2001 年 8 月 3 日

サイエンティフィック・システム研究会
科学技術計算分科会　第 1 回会合資料

タンパク質立体構造の予測問題

国立遺伝学研究所
西川　建

1.　DDBJ について

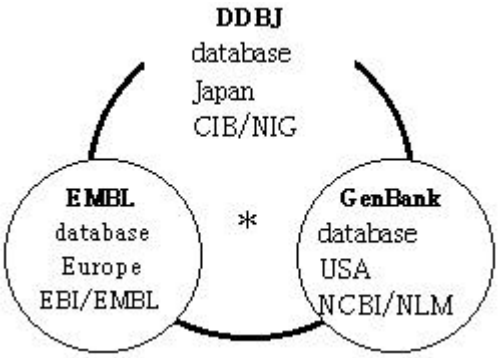
本日は、このような場で発表させていただく機会を得てたいへん光栄に思っております。

まず本題に入る前に、DDBJ (DNA Data Bank of Japan) の宣伝をさせていただきます。DDBJ とは、我々が所属する国立遺伝学研究所が運営している日本 DNA データバンクで、主に日本で決められたＤＮＡ配列を受け付け登録しています。

さらに、アメリカ、ヨーロッパと協同で、国際塩基配列データベース (**Nucleotide Sequence Database**) を構築しており、主に DNA の塩基配列データを収集・編集し、全世界に提供しています。

DDBJ ： **日本** DNA **データバンク**
- DNA　Data Bank of Japan
- **国立遺伝学研究所 生命情報・DDBJ 研究センターが運営する国際共同研究事業**
- GenBank ⇒ NCBI／NLM／NIH (**米国**)
- EBI／EMBL ⇒ **欧州**
- DDBJ／EMBL／**ＧｅｎＢａｎｋ 国際** DNA **データバンク**
- **特許データも含む。**(JPO／EPO／USTPO)
- **公的な登録番号発行権**

DDBJ／EMBL／GenBank International
Nucleotide Sequence Database



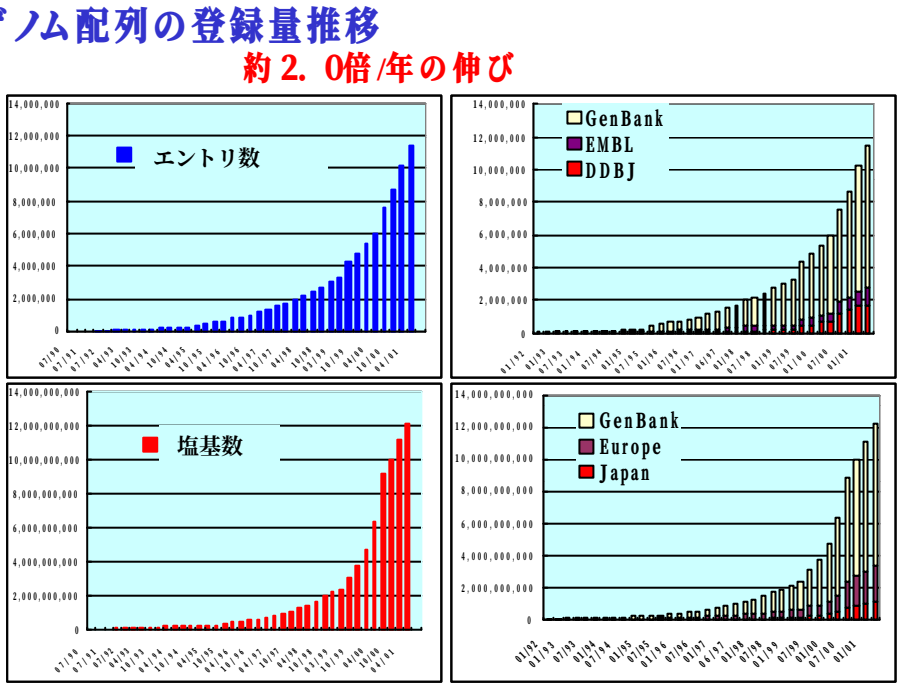
NIG:　National Institute of Genetics
CIB:　Center for Information Biology
DDBJ:　DNA Data Bank of Japan
EMBL:　European Molecular Biology Laboratory
EBI:　European Bioinformatics institute
NCBI:　National Center for Biotechnology Information
NLM:　National Library of Medicine

＊ International Advisory Meeting
＊ International Collaborative Meeting

画面 1

画面 2 は、国際塩基配列データベースに登録されたゲノム配列の登録量をグラフ化したものです。右側のグラフは、DDBJ(日本)、GenBank(アメリカ)、EMBL(ヨーロッパ)の内訳を示したものですが、GenBank の登録量が最も多く、DDBJ は全体のおよそ 1 割に貢献していることが判ります。

画面 3 は、DDBJ のホームページです（http://www.ddbj.nig.ac.jp/Welcome-j.html）。塩基配列の登録の仕方や、データベース検索、そして、これからお話する立体構造の予測などのサービスは全てこのホームページ上で運営しています。なお、ユーザが実際に DDBJ にデータを登録される際には、ホームページ内の「sakura」というシステムをご利用いただいております。

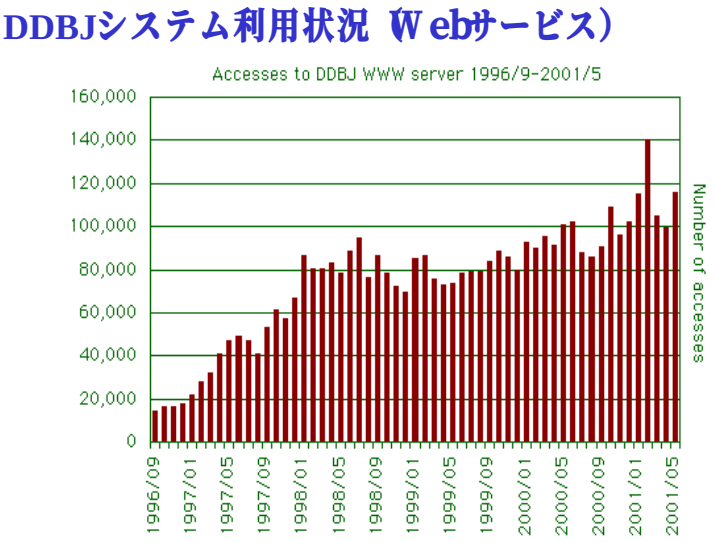


画面 2

画面 3

DDBJ ホームページの利用状況は、画面 4 に示す通りです。一月におよそ 10 万件ほどのアクセスがあることが判ります。

DDBJ には、実際のデータフローに携わる者と、内容をチェックするレビューワーと呼ばれている者、ソフトウェアを開発する SE、そしてシステムの管理者など、合計で 50～60 人以上がおり、ちょっとした会社並みの規模で活動をしています。ホストは、今年から富士通さんの VPP5000 を導入しています。今後も多くの皆さまに DDBJ をご活用いただきたく思っております。



画面 4

2. タンパク質立体構造の予測問題

それでは、本題に入らせていただきます。

タンパク質立体構造の予測問題ということで、本日は、以下の順序で話させていただきます。

1.「なぜ構造予測問題が難しいのか」

その理由についてお話しします。

2. 構造予測の研究史

研究の歴史もずいぶんと長いので、その経過をかいつまんでお話しします。

3. アブイニシオ予測の可能性

構造予測の研究方法の中で、特に注目されているアブイニシオ予測の可能性についてお話しします。

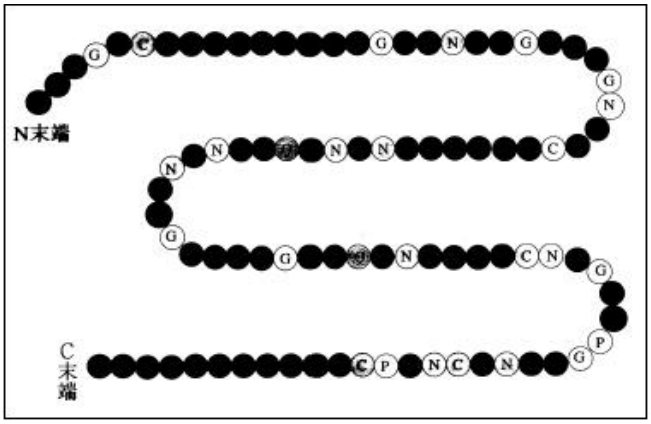
4. 予測問題が解決すれば・・!

最後に、タンパク質立体構造の予測問題が解決するとどうなるのか、その影響についてお話しします。

2-1．なぜ構造予測問題は難しいのか?

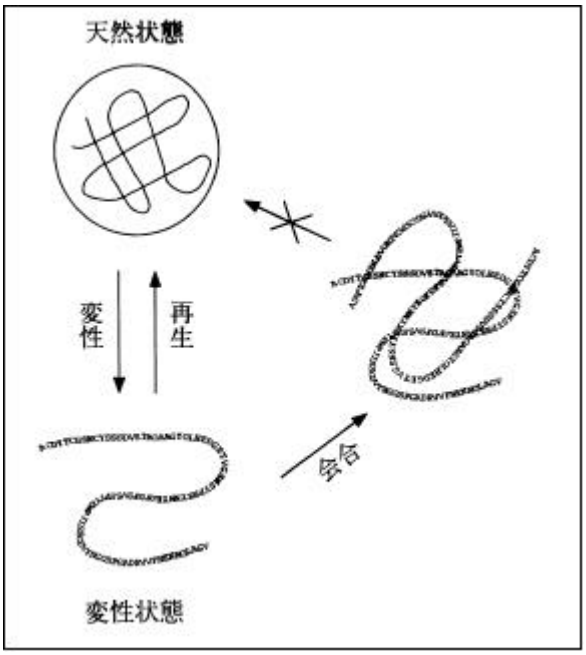
皆さんもよくご存知のタンパク質は、DNA からアミノ酸の配列として翻訳すると、画面 5 に示す通り「ひも状分子」（鎖状高分子）と呼ばれる 1 次元の高分子であり、N 末端から始まって C 末端で終わるという方向性を持っています。そして、アミノ酸の配列の順番が違うと、出来上がる立体構造も異なるという大きな特徴を持っています。

画面 6 は、今から 40 年ほど前の 1960 年に、X 線結晶解析で初めて解かれたミオグロビンの立体構造です。当時の人は、ミオグロビンはもう少しカチっとした結晶のようなものだと予想していたのに対し、結果はご覧の通り一見不規則に見えるものでしたから、非常に驚きました。そして、このことから、なぜアミノ酸の配列だけで別の構造になってしまうのかということを意識しだした人が非常に増えたわけです。そこから、現在に至る「予測問題」～コンピュータにアミノ酸の配列だけを与えて、計算の上で立体構造を予測する～ が意識されだしました。



画面 5

画面 6

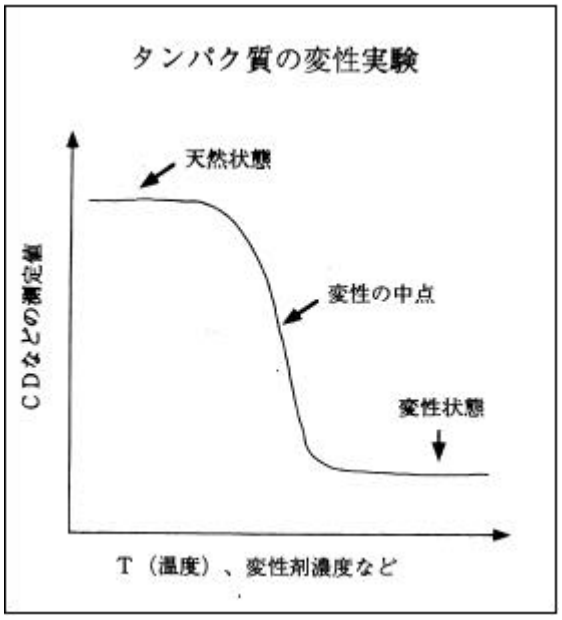


画面 7

画面 7 をご覧下さい。これは、温度を上げると天然状態のものが壊れてランダムコイルに戻り(変性)、逆に温度を丁寧に下げてやるともとの天然状態に戻る (再生) という関係を表した図です。この図からも分かるように、タンパク質は、このような可逆的な過程で「安定構造」であることが解明されています。ならば、コンピュータで安定構造を分子力学的にエネルギー計算すれば、立体構造は求まるはずだということは当初から分かっていました。しかし実際に解こうとすると非常に難しく、「予測問題」と

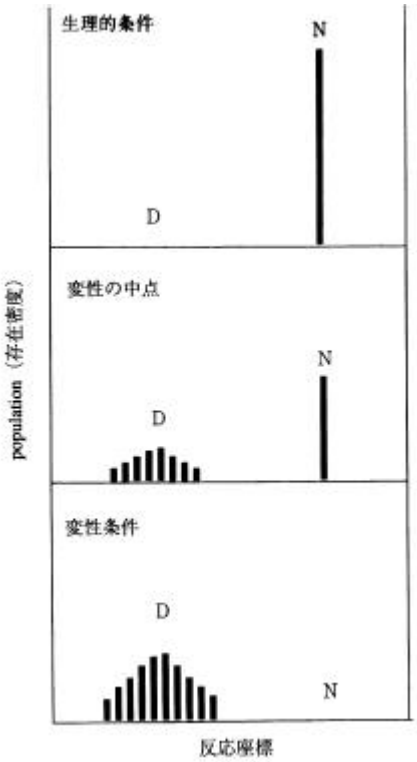
いうものが論文などで議論されだしてから 30 年以上経つ今でも、未だに完全な解決策はでていません。その理由は、タンパク質の特性に大きく関係しています。

画面 8 は、タンパク質の温度を上げていった際の変性曲線ですが、温度を上げると、天然状態からトランディションして変性状態に戻ります。別の見方をすると、画面 9 から分かるように、環境状態を生理的な条件にすれば、ポピュレーション (population：存在密度) は天然の状態(N)にあるが、逆に変性しやすいような環境状態にすれば、ポピュレーションは全て変性(D)に偏るということです。

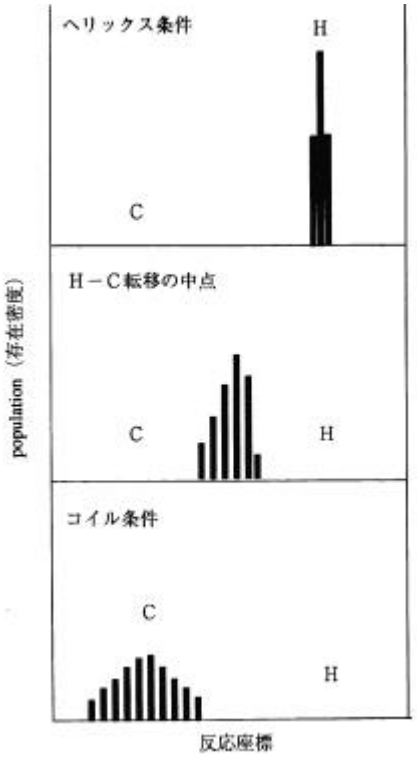


画面 8

途中過程の“変性の中点”といわれているところでは、実際に何が起こっているかを見てみると、ご覧の通り、ポピュレーションが天然状態と変性状態の 2 つに分かれています。重要なのは、N も D も両方あり、両者の中間状態には留まらないという点です。このように、半分が壊れて半分が出来ているというような、そのような例は他の多数の分子をみても見当たりません。N か D かどちらか一方である例が多数なのです。タンパク質のフォールディング (folding：折りたたみ過程) におけるこのような特徴は、「協同性 (cooperativity) 」と呼ばれています。



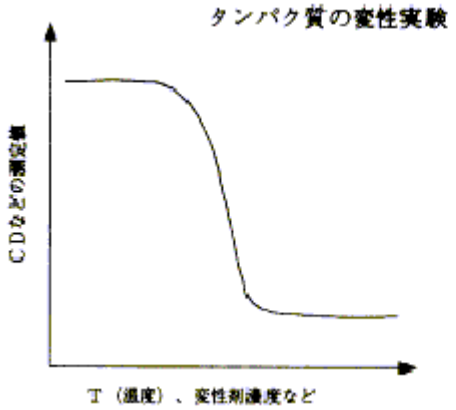
画面 9



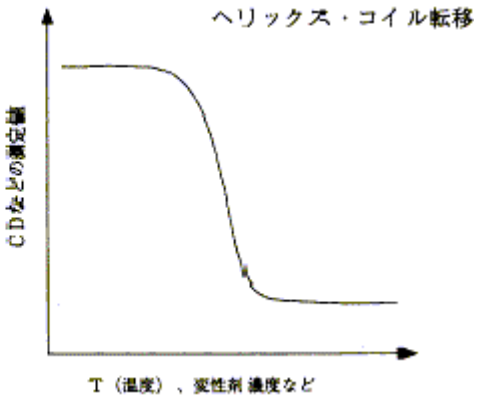
画面 10

少し話はそれますが、それでは、変性の中点では N か D かどちらか一方であることが当たり前かというと、違う例もたくさんあります。例えば DNA のヘリックス・コイル転移は、天然状態がヘリックス(二重螺旋)で、変性の結果コイル状になるのですが、この場合の中点は、画面 10 の通りポピュレーションが中間状態にあり、先ほどのタンパク質とは全く異なります。ある部分ではヘリックスが残っていて、ある部分ではほどけていて、それが入り混じっているという状態です。

このように、タンパク質には「協同性」という特性があるわけですが、画面 11 と 12 を見比べるとお分かりの通り、面白いことに、変性曲線では、タンパク質もヘリックス・コイル転移もどちらも同じように見えます。しかし、中身は全く違うということがお分かりいただけたと思います。



画面 11



画面 12

ミオグロビン
MYOGLOBIN (HORSE HEART)

画面 13

次に、タンパク質のもうひとつの特徴についてご紹介します。それは、タンパク質は、高分子としては非常に堅いという点です。画面 13 は、先ほどご紹介したミオグロビンの空間充填模型ですが、これを見ると、非常にまんまるで塊として見えるので、いかにも堅そうだということが想像できると思います。例えば、よい例かどうかは分かりませんが、バクテリアの鞭毛を例にとってみても分かります。バクテリアには、鞭毛をぐるぐるまわすモータのような部分があるのですが、そのモータ機能の部分は、いろいろな種類のタンパク質(いわば部品)が組み合わさって作られていて、個々の部品が非常に堅いのです。それでは、タンパク質はなぜ堅いのでしょうか。本日は詳しくはご説明しませんが、その由来は、先ほどご説明した、フォールディングのメカニズム (協同性) に深く関係していると私は考えています。

今までタンパク質の特徴についてお話ししてきましたが、ここで、なぜタンパク質の構造予測は難しいかという点について、まとめます。

立体構造の予測問題
与えられたアミノ酸配列から立体構造を予測せよ!

```
  1                                                  50
  KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGS
 51                                                 100
  TDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVS
101                        129
  DGNGMNAWVAWRNRCKGTDVQAWIRGCRL
```

chicken lysozyme

なぜ、タンパク質の構造予測は難しいのか?
1.フォールディングの「協同性」
部分が全体に影響し、全体が一体化する
線形論理(加算性)が成り立たない
従って分子全体を一度に考慮する必要がある
2.「多数の内部自由度(回転角など)をもつ系」を対象とせざるをえない

画面 14

力学的な考え方
「分子の形：エネルギー的にもっと安定な形になる」

エネルギー計算
1.分子内に働く力
静電相互作用
ファンデルワールス力
疎水相互作用 － 水分子の影響
水素結合　など
2."すべて" の形
アミノ酸 10　個のとき　：計算可能
アミノ酸 100 個のとき　：　？

アミノ酸 10 個のときの場合の数：
2^{10}　　$1000 = 10^3$
アミノ酸 100 個のときの場合の数：
2^{100}　　$= 10^{30}$

画面 15

まず、フォールディングに「協同性」があるという点です。協同性は、2 状態的 あるいは 相転移的とも言われていますが、部分が全体に影響して、全体が一体的にふるまうので、加算性がなりたっていないのです。つまり、部分の和を足していくと全体が分かるというわけではなく、分子全体を一挙に考える必要があるため、その点に、1 つめの難しさの原因があります。

また、もう 1 つの大きな原因に、「多数の内部自由度（回転角など）をもつ系」を対象とせざるを得ないという点があげられます。タンパク質は、ひとつのアミノ酸の主鎖の部分に、2 つの回転の自由度があります。それらがたくさんつながって、ペプチド面は平面なのですが、平面がくるくるまわるような形状になっており、アミノ酸に応じてそのそれぞれに側鎖が付いています。つまり、タンパク質は内部自由度を持っていますが、それを部分ごとに考えていって足しあわせれば全体の解がでるというものではなく、やはりそれら全ての分子を一度に考える必要があるので、その点が非常に難しい原因なのです。

それでは、「場合」の数はどのくらいになるか、という問題も、この世界ではよく耳にする難しい問題です。10 個のアミノ酸がつながったタンパク質の場合、現在の計算機を使えば、完全な計算が可能です。ところが、100 個のアミノ酸がつながった場合には、計算不可能となります。なぜ不可能になってしまうのでしょうか。実は、ひとつのアミノ酸でふたつの場合を許すだけでも、10 個がつながっている場合は、2^{10}　　1000 通りくらいしかありませんが、2^{100} の場合は、10^{30} 通りになるわけです。つまり、10 個と 100 個の差は、単なる 10 倍ではなく、10^{30} 倍もの差となるというわけです。このように、タンパク質は、アミノ酸配列の長さが長くなるにつれて、急激に問題が難しくなります。以上の点が、なぜ難しいかという基本的な理由だと私は思っています。

2-2．立体構造予測の 10 年史

2-2-1. 3D-1D（スレディング法）

立体構造予測の 10 年史

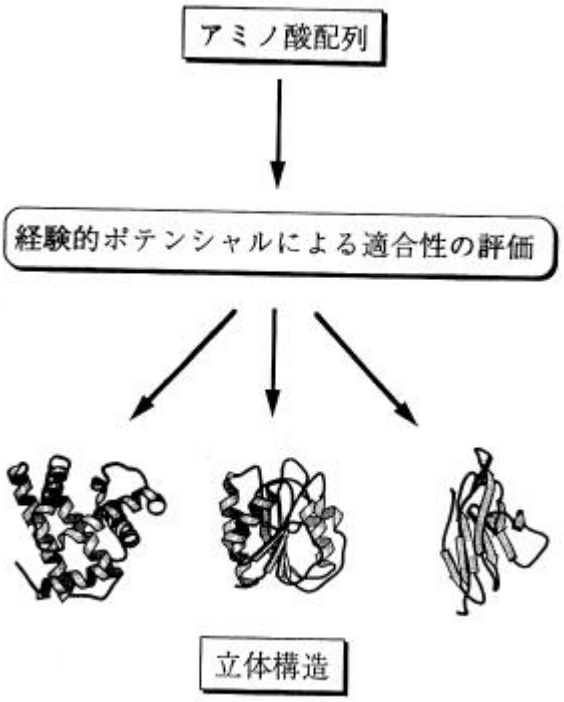
1．3D-1D（スレディング）法、1990～
2．構造予測コンテスト CASP、1994～
3．新しい相同性検索法 PSI-BLAST、1997～
4．アブイニシオ予測への挑戦、1998 年～

画面 16

次に、構造予測の歴史について簡単にお話しします。

先ほどは、この研究は 30 年以上前から始まったと申しましたが、簡単にいうと、その間は失敗の連続でした。大勢の人がチャレンジしては、やっぱりだめだという繰り返しでした。その原因のひとつには、明らかにコンピュータパワーの不足があげられますが、それだけではありません。状況が変わったのは 10 年前の 1990 年に「3D-1D 法」が考案されてからで、これを機にやっと一応まがりなりにもアミノ酸の配列から構造を予測できる確率が高くなりました。ですから、立体構造予測の歴史を実際に語るとなると、この 1990 年からの 10 年間に限られてきます。

3D-1D 法とは、当時はたいへん画期的な概念でした。普通は、与えられたアミノ酸配列から立体構造をまともに考えてしまいがちですが、この 3D-1D 法は逆に考えます。つまり、まず答えである、あらゆる既知の立体構造（現在 PDB(Protein Data Bank)では何千種類もの構造があげられます）を全て並べておき、調べたいアミノ酸配列をそのそれぞれに当てはめ、どの構造と適合しているか、計算結果のスコアを用いて数量的に評価するという方法です。最もスコアのよいものが一番適合しているだろうと予測するわけです。勿論、スコアにはある一定の基準を設けており、その基準に達していなければ解がない場合もあるので、この方法にはもともとそのような限界性もあるわけですが、アミノ酸の配列レベルでは全く違っていても、構造は似ているというケースが非常にたくさんあるので、当時、この方法は実用的に大変意味のあるものでした。



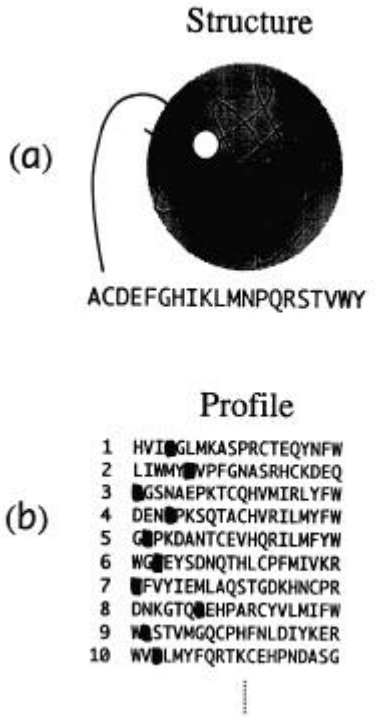
画面 17

スコアで評価する方法について、実際の具体例をすこしご説明したいと思います。画面 18 をご

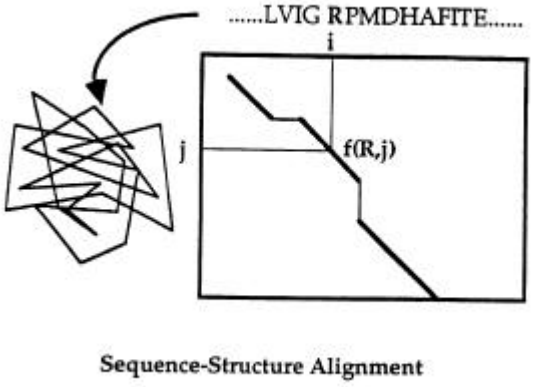
覧下さい。

(a)はタンパクの構造で、アミノ酸の部位を描いたものです。その中のひとつの部位に注目し、アミノ酸 20 種類（A：アラニン、C：システィン、D、E・・・）を順番に当てはめ、ある計量的な一種のポテンシャルファンクションを勘定して、まわりとの適合性を計算し、その結果を(b)のようにテーブルにまとめていきます。このテーブルは、縦が残基番号、横が 20 種類のアミノ酸で、一番適合(フィット)している順にアミノ酸を並べていきます。ここで黒く印をつけているのは、天然のアミノ酸の種類です。天然のものが必ずしもトップにくるわけではありませんが、高い確率でトップに近いポジションにくることが分かります。そして、(b)のテーブルを、画面 19 にある立体構造のイメージ(左側)の位置に並べると、入力配列と任意の配列とのいわゆる DP (Dynamic Programming)で最適の対応関係(アライメント)が計算でき、それにより、構造と配列の最適なアラインメントが得られます。

このように、プロファイル (profile) というテーブル構造があれば、スコアをすぐに勘定することができ、それによって一応解けます。これを “3D プロファイル型” と呼び、いわば、3D-1D 法のカナメとなる考え方です。



画面 18



画面 19

2-2-2. **構造予測コンテスト** CASP

構造予測コンテスト CASP　**1994 年～**

CASP1　(1994)
CASP1　(1996)
CASP1　(1998)
CASP1　(2000)

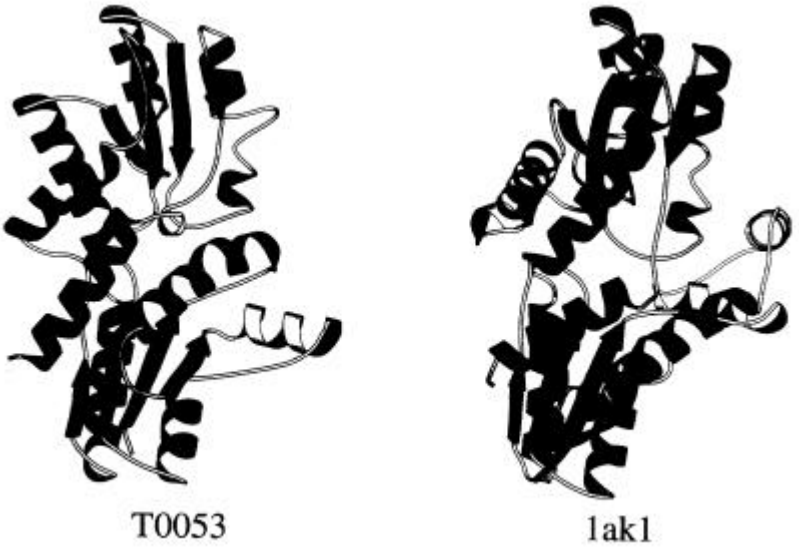
盲検法（blind test）

「スレディング法 = 構造予測法」公認
予測法としての性能(?)
ホモロジー(相同)とアナロジー(相似)

画面 20

当時、3D-1D **法が非常に反響を呼び、多くの研究者が「これはいけるぞ」ということで、我々も含め、いろいろなグループが予測法の開発に参入した結果、いろいろなバリエーションの方法論がでてきました。それではどのグループが一番よく当るのかということで、全世界レベルの「タンパク質構造予測コンテスト（**CASP**）」が** 1994 **年から始まりました。以降、**2 **年ごとに開催されています。このコンテストの一番の特徴は、インターネットを活用した完全なブラインドテスト** (blind test**：盲検法**) **であるという点です。参加者に対し、問題となる何十題ものアミノ酸配列（勿論未発表のもの）が電子メールで送られてきて、その構造を、問題ごとに設けられた期日以内に予測し座標データを送り返す、そして予め選定された審査員がそれを評価し順位を決めるというシステムです。ブラインドテストですから非常に厳しく、下手すると自分たちのグループの順位が悪く恥をかく、ということにもなりかねません。**

我々もこのコンテストに参加しましたが、うまくいった例をお見せしましょう。画面 21 **をご覧下さい。**



画面 21

左側のT0053が問題として出されたもので、勿論問題としては配列しか知らされていませんが、我々のグループは、先ほどご紹介した 3D-1D 法により、既知構造の中の 1ak1 に似ているだろうと予測し回答しました。結果は、ご覧の通り確かによく似ているので、この場合は花丸をもらったわけです。

しかし、全部がこのように正解するとは限らず、間違えることも多くあり、この方法では全体の 5 割が正解すればよいという状況でした。ですから、3D-1D 法とは、確かに構造予測として使えると公認はされましたが、予測の「性能」という点ではあまりよくないことが、このＣＡＳＰのコンテストにより非常にクリアに判ったわけです。

2-2-3. 新しい相同性検索法 PSI-BLAST

新しい相同性検索法 PSI-BLAST　1997 年～

配列比較
プロファイル、E-value 表示
検索結果が信頼できる

PSI-BLAST の登場によって構造予測法は一変

PSI-BLAST による立体構造予測
PSI-BLAST ＋ スレディング手法

画面 22

そのうちに、とんでもない方向から、「PSI-BLAST」という非常に優秀な予測法が出てきました。PSI-BLAST とは、既に世界中で使われていたホモロジー検索法「BLAST」の拡張版であり、同一の開発者によって 1997 年に発表された方法ですが、これはあくまでホモロジー検索のために開発した方法であり、構造予測のためのものではありませんでした。しかし、これを予測法に用いると、3D-1D と同じように非常に弱いレベルのホモロジーまで検知し、かつ、3D-1D に比べ性能が非常によく、信頼性が高いという性質をもっていたのです。

それでは、何がそれほど性能をよくしているかといいますと、従来の BLAST は問合わせ配列 (query) を使ってそのまま配列サーチするのに対し、PSI-BLAST の場合は、query のホモロジー配列の情報も加えた非常に大きなデータベースにかけ、似たものを全て引っぱってくるためです。つまり、まずホモロジー配列の「マルチプルアラインメント」を作り、全ての情報をプロファイル形式にまとめ～この点は 3D-1D のプロファイルと似たような形式ですが～、このプロファイルを用いてデータベースに対しサーチするわけです。ホモロジー検索の場合には配列データベースをサーチさせますが、構造予測をしたい場合には、PDB(Protein Data Bank)のデータベースを検索させます。PSI-BLAST はあくまで配列どうしを比較するホモロジー検索法ですが、配列が似ていれば構造も似ているであろうという論理により、構造予測法としても活用できるわけです。

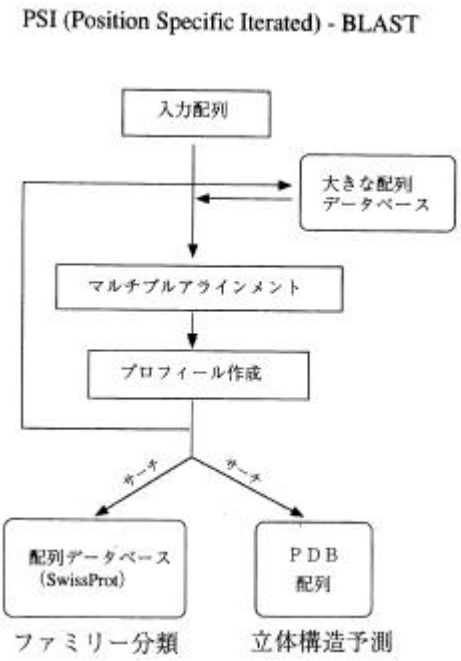
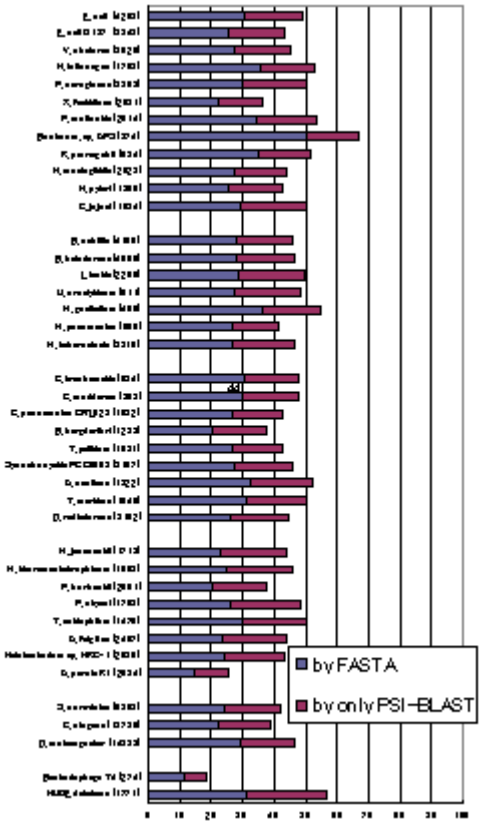


図 23

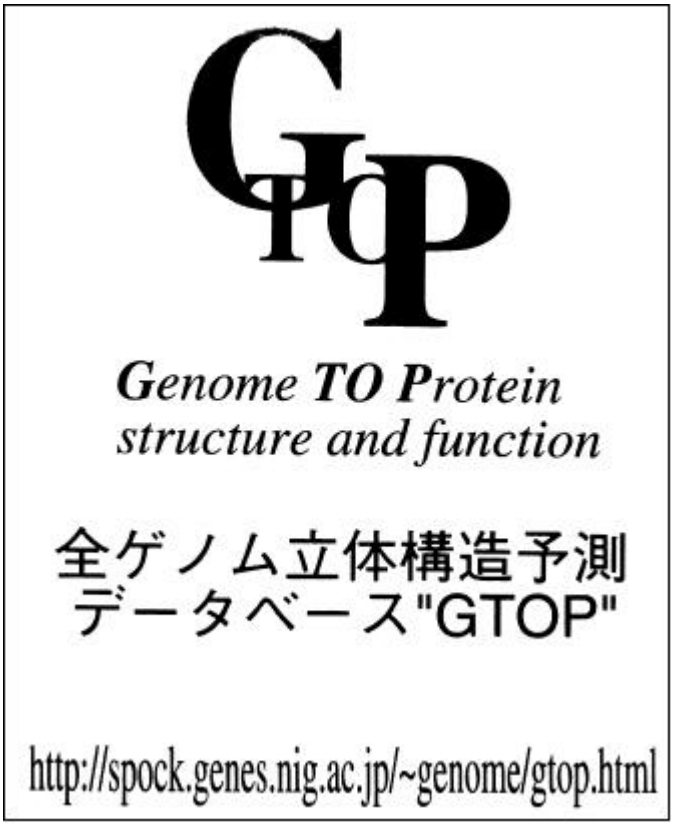
折りしもゲノムの時代が始まり、我々をはじめ多くの研究者が、PSI-BLAST をゲノムに応用するとどのくらいの威力があるかを検討しはじめました。画面 24 は、実際に我々が計測した結果ですが、大腸菌やバクテリア、ショウジョウバエなど、全ゲノム中の遺伝子の立体構造を、FASTA (BLAST と同様に当時広く使われていた相同性検索プログラム)、PSI-BLAST のそれぞれで予測を行った結果です。それぞれの生物種は 4000～20000 個以上の数の遺伝子をもっていますが、全体を 100 とした場合の割合(%)を単純に比較すると、PSI-BLAST が、全体の 45%の立体構造を予測できるという結果がでました。我々は、これらの解析結果を、すべて「GTOP」というデータベースに登録し、WEB にて公開しています(http://spock.genes.nig.ac.jp/~genome/grop.html)。

以上が、これまでのおおよその研究史です。このように、PSI-BLAST は実用的に非常によい方法ですが、それでもやはり 45%しか予測できないという限界があるわけです。そこで、現在注目されている方法である「アブイニシオ法」についてご説明しましょう。



立体構造予測できた ORF の割合

画面 24



画面 25

2-3．アブイニシオ予測の可能性

アブイニシオ予測の可能性
◆CASP での「成功」(1998 年～)
◆最終目標：「X 線構造並みの精度」の予測
◆アブイニシオ予測の２つの段階
　1．フォールドの予測 (fold recognition)
　2．真の立体構造への最適化 (refinement)
◆成功の条件
　1．力場
　2．最適化の理論的手法
◆「拡張アンサンブル法」
　1．マルチカノニカル法
　2．レプリカ交換法

画面 26

アブイニシオ法とは、ラテン語の「はじめから」の意味を持つ、本来は非経験的とでも訳すべき方法ですが、この研究分野では、3D-1D や PSI-BLAST のような類推に基づく方法以外の方法だと思ってください。つまり、データベースを用いた類推ではなく、計算だけで構造を予測する方法をさしており、現在大変注目されだしています。

そのきっかけとなったのは、1998 年に実施された第 3 回目の CASP であり、ここでアブイニシオ法での成功例が出たためです。2000 年に実施した第 4 回 CASP では、もっと成功例が増えており、国際的にも「これは使える」ということが判ってきたわけですが、ただしこれは、理由は後ほど説明しますが、少々カッコつきの成功であるといわざるを得ません。

アブイニシオ法の最終目標は、X 線構造並みの予測を計算機で出すということです。アブイニシオ法は､内容的に 2 つの段階にわけて考えることができます｡1 つめは 3D-1D 法と同じように、全体的にどのような形をしているのかという「フォールドの予測」の段階です。つまり、それぞれのトポロジーを当てるという段階です。そして 2 つめの段階は、さらにそこから最終的な真の立体構造へともってくる段階です。

さて、先ほど、”カッコつきの成功” であったと申し上げましたが、それは、ここでの第一段階でのみ成功しているためです。3D-1D 法では、既知のものと似たものがない場合は解がないという結論で終わりますが、アブイニシオでは、こういう形をしているであろうという解をだし、それが実際に正解であることが判った例が、第 3 回 CASP から出てきたというわけです。しかし、本当の最終目標である、真の立体構造への最適化までには、まだ辿りついていません。

我々が最近考えているのは、その予測が成功する条件は大きく分けて 2 つあるのではないかということです。

まず 1 つは、分子力学の問題である、「フォースフィールド (force field：力場) の選び方」についてです。力場をどうとるかの判定基準は、X 線構造のエネルギーが常にいちばん最小でないと困るのですが、計算していくうちに、全く違う構造なのにどんどん下がってしまうような力場では困るということです。

また、力場の問題が満たされたとして、2 つめに、「最適化の理論」が必要となります。これは構造探索法と呼びますが、特に日本の研究者たちが頑張っておられる分野です。「拡張アンサンブ

ル法」をはじめいくつかの手法もでており、非常に発達してまいりました。このような手法を使えば、第 2 段階の概ねの構造も分かることになり、そこから最終の天然構造にまで辿りつけるのではないか、と考えています。

ここで、先ほどお話しした 2 つめの段階 (真の立体構造の最適化) を求める実例について、我々の計算例をもとにご説明したいと思います。画面 27 は、力場はなにを使ったかということが書かれたもので、我々は、AMBER86 を使いました。その内容は、ローカルなターム(エネルギー項：bond-length,bond-angle,dihederal angle)にディペンドするローカルパターンと、van der Waals のターム、hydrogen bonding (水素結合)、Electrostatic (静電気)です。そして、周囲の水分子の影響を考慮するために、既に開発されている Hydration energy (水場のエネルギー)という表面積に比例する計算方法をプラスし、実際にいくつかテストをしてみました。使用したタンパクは、画面 28 にある７種類だけですが、いろいろなタイプの class のもの (all α、α＋β、α/β) を使用しました。これらのサイズは、およそ 100 残基 (100 アミノ酸) 前後です。ここにある 7 種類のタンパクで同時に計算をし、全てがうまくいくようでないといけないわけです。我々は、タンパクの X 線構造が分かっているものの主鎖構造だけを与え、そこに拘束をかけて強制的なポテンシャルを与え、しかし側鎖は自由に動かすことができるという条件で最小化 (ミニマイゼーション) を行いました。画面 29 がその結果です。

Total energy :

AMBER86 force field
Local (bond-length, bond-angle, dihedral angle)
van der Waals (Lennard-Jones)
Hydrogen bonding (Lennard-Jones, 10-12)
Electrostatic (ε = 2r)

OONS (Ooi et al.) Hydration energy
function of accessible surface area

After the minimization:

Continuum Electrostatic energy (given by Poisson eq.)
in place of the AMBER electrostatic term

画面 27

Target[a]	Nres[b]	class[c]	name	reference
1lmb4	92	all α	lambda repressor	Clarke et al. (1991)
1molA	94	$\alpha+\beta$	monellin	Tomic et al. (1992)
1plc	99	all β	plastocyanin	Guss et al. (1992)
1rgeA	96	$\alpha+\beta$	ribonuclease Sa	Sevcik et al. (1993)
1thx	108	α/β	thioredoxin	Saarinen et al. (1995)
2hmzA	113	all α	hemerythrin	Holmes and Stenkamp (1991)
3chy	128	α/β	CheY protein	Volz and Matsumura (1991)

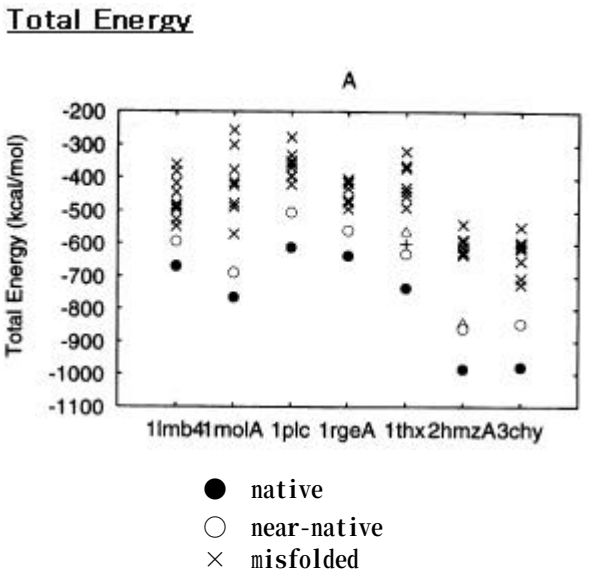
画面 28

model	RMSD (Å)		χ correct (%) [c]	
	backbone[a]	all[b]	χ_1	χ_2
1lmb4-1lmb4	0.136	1.62	78.6	80.0
1molA-1molA	0.116	1.90	76.5	72.7
1plc-1plc	0.121	1.45	70.8	22.2
1rgeA-1rgeA	0.104	1.51	82.4	83.3
1thx-1thx	0.130	1.57	64.3	88.9
2hmzA-2hmzA	0.117	2.01	65.5	79.2
3chy-3chy	0.123	1.55	75.0	85.7
Average	0.121	1.66	73.3	73.1

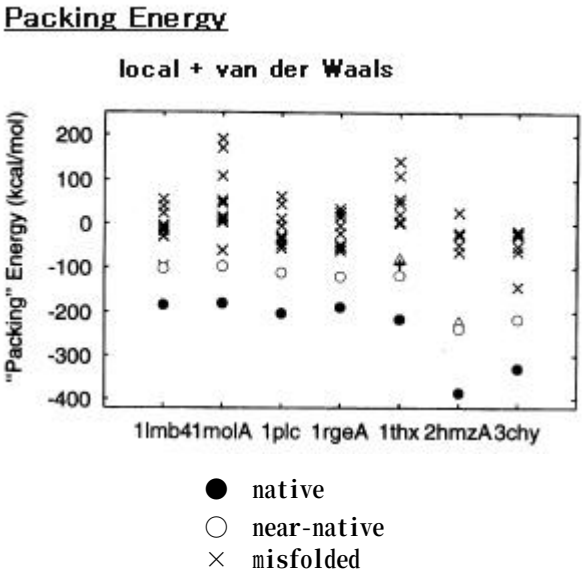
画面 29

7 種類のタンパクの自由度としては backbone も固定せず動けるようになっており、画面 29 を見ると、backbone だけの誤差も少々でています。右側が側鎖構造の比較で、X 線構造と比べると、7～8 割程度の角度は合っていますが、20～30%は違っているという結果が得られたことが分かります。このような、backbone(主鎖構造)だけコンストレーンをかけて得られた構造は、人為的な構造だということで、「ニアネイティブ・モデル」と呼んでいます。

次に、このニアネイティブ・モデルの数値と、実際の X 線構造並みの数値とを比較してみました。画面 30 をご覧下さい。これは 7 種類のタンパクのそれぞれのトータルエネルギーをプロットしている表ですが、●が X 線構造 native です(X 線構造から我々の力場で最小化した結果を用いています)。○は先ほどのニアネイティブ・モデル(near-native)ですが、画面 30 を見ると、やはり native よりは高いところに位置しています。×はミスフォールド(misfolded)といって間違った構造をわざと作りそれを最小化した結果ですが (3d-1d 法で検出された、スコアはよいが完全に間違っている構造を用いています)、それらは全て上方に位置しています。このグラフで、特に我々は、native と near-native との差に注目しました。



画面 30



画面 31

画面 31 は、Total Energy のローカルなタームと van der Waals のタームだけをプロットした結果です。これは「パッキング・エネルギー (Packing Energy) 」と言われていますが、サイドチェーン同士がお互いに噛み合ったときのエネルギーにあたるであろうと思われます。この点が、

先ほどの native と near-native の差にあたるのだということが分かります。次に、misfolded に注目してみましょう。トータルでみると、native と misfolded は差がありますが、パッキング・エネルギーでみると、この差が非常に小さい。しかし native と niar-native だけは非常に差があるということが分かります。

このような計算から、我々は、力場の問題は既に大よそクリアしているのではないかと思っています。現在は、拡張アンサンブル法などを使用して、near-native から本当の native の位置にまでもっていけるか実験を重ねているわけですが、クリアした次は misfolded で試してみるなど、計算を進めていきたいと思っています。そして、このような計算が結局はアブイニシオ予測の成功につながるのではないかと考えています。もちろん、それは成功してから言うべきことですが、概ねゴールは見えてきたなという気がしています。

2-4．予測問題が解決すれば・・！

予測問題が解決すれば・・！

◆計り知れない波及効果
　ゲノム情報解析への応用
　「X 線構造」の増大
　モデリング／ドッキング、など
◆予測可能 ＝「理解できた」
　フォールディングの理解
　　人工タンパク質の世界が開ける（医学、薬学、工業薬学.. ）
◆タンパク質「基礎研究」の終わり、
　「応用研究」の時代へ
　超大型マシーンによる理論計算への需要

画面 32

もし、タンパク質の構造予測問題が解決したとするとどうなるでしょうか。一言でいって、これは計り知れないほどの波及効果があるであろうと私は考えます。

まず、X 線構造並みの精度がよいモデルが計算でできるということです。これは、ゲノム解析の 100%近くがアブイニシオ法で予測できることを意味する筈で、現在、いわゆるモデリングだとかドッキングといわれるものが、革命的に改良されていくと思われます。

また、予測問題が解決されるということは、我々人間が予測問題を “理解できた” ということです。特に、非常に難しいと言われているフォールディングについても理解できたことになります。そのことにより、現在は実際には試行錯誤的なことしかやられていないタンパク質工学の分野にも、アミノ酸配列の設計やデザインなど、真の意味での研究ができるようになるわけで、それらは医学や薬学、工業薬剤などにも応用できます。

このように、タンパク質の構造予測問題が解決することにより、基礎研究の時代は事実上の終わりを告げ、応用研究の時代に入っていくものと思われます。新しい応用研究の時代には、超大型のマシーンが必要となり～それは大きければ大きいほどよい筈です～、今後、それらの需要は物凄い勢いで上がっていくことでしょう。