり多数の立体図を用意しておいても，いろいろな要求にぴったり答えることはなかなかむつかしいという難点がある。一方，PSPCS のように一連のプログラムを開発していつでも使用できるようにしておこうとする立場は，各種の要求にきめ細かく対応できる柔軟性は備えている。しかし，現状では，そのようなプログラムの利用の希望者が，気楽に計算機を使える所まではいっていない。AMSOM の立場は，現実には大いに有効である。PSRCS には，現在のところ次節に紹介するプログラムが含まれている。これは多くの人々に気軽に使っていただければ，大変に有用だと思われる。しかし，実際に使用する上での技術的・経済的障害は無視しえない。2次情報は必要に応じて1次情報から作ろうというPSPCS の本来の思想を生かすためには，プログラムが容易に使える体制を作る必要がある。

プログラム・システム作成上での第3の問題点として，システムの構造について考える。システム中のプログラムは，次の2つの種類に分類できる。第1は，他のプログラム中で頻繁に現われる部分をまとめたもので，サブルーチン形式にしておくのが適当なもの，第2は，特殊なひとまとまりの機能を持つもので，コンプリート・プログラム形式にしておくのが適当なものである。有機的にはたらくシステムとしての理想的な姿は，多くの第1のタイプのサブルーチンが存在し，その上にそれらを駆使した各種のコンプリート・プログラムが存在する形態であろう。PSPCS では，この形態を理想として追求しているけれども必ずしも充分は達成されていない。それは，PSPCS グループが，すでに各種の蛋白質原子座標処理プログラム開発の実績のある研究者からなり，既開発のプログラムを持ちよることから活動を開始した点にも一因がある。そのため PSPCS には，現在のところ，充分に相互に組織化されていないコンプリート・プログラムの集まりとしての性格と，組織化を目ざすいくつかのサブルーチンの集まりとしての性格が混在している。

システムとして考えた場合には，システム内の諸プログラムに有機的つながりを持たせねば価値がないが，特定の研究者が特定の機能を持つプログラムを作る場合にはあまり思いなやむ必要のない問題であろう。名大の坂部知平助教授・佐々木教祐博士・別府良孝博士のグループは，蛋白質のX線解析に必要なプログラムや立体構造のグラフィック・ディスプレイのプログラムを多く収集・開発しておられる（例，Diamond の実空間におけるリファインメント・プログラム[7]，Levitt のエネルギー・リファインメント・プログラム[8]，Maclachan のリボン模型作図プログラム；別府の立体構造のディスプレイ・プログラム NAMOD；等々）名大グループでは，これらのプログラムを別々のコンプリート・プログラムとして導入あるいは開発し，名大計算センターで使用できるように必要な改訂を加える以上のことはしなかったそうだ。この名大グループの経験は，プログラム・システムというものは，結局いくつかの独立なコンプリート・プログラムの集まりの形になっていくことを示しているのかもしれない。もしそうだとすれば，PSPCS の現状は，あるべき姿に近いのかもしれない（先に，PSPCS は，そ

のプログラムの利用が容易になるようにする必要があると述べたが，名大グループでは上述の諸プログラムを計算センターに登録し，簡単なマクロで呼び出して使えるようにしておられる[9,10])。

以上プログラム・システム PSPCS 開発にあたってきて問題となった点を論じてきた。いずれの点に関しても，議論がはなはだすっきりしない印象を読者に与えることとなったと思う。これはシステムのあるべき姿として追求しているものと現実との差が大きいからである。しかし，PSPCS 中のプログラムそのものは，かなり充実したものと思う。これはもちろん，作成グループがかなり実績のある研究者からなっていることによる。次節では，これら個々のプログラムを簡単に紹介し，最後の節で PSPCS を含めたわが国の今後の蛋白質の情報収集・管理・処理体制について考えたい。

V. PSPCS 中のプログラム

代表的な例をいくつか簡単に紹介する。

1. BNL フォーマットから PSPCS フォーマットへの変換（作成者：水野裕重） 変換と同時に BNL フォーマットで格納されている PDB データの欠落重複などの誤りを調べる。原子間の結合データも計算する。

2. 近傍原子の探索（作成者：武富敬） 種々の処理プログラム中で，特定の原子の近傍に存在する原子を探し出す必要がある。蛋白質分子の存在する空間をメッシュに分け，各原子のメッシュへの帰属を定めることにより，近傍原子の探索を高速で行なう。

3. 主鎖のボンド長，ボンド角および二面角の計算（作成者・郷 通子）

4. α-ヘリックスの探索（作成者：長野晃三（東大薬）） 主鎖原子の座標から2面角 ϕ, ψ, ω を計算し，(ϕ, ψ) がある範囲にはいり，かつ i 残基のO原子と ($i+4$) 残基のN原子が水素結合距離にある条件を満たすときに，α-ヘリックスが存在すると認定する。

5. 蛋白分子中の水素結合の探索（作成者：花田光弘，郷 信広） Donor 原子および Acceptor 原子となりうるすべての原子を拾い出す。その間の距離が一定範囲内にあるものについて，水素結合が形成された場合の結合角を計算する。結合距離・結合角がともに一定の範囲内にあるものを水素結合と認定する。

6. 蛋白質分子の水素原子の座標の計算と，水素結合の判定（作成者：中田吉郎（群馬大教養）） X線回折法からは決定できない水素原子の位置を推定する。次に水素結合に関与している水素の場合，経験的なポテンシャル・エネルギー関数の値を計算することにより水素結合形成を判定する。上記5番よりは水素結合判定条件が厳しいと思われる。

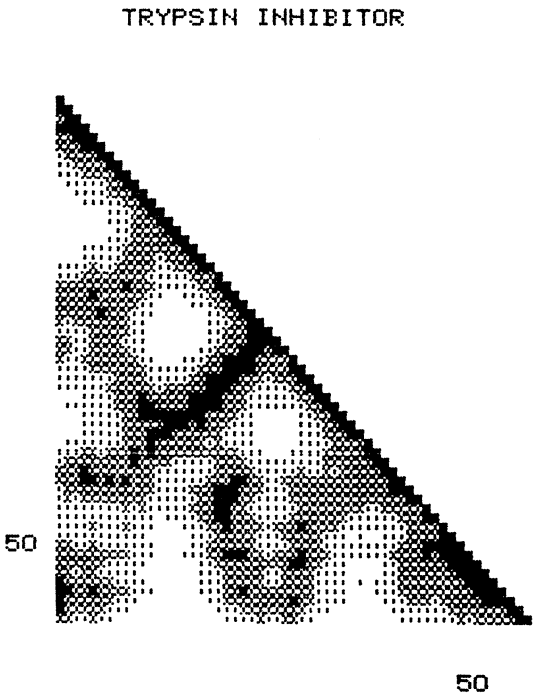


図 2. ドット・プリンターによって印刷された α-炭素原子間の距離を示す地図
蛋白質はトリプシンインヒビター（プログラム作成者のご好意により掲載）

7. α-炭素原子間の距離を示す地図（distance map）のディスプレイ（作成者：大畠玄久（京大化研）） 蛋白質の3次元空間における折れたたみ構造を，2次元の紙の上に表示するのに，α-炭素原子間の距離を，残基番号をたて軸，よこ軸にとった空間に地図として表示する方法がある。これは，その地図をグラフィック・ディスプレイで作る。

8. α-炭素原子間の距離を示す地図をドット・プリンターによって印刷する（作成者：輪湖博（早大理工）） 上記7番と同じ地図を，東大計算センターのプリントロニックスというドット・プリンターで印刷する。図2にその印刷例を示す。濃い所が，距離の短い α-炭素の組を示す。図2からトリプシン・インヒビターには，アンチ・パラレル-β-シートが1組と α-ヘリックスが1つ存在することがわかる。

9. 2つの立体構造の重ね合せ（作成者：西川建・大井龍夫（京大化研）） ある蛋白質の類似した2つの立体構造を比較をするために，対応する原子の間の距離の平方和を最小にするように重ね合せる。その結果として，2つの立体構造の類似度を表わす RMS(root-mean square) deviation が同時に得られる。

10. 蛋白質の表面の計算（作成者：水野裕重・郷信広） 水分子を球形で近似し，この球が外側から接触しうる点を蛋白質の表面と定義，その面積などを計算

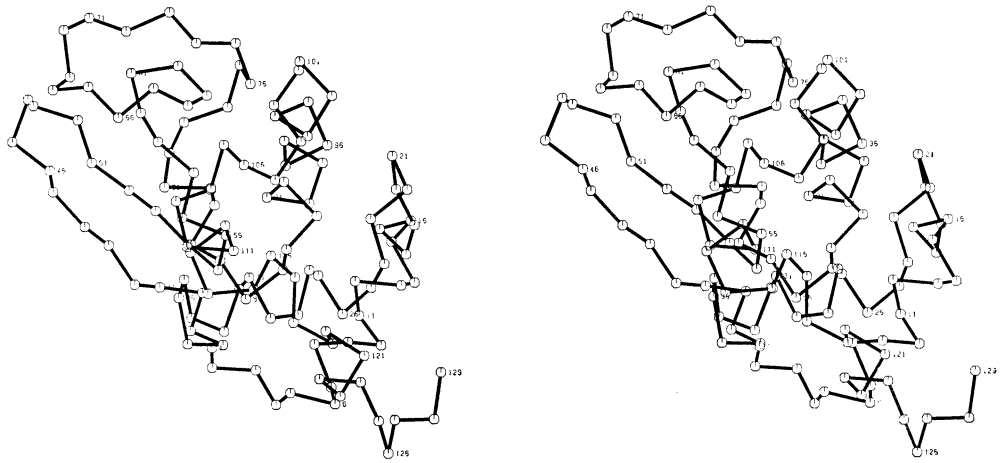


図 3. リゾチームのα-炭素原子間を棒で結んだ骨格模型の立体図
13 のプログラムでプロットした

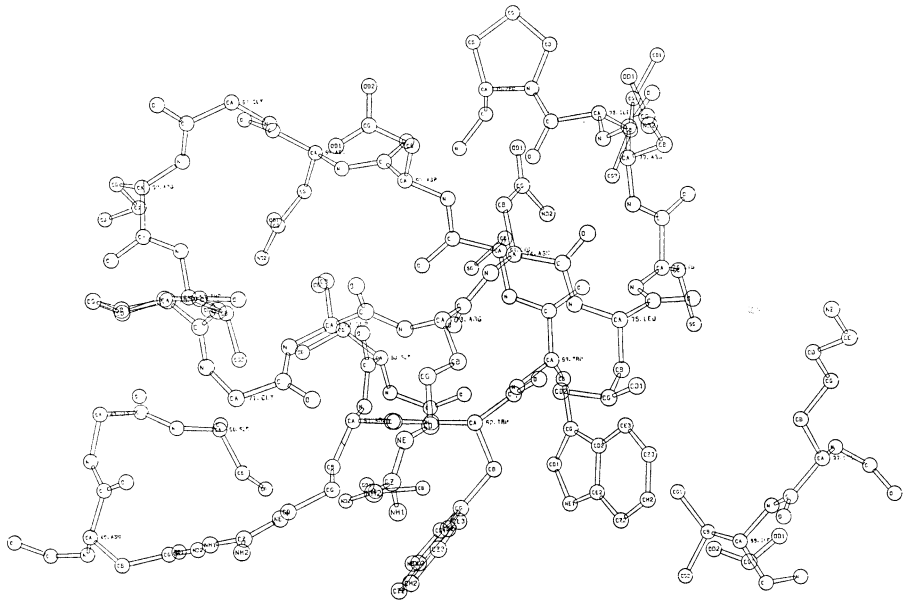


図 4. 15 のプログラムでプロットしたリゾチーム分子中の一部分の図

する。最近蛋白質の表面という概念は重要性を増してきた。表面に露出している残基の性格を調べたり，ある特定の部分が内部にうずもれているかを調べるのに用いることができる。

11. Voronoi 多面体の計算（作成者：宮澤三造・郷通子） 蛋白質分子中において構成原子がいかに空間的に詰っているかなどを調べる目的のために，蛋白質分子の占めている空間を，構成原子を真中に据えた多面体に分割する。ある原子を含む多面体としては，その原子とすべての近傍原子との間のすべての垂直二等分面が切り出す最小の多面体（Voronoi 多面体）を考える。これによって空間の各点は，一意的にいずれかの原子を含む多面体に帰属する。表面付近の原子については，垂直二等分面の集まりが外に開いてしまって多面体が構造できないので，これを避けるために，表面には溶媒分子を置く。

12. 球状蛋白質における解離基の解離状態と，それによる電気多重極能率の計算（作成者：中村春木・和田昭允（東大理）） まず与えられた pH，イオン強度に対して，各解離基の解離状態を種々の効果を考慮に入れて計算する。その結果をプロッターで図示し，また，蛋白質全体の電気多重極能率を計算する。

13. α-炭素原子間を棒で結んだ骨格模型の立体図のプロット（作成者：沼 道子・花田光弘・郷 信広） 図3に例示した図をプロットする。分子を見る方向を任意に指定できる。

14. 蛋白質分子の透視図（作成者：中田吉郎（群馬大教養）） 構成原子を van der Waals 半径を持つ球

で表わし，しかも他の原子によってかくされた陰線を消去した図を，透視図法でプロットする。表面の様子を観察するのに役立つ。

15. 蛋白質分子中の任意の部分の図をプロットする（作成者：合志貞夫・郷 信広） 図4にプロットされた図の例を示す。これはリゾチーム分子を 15 Å ごとの層に輪切りにした1つの層の一部分の図で，残基 60 Ser から残基 79 Pro までの部分が見えている。同じ部分を少し角を変えて描いた1対の図を実体鏡でみることにより，立体的に見ることができる。例示した図以外に，任意に指定した残基およびその周辺の主鎖などを選んでプロットできる。このプログラムは，図4のような図をプロットできるが，まだ完全にはでき上っていない。

PSPCS には，上記の他に直接原子座標は扱わないが，蛋白質の立体構造研究に用いられるプログラムがいくつかある。PSPCS 中のこれらのプログラムは，一般の利用者に使っていただく立場で開発している。しかし，利用していただく体制は充分に整ってはいないので，プログラムの利用の希望を持たれる方は，それぞれのプログラムの作成者に直接連絡をとっていただきたい。

前節および本節でふれたいろいろのプログラム以外にも多くの処理プログラムが，わが国内で書かれているものと思われる。たとえば，筆者の知る範囲では，東大の横山茂之氏・田隅三生教授らの，任意の蛋白質から特定の残基を選び出し，その側鎖の2面角を計算するプログラムなどがある。

VI. おわりに

PSPCS 作成グループのメンバーは，それぞれに自分の研究の必要上から蛋白質原子座標データの処理プログラムの開発に携わってきた人々である。その人々が集ってプログラム・システムを作ろうと考えたのは，IV節の始めにも書いたが，ばらばらに行なわれている努力を多少組織化することによって協力の効果が出ると考えたからである。さらにまた，次の期待もあった。既製の原子座標データの処理プログラムがあれば利用したいが，自分で作るのは大変だと考えておられる生化学方面の研究者は潜在的に多いのではなかろうか。PSPCS がシステムとして充分整備されれば，これらの研究者にも有効に使っていただけるのではないか。

PSPCS は，今まで約2年間のグループの活動を通して，前節に紹介したような多様なプログラムの蓄積を見た。しかし，システムとしての理想と比べると，VI節で論じたように，いくつかの問題点が指摘できる。これらが思ったほどには簡単に解決していかない理由の1つは，PSPCS 作成グループの活動が，メンバーのだれにとっても，自分の主研究テーマ追求のついでに行なわれている点にある。これはけっして非難すべきことではなく，第一線の研究者の集まりとしてはこの形態しかなく，これ以上を望むことはできない。しかし，その結果，システムとしては不充分な点が多いし，またでき上ったプログラムも手作りの機械に似て製作者以外には，あまり使いやすくはない。

今後は，ますます情報処理能力が研究の質を決めてゆく傾向が強まるであろう。標準的な実験機器は，手作りするよりは商品化された誰にも使いやすい製品を使うのが普通である。蛋白質研究に必要な情報処理プログラムも，誰にも使いやすい形のものが整備されることが望まれる。PSPCS 作成グループは，それをめざして活動してきたが，上述のようにその活動には限界がある。今後は，処理プログラム・システムの整備を，1つの独立したプロジェクトとしてとらえ，蛋白質研究者がそれをサポートしていく必要があると思う。蛋白質に関しては，原子座標の他にも，アミノ酸配列，文献情報など収集管理すべき情報は多い。このうち，原子座標は，田隅教授のご好意により東大で扱われているが，これも含めて，将来は協同利用研究所などに情報センターを設けて，情報の収集・管理・配布にあたる体制を作ることが望まれる。処理プログラム・システムの開発・整備事業にも情報センターがあたることが望まれる。これが筆者らの夢である。

本稿中でかなりの比重を占めることになった PSPCS の活動の紹介も，執筆の都合により九大のメンバーのみによってなされることになった。この機会に，PSPCS 作成の活動に参加いただいた他のメンバー（V節の作成者に名が出ている）の日頃のご協力に感謝します。また，角戸教授・田隅教授・佐々木博士・別府博士には，本稿執筆に際していろいろご教示をいただきましたことを感謝します。

文 献

1) Protein Data Bank : *Nature New Biology*, **233**, 223 (1971)
2) Protein Data Bank : *Acta Crystallogr. sect.* B., **29**, 1746 (1973)
3) Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. : *J. Mol. Biol.*, **112**, 535 (1977)
4) 角戸正夫，安岡則武：日本結晶学会誌，**19**，294 (1977)
5) PSPCS 作成グループ：タンパク質原子座標の情報処理プログラム・システム，1977 年3月
6) Feldmann, R. J. : *AMSOM, Atlas of Macromolecular Structures on Microfiche, Tracor Jitco Inc.*, Rockville, Maryland (1977)
7) Diamond, R. : *Acta Cryst.*, **A 27**, 436 (1971)
8) Levitt, M. : *J. Mol. Biol.*, **82**, 393 (1974)
9) 佐々木教祐：名古屋大学大型計算機センターニュース，**8**，354 (1977)
10) 佐々木教祐：日本結晶学会誌，**20**，186 (1978)