

構造ジェノミクスとバイオインフォマティクス

(株)ファルマデザイン 松尾洋

はじめに

ゲノム計画の進展により、各種生物種の遺伝子のレパートリーの全貌が明かされつつある。ヒトゲノムに関しても、その配列決定プロジェクトはいよいよクライマックスを迎えつつある。昨年(1999年)12月には、22番染色体の全配列が決定されたことが日米欧の共同チームにより発表された(文献1)。本年1月には、Celera Genomics社が、全ゲノムの約90%に相当する配列の決定を完了したことを発表した(<http://www.celera.com/>)。米国NIH・英国Sanger Centreを中心とするプロジェクトからも、本年3月までには全ゲノムの90%をカバーするrough draft配列が公開されることが予定されている(2000年1月時点; 例えば、

http://www.nhgri.nih.gov/NEWS/human_genome_defined_by_spring.html、等参照)。

このような状況の中でポスト・ゲノムの課題は、レパートリー中のそれら多数の遺伝子に関するより高次のデータ(産物の立体構造・機能、発現プロファイル、多型、産物と他分子との相互作用、等)をゲノム規模で高速に収集すること(genome-scale high-throughput data production)、そして、それら多種多様大量のデータを効果的に組織化し総合的に解析することにより、遺伝子型を表現型へ結び付けるような情報を抽出すること(bioinformatics)である。そこで近年、そのための様々なアプローチが開始されている(図1)。例えば、創薬・医療の観点から最近大きな注目を集めているのが、SNPs(一塩基多型)収集を中心とする薬理ジェノミクスである。完成しつつあるヒトゲノムの「参照」配列に基づき、薬剤応答性・疾患感受性の違いをもたらすDNA多型(その多くを占めるSNPs)を組織的に収集することは、ゲノム計画の自然な延長上にある。

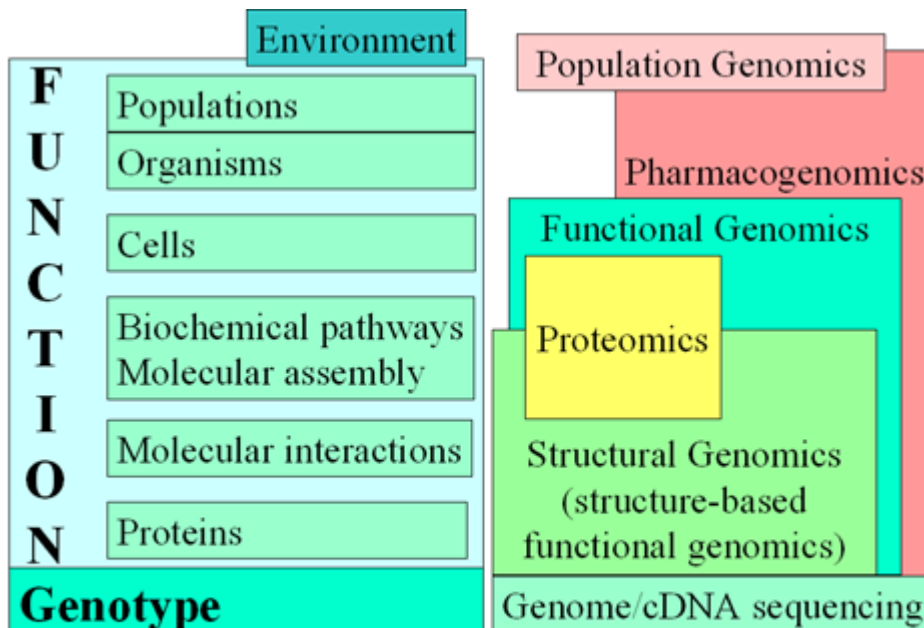


図1. 遺伝子型を表現型へ結び付けるための、様々なポスト・ゲノムのアプローチ。生体システムを構成する物理的階層における様々なレベルに応じて、多彩なアプローチが試みられている。

一方、こうして得られる配列・多型・遺伝子産物（蛋白質）機能に関する情報を創薬・医療へと結び付ける上で重要なのが、蛋白質の立体構造に関する情報である。疾患のメカニズムを分子レベルで理解し、薬剤の合理的設計を行なうには、立体構造情報が不可欠だからである。そのため、蛋白質立体構造情報をゲノム規模で組織的・大量に収集しようという「構造ジェノミクス」のプロジェクトが、日米欧でほぼ時期を同じくして本格的に始まるようとしている（2、3）。

そこで以下では、構造ジェノミクスに注目し、その現状を紹介する。その中で特に、私自身が担当している分野であるバイオインフォマティクスの役割にも注目する。さらに、構造ジェノミクスとSNPsデータ（薬理ジェノミクス）の結合が、医療・創薬に大きなインパクトを与え、ゲノム創薬・テラメド医療実現の基盤となることを論ずる。

構造ジェノミクスの概要

構造ジェノミクスの目標は、生物ゲノム中にコードされているすべての蛋白質に対して立体構造情報を与えることである。このような課題が現実的なものとなったのは、ゲノム計画の進展により大量の遺伝子が同定されクローンが得られるようになったこと、及び、技術的進歩によりX線結晶解析とNMRによる立体構造決定がかなりの程度ルーティン的に行なえるようになってきたことによる。

ただしそうは言っても、一つの立体構造を決定するには、遺伝子（cDNA）の塩基配列を決定するまでに較べると、はるかに多くの時間・コストを要求される。2000年1月26日現在、PDBには通算10,521の蛋白質立体構造座標（理論モデルを除く）が登録されており、その数は年間約2,000のペースで増加している（<http://www.rcsb.org/>）。このペースだと、例えばヒトゲノム中の約10万前後の蛋白質の立体構造をすべて実験的に決定するためには、単純に計算して最低何十年もかかることとなる。

そこで構造ジェノミクスでは、次のようなステップで立体構造決定を進める：

- （1）アミノ酸配列比較により蛋白質をクラスター（ファミリー）に分類する（同一残基率約30%程度以上ならば一つのクラスターに分類する）；
- （2）様々な観点（医学的な重要性、等）から、構造決定を行なう上でのこれらクラスターの優先順位を付ける；
- （3）各クラスターから1つずつ代表（ターゲット）を選ぶ；
- （4）選ばれた各ターゲットの立体構造を実験的に決定する。

こうして、全蛋白質クラスター（ファミリー）の立体構造を網羅的に決定していくのである（cf. 文献4）。

蛋白質の立体構造（フォールド）は、進化の過程でアミノ酸配列が大きく変化しても、全体として非常に良く保存されることが経験的に知られている。とりわけ、同一残基率が30%以上もあれば、それら蛋白質の立体構造が全体的に極めてよく保存されていることはほぼ確実である。従って、上記のようにして得られるクラスター（ファミリー）内の一つの蛋白質の立体構造を実験的に決定すれば、同じクラスター（ファミリー）内の他の蛋白質の立体構造は、比較モデリング（ホモロジー・モデリング）によってかなりの精度で予測することができるようになる。また、実験的に解くにしても、分子置換法を用いて比較的簡単に構造決定できるようになると期待される。

例えば、米国NIH（NIGMS）主導のプロジェクトでの計算によれば、35%の同一残基率を基準

にして蛋白質を分類していけば、得られるクラスターの数は最初の蛋白質の数の約10%にまで減る(2)。従って、仮にヒトゲノム中に約10万の遺伝子が存在していたとして、約1万の蛋白質の立体構造を実験的に解けば、ヒトゲノム中の全蛋白質クラスター(ファミリー)をカバーできることになる。先程述べたように、現在年間約2,000の構造座標が世界中から公表されている。この努力をすべてヒト蛋白質の立体構造の決定に割り当てたと仮定すると、約5年でヒトゲノム中の全蛋白質の構造情報が得られる計算になる。

当然、現在一年間に報告されている約2,000の構造座標のかなりの部分は、既に以前に立体構造が報告されている蛋白質の立体構造を異なる条件で解いたものや、近縁のホモログの立体構造を解いたものによって占められている。従って、1万蛋白質クラスターを5年でカバーするには、さらに立体構造決定のペースを加速させる必要がある。さらに、例えば膜貫通領域部分の構造決定に関して技術的なブレイクスルーが必要となるであろう。これらのことのためにこそ、日米欧で組織的な立体構造決定のためのプロジェクトが立ち上げられようとしているのである。

現在計画されているかまたは既にスタートしている大規模な構造ジェノミクスプロジェクトとしては、次のようなものがある。日本では、理化学研究所ゲノム科学総合研究センター(1998年10月発足)のタンパク質構造・機能研究グループ(横山茂之プロジェクトリーダー)が挙げられる。本プロジェクトは現在、和光市にある理化学研究所内で既に進行しているが、2000年10月に横浜市鶴見区のゲノム科学総合研究センターのキャンパスに移転し、本格的な量産体制を整える。また、昨年の本誌で紹介があったように、大阪大学・理研播磨ストラクチュローム(SPring-8)の倉光プロジェクト(高度好熱菌丸ごと一匹プロジェクト)が進展している(3)。米国では、6つのプロジェクトが、NIH(NIGMS)からそれぞれ年間300万ドル程度の予算を得て、2000年度よりスタートする予定である(2)。欧州では、例えば、ドイツ政府から5年間で約2,000万ドルの資金を得る予定のコンソーシアムProtein Structure Factory、がある。

さて、構造ジェノミクス全体としての目標は全蛋白質の立体構造情報を網羅することである。しかしながら、具体的に進行しようとしている個々のプロジェクトでは、それぞれ独自の観点から、構造決定の候補となるターゲット蛋白質の間に優先順位をつけて、それに従って実験的立体構造決定を進めることとなる。そこで、それぞれのプロジェクトがどのような観点から優先順位付けを行なうのかが興味の持たれるところだ。高度好熱菌丸ごと一匹プロジェクトは、全生物種に共通の基本的な蛋白質のセットを網羅することを目指していると考えられる(3)。一方で、後生動物に特異的な蛋白質のセットを優先的に網羅するという観点や、ヒト疾患関連蛋白質群を優先的に網羅するという観点などもあり得る。

構造ジェノミクスにおけるバイオインフォマティクス

構造ジェノミクスのプロジェクトにおいてバイオインフォマティクス(生物情報学)が重要な研究手段の一つとなることは、他のジェノミクスプロジェクトにおいてと同様である。極論すれば、ジェノミクスプロジェクト自体は、データの機械的大量生産とインフォマティクスの二本柱で成り立っているとさえ考えられる。

構造ジェノミクスにおけるバイオインフォマティクスには、特に次の2つの大きな役割がある。

1つは、網羅的な立体構造決定を効率的に行なうのを支援することである。具体的には当面、先に述べた手続きに従って多数の構造決定候補蛋白質群をクラスタリングして優先順位を付け、構造決定のターゲットを選択することが大きな課題となる。また、世界中の様々なプロジェクト

の間でできるだけ重複なく立体構造決定を行なうために、プロジェクト間での情報の共有を行なうことも重要である（例えば、PRESAGEプロジェクト； <http://presage.stanford.edu/>）。

もう1つは、決定される多数の立体構造データを機能的・進化的な特徴と関連付けながら分類・解析してデータベース化し、生物学・医学関連の研究において有効に利用することである。そのために、一般的な情報解析技術に加えて、立体構造予測（ホモロジー・モデリング）・ドメイン境界予測・立体構造比較などの多数の理論的手法が用いられる。既に述べたように、1つの立体構造が決定されれば、ホモロジー・モデリングによって他の多数のホモログの立体構造が予測できるようになる。構造ジェノミクスの進展は、そのような構造予測を適用できる範囲をこれまでと較べて飛躍的に拡大することになるであろう。その他にも、大量の立体構造情報が得られるようになれば、例えばHomologous Core Structure（配列モチーフの立体構造版のようなもの；文献5）を多数の蛋白質ファミリーに対して適用できるようになり、蛋白質の構造と機能・進化との関係をゲノム規模で体系的に理解することが可能になる。

構造ジェノミクス以後

構造ジェノミクスのプロジェクトの直接的な目標は、ゲノム規模で蛋白質立体構造情報を網羅することにある。それら構造情報は、さらなる生物学・医学研究や創薬その他の活動にとって有用なインフラストラクチャとなる。

構造ジェノミクスの進展の結果、配列・立体構造は既知だが機能は未知であるような蛋白質が多数もたらされることになるだろう。蛋白質に関して我々が最終的に関心を持つのは、大抵の場合立体構造自体ではなく機能である。立体構造情報は、機能メカニズムを分子レベルで理解するために必須である。しかしながら、立体構造情報が得られたからといって、一般には立体構造情報自体からその蛋白質の機能を演繹的に予測することはできない。従って、立体構造決定と呼応する形で機能を解明するためのハイスループットの実験体制（機能ジェノミクスプロジェクト）を整える必要がある。

蛋白質には、様々な物理的レベルと観点からの「機能」が存在する。どのような低分子を結合するかを知るためには、例えばHTS（High Through-put Screening）のようなアプローチが必要となるだろう。どのような蛋白質群と協調して働いているか（分子系の中での役割）を知るには、例えば、DNAチップを用いた発現プロファイル収集や、酵母2 ハイブリッドシステムやプロテインチップを用いた蛋白質-蛋白質相互作用の解析が必要となるだろう。生体の中での役割（高次の表現型との関係）を知るには、例えば組織的なノックアウト実験などが必要となるだろう。このように、構造ジェノミクスを有益なものとするには、機能ジェノミクスとの連携が必要となるのである。

さらに、個々の蛋白質の立体構造が大量・組織的に決定されるようになれば、構造生物学研究者の課題の重点は複合体の構造決定へと移ることだろう。複合体（蛋白質-リガンド複合体、蛋白質-蛋白質複合体、蛋白質-核酸複合体、multi-enzyme complex、等）の構造データが多数得られるようになれば、蛋白質の機能メカニズムについての理解はさらに深まるであろう。

構造ジェノミクスとSNPs収集プロジェクトの統合

ここで特に、最近大きな注目を集めているSNPsの大量収集と、構造ジェノミクスの成果とが結びつくことにより、創薬・医療に対してどのように有益な結果がもたらされるかを考えてみよう。

SNPsの大量収集は、体質・疾患感受性・薬剤応答性など多種多様な表現型の個体差と遺伝子型との連関を検出することを可能にする。特に、コード領域及び転写調節領域におけるSNPs (cSNPs) は、遺伝子産物の質的・量的変化を通じて疾患感受性・薬剤応答性の違いに直接的に関与する可能性があるため、重要である(6、7)。しかし多くの場合では、SNPsによる候補遺伝子群の同定は、創薬への出発点に過ぎないだろう。特に、多因子性疾患とSNPsは、多くの場合(あったとしても)弱い連関しか示さないだろう。そこで、候補遺伝子の多型が疾患感受性・薬剤応答性の違いをもたらす(あるいは、もたらさない)メカニズムを知り、合理的に治療法・薬剤を設計するには、立体構造情報に基づく蛋白質機能メカニズムの理解が重要となるだろう。

SNPsデータと蛋白質立体構造情報の統合がいかに有用な情報をもたらすかを、メチレンテトラヒドロ葉酸還元酵素(MTHFR)を例に検討してみよう。調査された世界中のどの地域集団においても、MTHFR遺伝子にはC677Tという対立遺伝子が大体30%程度の頻度で存在することが知られている(8)。この単一の塩基の置換は、産物の222残基目のアラニンを変化させる。この変化により酵素活性が低下することが知られている。MTHFRは、血中のホモシステインをメチオニンに代謝するために必須なメチルテトラヒドロ葉酸を供給するための唯一の反応を触媒する(図2)。従って、MTHFRの活性の低下は、血中のホモシステイン濃度を高める結果となると考えられる。一方、血中の高ホモシステイン濃度が循環器疾患発症リスクを高めることを示すassociation studyの結果が多数報告されている。そこで、MTHFR遺伝子のA222V変種は、循環器疾患感受性の上昇と関連していることが予想される。実際そのことが、association studyにより報告されている(8)。そこで、そのような循環器疾患感受性の上昇の分子レベルでのメカニズムを理解するためにMTHFRの立体構造を知ることが望まれていた。昨年、その大腸菌ホモログ(図3)の立体構造(図4)が決定された(9)。それにより、図5に示すように、MTHFR遺伝子の多型を表現型(循環器疾患感受性)の違いに結び付ける分子レベルでの機構が示唆されたのである(9)。

このMTHFRの例が示唆するように、今後得られるであろう大量のSNPsデータと蛋白質立体構造データとを統合して利用する中から、副作用が少なく個々人の体質により適合し得る新しい医薬品・投与量選択法の開発がもたらされる事例も多く出てくるであろう。

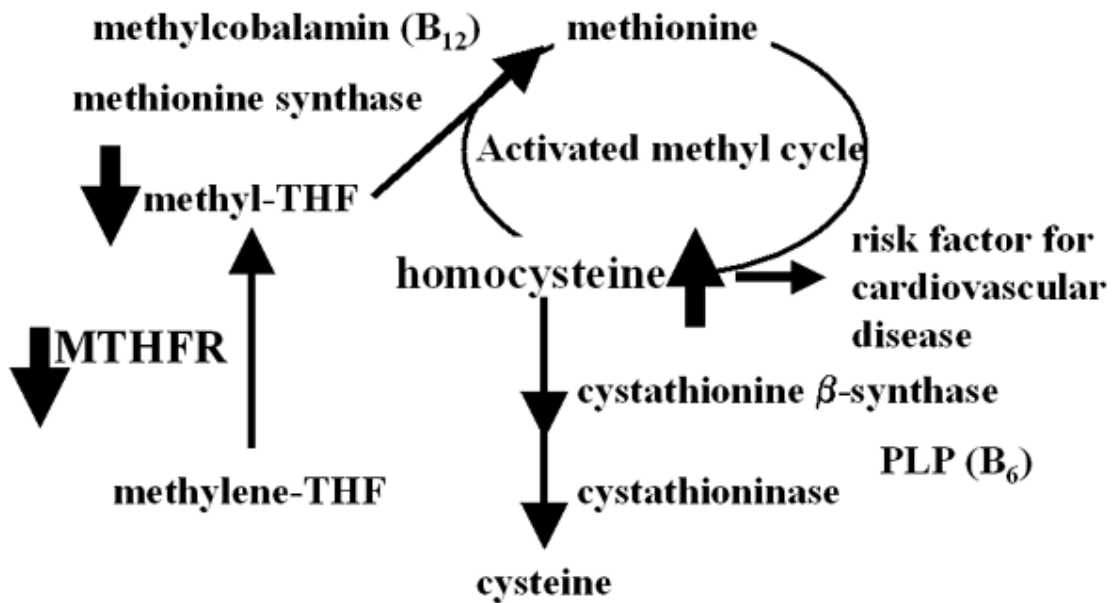


図2. MTHFR C677T→血中高ホモシステイン濃度→循環器疾患感受性への相関。

```

MVNEARGNSSLNPCLEGSASSGSESSKDSSRCSTPGLDPERHERLREKMRRRLLES--GDK
-----MSFFHASQRDALNQLAEVQGOI
      . * * : : : * . * :

WFSLEFFPPRTAEGAVNLI SRFD RMAAGGPLYIDVTWHPAGDPGSDKETSSMMIASTAVN
NVSFEFFPPRTSEMEQTLWNSIDRLSSLKPKFVSVTYG-ANSGERDR---THSIIKGIKD
. *:*****:* . * . :**:: * ::** : *.. * : : * . :

YCGLETILHMTCCRQRLEEITGHLHKAKQLGLKNIMALRGD-PIGDQWEEEEGGFNAYVD
RTGLEAAPHLTCIDATPDELRTIARDYWNNGIRHIVALRGDLPPGSGKPE----MYASD
***: *:** :*: :. : *::*:***** * ↓ * **

LVKHIRSEFGDYFDICVAGYPKGHPEAGSFEADLKHLKEKVSAGADFIIITQLFFEADTFF
LVTLLK-EVAD-FDISVAAYPEVHPEAKSAQADLLNLKRKVDAGANRAITQFFFDVESYL
** . : *..* ***.**:**: **** * :** :**.***.*** : **:***:::

RFVKA CTDMGITCPIVPGIFPIQGYHS LRQLVKLSKLEVPQEIKDVIEPIKDNDAAIRNY
RFRDRCVSAGIDVEIIPGILEVSNFKQAKKFADMTNVRI PAWMAQMF DGLDD-DAETRKL
** . *.. ** *:**:*:..... : : : : : : * : : : : : * ** * :

GIELAVSLCQELLASGLV PGLHFYTLNREMATTEVLKRLGMWTE D PRRPLPWALSAHPKR
VGANIAMDVKILSREGVKDFHFYTLNRAEMSYAICHTLGVRPGL-----
      . :** : * :***** : : : ** :

```

図3. ヒトMTHFR 触媒ドメインと*E. coli* MTHFRとのアミノ酸配列アラインメント (同一残基率33%)。矢印が、A→V多型の生じる位置。

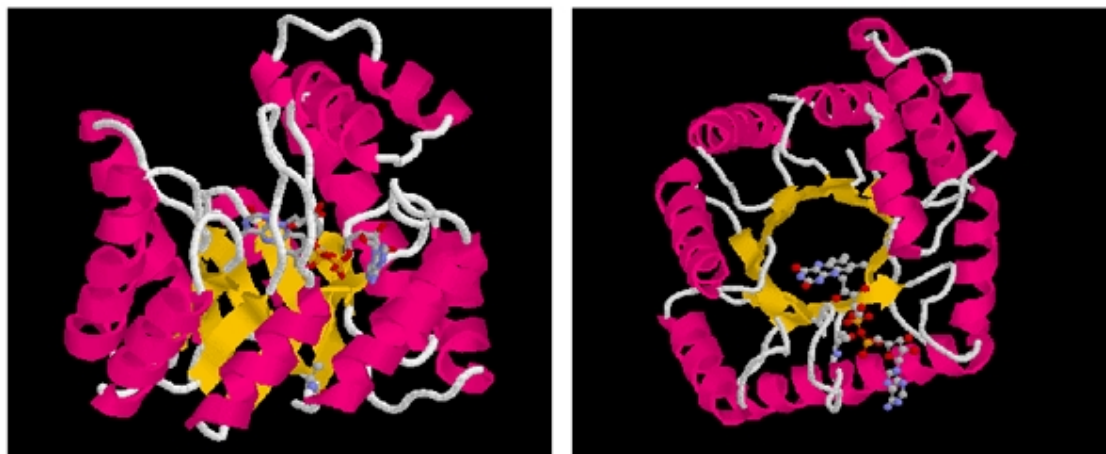


図4. *E. coli* MTHFRの立体構造 (Guenther *et al.*, 1999; PDB code 1B5T)。多型の生じるAla177は5番目の α ヘリックスと6番目の β ストランドの間(樽状構造の底面)に位置する。それとは逆に、基質・FAD結合部位は、樽状構造の上部の位置する。

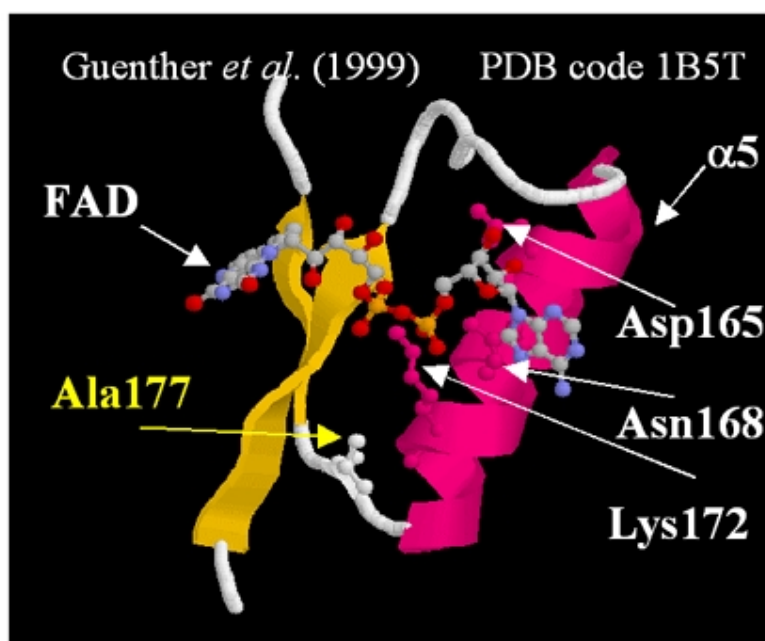


図5. A177V変異は5番目のヘリックスを乱す。そのことにより、FADが解離しやすくなり、従って酵素活性が低下すると考えられる。

おわりに

日米欧における構造ジェノミクスプロジェクトは、期せずして昨年辺りからほぼ同時期にスタートし、今年以降本格化する情勢にある。今後これらのプロジェクトが順調に進展すれば、配列決定を中心としたゲノム計画と同様に、生物学・医学研究の進め方に大きなインパクトを与える成果がもたらされるであろう。また一方で、創薬・医療等を通じて産業界・社会に与えるであろうインパクトも見逃すことはできない。例えば創薬・医療の観点から特に重要と思われる蛋白質群の立体構造情報は、早い時期（数年以内）にほぼすべて与えられてしまう可能性さえある。米国では、ハイスループットの蛋白質立体構造決定を目的とするベンチャー企業も現れ始めている。遺伝子配列決定の分野において起こった特許化競争が、蛋白質立体構造情報に関連して起こる可能性もある。今後の展開が注目される。

参考文献：

1. Dunham, I., *et al.* (1999) *Nature* **402**:489-495.
2. Protein Structure Initiative by NIGMS, NIH. <http://www.nih.gov/nigms/news/announcements/psi.html>
http://www.nih.gov/nigms/news/meetings/structural_genomics_targets.html
http://www.nih.gov/nigms/news/reports/protein_structure.html
<http://www.biodigm.com/science/sg-review.htm>
3. 倉光成紀 (1999) 構造生物、Vol.5, No.2, 16-23.
4. Burley, S.K. *et al.* (1999) *Nature Genetics* **23**:151-157.
5. Matsuo, Y and Bryant, S.H. (1999) *Proteins* **35**: 70-79.
6. Cargill, M. *et al.* (1999) *Nature Genetics* **22**: 231-238.
7. Halushka, M.K. *et al.* (1999) *Nature Genetics* **22**: 239-247.
8. Morita, H. *et al.* (1997) *Circulation* **95**: 2032-2036.
9. Guenther, B.D. *et al.* (1999) *Nature Structural Biology* **6**: 359-365.