

ヒトゲノム構造・機能アノテーションデータベース SAHG の公開

本野 千恵
創薬分子設計チーム
研究員



DNAにコードされた遺伝情報はアミノ酸配列に翻訳され、それが天然状態で固有の立体構造を形成した結果、分子機能という形で具現化します。ゲノム配列決定を受けて直ちにその生物由来のタンパク質の構造と機能をインフォマティクスの力で列挙するという展望のもとにアミノ酸配列から立体構造を予測・更に分子機能を類推する方法の基礎研究と、それを「タンパク質の構造・機能アノテーションシステム」という形で統合化するプロジェクト (JST BIRD、代表: 名大・太田元規教授) にCBRCの研究者も参画してきました。このシステムをヒトゲノム由来のタンパク質配列に適用し、その結果を「Structural Atlas of Human Genome: SAHG」というデータベースにまとめウェブ上で公開する運びになりました。私は立体構造予測の自動化システム構築と、他の参加メンバーによる構造・機能予測結果のデータベース化を担当しています。

タンパク質立体構造予測に関しては、CBRC富井健太郎研究員の開発したフォールド認識法FORTEを中心とする予測手法が、世界的な評価機構CASPの上位にランクされた実績があり、既に確立した技術だと感じていました。しかし、高等生物由来タンパク質はディスオーダー領域を多く含みマルチドメイン構造をとるため、立体構造予測精度が上がらないケースも見られました。ヒトゲノムへの応用を試みて初めて、高精度の立体構造予測に先立ち、正しいドメイン予測とディスオーダー領域の除去が不可欠であることを実感しました。試行錯誤の末、高等生物に特化した以下のような手法を開発しました。まず、予測対象配列にBLAST、PSI-BLAST、フォールド認識法(Smith-Waterman profile-profile alignment)による鑄型探索を順次適用して確実な構造ドメインから最適に同定し、残った予測困難な領域をディスオーダー予測を利用して緩やかに細分化した上でFORTEで鑄型探索をします。もう一つの特徴は、可能ならばホロ体・アポ体の両方を同時に予測し、リガンドの結合による立体構造変化をアニメーション表示することです。

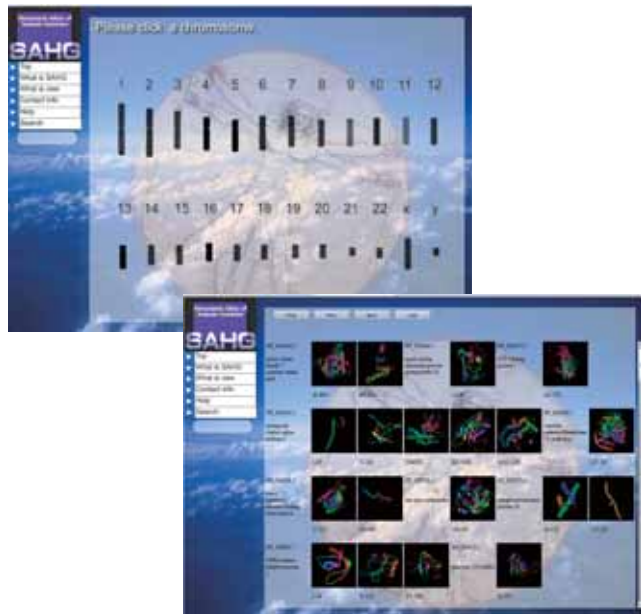
さらに、この立体構造予測結果と他のメンバーの研究結果とを統合したSAHGデータベースを作成しました。

CBRCで開発されたディスオーダー領域予測プログラム;POODLEの結果や酵素触媒機構データベース;EzCatDBをはじめ、タンパク質複合体予測(名大)、リガンド結合部位予測・リガンド予測(東北大)、リガンド結合に伴う構造変化予測(横浜市立大)等によるゲノムアノテーション結果が、外部データベースへのリンクと共に盛り込まれています。SAHGはマウスクリック主体のインターフェースを有し、立体構造上で構造変化、基質結合と反応がどう実現しているのかを視覚的に理解できます。現状で極めて高水準の予測結果を集結させて、一目瞭然の形で提供しています。プロジェクトでは、既に一部の予測の検証実験が行われています。

今後、核酸-タンパク質間相互作用や時間軸を意識した相互作用情報をSAHGへ取り込み、細胞機能のモデル作成にも寄与できる、より高次で複雑な機能アノテーション提供の基盤としたいと考えています。また、ゲノム中のタンパク質の立体構造のアニメーションがずらりと並ぶリスト画面をたまたま見た子ども達が、自分の体の中にこんなに多様で美しい物質がある!と感じてくれないか、とひそかに期待しています。

SAHGは<http://bird.cbrc.jp/sahg/>で一部公開中ですが、2009年度内に全染色体について公開していきます。

SAHGデータベースのトップ画面。閲覧したい染色体をクリック!もしくは検索を。



SAHGの検索結果リスト。アニメーションで構造変化を表示しています。閲覧したい構造をクリックすると、詳細なアノテーション画面へ。