

Webで実践 生物学 情報リテラシー

BIOINFORMATICS
LITERACY

広川貴次／美宅成樹 ― 著

序

ほぼ 10 年前にヒトゲノムの解読が完了しましたが、その後の生物ゲノム解析の勢いには目ざましいものがあります。次世代シーケンサと呼ばれる高速 DNA 塩基配列読みとり装置が開発されたことで、個々人のパーソナルゲノムも現実的な時間と現実的な費用で解析できるようになり、医学応用にもう一歩で手が届きそうです。10 年前には、このようなスピードで技術的な発展が起こるとは考えられませんでした。前著『できるバイオインフォマティクス』から 10 年が経ち、新たに『Web で実践 生物学情報リテラシー』と題して、この時代に即応した本書を出版することにしました。生物のゲノム情報の基礎は変わらないので、1 章（生物情報データベース）、2 章（配列解析）、3 章（立体構造予測）および 4 章（文献データベース）は前著の内容をある程度継承しています。これに対して、医薬応用に近い活性部位の話題を 5、6 章で、時々刻々増大していくゲノムデータについての話題を 7 章で紹介しました。まったく新しい話題として 8 章（相互作用ネットワーク、システム生物学系）では、大量の遺伝情報をお互いに関連付け、生物体内のシステムを解きほぐそうという試みを紹介しました。さらに、9 章では、創薬研究を支援するインシリコスクリーニング関連ツールを紹介しました。

次世代シーケンサは、ゲノム情報から生物学的意味を解読するプロセスのハードウェアに当たります。そして、それに対応するソフトウェアが完成して始めて、社会的にも意味のあるものとなります。しかし、これらのソフトウェア部分は、まだ次世代シーケンサや構造解析装置などが生産する生物情報が持つ潜在的な生物学的意味の一部を抽出しているにすぎません。ゲノム情報を本当に理解するためには、さらに高度なソフトウェアを開発していかなばなりません。

生物も自然の中にあり、すべて分子でできています。確かに生物らしい分子（生体高分子）は大きく複雑です。しかし、生物が地球上で大きく進化した理由は、生物らしい分子が持つ一見複雑な配列の裏には、非常に単純な物理的側面があるという事実があるからです。これまで見かけの複雑さに隠れて、配列の物理的単純さは注目されてきませんでした。しかし、近年ハードウェアが生産するゲノムデータはあまりにも大量で、ゲノム情報解析のパラダイムシフトへの圧力は年々高まっています。この方向での残された課題と新しい情報リテラシーのあり方を、10 章で紹介します。

さて、初めて生物科学のデータベースを使う人にとって、ある 1 つの仕事の流れを何のナビゲーションもなく、最短の道筋で進むことは不可能です。ある程度慣れてきた人にとっても、別の流れの解析を行おうとすると同じことが起こります。それだけ生物情報は階層的で奥が深いのです。本書は生物科学の研究を志しているが、データベースの利用に関してはまったくの初心者というような人たちに対して、できるだけ親切に解説を加えることを目的として企画されました。本書の最も大事な部分はインターネットのホームページの画面で構成されています。公共のデータベースはすべてインターネットで公開され、そこでは付属のツールによっ

て、ある程度情報抽出ができます。しかし、データの情報が膨大で、ツールが高機能であればあるほど情報抽出のためのコマンドが多く、その条件設定も複雑になります。データがどのような形をしていて、ツールのどのコマンドを使えば目的の情報にたどりつけるかを知るには、実際にホームページの画面を見て試してみることが最も簡便で、確実です。本書で生物情報の典型的な仕事で出てくるデータベースの画面の流れに沿ってたどることによって、自分のやりたい仕事が自然に覚えていただけることでしょう。そういう意味で本書は講習会に適したテキストになっています。

本書ではタイトルに「情報リテラシー」という言葉を用いました。情報リテラシーは一般には「情報を使いこなす力」というような意味ですが、生物科学の分野では二重の意味があります。生物自体が一種の情報処理機械ですので、私たち人類が誕生する以前から、生物は自分を構成する情報（ゲノム配列情報を中心とした多様な情報）を完全に使いこなし、生き、世代を重ねてきました。つまり、生物は内部に完全な情報リテラシーのシステムを持っているのです。これに対して、20世紀後半以降、人間は生物についての多様な情報を解読・操作する技術を開発してきました。そのプロセスは、人間が生物を理解するための情報リテラシーを向上してきた歴史と言うことができます。しかし、現状を客観的にみると、後者の意味での情報リテラシーは、大量の生物情報の解析にもかかわらずまだまだ未熟です。そこで本書では、生物理解への情報リテラシーの最前線に、できるだけ多くの人々が触れられるようにし、最終的な生物理解への道を太くしたいと考えたのです。

生物情報の講習会を行ってみると、より広い範囲の人たちが受講してきて、実は生物のデータベースから利益を得ることができるのは、研究をすでに行っている人たちばかりではないということに気づきます。大学では、生物系の講義がたくさんあります。必ずしも生物科学を志す人たちだけではありませんが、レポートを書くため、あるタンパク質に関係する文献を検索することが必要になるかもしれません。また、医学部・看護学部などの学生は、ある遺伝子病について発表するために関連の情報を短時間で収集しなければならないということがあるでしょう。企業の人で、それまで生物学の知識はないが、バイオテクノロジーの部門に回ったので、すぐに遺伝情報を使えるようになりたいと講習会を受けに来ることもあります。そうした人たちの自習用としてもこの本は役立つに違いありません。

本来、生物分野のデータベースは非常に専門化した情報の集まりですから、それを利用するには高度な知識が必要だと思われがちです。しかし、少し親切的なインストラクターがいれば、初心者でも使ってみることも自体は意外と簡単です。生物科学の普及には、専門家にとっては親切すぎるかもしれない本書のような本が、実は最も必要とされているのだと思います。「これまでなんとなく敷居が高くて、敬遠していた生物情報のデータベースとソフトウェアツールをとにかく利用してみようという気になった」という人が少しでも現れたら、この本の意図が達成されたということになります。より多くの人たちに使っていただけることを期待しています。

本書は、1章から9章を広川がまとめ、未来に向けた解説の10章と全体の監修を美宅が行いました。ご協力いただいたすべての関係者に深く感謝いたします。

2013年7月 美宅成樹

CONTENTS

CHAPTER

1

分子生物学データベースを利用してみよう

実習の前に

- 2 分子生物学データベースの基礎知識
- 3 分子生物学データベース
- 3 パスウェイデータベース

実習

- 4 **実習 1** タンパク質配列データ取得 (Genome Net)
- 7 **実習 2** タンパク質立体構造データ検索 (PDB)
- 12 **実習 3** 遺伝子情報の検索 (KEGG GENES)
- 15 **実習 4** ゲノム種別配列の取得 (NCBI FTP)
- 17 文献および関連サイト

CHAPTER

2

配列を比較してみよう

実習の前に

- 20 配列を比較する意味
- 20 配列を比較する方法
- 21 生物学的意味を持つ文字 (アミノ酸残基) 置換
- 22 相同性配列検索
- 23 多重配列アラインメント
- 25 系統樹解析

実習

- 26 **実習 1** 相同性配列検索 (BLAST)
- 31 **実習 2** 多重配列アラインメント (Clustal X, Phylodendron)
- 33 **実習 3** プライマー設計 (Primer3)
- 36 文献および関連サイト

タンパク質の立体構造を予測してみよう

実習の前に

- 40 タンパク質立体構造予測
- 42 予測構造の評価
- 42 タンパク質-タンパク質複合体予測

実習

- 43 **実習 1** ホモロジーモデリング法 (SWISS-MODEL)
- 47 **実習 2** タンパク質-タンパク質ドッキングによるタンパク質複合体予測 (ClusPro)
- 50 **実習 3** タンパク質折りたたみ体験 (Foldit)
- 53 文献および関連サイト

文献データベースを活用してみよう

実習の前に

- 56 Entrez：分子生物学データ統合検索システム

実習

- 57 **実習 1** 指定した論文の検索 (PubMed)
- 60 **実習 2** Advanced 機能による検索 (PubMed)
- 63 **実習 3** MeSH 項を利用した広域論文の検索 (PubMed)
- 66 **実習 4** My NCBI による個人設定 (Entrez)
- 70 文献および関連サイト

配列情報からタンパク質の機能を予測してみよう

実習の前に

- 72 遺伝子オントロジー
- 73 モチーフデータベース
- 74 モチーフの表現方法
- 75 膜貫通領域の推定
- 76 シグナルペプチドの予測
- 77 細胞内局在予測

実習

- 78 **実習 1** モチーフ検索 (InterPro)
- 80 **実習 2** シグナルペプチド予測 (SignalP)

- 81 実習 3 細胞内局在予測 (PSORT)
- 83 実習 4 膜タンパク質予測 (SOSUI)
- 87 文献および関連サイト

CHAPTER

6

立体構造情報からタンパク質の機能を予測してみよう

実習の前に

- 90 タンパク質の立体構造情報と機能
- 91 立体構造モチーフ
- 93 タンパク質の折りたたみ様式

実習

- 94 実習 1 活性ポケット候補部位探索 (CASTp)
- 97 実習 2 リガンド結合および活性部位予測 (PINTS)
- 100 実習 3 静電ポテンシャル解析 (eF-surf)
- 103 文献および関連サイト

CHAPTER

7

ゲノムデータを閲覧してみよう

実習の前に

- 106 ゲノム地図の作成と塩基配列の決定
- 107 ゲノム配列を調べる
- 108 ゲノムデータの閲覧ツール：ゲノムブラウザ
- 110 これからのゲノム解析

実習

- 110 実習 1 ゲノムプロジェクト検索 (GOLD)
- 112 実習 2 遺伝子検索 (Map Viewer)
- 120 文献および関連サイト

CHAPTER

8

生物情報をネットワークで眺めてみよう

実習の前に

- 122 生体ネットワークの種類と表現方法
- 123 ネットワークの解析方法：次数分布とスケールフリー性
- 124 ネットワークの解析方法：クラスター係数

実習

- 124 実習 1 代謝データベース検索 (KEGG)
- 128 実習 2 化合物、遺伝子情報検索 (DrugBank)

- 134 **実習 3** ネットワークの可視化と編集 (Cytoscape)
- 140 文献および関連サイト

CHAPTER

9

創薬研究に情報科学を活用してみよう

実習の前に

- 142 LBDD と SBDD
- 143 LBDD によるインシリコスクリーニング
- 144 SBDD によるインシリコスクリーニング

実習

- 145 **実習 1** 化合物類似性検索 (PubChem)
- 149 **実習 2** タンパク質と化合物のドッキング計算 (SwissDock)
- 154 文献および関連サイト

CHAPTER

10

生物情報リテラシーの残された課題

- 156 生物情報に対する研究の歩み
- 157 バイオインフォマティクスの残された課題
- 162 問題解決への仮説
- 166 仮説を裏付ける若干の証拠
- 171 おわりに
- 172 文献および関連サイト

- 173 付録：バイオインフォマティクス関連 Web サイト
- 187 索引

CHAPTER

1

BIOINFORMATICS

分子生物学データベースを 利用してみよう

概要

現在では、多くの分子生物学データにインターネットを通じて自由にアクセスでき、必要なデータを取得することができます。本章では、核酸配列や、タンパク質配列・立体構造情報を中心にデータベースの内容やデータベース間の関連付けを実習します。



実習の前に

▼ 分子生物学データベースの基礎知識

分子生物学の概念を理解する上で重要な遺伝情報は、DNA の複製、転写、RNA による翻訳（RNA から DNA への逆転写経路もある）、タンパク質合成と伝達されます。そして、この基本となる過程は、セントラルドグマと提唱されています。DNA 配列は、生物が持つ最も基本的な情報で、A（アデニン）、T（チミン）、G（グアニン）、C（シトシン）と呼ばれる構成要素（塩基）が一次元的な配列として連なったものです。各塩基は A-T、G-C の相補対をつくり、二重らせん構造を形成します。二重らせん構造をとることが、DNA から DNA への複製や DNA から RNA への転写など生合成過程で重要な役割を果たしています。遺伝子は、この DNA 配列の中で機能単位となっている領域です。遺伝子領域は、実際にはタンパク質のアミノ酸配列を指定しています。遺伝子領域から転写されたメッセンジャー RNA（mRNA）の情報はトランスファー RNA（tRNA）とリボソームの働きによりアミノ酸配列（20 種類）へと変換されますが、これがタンパク質です。DNA 配列は二重らせん構造を形成していますが、タンパク質を構成するアミノ酸配列は特定の立体構造を持ちます。ただし、二重らせん構造よりも複雑です（図 1-1）。さらに、DNA 配列、遺伝子（タンパク質）といった情報を、ある特定の生物種の枠組みで包括的に塩基配列・遺伝子・タンパク質情報を扱う際には、ゲノ

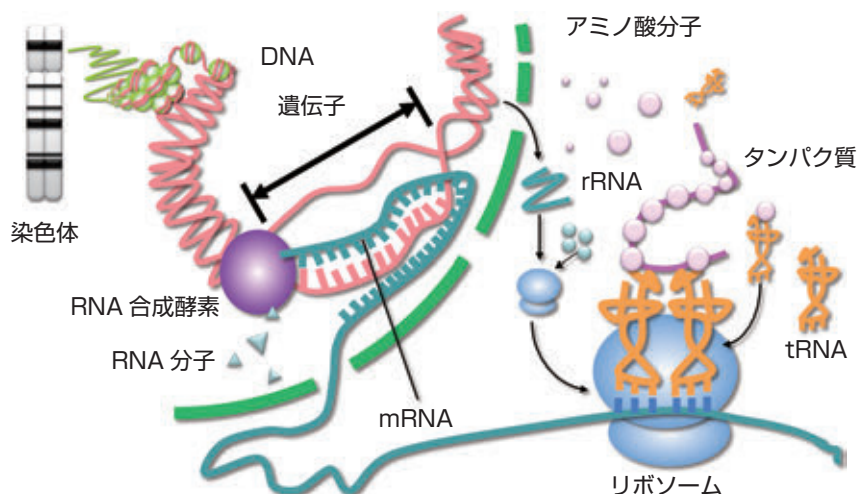


図 1-1 DNA 情報からタンパク質が合成される過程

ムという表現を用います。一般には、ヒトゲノム、マウスゲノム、イネゲノムのように「生物種名+ゲノム」という具合に使われます。

▼ 分子生物学データベース

近年、生物情報データベースは多様化してきていますが、その中でも① DNA/RNA 塩基配列データベース、②タンパク質配列データベース、③タンパク質立体構造データベース、④文献データベースは、分子生物学研究において最も基本であり、利用頻度の高い代表的なデータベースです。実際にデータベースを利用する際には、タンパク質配列データベース 1 つをとっても、米国、スイス、日本と各国の研究機関や大学で独自に構築されてきたデータベースが共存しています。これは歴史的な経緯によるもので、ほとんどの場合、それぞれのデータベースに登録されている配列は重複しています。ただし、データベースのフォーマットや提供されている情報には、それぞれに特色があります。

また、これらのデータベースは、注目する遺伝子（タンパク質）を軸にお互いが関連付けられています。これは、分子生物学データベースの特徴でもあります。例えば、タンパク質配列データベースの検索で見つけたタンパク質は、アミノ酸配列の設計図である DNA 塩基配列が存在しており、同じタンパク質が DNA/RNA 塩基配列データベースにも登録されています（ただし、情報は塩基配列）。また、タンパク質配列は一般にユニークな立体構造を形成します。タンパク質立体構造データベースでは、この立体構造を表現する原子座標を中心とした内容のデータ形式で登録されています。文献データベースでは、そのタンパク質の発見に由来する論文が関連付けられています。

基本となるデータベース以外にも、生物を理解する上で非常に重要な情報データベースが多く存在しますので（表 1-1）、研究の目的に応じて幅広く活用することをお勧めします。

▼ パスウェイデータベース

遺伝子やタンパク質のデータベース登録数の急増と、世界中で進められている各種生物のゲノム解読により、ある生物が生命活動を維持するために必要な「役者（遺伝子やタンパク質）」を把握できるようになってきました。これらのデータベース情報を基盤として、最近のゲノム解析では、遺伝子・タンパク質間での相互作用に注目したパスウェイ解析が行われ、反応メカニズムの解明や生物種ごとの比較に用いられています。パスウェイ自体は、生化学の分野で古くから行われている研究ですから、代表的な代謝についてはいくつかのパスウェイマップが既に構築されています [2]。そこで、これまで蓄積されてきたパスウェイマップをゲノム解析の結果にあてはめ、より高度な予測を行うことがパスウェイ解析の 1 つのアプローチとなっています。生物種ごとのゲノムが明らかになってきていますので、種の違う機能的・進化的に等価な遺伝子の関係（オーソログ）をパスウェイ解析上で比較することで配列解析のみでは困難な遺伝子の機能を予測することもできるのです [3]。

表 1-1 代表的な分子生物学関連データベース [1]

データの種類	データベース名	管理・運営
DNA/RNA 塩基配列	GenBank	National Center of Biotechnology Information (米国)
DNA/RNA 塩基配列	EMBL	European Bioinformatics Institute (英国)
DNA/RNA 塩基配列	DDBJ	国立遺伝学研究所 (日本)
タンパク質アミノ酸配列	PIR	Georgetown University (米国)
タンパク質アミノ酸配列	UniProt	European Bioinformatics Institute (英国), the SIB Swiss Institute of Bioinformatics (スイス), the Protein Information Resource (米国) の共同体制
タンパク質アミノ酸配列	PRF	蛋白質研究奨励会 (日本)
タンパク質・生体高分子立体構造	PDB	Research Collaboratory for Structural Bioinformatics (米国)
タンパク質立体構造分類	SCOP	Medical Research Council (英国)
タンパク質立体構造分類	CATH	University of College London (英国)
タンパク質配列モチーフ	PROSITE	University of Geneva (スイス)
アミノ酸指標	AAindex	京都大学化学研究所 (日本)
生命システム情報統合	KEGG	京都大学化学研究所 (日本)
ヒト遺伝子地図	GDB	John Hopkins University (米国)
ヒト遺伝病	OMIM	John Hopkins University (米国)
リンク情報	LinkDB	京都大学化学研究所 (日本)



実習

実習では、ヒト上皮成長因子受容体遺伝子を検索の対象として、代表的な分子生物学データベースを検索し、タンパク質立体構造表示操作等を行います。また、ゲノムデータ全体の取得方法についても実習します。

▼ 実習①：タンパク質配列データ取得

基本的に遺伝子は、核酸配列→タンパク質配列→タンパク質立体構造と階層的に情報を持っていますが、対象とする遺伝子によっては、核酸配列のみが明らかになっているもの、タンパク質配列まで明らかになっているが立体構造は未だ解かれていないものなど、現時点では情報量がさまざまです。難しさ（言い換えれば、情報量が不足している）という点では、タンパク質立体構造を解明することが最も困難とされています。それに対して、タンパク質配列情報はよく知られていますから、タンパク質配列情報から調べてみて、核酸配列や立体構造情報があるかどうかを確認するのがポピュラーな手順です。データの検索には、日本で構築された分子生物学統合データベース GenomeNet を利用します。

- ① Web ブラウザを利用して、以下の URL より GenomeNet にアクセスします。代表的なデータベースのデータ登録数の状況も“DB growth curve”より参照できます (図 1-2)。

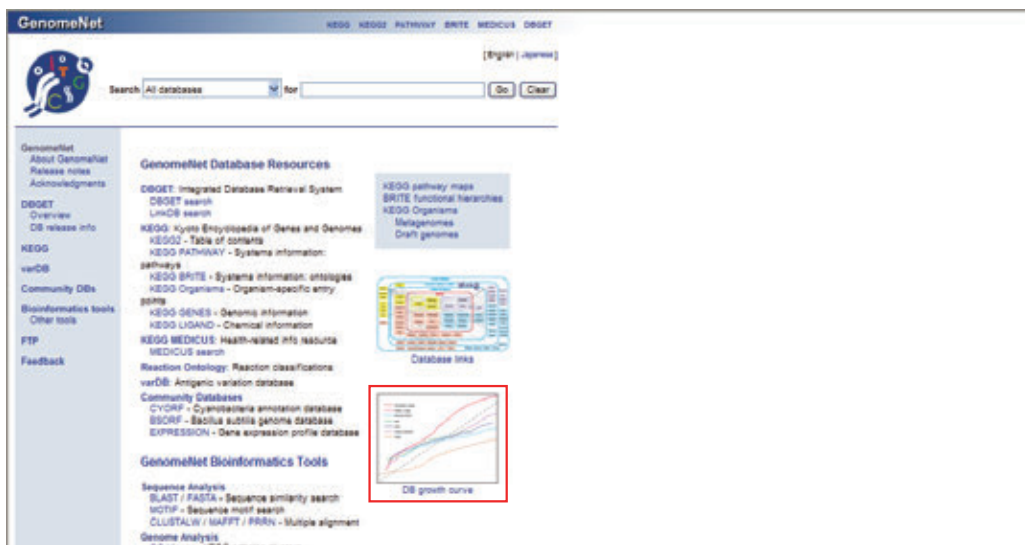


図 1-2 GenomeNet のトップページ

- ② GenomeNet システムを使って配列情報を取得します。例題としてヒト上皮成長因子受容体遺伝子を用います。最初にタンパク質アミノ酸配列データベース UniProt で検索します。GenomeNet のホームページ上部にある Search リストからタンパク質アミノ酸配列データベース“UniProt”をクリックします。上皮成長因子受容体遺伝子のキーワードである「EGFR」と生物種を特定するための「human」を検索フォームに入力し、**Go** をクリックします (図 1-3)。



図 1-3 検索データベースおよびキーワードの入力

- 3 通常、限定したキーワードを用いない限り、検索結果は、いくつかの候補を出力します。例えば、今回の場合は VEGFR（血管内皮増殖因子受容体）が含まれていますので、タイトルを確認して目的の遺伝子を示すエントリーを選択することが重要です（図 1-4）。



図 1-4 キーワード検索結果

- ④ データベース内容を確認します。ここでは、検索結果より“EGFR_HUMAN [UniProt]”を選択します。データベースには、登録日や配列、アノテーション更新日、酵素番号や関連論文などが記載されています。

アミノ酸配列情報は、ページの最後にあります。先頭にある“SQ”をクリックすると 1 行目に ‘>’ で始まるタイトル行と、2 行目以降にアミノ酸配列の一文字表記が連なったテキストが表示されます。この書式を FASTA 形式と言います。いろいろな解析の基本となる重要な入力形式です（図 1-5）。

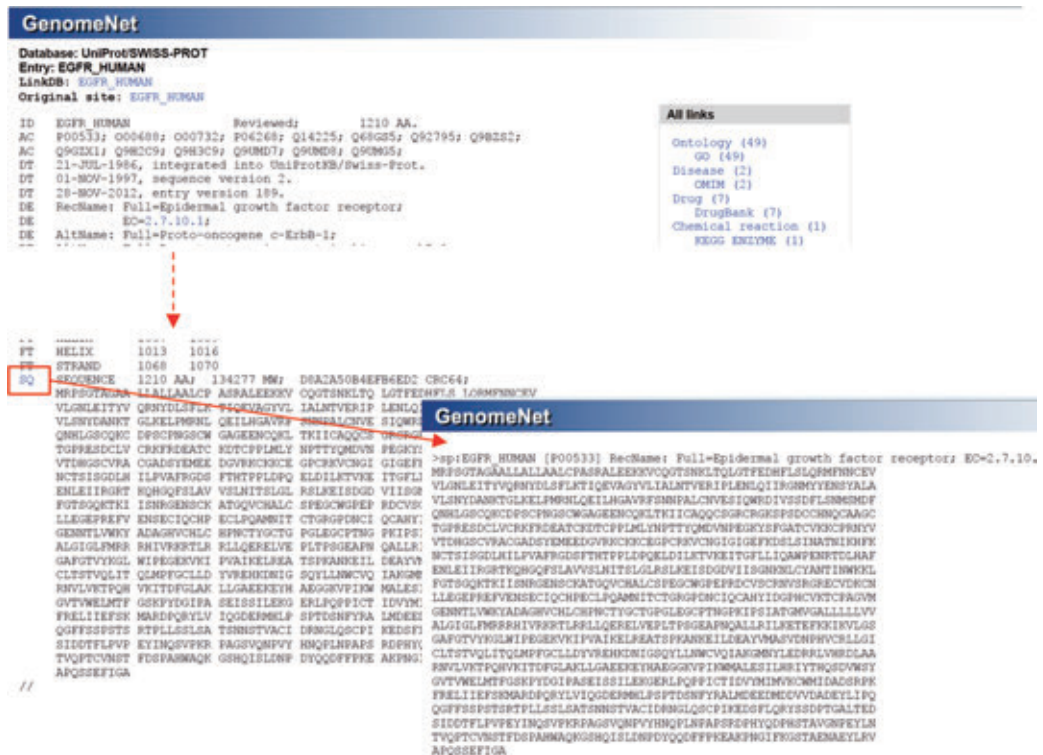
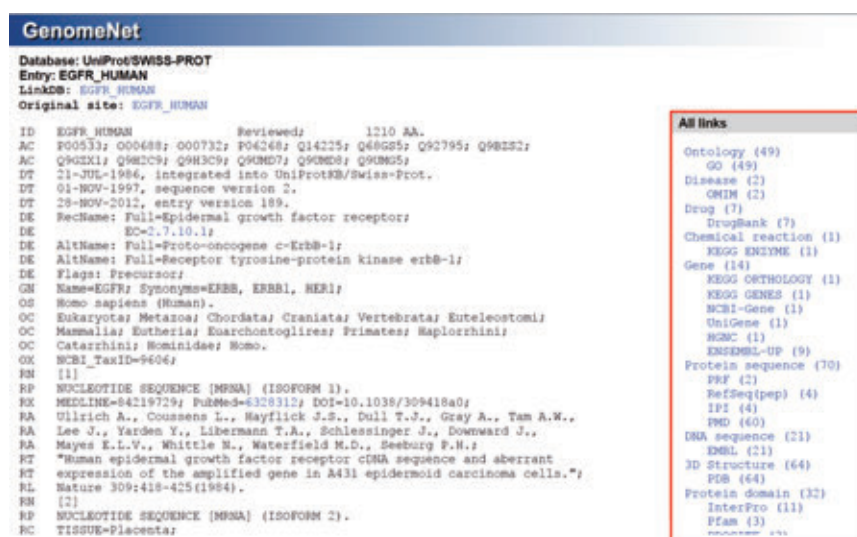


図 1-5 ヒト EGFR の UniProt エントリーと配列情報の FASTA 形式（右下）

- ⑤ GenomeNet では、データベースの DR 行に記載されている他のデータベース上の関連エントリーを “All links” サブメニューを通じて調べることができます。塩基配列情報は、Gene や EMBL、立体構造情報は PDB をクリックします (図 1-6)。

関連エントリーは、単一でない場合があります。塩基配列の場合、mRNA としてのエントリーもあれば、任意のエクソン領域のエントリーもあります。立体構造情報でも、実験手法の違いやリガンド結合状態、変異体などによって同一遺伝子でも複数のエントリーが存在します。各エントリーの記述を確認して利用目的に合ったエントリーを選択することが重要です。



GenomeNet

Database: UniProt/SWISS-PROT
Entry: EGFR_HUMAN
LinkDB: EGFR_HUMAN
Original site: EGFR_HUMAN

ID: EGFR_HUMAN Reviewed: 1210 AA.
AC: P04268; O00688; O00732; P04268; Q14215; Q46GE5; Q92795; Q981S2;
AC: Q981S2; Q981S2; Q981S2; Q981S2; Q981S2; Q981S2;
DT: 21-JUL-1986, integrated into UniProt/SWISS-Prot.
DT: 01-NOV-1997, sequence version 2.
DT: 28-NOV-2012, entry version 189.
DE: RecName: Full=Epidermal growth factor receptor;
DE: EC=2.7.10.1;
DE: AltName: Full=Proto-oncogene c-ErbB-1;
DE: AltName: Full=Receptor tyrosine-protein kinase erbB-1;
DE: Flags: Precursor;
GN: Name=EGFR; Synonyms=ERBB, ERBB1, HER1;
OS: Homo sapiens (Human).
OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC: Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhina;
OC: Catarrhini; Hominidae; Homo.
OX: NCBI_TaxID=9606;
RN: [1]
RF: NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 1).
RX: MEDLINE=6421972; PubMed=328312; DOI=10.1038/309418a0;
RA: Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA: Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
RA: Mayes E.L.V., Whittle M., Waterfield M.D., Seeburg P.H.;
RT: "Human epidermal growth factor receptor cDNA sequence and aberrant
expression of the amplified gene in A431 epidermoid carcinoma cells."
RL: Nature 309:418-425 (1984).
RN: [2]
RF: NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 2).
RC: TISSUE=Placental

All links

- Ontology (49)
 - GO (49)
- Disease (2)
 - OMIM (2)
- Drug (7)
 - DrugBank (7)
- Chemical reaction (1)
 - KEGG ENZYME (1)
- Gene (14)
 - KEGG ORTHOLOGY (1)
 - KEGG GENES (1)
 - NCBI-Gene (1)
 - UniGene (1)
 - WGC (1)
 - ENSEMBL-UP (9)
- Protein sequence (70)
 - PF (2)
 - RefSeq(pep) (4)
 - IP (4)
 - PM (60)
- DNA sequence (21)
 - EMBL (21)
- 3D Structure (64)
 - PDB (64)
- Protein domain (32)
 - InterPro (11)
 - Pfam (3)

図 1-6 関連エントリーを調べる All links サブメニュー

⑦ 実習②：タンパク質立体構造データ検索

タンパク質立体構造データベースは、PDB (Protein Data Bank) にて管理されています。PDB は、X 線結晶解析法、NMR 法 (核磁気共鳴法) によって決定されたタンパク質や核酸等の生体高分子の三次元構造の構造座標 (立体配座) を管理している国際的な公共データベースです。PDB データは、構造生物学やバイオインフォマティクス研究に欠かせない情報です。
PDB <http://www.rcsb.org/>

- ① 実習①の検索結果から GenomeNet の “All links” メニューより PDB データにアクセスします。PDB では、1 つの対象タンパク質でも、構造決定手法 (X 線結晶解析法、NMR 法) や変異体、低分子化合物、他のタンパク質との複合体結晶、特定の領域構造 (ドメイン構造等) など結晶化条件の違いによって、複数の構造情報が登録されていることがほとんどです。ヒトの EGFR については、80 構造が登録されています (図 1-7)。

DE	AltName: FullReceptor tyrosine-protein kinase erbB-1;	Gene (14)
DE	Flags: Precursor;	KEGG ORTHOLOG (1)
GN	Name:EGFR; Synonyms:HERB, ERBB1, HER1;	KEGG GENES (1)
OS	Organism: Homo sapiens (Human);	NCBI-Gene (1)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	UniGene (1)
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;	NCBI (1)
OC	Catarrhini; Hominoidea; Homo;	ENSEMBL-UP (9)
OK	NCBI_TaxID=9606;	Protein sequence (70)
SN	[1]	PSP (2)
SP	NUCLEOTIDE SEQUENCE [mRNA] (120F08M 1);	RefSeq(pep) (4)
SK	PubMed:6328312; DOI:10.1038/309418a0;	EPI (4)
SA	Ullrich A., Coussens L., Mayhew J.B., Dull T.J., Gray A., Tam A.W.,	FM2 (40)
SA	Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,	DNA sequence (21)
SA	Hayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.M.;	EMBL (21)
ST	"Human epidermal growth factor receptor cDNA sequence and aberrant	3D Structure (80)
ST	expression of the amplified gene in A431 epidermoid carcinoma cells.";	PCB (80)
SL	Nature 309:418-425(1984).	Protein domain (33)
SN	[2]	InterPro (11)
SP	NUCLEOTIDE SEQUENCE [mRNA] (120F08M 2);	Pfam (4)
SN	120F08M 2	

図 1-7 タンパク質立体構造データベース PDB へのリンク

- ② 80 構造の中から化合物名 TAK-285 と複合体を形成している EGFR のキナーゼドメイン構造の立体構造を選択します。PDB-ID は 3POZ です。PDB では、数値とアルファベットからなる 4 文字の ID で識別されています (図 1-8)。

2J3F	CRYSTAL STRUCTURE OF EGFR KINASE DOMAIN IN COMPLEX
2J4M	CRYSTAL STRUCTURE OF EGFR KINASE DOMAIN IN COMPLEX
2J1T	CRYSTAL STRUCTURE OF EGFR KINASE DOMAIN T790M MUTA
2J10	CRYSTAL STRUCTURE OF EGFR KINASE DOMAIN T790M MUTA
2J1V	CRYSTAL STRUCTURE OF EGFR KINASE DOMAIN T790M MUTA
2K81	HETERODIMERIC ASSOCIATION OF TRANSMEMBRANE DOMAINS
2K79	CRYSTAL STRUCTURE OF THE COMPLEX BETWEEN THE EGFR
2K7D	CRYSTAL STRUCTURE OF THE COMPLEX BETWEEN THE EGFR
2K7E	CRYSTAL STRUCTURE OF THE COMPLEX BETWEEN THE EGFR
2K6P	STRUCTURE OF EGFR IN COMPLEX WITH HYDRAZONE, A POT
2K2U	CRYSTAL STRUCTURE OF ISOLATED DOMAIN III OF THE EC
2K2V	CRYSTAL STRUCTURE OF THE EXTRACELLULAR REGION OF T
2K2L	X-RAY STRUCTURE OF EGFR IN COMPLEX WITH OXIME INHIB
2K00	CRYSTAL STRUCTURE OF C-CBL-TKB DOMAIN COMPLEXED WI
2C09	CRYSTAL STRUCTURE THE FAB FRAGMENT OF MATUZUMAB (F
2C0V	ANTIBODIES SPECIFICALLY TARGETING A LOCALLY MISFOL
2C0Y	ANTIBODIES SPECIFICALLY TARGETING A LOCALLY MISFOL
2C0P	CRYSTAL STRUCTURE OF THE EGF RECEPTOR JUXTAMEMBRAN
2K7S	CRYSTAL STRUCTURE OF THE INACTIVE EGFR KINASE DOMA
2J2A	CRYSTAL STRUCTURE OF EGFR 696-1022 T790M MUTANT CO
2L2B	EGFR KINASE DOMAIN COMPLEXED WITH AN IMIDAZO[2,1-b
2K7P	THE EXTRACELLULAR AND TRANSMEMBRANE DOMAIN INTERFA
2K82	CRYSTAL STRUCTURE OF C-CBL TKB DOMAIN IN COMPLEX W
2K00	CRYSTAL STRUCTURE OF CBL-C (CBL-3) TKB DOMAIN IN C
2K0Y	ANTI-EGFR/HER3 FAB DL11 IN COMPLEX WITH DOMAIN III
2K7V	CRYSTAL STRUCTURE OF CBL-3 TKB DOMAIN IN COMPLEX W
2K0Z	EGFR KINASE DOMAIN COMPLEXED WITH TAK-285
2K0Q	CRYSTAL STRUCTURE OF THE EXTRACELLULAR DOMAIN OF T
2K0I	CRYSTAL STRUCTURE OF THE MUTATED EGFR KINASE DOMAI
2K0G	CRYSTAL STRUCTURE OF THE MUTATED EGFR KINASE DOMAI
2K0N	CRYSTAL STRUCTURE OF THE MUTATED EGFR KINASE DOMAI
2K0O	CRYSTAL STRUCTURE OF THE WILD-TYPE EGFR KINASE DOM
2G5J	CRYSTAL STRUCTURE OF EGFR KINASE IN COMPLEX WITH B
2G5P	CRYSTAL STRUCTURE OF EGFR KINASE T790M IN COMPLEX
2K00	CRYSTAL STRUCTURE OF THE INACTIVE EGFR TYROSINE KI

化合物名 TAK-285 との複合体結晶構造データ

図 1-8 ヒト EGFR の PDB エントリーリスト

- ③ GenomeNet 上で公開されている PDB 情報は、登録情報の概要を把握するのに適しています (原子座標行は割愛されています)。実際に立体構造の可視化や、詳細な情報を確認したい場合には、PDB オリジナルの Web サイトをクリックします (図 1-9)。

GenomeNet

Database: PDB
Entry: 3POZ
LinkDB: 3POZ
Original site: 3POZ

PDBのオリジナルWebサイトへのリンク

```

HEADER    TRANSFERASE/TRANSFERASE INHIBITOR      23-NOV-10   3POZ
TITLE     EGFR KINASE DOMAIN COMPLEXED WITH TAK-285
COMPND    MOL_ID: 1;
COMPND    2 MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR;
COMPND    3 CHAIN: A;
COMPND    4 FRAGMENT: RESIDUES 604-1022;
COMPND    5 SYNONYM: PROTO-ONCOGENE C-ERBB-1, RECEPTOR TYROSINE-PROTEIN KINASE
COMPND    6 ERBB-1;
COMPND    7 ECI 2.7.10.1;
COMPND    8 ENGINEERED: YES
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM SCIENTIFIC: HOMO SAPIENS;
SOURCE     3 ORGANISM COMMON: HUMAN;
SOURCE     4 ORGANISM TAXID: 9606;
SOURCE     5 GENE: EGFR, ERBB1;
SOURCE     6 EXPRESSION_SYSTEM: SPODOPTERA FRUGIFERA;
SOURCE     7 EXPRESSION_SYSTEM TAXID: 7119;
SOURCE     8 EXPRESSION_SYSTEM STRAIN: SFG;
SOURCE     9 EXPRESSION_SYSTEM VECTOR TYPE: BACULOVIRUS;
SOURCE     10 EXPRESSION_SYSTEM PLASMID: PFASTBACHT
KEYWDS     KINASE DOMAIN, ANTI-ONCOGENE, ATP-BINDING, CELL CYCLE, DISEASE
KEYWDS     2 MUTATION, GLYCOPROTEIN, MEMBRANE, NUCLEOTIDE-BINDING,
KEYWDS     3 PHOSPHOPROTEIN, RECEPTOR, SECRETED, TRANSFERASE, TRANSMEMBRANE,
KEYWDS     4 TYROSINE-PROTEIN KINASE, TRANSFERASE-TRANSFERASE INHIBITOR COMPLEX
EXPDTA     X-RAY DIFFRACTION
AUTHOR     K.AERTGEERTS, P.SKEENE, S.SOGABE
REVDAT     2 15-JUN-11 3POZ 1 JENL
REVDAT     1 30-MAR-11 3POZ 0
JENL       AUTH K.AERTGEERTS, P.SKEENE, J.YANO, B.C.SANG, H.SOU, G.SHELL.
  
```

All links

- Chemical substance (2)
- PDB-CCD (2)
- Chemical reaction (1)
- KEGG ENZYME (1)
- Gene (1)
- KEGG GENES (1)
- Protein sequence (2)
- UniProt (1)
- PDBSTR (1)
- Protein domain (1)
- Pfam (1)
- Literature (1)
- PubMed (1)
- All databases (8)
- Download PDF

図 1-9 PDB データ 3POZ の概要

- ④ PDB の Web サイトでは、さまざまな情報がタグ付きの項目を用いてわかりやすく整理されています。立体構造を可視化したい場合は、立体構造画像がある右側の項目内の、“View in 3D” をクリックします（図 1-10）。Java 機能を用いて回転、拡大等が自由に調整できる立体構造のビューアー画面が起動します。※実行の確認のための警告ウィンドウが表示されることがありますが、実行ボタンをクリックし実習を続けます。

PDB
PROTEIN DATA BANK

Search: All Categories: 3POZ e.g. PDB ID, molecule name, author

タグ項目

3POZ

EGFR Kinase domain complexed with tak-285

DOI: 10.2210/pdb/3poz/pdb

Primary 3D Structure

Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein.

Aertgeerts, K., Skeene, P., Yano, J., Sang, B.C., Sou, H., Smith, G., Jennings, A., Demetri, G., Kubota, N., Hoshino, A., Johnson, T., Tanaka, T., Hsu, M., Saka, Y., Sogabe, S.

Journal: (2012) J Biol Chem. 286: 18779-18785

PubMed: 22454502

PubMed Central: PMC3090692

DOI: 10.2210/pdb/3poz/pdb

Search Related Articles in PubMed

Published Abstract:

Allosteric signaling of ERB family members human epidermal growth factor 2 (HER2) and epidermal growth factor receptor (EGFR) is implicated in many human cancers, and HER2 is predictive of human disease recurrence and prognosis. Small molecule kinase inhibitors of... [Read More & Search Published Abstracts]

1. Multimeric Sequences

Classification: Transmembrane/Transmembrane Inhibitor

Structure Weight: 36.65 (4.9)

Molecule	Polypeptide	Epidermal growth factor receptor	protein	Length
1	Protein	Epidermal growth factor receptor	protein	327
2	Protein	Epidermal growth factor receptor	protein	327

立体的構造表示

View in 3D

More Images...

Downloadable images:

- Simple Viewer
- Protein Workshop
- Block Viewer

図 1-10 PDB エントリー 3POZ のトップページ

- ⑤ リボンモデルと呼ばれるタンパク質の二次構造に着目した表現方法（ピンクは α ヘリックス構造、黄色は β シート構造）で立体構造が表示されます。今回のエントリーのように低分子化合物との複合体構造の場合は、低分子化合物が原子種ごとに色分けされたボール&スティックで表示されます。分子表示領域内にて、マウススクロールボタンで拡大縮小、

マウス右ボタンで詳細メニューが表示されます。基本表示は、分子表示下にあるオプションメニューで調整できます (図 1-11)。

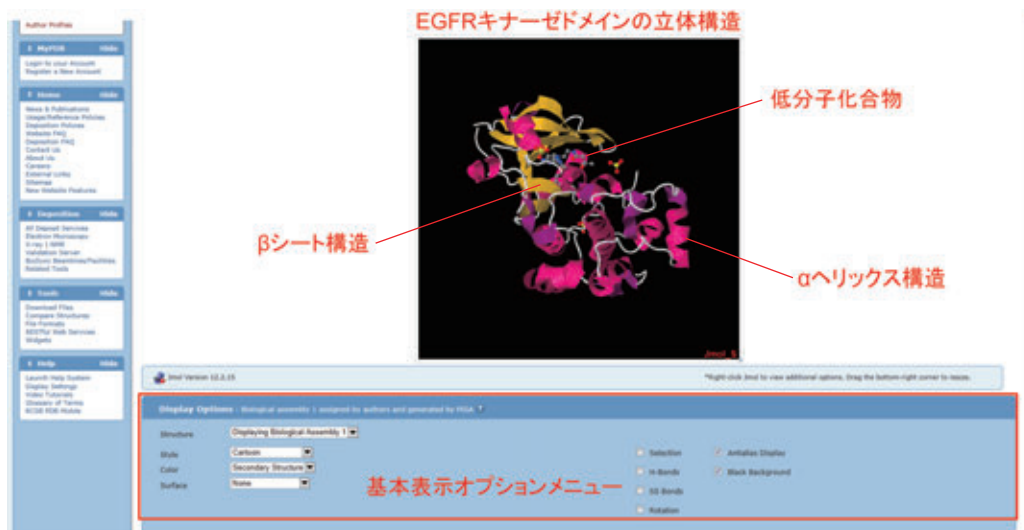


図 1-11 立体構造表示ウィンドウ

- ⑥ オプションメニューの活用例として、化合物結合部位に着目した表示に変更します。Style 項目から “Ligands and Pocket” を選択します。化合物と結合に関与する周辺アミノ酸残基が表示されます (図 1-12)。

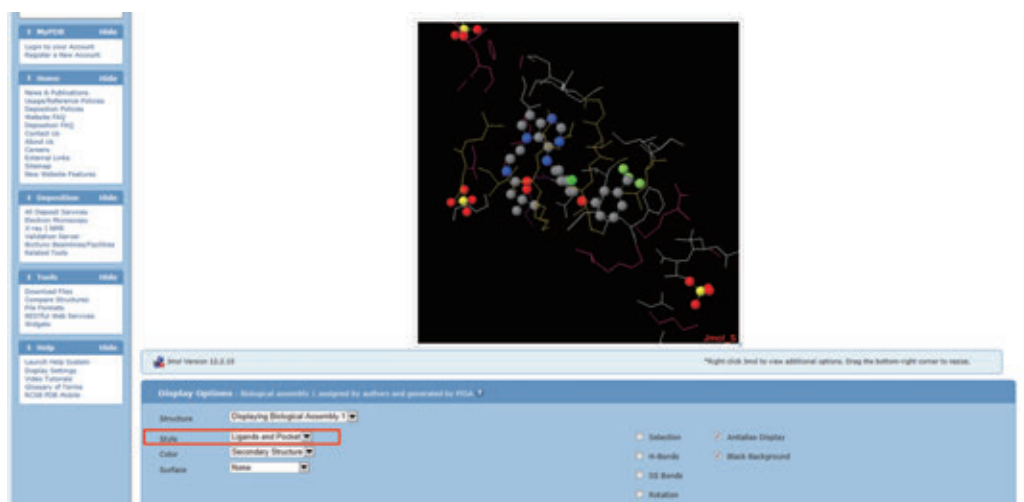


図 1-12 表示スタイルの変更

- ⑦ 構造決定のための実験情報は “Experimental Details”、化合物情報は “Ligand Chemical Component” や “External Ligand Annotations” 項目が参考になります (図 1-10)。

著者略歴

広川貴次（ひろかわたかつぐ）

沖縄県出身。

東京農工大学工学部物質生物工学科（現 生命工学科）卒業。同大学大学院工学研究科博士課程修了（工学博士）。

（株）菱化システム計算科学部勤務を経て、2001年に産業技術総合研究所に入所。生命情報科学研究センター、生命情報工学研究センターを経て、2013年より創薬分子プロファイリング研究センター理論分子設計チーム長。バイオインフォマティクスおよびインシリコ創薬に関する方法論の開発と応用研究を目指している。

東京大学大学院新領域創成科学研究科客員准教授および北海道大学大学院生命科学院客員教授を併任。日本生物物理学会、日本蛋白質科学会、日本バイオインフォマティクス学会、CBI学会会員。

趣味はスポーツと料理（主に洋食）。

美宅成樹（みたくしげき）

三重県出身。

東京大学理学部物理学科卒業。同大学大学院を終了後、東京大学工学部物理工学科助手、東京農工大学工学部生命工学科教授、名古屋大学院工学研究科マテリアル理工学専攻応用物理分野教授を経て、現在は名古屋大学名誉教授、豊田理化学研究所客員フェロー。

日本生物物理学会、日本物理学会、日本分子生物学会、日本バイオインフォマティクス学会会員。

著書・訳書は『できるバイオインフォマティクス』『即活用のためのバイオインフォマティクス入門』（以上、中山書店）、『ヒトゲノム計画とは何か』『科学101の未解決問題』『子供にきちんと答えられる遺伝子Q&A100』（以上、講談社ブルーバックス）、『分子生物学入門』（岩波新書）、『生きもの探検！親から子へ伝わる遺伝のしくみ』（数研出版）、『はじめての科学英語論文（第2版）』（丸善）、『生物とは何か？』（共立出版）など。

1990年から始まったヒトゲノム計画にはゲノム情報の側面から関わり、プロジェクトの推移を内部で体験。当初からゲノム科学における教育の問題と社会との接点の問題が重要と考え、バイオインフォマティクス講習会や市民講座などを継続的に行ってきた。

趣味は観劇（宝塚歌劇）、映画鑑賞、囲碁、スノーボード、水泳、アイロンがけなど。

Web で実践 ^{じっせん} ^{せいぶつがくじょうほう} 生物学情報リテラシー

2013 年 9 月 2 日 初版第 1 刷発行 ©

〔検印省略〕

著者 ——— ^{ひろかわたかつぐ} 広川貴次 ^{みたくしげき} 美宅成樹

発行者 ——— 平田 直

発行所 ——— 株式会社 中山書店
〒113-8666 東京都文京区白山 1-25-14
TEL 03-3813-1100 (代表) 振替 00130-5-196565
<http://www.nakayamashoten.co.jp/>

DTP・印刷 ——— 新日本印刷株式会社

Published by Nakayama Shoten Co., Ltd.

Printed in Japan

ISBN 978-4-521-73772-0

落丁・乱丁の場合はお取り替え致します

本書の複製権・上映権・譲渡権・公衆送信権（送信可能化権を含む）
は株式会社中山書店が保有します。

JCOPY 〈(株)出版者著作権管理機構 委託出版物〉

本書の無断複写は著作権法上での例外を除き禁じられています。

複写される場合は、そのつど事前に、(株)出版者著作権管理機構

（電話 03-3513-6969, FAX 03-3513-6979, e-mail: info@jcopy.or.jp）の許諾
を得てください。

本書をスキャン・デジタルデータ化するなどの複製を無許諾で行う行為は、著作権法上での限られた例外（「私的使用のための複製」など）を除き著作権法違反となります。なお、大学・病院・企業などにおいて、内部的に業務上使用する目的で上記の行為を行うことは、私的使用には該当せず違法です。また私的使用のためであっても、代行業者等の第三者に依頼して使用する本人以外の者が上記の行為を行うことは違法です。
