

# ホヤプロテイン統合データベースの構築

筑波大学下田臨海実験センター  
稲葉 一男

## Development of An Integrated Protein Database of The Ascidian *Ciona intestinalis*

Kazuo Inaba

Shimoda Marine Research Center, University of Tsukuba

The marine tunicate species *Ciona intestinalis* has been an attractive research organism for biologists for more than a century. Because of its mosaic development, small numbers of cells in the tadpole larvae, and its phylogenetic position, *C. intestinalis* is a suitable model for understanding chordate development and vertebrate evolution. Here we developed an integrated protein database, named CIPRO. The database includes original large-scale transcriptomic and proteomic data, as well as information on its development and morphology, which are linked to gene and protein information. The CIPRO database is aimed at collecting those published data, as well as providing unique information from unpublished experimental data, such as 3D expression profiling, 2D-PAGE and MS-based large-scale proteomic analyses. For medical and evolutionary research, homologs in humans and major model organisms are intentionally included. CIPRO is now an international standard of tunicate protein database in the tunicate research community.

### 1. はじめに

海産生物はその多様性から生物学におけるさまざまな分野で重要な役割を果たしてきた。その中でホヤは、無脊椎動物でありながらヒトと同じ脊索動物の一員であることや、遺伝子構成がコンパクトであり遺伝子の重複が少ないことから、脊椎動物の進化の研究において不可欠な研究対象とされてきた。さらに、初期発生における細胞の運命が早くから決定され、その系譜も詳細に記載されていることや、オタマジャクシ型の幼生の細胞数が約 2,600 と少ないにもかかわらず、脊索、背側の中枢神経系、筋肉など、脊椎動物の体制を整えていることから、脊椎動物にもっとも近い無脊椎動物として、発生学や進化発生学の研究において大きく貢献してきた (図 1)。

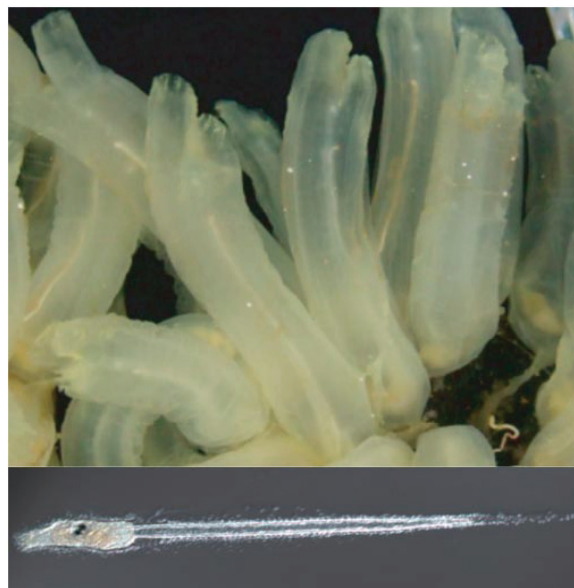


図1 カタユウレイボヤ(成体)とオタマジャクシ幼生(下)

ホヤの1種で日本を含め世界中の海に生息するカタユレイボヤ (学名 *Ciona intestinalis*) のゲノム配列が2002年に決定され、ホヤの研究は大きく前進した[1]。ゲノムサイズは160 Mb (遺伝子数は約16,000) とヒトの1/20のサイズである。さらに、*in situ* ハイブリダイゼーションやマイクロアレイによる発生過程や成体の各器官の遺伝子発現情報が整備された。これらの情報をもとにしたゲノム/cDNA データベースは多くの研究者のプラットフォームとして機能してきた[2]。また、遺伝子から予測されるタンパク質の情報やプロテオミクスを中心としたタンパク質同定データ[3]、発生にともなう細胞分化情報やタンパク質局在情報[4]などが次々と得られてきた。本研究開発では、これらタンパク質の情報やデータを基盤とした既存のゲノムデータベースの整備や細胞情報の収集、さらには遺伝子発現、発生や細胞機能へのリンク情報も盛り込み、国内外のホヤ研究者、ならびに他の生物を扱っているライフサイエンス研究者に汎用されるホヤ統合プロテインデータベース *Ciona intestinalis* Protein Database (CIPRO、サイプロ) を構築することを目的として行った。

## 2. 研究開発の成果

### 2.1 データベースを構築するデータセット

本開発にあたり、まずは遺伝子モデルの見直しを行った。2002年にゲノム配列が決定されて以来、そこから予想される遺伝子モデルに関しては数回の改良がなされてきた。しかしながら遺伝子予測モデルの問題やアセンブリが断片化している可能性があった。cDNA 全長配列情報や追加の EST データが蓄積してきている状況を踏まえ、開発チームではこれらの情報をもとに新たな遺伝子モデル (KH-model) を作成した[5]。このモデルから予測されるタンパク質情報に基づき、CIPRO で用いるデータセットを再整備した。

タンパク質のアミノ酸配列から得られる情報として、分子量、等電点、相同性検索、近縁種である *C. savignyi* におけるオルソログ、機能モチーフ、分子系統樹、二次構造予測、立体構造予測、細胞内局在、遺伝子発現データ、文献データ、これまで利用されてきた遺伝子モデルとの対応表などをリンクした。さらに、ヒトやショウジョウバエなどのモデル生物の遺伝子/タンパク質のホモログ/オルソログ情報や疾患遺伝子情報 (MIM) を引き出すためのウィンドウを充実させることにより、ホヤが有するユニークな研究成果と情報を、ホヤのみでなくライフサイエンスに携わる広範な研究者に提供している。将来的にホヤの有する実験系のメリットを用いて、特に機能未知遺伝子/タンパク質の機能解明や、疾患遺伝子の機構解明に貢献すると期待している。

CIPRO では、公開されていないものも含め、タンパク質に関する多くの情報の収集と有機的なリンクを行った。特に、発生過程ならびに成体の二次元電気泳動/MS 解析および MS/MS 解析によるプロテオームデータ、および受精卵から成体に至る共焦点レーザー顕微鏡像とタンパク質細胞内局在情報に関しては、研究者にビジュアルに有用な情報を提供する上でユニークである。プロテオームデータに関しては、実際に遺伝子モデルから予測されるタンパク質が存在するか否か、ならびに各組織における存在量と組織間の発現量の比較に関して情報が得られる。一方、共焦点顕微鏡像に関しては立体像やセクション像により、ホヤ胚が発生していく様子や、細胞の空間的配置を視覚的に理解する上で役に立つ。

さらに細胞内局在情報は、細胞内におけるタンパク質機能と、胚全体の細胞配置を統合的に捉える上で有用である。

## 2.2 実験系研究者とバイオインフォマティクス系研究者の協調を目指して

本研究開発のチームは北海道大学のシステム開発のメンバーを除いてすべて実験系研究者である。CIPRO 作成の基本方針としては、実験系およびバイオインフォマティクス系の研究を協調的に促進することを目指した。そのために、「実験研究者が有用な情報を即座に取得できる」ことを第一の優先事項とした。

まず、実験研究者が検索したい項目を想定し、検索画面としてはキーワード、遺伝子 ID、タンパク質 ID、タンパク質の分子量/等電点、ホヤ組織名、データ種類、H-inv アノテーション番号、ホモログ、アクセッション番号、疾患遺伝子 (MIM) など、幅広いカテゴリーでホヤタンパク質を検索できるように設定してある (図 2)。実験データとしては、二次元電気泳動像、細胞内局在、および発現情報を提供している。各スポットの情報や発現時期に関する詳細についてはリンクにより取得できる。

実験研究者は、自らが対象とするタンパク質、新規に直面したタンパク質、興味のある組織に発現しているタンパク質に関する必要最低限の有用情報を短時間に取得し、サーベイしたいと考えている。本データベースではこれらの情報のエッセ

図2 タンパク質情報の表示画面。(1)タンパク質名と簡易情報、(2)アミノ酸配列、分子量、等電点、(3)BLAST、(4)ホモログ及びモチーフ情報、(5)疾患遺伝子、ホモログ情報その他、(6)他のホヤデータベースのID情報、(7)各種実験データ、(8)立体情報その他のグラフィック表示、(9)ユーザアノテーション、(10)ユーザコメント

ンスのみを1ページで表示できるように工夫した。実験研究者の目にとまった項目からは、リンクによりさらに詳細な情報を取得できるようになっている。

### 2.3 その他の提供情報

我が国において、ホヤは食用としての知名度はわりと高いが、研究対象としての知名度は残念ながらそれほど高くない。しかしながら、ホヤを用いた研究の有用性は、最近特にライフサイエンス全体に浸透しつつある。そこで、データベースでは、ホヤという生物の紹介や体の構造なども紹介している。さらに、ホヤ関係者のみならず、広く農学、医学関連の研究者にも利用しやすいように、ユーザマニュアルを日本語と英語で提供している。

## 3. まとめ

CIPRO はすでに公開され、ホヤ研究者をはじめ多くの方々に利用して頂いている (URL: <http://cipro.ibio.jp/2.5>)。データベースの概要は論文としても公開した[6]。昨年(2019)の11月にホヤおよび関連生物の研究者がフランスのニースに一同に集まり、データベースに関する今後の国際協調について議論しあった。その中で、CIPRO はタンパク質と細胞、発生情報を提供するデータベースとして多くの研究者に利用され、今後も国際標準として基軸となるデータベースであると合意された。また、タンパク質解析ツールやリソース、局在に関する過去の文献データなど、更なる情報発信のプラットフォームとしての要望もあった。さらに、研究材料として有用なホヤが他にも複数種あり、多くの研究成果を生み出してきたが、特に日本産マボヤ(食用)、ヨーロッパ産ホヤ、オタマボヤ、群体ボヤなどに関する種々の情報は、カタユウレイボヤの情報を基盤として、データベースとして統合すべきである。既存の他のデータベースとの有機的な統合も検討すべき課題である。国際的な視点を重視しつつ、ホヤ研究者コミュニティのニーズに合ったデータベースの書式整備とデータ提供の体制を今後早急に整えていかななくてはならない。

## 4. 研究開発実施体制

代表研究者 稲葉 一男 (筑波大学下田臨海実験センター)

研究開発題目

- (1) ホヤプロテインデータベースの開発  
グループリーダー 稲葉 一男 (筑波大学下田臨海実験センター)
- (2) ホヤプロテインデータベースのシステムの開発と構築  
グループリーダー 遠藤 俊徳 (北海道大学情報科学研究科)
- (3) ホヤゲノム・プロテインデータベースの統合  
グループリーダー 佐藤 ゆたか (京都大学理学研究科)
- (4) ホヤタンパク質の情報収集と検索ツールの開発  
グループリーダー 小笠原 道生 (千葉大学融合科学研究科)
- (5) ホヤタンパク質の情報収集と細胞機能情報との統合  
グループリーダー 堀田 耕司 (慶應義塾大学理工学部)

- (6) ホヤ発生関連タンパク質の情報解析系の開発  
グループリーダー 谷口 寿章 (徳島大学疾患酵素学研究センター)
- (7) ホヤタンパク質の局在情報の収集と解析  
グループリーダー 高橋 弘樹 (基礎生物学研究所)

## 5. 参考文献

- [1] Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*. 2002.298:215721-215767.
- [2] Satou Y, Takatori N, Fujiwara S, Nishikata T, Saiga H, Kusakabe T, Shin-i T, Kohara Y, Satoh N. *Ciona intestinalis* cDNA projects: expressed sequence tag analyses and gene expression profiles during embryogenesis. *Gene*. 2002. 287:83-96.
- [3] Nomura M, Nakajima A, Inaba K. Proteomic profiles of embryonic development in the ascidian *Ciona intestinalis*. *Dev Biol*. 2009. 325:468-481.
- [4] Hotta K, Mitsuhara K, Takahashi H, Inaba K, Oka K, Gojobori T, Ikeo K. A web-based interactive developmental table for the ascidian *Ciona intestinalis*, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Dev Dyn*. 2007. 236:1790-1805.
- [5] Satou Y, Mineta K, Ogasawara M, Sasakura Y, Shoguchi E, Ueno K, Yamada L, Matsumoto J, Wasserscheid J, Dewar K, Wiley GB, Macmil SL, Roe BA, Zeller RW, Hastings KE, Lemaire P, Lindquist E, Endo T, Hotta K, Inaba K. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. *Genome Biol*. 2008 9:R152.
- [6] Endo T, Ueno K, Yonezawa K, Mineta K, Hotta K, Satou Y, Yamada L, Ogasawara M, Takahashi H, Nakajima A, Nakachi M, Nomura M, Yaguchi J, Sasakura Y, Yamasaki C, Sera M, Yoshizawa AC, Imanishi T, Taniguchi H, Inaba K. CIPRO 2.5: *Ciona intestinalis* protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. *Nucleic Acids Res*. 2011. 39(Database issue):D807-814.