

# 談話室

## CASP6会議報告—日本勢がんばる—

神戸大学理学部 千見寺浄慈, 高田彰二  
産業技術総合研究所生命情報科学研究センター  
本野千恵, 富井健太郎

### 1. CASPとは (高田)

“ナンバー1よりオンリー1”という心地良いフレーズが一時流行した。科学研究は本来オンリー1を目指すものであるけれど、時として一つの問題解決に対する競争の中でナンバー1を狙ってしまうこともある。その典型例がCASP<sup>\*</sup>(タンパク質構造予測技術評価会議)かもしれない。

アミノ酸配列情報からタンパク質の天然立体構造を予測する、いわゆる構造予測問題<sup>\*</sup>は、生物物理の古典的未解決問題として、多くの理論家の間で熾烈な競争の最中である。予測技術を正しく評価することは意外に難しい。研究者は普通、PDBに登録されている構造を“予測”してみて“よくできました”という論文を書く。答えを知っていて予測するから、予測過程でズルをしているか否か、論文を読んで看破することは困難である。

そこでCASPが登場した。CASPでは、実験的に近々構造決定が終わる予定のタンパク質配列情報を実験グループからもらい、主催者のウェブページ上で初夏に“問題”として公開する。予測家は夏の定められた期間内に、これら“問題”に対する“解答”(予測構造)をアップロードする。実験構造決定後の10月頃、第三者が各予測構造の良否を数値スコアによって評価する。各問題のスコアを総合して予測チームの成績が決まる。12月に結果発表の会議(2004年はイタリアのガエタ)が行われ、スコア上位者が講演を行いフィナーレとなる。1994年以来隔年開催されており、2004年のCASP6では、76の問題に200を超える予測チームが挑んだ。

予測法の違いを反映して、CASPには問題となるタンパク質による部門分けがある。問題タンパク質の分類

は、問題タンパク質の配列・構造とPDB中の配列・構造が、どれくらい類似しているかで決まる。CASP6では類似度の高いほうから、CM/easy, CM/hard, FR/H, FR/A, NFの5分類があり(CM = comparative modeling, FR = fold recognition<sup>\*</sup>, H = homologous, A = analogous, NF = new fold)、このうち前3者をまとめてホモロジーモデリング<sup>\*</sup>(HM)部門、後2者を非ホモロジーモデリング(非HM)部門と呼ぶ。また予測チームには、全自動計算によって48時間以内に解答を出すサーバと、数週間くらいの期間で人が介在して行うヒューマンチームとがある。

CASP6において日本の寄与は、これまで以上に高かった。国別参加チーム数が米国について2位であった点は、日本のこの分野におけるアクティビティを反映している。主要成績は、FR/H部門で産総研CBRC(Computational Biology Research Center)が3位、北里大が4位に入りそれぞれ講演を行った。NF部門では、神戸大がElite Club Membershipに入り、NF部門サーバ中では神戸大Rokkyが2位であった。また特記すべきこととして、理研構造ゲノムRSGIから大量の問題タンパク質を提供していただき、一機関からの提供としては最大級のものとなった。

以下、HM部門、非HM部門それぞれについて述べる。

### 2. ホモロジーモデリング(HM)部門(本野, 富井)

HM部門に関する評価と成績上位者の発表は会議2日目の12月5日に行われた。今回は、CMとFR/Hの2部門に分けられ、それぞれA. ValenciaとR. Dunbrackによる評価結果が報告された。彼らによる評価では、CM部門の成績上位に、K. Ginalski(Univ. of Texas, 米国)、C. Venclovasら(Inst. of Biotechnology, リトアニア)、A. Kolinskiら(Warsaw Univ., ポーランド)と、J. Skolnickら(Univ. at Buffalo, 米国)のチームが選ばれた。FR/Hでは、同じくGinalski, M. Federら(Intl. Inst. of Molecular & Cell Biology, ポーランド)、私達CBRC-3D(富井, 広川, 本野, 産総研CBRC)とCHIMERA(竹田-志鷹ら, 北里大)チームが選出された。

これらのチームのほとんどが、各自予測過程を極力自動化し、ほぼ全問に回答していた。ただ、CM, FR/H共に上位であったGinalskiの優れた予測モデルは、自動的に生成されたものを、タンパク質立体構造への深い造詣や丹念な調査に基づいて選別し磨きをかけたも

### CASP6 Meeting Report—Japanese Predictors Achieved Good Scores—

George CHIKENJI<sup>1</sup>, Shoji TAKADA<sup>1</sup>, Chie MOTONO<sup>2</sup> and Kentaro TOMII<sup>2</sup>

<sup>1</sup>Faculty of Science, Kobe University

<sup>2</sup>Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology



FR/H部門3位に入り講演する富井(撮影は鬼塚氏(松下電器))

のであった。今回参加したチームの多くの予測が、3D-JuryやRobettaなどのメタサーバに依存していたが、私達CBRC-3Dのプロファイル比較法を利用した網羅的モデル構築/検証法やSkolnickらのthreadingのような独自手法の健闘もみられた。有効な新規予測技術の開発を促すというCASPの目標からは、他者の予測結果を利用するメタサーバへの賛否もあるが、予測精度向上への貢献度は大きく、当面主流であると思われる。今回、私達の予測結果を通して、信頼度が低い予測しか存在しない場合に顕在化するメタサーバの弱点を乗り越える可能性を示せたことが、予測技術向上への足がかりとなることを期待している。今後は、より強力かつ有効な手法の構築を目指して行きたい。

結果発表とは別に興味深かった点は、米国NIHのグラント (high-accuracy protein structure modeling) の担当者が会議に参加し、発表枠を取って趣旨を解説し応募を呼びかけていたことである。これに象徴されるように、HM部門では、今後一層モデル構造の精密化のレベルでの競争が激しくなり、その技術が革新されていくはずである。医薬など他分野での応用に耐えるモデル構築を視野に入れた研究が望まれる段階にきていると思う。

もう一点、HM部門の成績上位には、米国の他にポーランド勢が目立った。今は米国のGinalskiも、かつてはポーランドから参加していた。ポーランドでは、CASPでの成績とグラント獲得とに強い相関があるという。この分野に資金的に手厚いサポートがあることが、予測技術開発の動機付けの牽引力にもなっていることをよく示していると思われた。

### 3. 非ホモロジーモデリング (非HM) 部門 (千見寺)

その日、来るべき (だと思い込んでいた) メールは来なかった。その日は、CASPの上位入賞者にCASP

会議での講演依頼のメールが来る日のことである。次の日も、またその次の日もメールは来なかった。「おかしい、何か間違っている。」確かに間違っていた (我々の認識が)。結局、我々は上位入賞者ではなかったのである。我々 (藤埜、千見寺、朴、高田) は自慢のエネルギー関数、構造探索法をひっさげてCASP6の非ホモロジーモデリング部門に参加した。実をいうと、我々はこの部門で3位以内に入ることを目標に準備を進めてきた (つもりでいた) のだ。以下は失意のままに参加したCASP会議の報告である。

FR/Aの評価はR. Dunbrack氏が行い、上からD. Baker, K. Karplus, K. Ginalski, D. Jonesというランク付けがなされた。ここではBakerがずば抜けて強く、2位以下にダブルスコア以上の差をつけるほどであった。Bakerの強さの理由として、やはりフラグメントアセンブリ法は強力だった、ということに加えて (i) コンタクト予測がうまくいった、(ii) 多数のホモログのシミュレーションを行い信頼度を上げた、(iii) 全原子モデルによる構造精密化がうまくいった (ものもあった)、といったことが挙げられる。

NFの評価者はB. K. Lee氏であった。Lee氏は (他のカテゴリーとは異なり) 明確な順位付けは行わずに上位約10チームを発表し、そのチームに“Elite Club Membership”を与える、という発表を行った。そのエリートクラブ会員は (厳密ではないが大雑把にいて成績順に)、D. Baker, A. Kolinski, K. Ginalski, K. Karplus, J. Skolnick, D. Jones, D. Baker (server), S. Takada, S. Takada (server), J. Lee, Y. Azaria, といったグループである。我々が高田グループも講演こそ逃したが、ヒューマン、サーバともにエリートクラブ会員に入ることができ、そこそこ健闘したといえる (とはいえ、本当は3位以内を狙っていたので個人的には悔しさを禁じ得ないが)。我々よりも上位チームはスレッディングサーバなどの情報を積極的に使い、コンタクト予測等に活かしていたのに対して、我々は正攻法すぎた感がある。我々は自分たちのエネルギー関数を信じてシミュレーションしただけなのだ。勝つためには他人のサーバをカンニングするなど、もっと泥臭いことをすべきだったかもしれない (不本意だが)。NF部門の全般的なパフォーマンスは悪かったというべきだろう (Bakerといえども)。全体の傾向として、NFターゲットといえども部分構造は既知構造データベースに存在したので、答えの構造が既知構造に類似していればいるほど予測精度が上がっていたという点は興味深い。なぜこのようなことになっているのかを明らかにすることがNF予測のブレイクスルーの鍵になるような気がする。

最後に生物物理学者(だと勝手に思いこんでいる者)の立場から見た CASP の感想を述べたい。生物物理の立場から見た立体構造予測というものは「人間がタンパク質の立体構造構築原理についてどれくらいわかっているかを確かめるための尺度」という位置づけにあると考えられる(どうでしょうか)。ところが CASP ではあてること至上主義になっており(あたるようになること自体はとても大事ですが)、何もわからなくても予測できるようになればよいことになっている。その象徴ともいべきものがメタサーバの君臨であろう。生物物理学者たるもの「わかってあてられる」ようになりたいものである(そのような研究も暖かく見守ってほしいものだ)。そしてその上で、勝ちたいもので

ある。

#### 4. 最後に (高田)

CASP6では、上述の構造予測のほかに、disorder 部位予測、ドメイン分割予測、機能予測も行われた。CASP はタンパク質モデリングの周辺問題もとり込みながら変化を続けている。

CASP 参戦者たちは皆ナンバー1を狙って凌ぎを削っているけれど、まだ誰もが、任意の配列に対して高確率高精度で構造予測するという最終目標には遠くおよばない(とくにNF)。そういう意味では勝者なんていない。最終目標を最初に実現する人だけが、“オンリー1”になれるのである。

**千見寺浄慈（ちけんじ じょうじ）**

神戸大学理学部博士研究員

連絡先：〒657-8501 神戸市灘区六甲台町

E-mail: [chikenji@theory.chem.sci.kobe-u.ac.jp](mailto:chikenji@theory.chem.sci.kobe-u.ac.jp)

**高田彰二（たかだ しょうじ）**

神戸大学理学部助教授

連絡先：〒657-8501 神戸市灘区六甲台町

E-mail: [stakada@kobe-u.ac.jp](mailto:stakada@kobe-u.ac.jp)

**本野千恵（もとの ちえ）**

産業技術総合研究所生命情報科学研究センター研究員

連絡先：〒135-0064 東京都江東区青海2-42

E-mail: [c-motono@aist.go.jp](mailto:c-motono@aist.go.jp)

**富井健太郎（とみい けんたろう）**

産業技術総合研究所生命情報科学研究センター研究員

連絡先：〒135-0064 東京都江東区青海2-42

E-mail: [k-tomii@aist.go.jp](mailto:k-tomii@aist.go.jp)