

BIRD

Institute for
Bioinformatics
Research and Development

バイオインフォマティクス推進センター事業 成果集



バイオインフォマティクス推進センター事業
成果集



BIRD 成果集 目次

はじめに	4
統括からのメッセージ	5
事業の概要	6
発足の経緯	
事業概要	
推進体制	
予算推移	
座談会	8
BIRD が拓いたもの 引き継ぎたいもの	
バイオインフォマティクスとは	22
研究開発課題紹介	
研究開発課題一覧	24
金久 實（京都大学化学研究所バイオインフォマティクスセンター 教授）	26
菅原 秀明（国立遺伝学研究所生命情報・DDBJ 研究センター 特任教授）	28
高木 利久（東京大学大学院新領域創成科学研究科 教授）	30
中村 春木（大阪大学蛋白質研究所 教授）	32
森下 真一（東京大学大学院新領域創成科学研究科 教授）	34
稲葉 一男（筑波大学下田臨海実験センター 教授）	36
西岡 孝明（奈良先端科学技術大学院大学情報科学研究科 特任教授）	37
伊藤 啓（東京大学分子細胞生物学研究所 准教授）	38
伊藤 隆司（東京大学大学院理学系研究科 教授）	39

大浪 修一（理化学研究所 生命システム研究センター 発生動態研究チーム チームリーダー）	40
岸野 洋久（東京大学大学院農学生命科学研究科 教授）	41
肥後 順一（大阪大学蛋白質研究所附属プロテオミクス総合研究センター 客員教授）	42
矢田 哲士（京都大学大学院情報学研究科 准教授）	43
内山 郁夫（自然科学研究機構 基礎生物学研究所 助教）	44
太田 元規（名古屋大学大学院情報科学研究科 教授）	45
川戸 佳（東京大学大学院総合文化研究科 教授）	46
塩田 浩平（京都大学 理事・副学長）	47
白井 剛（長浜バイオ大学バイオサイエンス学部 教授）	48
宮川 剛（藤田保健衛生大学総合医科学研究所 教授）	49
石井 信（京都大学大学院情報学研究科 教授）	50
黒川 顕（東京工業大学大学院生命理工学研究科 教授）	51
榊原 康文（慶應義塾大学理工学部 教授）	52
清水 謙多郎（東京大学大学院農学生命科学研究科 教授）	53
中井 謙太（東京大学医科学研究所 教授）	54
長野 希美（産業技術総合研究所 生命情報工学研究センター 主任研究員）	55
馳澤 盛一郎（東京大学大学院新領域創成科学研究科 教授）	56
馬見塚 拓（京都大学化学研究所 教授）	57
水口 賢司（医薬基盤研究所バイオインフォマティクスプロジェクト プロジェクトリーダー）	58

参考資料	
データベース、ツール一覧	62
主要なデータベースに関するデータ	64
ゲノムリテラシー講座	68
BIRD 年表	70

はじめに



独立行政法人科学技術振興機構
理事長 中村 道治

科学技術振興機構(JST)ではバイオインフォマティクスの推進、およびそれを基盤とした新しい生物科学の創造を目指して、平成13年度にバイオインフォマティクス推進センター(BIRD)を設置いたしました。

バイオインフォマティクスは、ゲノム情報を始めとした膨大かつ多種多様な生物情報を整理統合し、そこから有用な知識を見出す研究分野です。生物科学の発展に伴い研究の成果として得られるデータの種類や量が飛躍的に増えている現在において、バイオインフォマティクスは現代の生物学や医学分野の基盤であるのはもちろんのこと、創薬等の応用研究への利用も進んでいます。BIRDではバイオインフォマティクスの研究開発や、データベースの高度化や標準化を支援し、またこれに携わる人材の育成を図るなど様々な事業を展開してまいりました。

一方で、我が国において生物科学分野の一元的かつ恒久的な統合データベースを整備するための検討が内閣府において行われました。その結果、文部科学省の統合データベースプロジェクトと当機構のバイオインフォマティクス推進センター事業を平成23年度に一本化して、統合データベースのための新しい組織を作ることとなり、平成23年4月に当機構にバイオサイエンスデータベースセンター(NBDC)が設置されました。これに伴い、バイオインフォマティクス推進センター事業は平成23年度末を以て終了いたします。

本書では、これまでの11年に亘るバイオインフォマティクス推進センター事業の歩みを振り返ると共に、研究開発の成果を今後の展望も含めた形でまとめました。さらにBIRDが日本のバイオインフォマティクスの発展に果たした役割についてご議論いただいた座談会の記事も収録しております。これらは今後のバイオサイエンスデータベースセンター(NBDC)の運営にも活かして参りたいと思います。

関係各位のこれまでのご協力に深く感謝申し上げますとともに、今後とも皆様方の一層のご指導、ご鞭撻を賜りますようお願い申し上げます。

平成23年11月

統括からのメッセージ



バイオインフォマティクス推進センター
統括 勝木 元也

バイオインフォマティクス推進センター(Institute for BioInformatics Research & Development; BIRD)は、科学技術会議ライフサイエンス部に設置されたゲノム科学委員会の報告に基づき、JSTの事業として平成13年度に発足したプロジェクトです。

種々の生物種でゲノムの全塩基配列の決定が相次ぎ、ヒトゲノムの全塩基配列決定も間近に迫っていた時期に、報告書の中に「ゲノム科学は情報科学である」とする文言が書かれ、我が国にとっては未開に近い領域開発の必要性が指摘されました。

バイオインフォマティクスの開発には、二つの方向からのアプローチが必要です。それは、すでに我が国に存在するデータベースの高度化と、生物学と情報学のインフォマティクスを介しての融合の二つです。後者は、生物における膨大で無秩序に見えるゲノムやタンパク質や形態のデータをコンピュータ上で整理し、生物学的に意味のある知識をその中から発見していくというものです。それには、インフォマティクスを駆使する情報学の専門家とその意味を発見していく生物学の専門家とが、一体となって取り組む必要がありました。それは新しいコンピューティショナルバイオロジーの確立を目指す人材の発掘でもありました。

私が統括をお引き受けすることになったとき、いち早くバイオインフォマティクスの重要性に目をつけておられた国際高等研究所の松原謙一副所長に相談したところ、その試みはすでに「情報生物学適塾」として種を播いてあるから参考にしてはどうかと教えられました。これが引き金となって、ドライとウェットを一組とするテーマを募集しました。人によってはすでにその両方の方法論を身につけている応募者もいましたが、新たな

分野を開こうとする創造的発想と実現性を念頭に置きながら、松原謙一、吉田光昭、堀田凱樹、榊佳之および私の5名からなるバイオインフォマティクス委員会委員と数名の分科会委員とを審査委員にして、申請課題の審査を行い、高木利久副統括と共にプロジェクトを推進しました。

選ばれた方々は、きわめて創造的で、それぞれの報告をご覧いただければおわかり願えると思いますが、情報からの人たちが生物学者といっても良い進化を遂げ、生物学からの人たちが、コンピュータを駆使する現代のコンピューティショナルバイオロジーの担い手になりました。

委員会の方針は、活動は見えていくけれども、結果を急がせないこと、横道の小さな成功を評価しないこと、出来る限りの経費を活動に当てること、そして本人が納得できる結果だけを評価する姿勢を貫きました。

BIRD事業の成果は、数々の優秀な論文や新しいプログラム、そして高度化されたデータベースによってのみ現れているのではなく、新しい領域開拓への創造的挑戦と、いくつもの失敗を通して問題のありかを発掘しながら戦ってきた研究者達の変貌とも言うべき成長にあります。

ここでなされた活動は、必ず自信となり、次代の生物学の形成につながるものと確信します。

私にとってはこんなに楽しい仕事の統括をさせて頂いたことに感謝いたしますとともに、創造に明け暮れた参加者達を誇りに思います。本当に楽しかった。ありがとうございました。

そしてすべての方々のご協力にあらためて感謝いたします。

平成23年11月

事業の概要

1. 発足の経緯

平成12年にヒトゲノムの概要配列が決定され、また同じ頃、その他数多くの生物のゲノム配列が決定されたことを受け、遺伝子発現解析、タンパク質の構造決定、プロテオーム解析、分子間相互作用解析を始めとした、いわゆるポストゲノムシーケンス研究が激しい国際競争の中で本格化してきた。こうした国際的な動きの中で、我が国のゲノム情報科学の研究開発課題として以下の3点が挙げられていた。

- (1)ゲノム配列データを始めとする多種多様で膨大な生命情報をどのように整理統合し、データベース化するか(データベース構築)
- (2)そのデータベースから生物学医学に関する有用な知識を導き出すとともに、それを検証するための情報解析技術をいかにして開発するか(情報解析技術)
- (3)それらを担う人材をどのようにして確保するか(人材養成)

2. 事業概要

バイオインフォマティクス推進センター事業では膨大かつ多種多様な生物情報を整理統合し、そこから有用な知識を見出すことにより、新しい産業の創出、新しい医療の開拓、新しい農業の構築へと発展することを可能とする情報生物学(バイオインフォマティクス)の発展の推進及びそれを基盤とした21世紀の新しい生物学の創造を目指し、具体的には以下の3項目を実施した。

- (1)生命情報データベースの高度化・標準化
膨大な生物情報からの新しい知識の発見等に不可欠なデータベースの構築や高度化を行った。また、これらのデータベースを有機的に統合化し、生物をシステムとして理解することを可能とするともに、それらが広く社会に活用されるような環境を整備した。平成13年～17年度に第Ⅰ期研究開発を、平成18年～22年度に第Ⅱ期研究開発を実施した。
- (2)創造的な生物・情報知識融合型の研究開発
情報科学と生物学を融合したアプローチにより、多彩な生物情報から生物現象の原理や法則を発見し体系化することを目指す研究開発を推進し、これらの成果がバイオインフォマティクスの発展及び新しい

そこで、我が国におけるゲノム情報科学に係る戦略、特に上記の3課題について、今後の在り方や支援方を議論するために、科学技術会議ライフサイエンス部会ゲノム科学委員会にゲノム情報科学ワーキンググループが設置され、ワーキンググループで議論された結果が平成12年11月に、ゲノム科学委員会報告書『ゲノム情報科学における我が国の戦略について』として取りまとめられた。本報告書を受けて、「データベース構築」及び「情報解析技術開発」を推進するために、科学技術振興事業団(現・科学技術振興機構)は平成13年4月にバイオインフォマティクス推進センター事業を開始した。

なお、「人材養成」に関しては科学技術振興調整費により大学や研究所での人材養成に対する支援が実施された。

情報生物学創造のインセンティブとして機能し、これを基盤として新たな生物学が樹立されることを目指した。新規採択は、平成13、17、18、19年度に行い、それぞれ3年間の研究開発を行うとともに、継続のための評価を経て、2年間の継続研究開発を実施した。

- (3)研究成果情報の発信
バイオインフォマティクスに関連するソフトウェアの実習を伴う研修会を開催した。また、上記(1)(2)で開発されたデータベース、ツール類等を公開、提供するとともに、平成14年度には、それまで「高性能生体データベース開発事業」及び「多型情報データベース開発事業」として開発、提供してきたHOWDYやJSNP等のデータベースの運用や公開もバイオインフォマティクス推進センター事業に一本化し、これらの成果の公開を継続して行った。
- また、GBIF(地球規模生物多様性情報機構)に関し、GBIF拠出金の支払いと、日本における生物多様性データの取り扱いを議論するGBIF技術専門委員会の事務局をJSTが担当した。さらにバイオインフォマティクス推進センターのWebサイトからGBIF関連情報を公開した。

3. 推進体制

バイオインフォマティクス推進センターは、新しい生物情報の研究開発によるデータベースの整備等を推進するとともに、当該データベースの普及を促進するための拠点であり、研究開発の推進に際しては、統括をプログラムオフィサーとし、バイオインフォマティクス委員会の有識者から助言を得る体制で事業を運営した。

統括 勝木元也(自然科学研究機構 基礎生物学研究所 所長)
副統括 高木利久(東京大学医科学研究所 教授)(平成20年度まで)
(所属機関、役職は平成13年当時)

統括は生命情報データベースの高度化・標準化、及び創造的な生物・情報知識融合型の研究開発における研究開発課題の評価のとりまとめ、研究者に対する指導・助言、研究開発の進捗状況の総合的な把握・調整等を職務とし、副統括は研究者に対する指導・助言、普及活動の調整、計算機資源活用の調整及び統括の補佐を職務とした。

また、バイオインフォマティクス推進センター事業に関し、次に掲げる事項を調査審議するため、外部の学識経験者からなるバイオインフォマティクス委員会を組織した。

- (1)研究開発課題及び代表研究者の選考に関すること。
- (2)研究開発の実施及び評価に関すること。
- (3)その他事業の推進に必要な事項に関すること。

バイオインフォマティクス推進センターでは、研究開発経費の執行及び管理、研究員等の雇用、知財取得の支援、科学技術計算サーバの運用と利用支援など研究開発課題の進捗を支援する多様な業務を担っていた。大学等とは共同研究契約を締結し、消耗品や大学等に雇用される研究者の旅費等は大学等に納入した。

平成17年度採択課題より、それまでの共同研究契約に基づくバイオインフォマティクス推進センターの事務所による研究開発支援(研究開発経費の執行、研究員等の雇用、特許等知財取得の支援など)に代わり、委託研究契約に基づき研究開発を推進することとした。これにより、研究開発費の執行や研究員の雇用は基本的に大学等の研究機関によって行われることとなった。

4. 予算推移

平成13年度	20.0億円	
平成14年度	21.5億円	JSNP等の予算追加
平成15年度	20.2億円	10月より独立行政法人化
平成16年度	17.7億円	
平成17年度	18.6億円	新規研究開発課題募集の予算追加
平成18年度	18.1億円	
平成19年度	16.8億円	科学技術計算サーバの整理、事務合理化
平成20年度	16.8億円	
平成21年度	18.4億円	文部科学省統合データベースプロジェクト一部移管
平成22年度	17.6億円	文部科学省統合データベースプロジェクト中核機関分の移管
平成23年度	0.6億円	バイオサイエンスデータベースセンターの発足 (ライフサイエンスデータベース統合推進事業の開始)

座談会

BIRDが拓いたもの 引き継ぎたいもの

BIRDでは、創造的な生物・情報知識融合型の研究開発や、生命情報データベースの高度化・標準化を支援する一方で、これに携わる人材の育成を図るなど様々な事業を展開してきました。11年間に及ぶBIRDの成果を総括する目的で、統括・副統括・バイオフィーマティクス委員会委員として、設立時からBIRDの推進にご尽力いただいた4人の先生方にお集まりいただき、創立から黎明期に至るBIRDの歴史・当時の研究環境、この領域の最近の進歩、今後の発展について語っていただきました。



BIRDはこうして始まった

勝木元也先生(以下、勝木) 平成13年度から始まったバイオフィーマティクス推進センター事業(BIRD)が、十年目を迎え、事業の事後自己点検をすることになりました。これまでBIRDに採択された方々の成果を通常の方法で評価し、事業の意義を点検するののも一つのやり方です。しかし、データベースの高度化に貢献した金久さんや中村春木さんが、このシステムのもつ創造的研究開発は続けなくては、と言ってくれたので、BIRDの成長をご存じの方を中心に、座談会を通して自己点検をしようと思いました。デジタルな評価ではなく、アナログな評価をしたいと思っています。時の流れのなかで、変化し、成長してきた、そして世の中の状況も変わってきたことを踏まえないと、この事業の意義があまり浮き彫りにされないだろうと思うからです。中村さんや金久さんは、「学びて習わざる」ことの多い現代の若者たちを、こういう形で育てるのは、卓越した研究者として育てるのにリアリティのある、とても良いシステムだと評価して下さっているようです。今日は、最初から副統括として一緒にやってきた高木さん、十年前からバイオフィーマティクス委員会委員として助けていただいた松原先生、吉田先生にお集まりいただきました。個々の成果はさておき、事業総体の意義について、BIRDが始まる時

の状況や、第一期・第二期を振り返って、何が育ってきたのか、このシステムを研究開発とどうつないでいくか。まあ終わったほうがいい場合もあるでしょうけれども、一方でどのようにリニューアルしていったらいいのか。そんなことを話していこうと思います。また、今年度から始まったバイオサイエンスデータベースセンター(NBDC)へどうつながるかということも残っています。これはNBDCの高木さんが中心におやりになることではありますし、BIRDそのものを続けることはないでしょうが、その精神を他の事業に繋げるべきかどうか議論していただきたい。全体としてはBIRDが開拓したものを中心に、お話を伺いたいと思います。まず、平成13年度に、どのようにしてBIRDが始まったのか、が一番のポイントだと思います。平成12年に科学技術庁(当時)のゲノム委員会で、日本の貢献が期待されたヒトゲノムの完全配列決定をどこでやるか、ヒトゲノムが全部分かったときにその後のゲノム科学をどう進めるか、という議論がされました。ヒトゲノムの配列の決定は、榊さんがおられた理研で行うことになったのですが、同じ理研で、横山茂

之さんのタンパク質のフォールディングの法則を見つけるというプロジェクトを作ろうという議論もありました。議論の途中でベンダーがショウジョウバエの全配列をショットガンで決めるとの情報が入って、とても騒がしい状況でした。今後のプロジェクトにどれくらい予算をつけて、どこでやるか、ということを議論しました。そこに、当時の科学技術会議ライフサイエンス部会の専門委員だった私が参加することになったのです。

しかも、なぜか、私に作業部会長の役割が廻ってきて、それで書いたのが、一言でいうと、「ゲノム科学は情報科学である」という報告書素案（A4で1枚）です。これが受けた。受けなかったのは医学系の人達です。そもそも言い出した松原さんにも当然受けた。

松原謙一先生（以下、松原） 僕は日本のヒトゲノム計画を始めた最初からこれは情報科学だと言ってきました。それで、日本のゲノムの立ち上げの最初から、全部記録をとってあります。

勝木:私は多分そういうことを、日頃の議論の折りにお聞きしていたから、松原さんに無断で、書いたんだと思います。

それで、この素案が、報告書の下敷きになりました。それでこれを膨らませましようということになった。全ての生物現象がそうなるとは全然思っていませんでしたが、個別の生物現象をきちんと調べるのに、インフォマティクスなりデータベースが必要だということは、シークエンスをやる以上は自明なんだという主張をしたかったんです。みんなが賛成してくれてヒトゲノム配列を全部決める必要があるし、他の生物も調べる必要があるという議論になって、プロジェクトとして実現していきますというのが、平成12年度の状況です。当時は、タンパクの基本構造である、異なる1,000個のフォールディングが判れば、それから周期律のような法則が見えて、基本構造を組み合わせれば、タンパクの3次元構造の予測ができる。この横山提案は、面白いなと思いました。

私も何かの会議に招かれたときに、ゲノムシークエンスの中に、ある種の周期律があるに違いないという目で見たら面白い。そのためには塩基配列を生物機能と結びつけなければならないという議論をしていたので、横山さんの話が出たときに面白いと思ったんです。私はそのときには予算は別に出すべきで、ゲノム情報とはテイストが違くと発言して、科学的でないと横山さんから顰蹙を買いました。しかし、その意味は、常在するゲノム情報と、そこから読まれ、変化して発現してくるタンパク質とでは、その過程で情報に様々な文法上の変化があるので、ゲノム情報とは違うものだ。だから、分けたほうがいいという議論をしました。しかし結果的には一緒に理研でという話になりました。そこからタンパク3000が今度是小委員会に移って、タンパクの構造科学になりました。

ちょっと激しい議論のあった会議だったんですよ。あとで考えると、とて

も重要な会議だったと思いますね。その結論として、どこで何をやったかということは今から検証されるべきだし、なんであそこで、あれだけついたゲノム関連の予算が、たちまちのうちに無くなったのかということも検証すべき大事な問題です。

その後、2月の半ばに突然BIRD事業で統括をやってくれないかという話が私にきたんです。高木さんと組むなら、生物学としても面白そうだし、とくにインフォマティクスをやる人がバイオリジストになってくれればと思い引き受けました。それが平成13年の初めのころで、突然の話だったんです。

吉田光昭先生（以下、吉田） ゲノムのプロジェクトからいうと、ミレニアムの途中ですか。

勝木:途中です。ただし、省庁再編で文科省になった年です。

吉田:わたくしは2000年に大学の定年後に会社に移りましたので、そのころの情報は完全にブランクになっているんです。今の話も、ああ、そうだったのかと思って。

勝木:こういう報告書の類を書いたら、必ず吉田先生に見せて、ご意見を仰ぐんですけど、当時、まだ医科研に居られた吉田先生に持って行ったら明快で分かりやすいと、かなり褒められました。覚えていらっしゃるんですか。

吉田:ゲノム科学委員会の報告書の素案は、よく覚えていないのですが、覚えているのは、BIRDは一体何のために何をする予定なのかと質問したことです。たとえば医科研のヒトゲノム解析センターとか、金久さんがやってるKEGGとか遺伝研のDBJなどがちゃんと動いている。それらとの関係を疑問に思ったわけですが、勝木さんが、このプロジェクトを使って人を育てたいんだとおっしゃったんですね。それで大いに納得してバイオインフォマティクス委員会委員を引き受けたんです。

勝木:そこに至るまでに、実は松原さんと電話での議論があったんですね。そういう人たちを育てるためにはどうしたらいいか。毎週金曜か土曜か忘れましたけど、高木さんや榊さんや、ときには、ヒトゲノム解析センターの人や、文科省の補佐や課長も参加してくれて、医科研の私の部屋に来てもらって、ワインをあげながら楽しい議論をしました（笑）。そこで今後益々盛んになる膨大なデータの蓄積を利用できる日本の情報科学、とくに生物学を支えてくれる人がどうしたら育つか、そのためにはどういう方策があるのだろうか、という議論をしていたんです。そのときに“松原適塾”のことを聞いたんです。適塾というのは、幕末の文化と文明とが沸騰する時代に、今の阪大の一部にあった蘭学を勉強する私塾でした。そこからは、福沢諭吉、大村益次郎など、司馬遼太郎氏が多くの才能を見出して小説を書いておられますが、とにかく自由で、お互いが切磋琢磨する理想の塾と言えるものではないでしょうか。その“松原適塾”の話をしていただければいいかな。

松原:2000年ころでしたか。国際高等研究所で、情報生物学適塾を二

年連続でやりました。ゲノムを始めた当時はバイオインフォマティクスをやってる人をほとんど知りませんでした。まだ高木さんもういらっしゃらなかったような気がするんですが。

高木利久先生（以下、高木） いいえ、もうおりました。医科研にヒトゲノム解析センターができたのが平成3年なんです。BIRDができるちょうど十年前、吉田先生がご尽力されてヒトゲノム解析センターができたんですが、私は平成4年に移りました。金久さんが最初の教授で、わたくしがその助教授で。そのとき医科研には榊さんがおられましたし、しばらくして中村祐輔さんが米国から帰国されました。

松原:その前は？

高木:その前はデータベースとかコンピューターのほうにいました。

松原:それは何のデータベースですか。

高木:データベースに推論機能をつけるということをやっていました。その関係で、その当時、九州大学の榊さんがやっていた、データベースの手伝いに行きました。ゲノムのデータベース開発のお手伝いをしたんです。それが1980年代中頃です。その後、ヒトゲノム解析センターができてヒトゲノムプロジェクトが始まるということで、平成4年に移りました。

松原:先ほども言いましたが、ゲノムプロジェクトはDNAのシークエンスをどんどんして、そこから情報をどういう具合に引き出して活用してするかでした。まずヒトでしっかりやる。後でヒト以外のものに金を使ったのはけしからん、という人も現れたけれども、実はそうではなくて、それで現在のゲノム研究の基礎が確立できました。はじめの頃はシークエンスを一生懸命やっている人がいたけれども、バイオインフォマティクスってどういうふうにやるものかというのがほとんど分からなかったんですね。

当時、僕は本当はインフォメーションバイオロジーでやってほしいと思ったんです。インフォマティクスじゃなくてバイオロジーでやってほしいと。だけれどもそういうことのできる人は金久さんと高木さんくらいしかいなかった。ほとんどいなかったんです。この二人にどういう具合にやってもらおうか、ということから話が始まりました。一番苦労したのは、やはり適正な額の「事業費」を確保することで、協力的な雰囲気ができるまでに何年もかかりましたね。もう一つ生命倫理に絡めるもありましたが、インフォマティクスは難しかったですね。生物の情報を整理したり読んだり、次にどういうことを発展させたらいいかというようなことも、今みたいにきちんと出てきませんから。DNAのデータベースは遺伝研にあり、金久さんはKEGGを始めたときでした。でも、日本全体をどうするのかという話になると、まとめるのは難しかった。とくに、教育システムがなかったんです。教育システムについては文科省に行ってそういう話をする、生物ならば理学部で、情報ならば工学部で、そういうことを言って逃げるわけです。結局、どっち

でもなく両方にまたがるものは作れないと。待っていたら教育が始まらないので、じゃあ自分でやろうかということで始めたのが情報生物学適塾だったのです。完全にタダでした。

高木:松原先生がおられた国際高等研に無償で場を提供していただいたんです。

松原:20人の第一線の先生方をお願いしてテキスト作りから実習のコースまで、お一人一日の担当で皆さん助手を連れて来て、午前は講義、午後は実習というのを20日間やりました。しかし、この先生方は全てボランティアをお願いして、謝礼は出しませんでした。それから生徒は募集して、これも自費で全国から熱意のある人々が集まり熱気に溢れていました。コンピューターとかそういうものの世話は日立ソフトに、これも無償で全部世話してもらいました。要するに国のお金は一切使わず、二年間やりました。その間に東大にコースができることになったのですかね。

高木:ちょうど松原先生が適塾を企画されているころは何も教育組織がなかったんです。勝木先生が先ほどおっしゃっていた土曜日にワインを飲みながら議論をしていたころと同時期に、BIRDの話と、もう一つ人材養成プログラムを目指しようという話がありました。そのころは大学の自助努力を待っていても学科は作れなかったんですが、それなら何とか国からの寄付講座みたいなものを作れないかということがもちあがりました。当時、慶應義塾でやはり情報学の講座を作るべきだという構想をお持ちだった安西祐一郎さんがたまたま、ワイン懇談会に来てくれて、私学としても何か手立てがないか考えておられるようでした。そこで、文科省にバイオインフォマティクスで学科は作れないかということをお願いしたら、BIRDと同じく平成13年度から振興調整費による人材養成プログラムが始まったんです。それで東大にそういうものをつくりました。それが適塾の前後でしょうか。

松原:それができたので、もういいよと（笑）。ニーズは高いけど、先生の負担が大きいから。みんなタダで聞くんだけれども。

勝木:何日間ですか。

松原:二十日間だったと思います。先生は入れ替わり立ち替わりです。

勝木:それはスタンダードなテキストがあったわけじゃないんですね。

松原:テキストなんて何も無い。先生ごとにテキストがあったわけです。

高木:終わった後にそれをテキストにしました。





松原:先生も生徒も手弁当でした。でもみなさんやってくださいましたね。あのころ、教える場もありませんでした。

吉田:素晴らしいですね。

勝木:それを聞いて私は人を育てるしかないと思ったんです。バイオインフォマティクスとは何かということも、何を解くための手段かもはっきりしていないのに、「ゲノム科学は情報科学である」と言い切ったのですから、具体的なことは何もなかった。ただ、高木さんとよく話していたのは、当時の情報生物学の議論が、あまりに浅いのではないかな。そこで、とにかく、浅瀬で騒がしく、生き残る為だけの活動を封じて、「静かに、落ち着いて、深く考える”場”が必要ではないか」。そういう場に相応しい人を見つけて育てようと、よく話していたのを思い出しました。

高木:報告書の中に場を作ることが書かれているんですね。勝木先生は当時ずっと、そういうものが育つ場を作らなければいけないとおっしゃっていました。そのころは大学があてになりませんでしたから。

勝木:“松原”適塾がヒントになりましたね。

松原:国に要求しても駄目ですよ。やれる人とやりたいことがマッチする場ができないと。僕は国際高等研がよくあれだけの協力をしてくれたと思いますよ。

勝木:それは松原さんの熱意からじゃないですか。

松原:ちょうど僕は副所長だったんです。だけど高等研はあの後、あのやり方はいいということで、合宿や勉強会もやっています。僕が高等研に一番貢献したとしたら、あそこの立派な設備を泊まり込みで交流の場に使うというコロンブスの卵みたいなことでしょうかね。日本の今のバイオインフォマティクスの元祖みたいな人が、だいたい全部あそこにいましたよね。駄目になった人は一人もないんじゃないかな。

勝木:錚々たるメンバーですね。

松原:今になってみればね。当時はやっぱりそれぞれの先生は結構自分でこういうふうにやりたいんだけど、知名度もない、学生をとる場もない状況だったんです。そういうところに喜んで来てくださったということもあったのですね。

勝木:バイオロジーに情報学の人たちの目を向けさせたというのは、

今思うとすごいことだと思います。今だと東大の森下さんがそういうふうにいるというけど、大学院生はあそこに行かないそうなんです。バイオロジーの教授のポストに情報出身者がまだいないので、キャリアパスにならないと。これは学問を楽しむというものを越えた話なので、新しいものを作ろうというBIRDが、よく実現できたと思います。あの当時の人は、そんな計算がなかったんでしょうね。

松原:先生は今より貧乏だったけど、みなさん若かったし、何かいいこと、おもしろいことをやりたい、そういう気合が入っていました。

吉田:研究者全体としても夢があったと思います。一人ひとり単独の夢というのではなくて、全体として、あるいは時代の夢ですね。ゲノム解析をやるかやらないかと大騒ぎをして、やることに決めてやった。ところが、百年もかかるといわれていたのがあっという間に全配列が決まったでしょう。山ほどデータが出てくる、さあ、どうするんだと、やっている人も横にいる人もそう思っていました。BIRDは必然的に生まれてきたような面もありますよね。そういう時代の夢を背景に。考えてみれば、自分たちの仕事は多かれ少なかれ、ゲノムに関わっているので、ゲノム解析からどういうものが見えてきて、どうするんだとみんなそう思っていた背景がものすごく作用したと思いますね。

松原:まさにサイエンティストとしての、一番わくわくする夢があった。

吉田:誰もどうしていいのかわからない、でも何かが来る、そういう時代だったんですね。

勝木:その後、松原さんに聞いて印象的だったのは、バイオロジーにインフォマティクスが入った重要な論文がいくつか出始めたときに、こういう人を育てなければいけないとおっしゃっていたんですよ。要するに、バイオロジーの良い論文に、ゲノムが出てくる、しかも新しいものには必ずインフォマティクスのすごいのが付いているねえ、と。言い方を代えれば、そういう人を育てていかなければいけないあと。始めてすぐだったんです。それでドライとウェットの融合体を作ろうというのが一つのアイデアでした。しかし、ドライとウェットがこれほど水と油のように境界がはっきりした混ざり合わない分野とは思っていませんでした。考えてみれば、ゲノム配列が判ってどうしようと考えていたの

か不思議な気持ちになりました。しかし、その中から、お酢を混ぜて境界をなくして、ドレッシングにするのが、このBIRDのなかから何人か出てきた。それがこの一番面白い成果だと思います。

私自身が非常にアナーキーな性格なので、こういう学問分野を新しく作るのに最初から組織的なこととか、テーマの妥当性だとかを取って一切考えない。とにかく直観でいいから提案したものと、才能のある人を選ぼうと思っていました。

吉田:それをはっきり言っておられましたよ、先生は。人を育ててはいけない、人を育てるんだとおっしゃっていたのをよく覚えています。

勝木:先ほどの適塾じゃないですけど、育てる以外に人がいない、いないから逆に人を育てるチャンスだと思ったんですね。ただ、それをうまく誘導しなくちゃいけないし、場を賑やかにしなくちゃいけない。それについて私はほとんど貢献してないんですけど、高木さんが貢献してくれた。私は不幸にして基礎生物学研究所の所長になっちゃったから、距離的に遠くなったんですね。でも最初の仕掛けが良かったらしくて、最初に応じた4人は面白いと思います。時系列でいいますと、BIRDができるころのゲノム科学委員会と情報生物学適塾の話が一番の肝だと思いますので、何かこのことで、一言二言ありましたらお願いします。

松原:みなさん、俺が俺がと言わない人たちだと思うんですよ。だけど、こういう大事なことはやりましょうよ、という人たちでしたよね。阪大の細胞工学センターにいたころからずっとそれで通しましたから。俺が俺がという人は適塾をやるときからほとんど排除されましたが、幸いに、このバイオインフォマティクスをやる人には、ほとんどそういう人がいなかった。みなさんも大昔にアリューシャン列島を渡ってアメリカ大陸にきた後のアメリカ先住民の先祖と同じですよ。けんかする必要がどこにもないわけ。目の前に広々とした大陸があって、どうやって生きていくかのほうが重大なんです。時代も良かったんじゃないですか。

勝木:たぶん、そういう人でなくてはそこにも集まらないんですね。でもよく見つけましたね。

松原:今だったら毎日高木さんのところに行って、この人どう、あの人どう、とうるさく聞いていたでしょうね。当時高木さんはまだ若かったから。金久さんをはじめとせずいろいろな人に聞きまわりました。

勝木:やっぱり時代もそうですし、分野もそうですけど、人ですね。

松原:もう一つ、ヒトゲノムの立ち上げから世話した人がみんな熱心に協力してくれたんですね。あそこでこんなことをやっている、など熱心に言ってくださいました。

吉田:松原先生は大変な熱意を持って努力されていましたから。日本のゲノムがあるのは松原先生のおかげですよ。

勝木:困難な時代もすごくあったんだろうなと思いますね。

松原:だから腕まくりして(笑)。

勝木:腕まくりに関しては、いろいろおもしろそうですが、番外に回しましょう(笑)。

吉田:人を育てるということは非常に難しいですね。人を育てるのにお金を使うと勝木さんがおっしゃったときに、ものすごく納得した記憶があるんです。たとえば医科研でヒトゲノム解析センターを立ち上げたとき、生物情報をコンピューターでやれる人を呼ぼうという話になったわけですが、誰が良いか分かる人が一人もいなかった。他所からこの人が良いと紹介されたら、その人を選ぶ以外はまったく手がない。つまり、バイオロジーの側には人を選ぶことが出来なかった。そこで、人を育成しなければとなるわけですが、バイオインフォマティクスやゲノム情報を教えるということができる人がいなかった。

そこで、両者を同じ釜の中に入れて込んで見たのですが、コンピューターをやっている人は半年もするとコンピューター屋に戻って生物学者がまったく分からないことを口走るんですね。そういう状況でしたから、両者を跨ぐ新しい着想を持った若者がいても、それを評価し、そこにお金が出るというシステムがなかったんですよ。お互いに理解できない、そこを評価するシステムが日本にはなかった。そこにBIRDが出来て、遺伝研や京都大学をサポートするだけでなく、半分は人を育てるといわれたので、僕はこれだと思ったわけです。現在の人を育てる難しさとは質の違う難しさがあったんですよ。

勝木:それが一番面白いですね。答えは探すものじゃなくて、答えは創るものですからね。今、分子生物学は特にそうですけれども、短期的な競争社会になっていることにもちょっと関係ありますが、学生はみんな既存のパラダイムの中から答えを探すんですよ。点数が取りやすい。やっぱり場は大事ですね。いいチャンスが私に回ってきたので、点取り主義を排除するアナーキーなことができた。普通だと点数を気にします。しかし、おもしろそうで、自由に生き生きしている人を選ぶことにしたから良かったのかもしれない。当時のJSTの担当者は、私がかかりアナーキーで暴れまわるのを、口をあんぐりとして茫然と見ていた。私のほうには来なくて、高木さんにたらく当たっていたようです(笑)。私はその状況を知りながら「美は乱調にあり。諧調は偽りなり」でやっていましたけれども。

松原:勝木さんと高木さんはいつでも二人ペアで組んでいて、非常にいいなあと思っていました。

勝木:そうですね(笑)。

松原:暴れまわるのの後始末と、いつも一緒になってますよね。高木さんが暴れまわることはまずないから面白い。このペアは最後にどこにいくのか、僕はよく分かりませんでした。始めはとにかく、おとなしくしていると何も起こらないという意識が非常に強く出ていたし、特に暴れたり収めたりしていた時代ですからね。



BIRD第Ⅰ期(平成13年～平成17年)

勝木: (応募者の名簿を見ながら) それで第一期に入っていくわけですが、様々な分野から、生物学を目指すとはとても思えないものから、すでに小さく完結しているものまで、たくさんの応募があって、その中から選んでヒアリングをしました。今から考えると、ずいぶんバラエティに富んだ人がいますね。このときはまだ若くて実績もアイデアもなく、という人もいました。少なくとも第一期で、関心を持った人が100人以上もいたということです。科研費の分野でもないし、コンピューター系の話でもない。松原さんがおっしゃったけども、インフォメーション・バイオロジーというバイオロジーに主体がある情報学を手法とするものは少なかった。後に、高木さんから教わったのですが、バイオロジーになるには、情報が生物由来であるだけでなく、その情報が時空間的に制御されている生物現象のメカニズムを計算で明らかにする、新しい学問としてコンピューティショナル・バイオロジーという言葉を紹介されて、これが、私には非常に新鮮で、ゲノムとバイオロジー、情報とバイオロジーの融合や創造ということが腑に落ちたように思いました。

それで第一期の肥後さんは、タンパク質構造を、第一原理から高速計算機を使って決めようとしている人で、これは本当にオーソドックスなタンパクの構造の方です。

松原: 高木さん、この頃の高速計算機ってどのくらいのレベルなんですか。

高木: 今は本当にパソコンや小さなサーバクラスではないでしょうか。十年前ですからね。

勝木: 計算のソフトと一緒に組み込まれたMDと言うハードに投資していましたよ。五つくらいのアミノ酸をやっただけで、水の影響やらが

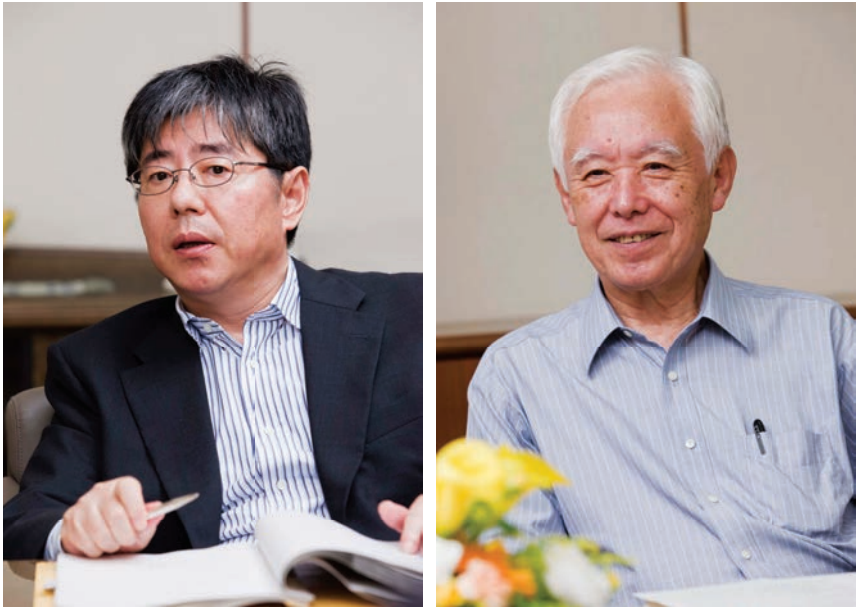
計算に含められないから、みんなつまこんだんです。結局三年で終わりましたが、当時は最先端の研究領域だったんですね。

岸野さんは数学のモデルでして、発想は極めておもしろかったのですが、バイオロジーとはちょっとかみ合わなかった。矢田さんは、まさに当時の正統派の遺伝子のデータベースを作るという方です。ある塩基配列のルールに従ってゲノム配列のなかから遺伝子を見つけ出すための計算手法を提案された。私が不満だったのは常に94か95パーセントなのに、正解率は世界一だと。

高木: ゲノム上の遺伝子領域を、コーディング領域を予測するんです。

勝木: その計算手法を使って100%になるのか、もしかしたら9.5合目に渡れないクレパスがあって、実はそれ以上は不可能かもしれないというのが私には不満だったんです。別の登山道か、新しい道を作って登らないと頂上には到達できないんじゃないかと。そのためには何かゲノムのなかに別の法則を見つけ出さないとなかなかそうはいかないのではないかと思ったんです。

一方、ドライとウェットとを一緒にやるという融合のほうですが、これは本当に面白い人たちが育ちました。ここで採択された人たちが思いもよらない人たちだった、ということがまず一つ。大浪さんは今も線虫をやっておられる理研の方ですが、北野宏明さんと一緒にやっていた。企業の人と組んだ伊藤隆司さんは金沢大学でした。森下真一さんが組んだのは酵母を研究していた東大の大矢禎一さんです。彼らはすごいなあと思うくらいエネルギッシュでしたね。伊藤啓さんは、当時基礎生物学研究所にいて、堀田研(東大生物物理)の出身なんです。ショウジョウバエについての神経やグリアのネットワークを徹底的に解析する、国際的に非常に重要な人物として評価されています。この五年間



きわめてユニークな、いい仕事をしてくれました。

育つという意味でいうと、ユニークな人が育ったなと思いますね。選ばれた人達が良かったので、自由にさせておけばいいという感じでした。適塾の人たちは、ここには出てきてないですね。

松原: 森下さんは適塾の先生ですね。

高木: 落ちたメンバーも錚々たる方々ですよ。

松原: 落ちたメンバーを見ると、頑張ってる人がずいぶんいて、申し訳なかったなあと思います。

吉田: この応募のタイトルを見ると、この時代はまだバイオインフォマティクスというのがよく理解されていなかったんだなあと感じますね。タイトル見ただけで、これは違うっていうのがいっぱい並んでいて、すごく時代を感じますね。

勝木: うまくいった人たちを見ると、みんな相棒がいいんですよ。さらに、森下さんは、第二期にも採択されて、メダカの武田洋幸さん(東大動物学)と組んで、素晴らしい成果を上げました。本人はさして喜ばなかったけど、見事な生物学者になって、おめでとくと申し上げたほどです。伊藤啓さんは一人でやっていましたが。

今、当時の応募者を見ると、まだインフォマティクスにならない、データを集めたり、バイオロジーだけのものが多かったですね。ただ、うまい選び方をしたので。しかも、面白いことに文科省の都合で平成14～16年は採用していないんです。当時は何故やらないのかという声もありましたし、私もやるべきだと思っていると折衝はしたんですけども、予算がつかなかったか何かで、三年間募集をしませんでした。これがかえて良かったかも知れない。最初のを徹底的に見ることが出来、ある程度形ができて、もし翌年新しい人を選べば、同じような応募が来たでしょうから、薄まっていたかも知れない。良い方に取った場合の話ですが。

これが第一期の経緯ですけども、この三～五年後の成果については、高木さん、どうですかね。感じとしては大成功という印象を、私は受けましたが。

高木: この融合の方々は確かにみなさん大活躍していらっしゃいますね。

吉田: この分野の評価には、特別な難しさがあると思うんですよ。通常のサイエンスでは、何かを発見する、何かを明らかにするという絶対的な成果が残り、それが評価される。しかしデータベースとか解析法とかは、次々と更新され、あるいは新しい方法に置き換わって行きますね。そして残りませんよね。しかし、今、残ってない、今、誰も知らないものでも、その時代には大きな意味があり、次のものが生まれるためにものすごく貢献したかもしれないという歴史的な視点から検証する必要があると思います。これがなかなか難しい。BIRDの活動が、統合データベースセンターにどういうふうに組み込まれて、残っているかを見てみる必要があると思います。

高木: まず具体的な成果の前に、先ほどの人を育てるということについてですが、たとえば森下さん、彼はもともとコンピューターの人で、今は自分でバイオロジーができるんですね。極論すれば生物学者と一緒にやらなくてもシークエンスを外注すればできてしまうんです。一人でできるようになった。伊藤隆司さんは、彼はもともとウェットですが、ほとんど自分でバイオインフォマティクスができるんです。ですから彼は自分のラボの中でデータベースを作ったり、解析できるまでになっているんです。大浪さんは小原さん(遺伝研所長)のところの出身ですが、彼も、もともとはウェットなんだけれども自分でバイオインフォマティクスができる人間になっている。そういう意味ではパイリンガルというか、ここで選ばれた四人はそれぞれ、伊藤啓さんも、もともとは実験手法だったわけですけども、今ではきちんとデータベースを作ってそれを解析するという方向に来ているので、みんな勝木先生の意図したように人材が育ってきています。二人組で始めたけれども、今は一人で全部完結できるようになってきたということが、人を育てるという意味では大きな成果かと思っています。

勝木: しかもバイオニアとして一流の業績を挙げている。四人とも当たりましたね。インフォマティクスのほうからバイオリジストが出たと



思っていますけど、伊藤隆司さんは逆ですね。

高木: 大浪さんも逆なんです。

松原: そしてようやくバイオインフォマティクスを教える人になったんですね。そう考えるとコントリビューション高いですよ。

勝木: しかしその人のところでは、新しいキャリアパスはまだできてないから、そこに行って何かをやるうという大学院生が極めて少ない。むしろキャリアパスの正道と見なされる情報なり生物学のプロパーに行きたいわけです。それで森下さんや伊藤啓さんなどが、大変なんです。

吉田: その考えは完璧に間違っていますね(笑)。実際にそういう人は企業でも要求されているんですけどね。

松原: ちょっと違うんだけど、場所が本郷にあるとか、そういうもっとつまらない理由もあるのじゃないかと。高木さんもね、柏に行ったらもっと大変だったと思います。

高木: いや、私は柏ですよ。本郷には学科を作りましたけれども。

ちょっと話がそれますが、本郷にあるということも影響しているのか、学科になったら進学振り分けの点数がものすごく高いんです。ですから、生物情報科学科とありますが、非常に学生の人気が高い。柏はなかなか難しい。ただ先ほど勝木先生が、答えばかりを探してしまうとおっしゃっていましたが、今の若い人の代弁をすると、今は助教でも任期付きなんです。だから三年とか最大五年くらいしか任期がないんです。ですから、おもしろそうでも時間の掛かる融合研究はできませんし、すぐに答えの出る論文を書かないと次のポジションがないんです。われわれの時代の助手といえばパーマメントの助手ばかりでしたが、現在はそういうことがないものですから、プロジェクト雇用で成果を出さないと三年で首を切られてしまうのが現実なんですね。そうすると勢い、バイオインフォマティクスなどの新しい分野をやらなければいけないとなると時間がかかりますし、やっていられないと。だから、自分の専門分野ですぐに論文が書ける、成果が出るものになってしまう。もっというと、今の若い人は卒論で研究室に入ったらマスターも同じ研究室に行くんです。なぜかと思ったら、学振のPDを取ったり、DCを取ったりするためには、論文一本書かないと取れないから。四年生の卒論でやった研究を、マスターの一年生で何とか頑張って論文にして、それでドクターコースのときに学振の奨学金をもらう、それをみんな狙ってますから、本当に脇目も振らずにいくわけです。だから先生方がお考えになっていることはなかなか実現できないんです。

勝木: そうですね。情熱のありようが違う。新しいものを作ろうという、こんな楽しいことは貧乏でもいいという時代と。食えない、死にそうだというのがみつかざるを得ないという現状がね。

吉田: 若い人たちを責めるというよりはシステム、体制だと思いますね。

勝木: そうですね。こういう科学研究に必要なのは競争的資金ではなく、パトロンのお金ですね。スポンサーになって結果を要求するんじゃない駄目で、「知を愛する人を愛する」パトロン方式でやるべきだと思うんですよ。

松原: インフォマティクスは、まだ技術として扱われている部分が大部分じゃないですか。だから技術としては、たとえば製薬企業に就職すると、生物はどうでもいいよと、コンピューター使えればいいよといわれる。大学でいうと産総研なんか行くとやっぱりインフォマティクスで論文書かないとドクターが取れない、バイオロジーとインフォマティクスの融合でやるとドクターに認めてもらえないから、情報学はできてバイオとの融合となると時間がかかる、そういうところが自然ではないですよ。

勝木: 第一期は少なくともそれを打ち破るような、そこで育った人たちが教授になるなりしていますが、そういう点で苦労されている現実があるということですね。

勝木 元也 Katsuki Motoya

基礎生物学研究所名誉教授、日本学術振興会学術システム研究センター副所長、
自然科学研究機構理事
JSTバイオインフォマティクス推進センター統括(平成13～23年度)



松原: 十年の間に日本でどういうエポックメイキングな仕事が出てきて、国際的にはどうなのか、その流れの記録を作ったらだいたい変わるんじゃないかと思いますね。それがどういうふうに波及効果があるか。今苦労に苦労を重ねているような人が教授になって、先ほど高木さんがおっしゃっていましたが、いい人を育てたいけれども寄って来ないような人がどうだったのか。何が問題だったかではなく、どういうことができたのか、そういうものはあったほうがいいのではないかと思うんです。

勝木: それにこういうアナログな議論が結びつくと。

松原: 僕らの専門分野である分子生物学だと、どういう良い技術開発をしたのか、どういうコンセプトを出したのか、それがどういう流れの中で出てきたのか、見れば分かるじゃないですか。でもこれはね、なかなか分からない。

勝木: 直観的にすごくいいなと思うけども。

松原: 流れがつかめるようになるといいなと思いますね。

勝木: それは非常に良いご提案なので事業評価の中にぜひ入れたいですね。

松原: でも高木さん以外にできないから、高木さんが苦労すると。

勝木: そこは高木さんに頑張ってもらいましょう(笑)。

吉田: こういう分野は論文にならないと前はよく聞きましたが、論文にならない時にインターナショナルにどう評価を受けるかということ、どうやって判断するのかということはずごく難しい。

勝木: 人を育てるというコンセプトでやったからできるけれども、統合データベースのような枠組みでやると、人を育てるというよりも、モノを、情報をたくさん集めるということになるから、そういう意味では使い道や業績がはっきりするわけです。しかしこういう新しいものが一個出たというだけではなかなか。まあ我々は人を育てて、その人がまた誰かを育てていくということがあるので大成功だと思います。そのとき新しい人たちが入りやすくするために、第一期の人たちはこういう素晴らしい、ほとんど誰も手をつけなかったことをいくつもやっていると思うんですけど、それが本当に大事だと思いますね。

松原: 何か運に恵まれたね。どんどん増殖して今非常に栄えているのと、運が悪くて終わりかけているもの、そういうものに対して正しい評価と、周りからの影響というのが分析できると良いと思います。一生懸命みなさんやってきたわけですし。

勝木: それはぜひ高木さんお願いします。高木さんにとっても統合データベースをやるときにすごく参考になると思いますので。

高木: 統合データベースはここを出ているようなコンセプトとは違って、新しい生物学の分野を作るとか、そのための人を育てるということではないので、代わりになるものではないです。

勝木: さて、データベースの高度化の話ですけれども、データベースは四つ。遺伝研と金久さんと中村春木さんと高木さん。四つのグループがゲノムの議論から出てきて、ある程度安定していて、これから継続しなくちゃいけないと。そしてこの人たちが当時本当にびっくりするような綱渡りのような短期的な資金でやっていた。とくに中村春木さんはひどい状態だったのですが。

松原: 生き延びたのが不思議なくらいですよ。奇跡です。

勝木: 私のところに、こんなものを出すから読んでくれと論文集が来たんです。短期間だったので、結局は十分に読み込めませんでしたけれども、熱意は非常に伝わってきて、それを選んだことは、私は良かったと思います。彼ほどメタモルホーシスした人はいませんね。芋虫がさなぎになる、さなぎから脱皮して、ちょうど蝶々になるぐらいのときにBIRDで採択して、一挙に国際的な日米欧三極の中心になったのは、いってみればBIRDのお陰だと中村さんはよくいってくれます。

高木 利久 Takagi Toshihisa

東京大学大学院新領域創成科学研究科教授、国立遺伝学研究所生命情報・DDBJ研究センター教授、
JSTバイオサイエンスデータベースセンター副センター長
JSTバイオインフォマティクス推進センター副統括(平成13～20年度)



BIRD第II期(平成18年～平成23年)

勝木: 少なくともデータベースというものの感覚は生物学者にとって真っ当になってきたというのが第二期だと思います。そうこうしているうちに、BIRDは五年が研究開発期間の区切りなので、十年目で一旦、研究開発が終了になるということが大問題になりました。高度化のほうは切れると困ったことになるので、だいたい二年ぐらい前に継続するにはどうしたらいいかという議論が始まり、それが第二期から統合データベースへ向かっていったというのが流れです。

第二期のことでいいますと、融合のほうも創造のほうもやっぱり少しルーティン化してきたと感じました。第一期から見ると広がりが少なくなって、まともに、分かりやすくなってきたんですね。これだけ良い仕事をしているのに申し訳ないけれども、最初の人たちに比べれば「So, what ?」(だからどうした?) というところがすごくあって、課題評価のときも質問が小さくならざるを得なくなっているというのが私の偽らざる感想です。逆に、ちょうどいい止め頃だったかもしれませんね。一度リニューアルしないと最初のスピリットが失われる可能性があるな、と統合データベースに引き継ぐ議論のときに思っていました。高木さん、どうですか、何か。

高木: 第二期はデータベースの高度化はともかくとしましても、いろいろな研究開発課題のところがちょっと小粒になってきましたね。解けそうな問題を解こうというような感じというんでしょうか。問題がはっきりしてきたこともあると思うんですが、ゲノムもいろいろ決まって、どんな問題を解かないといけいいのか、というようなことで、まったく新しいものを切り拓く感じはなくなってまいりましたね。また、人が育つという意味でも第一期のような大きくブレイクするような感じはいまのところないですね。

勝木: メタゲノムには大きな期待をしています。これこそ暗闇の中に、インフォマティクスという光を入れて、生物の存在を予測する探検的な学問だからです。しかし、まだ準備段階のような感じが少ししました。これは大きな期待があることの裏返しです。見えないほど大きな展開が期待できる鍵を掴んで貰いたい。

松原: 問題が細かくなってきたんですね。どれを選ぶかといったような発想のものが多くなってきた。でも大きなグループはそんなに多くないじゃないですか。高木さんのいうように、ちゃんと堂々と分野のリーダーとして裾野を広げてやってくれるような環境をなかなか作れなかった。言い方は悪いけれども、中途半端な予算を使いながら一部でバイオインフォマティクスをやっている、そうすると何かのデータベースを作るとか、何かの関連を調べるとか、割合やればできることも増えてしまったんですね。

勝木: 第三期はないなと思っていましたが、第三期やるならみなさんとご相談して、大幅にリニューアルしようと思っていました。今までは、どういう人か初めは分からないなりに直観で選んだ人が自然に育ってくれました。邪魔しなかったから良かったのかもしれませんが、この積み上げた上に新しい人を育てるために、今度は少し伸び伸びさせて、適正規模を計りながら、巨大化し超領域化する情報の粹を少しでも突破するようなバイオインフォマティクスの問題設定が必要だろうと思っていました。

松原: 日本のサイエンスの一般的な傾向だと思うんですけど、後ろ向きに進むのが得意な人が多いのが難しいですね。自分はこういうことをやった。その結果を使って、こういうことを作るんだと。だけど、過去の成果を強調していると、結局、後ろ向きに歩いて行くものだから、自分のやっていることが本当に良い方向に行くかどうかは振り向かない限り分からない。後ろ向きですから、やることは必ず成功する。多分とんでもないことはやらず、やればできることのほうに、とい

吉田 光昭 Yoshida Mitsuki

公益財団法人がん研究会がん化学療法センター所長、
東京大学名誉教授
JSTバイオインフォマティクス委員会委員(平成13～23年度)

う人がどんどん増えてきた。そんな感じがします。

ある会議などに出て話を聞くと、自分たちはこういうことができた、だから次はこういうことがしたい、たとえば今までできなかった結晶が膜の中でできるようになって、相互作用の研究がもっと細かい水の分子まで含めた計算ができるようになったとか、いろいろな相互作用の、今までできなかったもの、だからこういうことをしたい。それはサイエンスとしては正しいですよ。だけど、今の日本の科研費の状況で、自分たちがそれだけの範囲内で考えて、新しい提案をしたら本当にそれは正しいのかという判断がその議論の中にある。それを後ろ向きに進んでいると言うのです。これからどういふにやらなくてはいけないか、どういふことが自分たちのためにも日本のためにも、おそらく世界のためにも必要なのか。たとえば今の日本の社会だったら、もっと医学系の人の意見を聞きなさいと僕は思うんです。薬はもうどうでもいいから、ほっといても彼らはやるから、医学系の人の意見を聞きなさいと強く言いたい。ものを考えるときに自分らの成果から考える、そうすると細かいものがいっぱい出て、やればできることの提案になる。

勝木: 高木さんの先ほどの分析ではないですが、そういう環境になると、若い人がそういう問題でしか生き残れない。

高木: 少なくとも今みなさんそんな感じですよ。次のポジションのことを考えている。

松原: ぶっ壊す人がいてもいいと思うんですけどね。

勝木: そう思いますね。

松原: 五年間くれたらぶっ壊すけど、五年間でできなかったら僕は消えてやるという意気込みが見られない。

BIRDの精神を次につなぐには

松原: この次にバイオインフォマティクスという旗を掲げる人を選ぶとき、バイオインフォマティクスができて、自分はこういうデータベース作ってこういう方法論開発して、だからこれだという以上のものを言えるような、将来こういうことが必要だから今からそれを始めようという人が出てこないものかなと思いますね。どういう流れで動いていくかというものが、国際的にはどうなのか、というような。

吉田: 今日の話題とは別に、それはものすごく大事な問題ですね。ある意味では今の若い人たちが良くないというよりは、研究分野、サイエンスの進展の度合い(状況)が反映されているような気がするんですね。例えば、個人の力を超えたゲノム情報処理というものは、お金があっても自分単独ではできないわけです。連携が必要になるのですが、日本ではなかなかうまくいかないという中で、夢を持つチャンスが少なくなっているように思います。新しい技術なり、新しい現象の発見があって、そこから更に新しい世界が広がるということを示すことが出来れば、今の若者でもチャレンジするんだと思いますよ。

松原: 自分が与えられた、閉じたかまくらみたいな中でやろうとするのとは違う。

吉田: 今はそうなっているんですね。それをお金のせいにするんだけれども。

勝木: ですから、十年間やったBIRDでは“人”に注目したことで、ある程度は“人”が育ってきました。ドライとウェットでやるやり方は一つのいい方向だったけれども、今度はCRESTやさきがけという人たちの大將に、高木さんと森下さんとか場合によっては伊藤隆司さんとかが固有の領域を展開して、そういう人を育てるというようなプロジェクトに変身していくほうがいいのかなと思ったりもします。

松原: こういう感じで個々の人がやるのは、それなりにそれぞれ成果が出るからいいけど、大き

松原 謙一 Matsubara Kenichi

株式会社DNAチップ研究所取締役名誉所長、
大阪大学名誉教授
JSTバイオインフォマティクス委員会委員(平成13～19年度)





な流れになるかという、この中からだけでは出せません。たとえば高木さんが牧場経営的に大規模にやるのは良いと思う。CRESTなんかはとてもいいですね。

勝木: BIRDにこだわるんじゃなくて、BIRDの本質はね、そういうものを開拓した人を大将にして、牧場を。

松原: 自分の境界があまり強固でない人。

勝木: そうですね。放牧場。

松原: 高木さんは牧場をやったらいいなあとします。パイオインフォマティクス牧場。

勝木: 高木さんは数年前から統合データベースを背負っているから。この話はのちほど。

吉田: 統合データベースからどうやって新しいものを作っていくかというのは難しい問題です。ローマ人の言葉に『何事も最初は良かったのである』という言葉があるんですね。最初はいいんですよ。でもある時間が経つとそれは良いもので無くなるし、場合によっては害になることもあるわけです。BIRDはかつては大いに機能した。でも今やいろいろなことが分かっているいろいろなことが考えられる時代になったから、この体制ではBIRDは機能しないんだと考える必要がありますね。新しいメカニズム、牧場っていうのはどんなものにするかというのは非常に難しい問題ですよ。

松原: 今の原子炉の問題と同じで、現状を良く見ないで手当てをしようとする必ずロジックで陥るあるパターンがあるんですよ。半素人で議論して、良さそうなものの組み合わせで物事をやらそうとすることは、絶対にやってはいけな。上から考える発想が今横行している。

勝木: それもまたチャチだね。

松原: 組み合わせというのは一見いいですよ。ボーダーレスになりそうな感じがします。組み合わせで上からやらせようすると目的を設定して、この目的というのは社会のための学問になって良さそうだ。ところがそれ以外の自由度がゼロになってしまいます。

吉田: ゲノムを始めたとき松原先生が非常に推進派でしたが、僕はゲノムをやったってしょうがないという考えだったんです。しかし菅野先生(元癌研所長)が、ゲノムをやったからってすぐ役に立つと思うからそう言うんだ、日本がゲノムをやらなかつたら将来世界でどういう位置づけになるか考えろと。どうしろはなくて、そこでほんと放り出されました。そういうことを思い出すと、インフォマティクスは、普通は必要性や重要性は当然として、新しいインフォマティクスで新しい世界をどうやって拓くかは次の世代に任せるフェーズに入っている気がしますね。難しい問題を投げ出すようで無責任な気もしますが、既成の人が考えたんじゃ駄目なんですよ。若い人たちが、自分たちでやらないか、と思う状況を作ることが必要なんですよ。老兵は去るべし、と自分に向かって言っているわけだけど(笑)。われわれはそういうフェーズに来てるんですよ。

勝木: それにはね、自由に使って冒険が出来るお金が必要です。それが出来た先生方は、リアリストだったんだと思います。パトロンでないと、そんなこと言えません。だからCRESTでもいいし、新しいシステムでもいいですから、そう言える人に裁量権を渡して、託してみるんです。

松原: だから牧場経営ですよ。

勝木: そういう意味では試みとしてのBIRDの最初のころ、少なくとも時間経過をたどりながら、実態を理解したうえで、成果がどうだったかということを引きちんと評価してもらえれば建設的な次の展開が議論できと思います。事業としてはある意味では同じ立場にはもう立っていないし、次のフェーズに入るべきです。牧場みたいなパトロンのオーナーを作るシステムにして。

松原: 集まった人たちの顔ぶれを見て、どういう成果が上がるんだろうかというのはやればできると思うんですよ。だけどそういう中から今、話したように、やればできることだけではなく、日本はこれからどういう具合にやったらいいか、というイメージを作って提案できたらいいんじゃないでしょうか。今は薬に役に立つようなバイオインフォマティク

スセンターを作れとかそういう話がありますけども、次の世代にあるべきものは、漠としたものでいいと思うんです。やればすぐ見える成果が出るということばかりをそこに集めてもしょうがない。

勝木: 次の檄文を書くことを高木さんに託しましょう。第一の檄文は私が書いたつもりだから(笑)。

松原: ここに名前が連なった人の意見を聞くだけでは駄目で、インフォマティクスの外にいる、たとえばお医者さん、理学研究者や生物研究者の心ある人たちの意見も聞かなくてはいけない。

勝木: ただ、今の段階でうまくやらないと小さな意見に集中してしまいますね。

松原: それは、最後の責任はこちらで取らせていただきます。

勝木: 最近ステレオタイプな議論に圧倒されるようなことが多くありませんか。いろんな審査会とか審議会とか。見識は、みなさん高いんですよ。成績もいいんです。だけどアナーキーでないから、自由度が少ない。サイテーションがどうか、あんな馬鹿げた他人が作って出版社の商業主義でやってるものをなんで参考にするのかといつも思うのですが、ものすごくレベルが高いと思っている人たちからステレオタイプなレベルの話が出てくる時代なんですね。これは若者の持つ創造性を一挙につぶしかねませんからね。創造性は人にしか存在しませんから、多様で多面的な人の属性を見つけるこそこそがすべてなのに、成果ばかりが強調されると、新しいものへの挑戦が出来なくなりますね。今それが心配です。

松原: 世の中の流れがそういうふうになっているから。でもどういう話をしたんですかというときにこれを見て、こういう具合に成果が出てきたから次にこういうふうにしましたと。それだけで固めてしまうとまずいと思う。そこに見識が入らないと。

勝木: 新たな適塾です。

松原: でも今のCRESTのやり方を見たら、どうやって決めるか僕は知りませんが、始めに一番適切な人を決めて、その人が四、五匹の馬



を育てるわけね。しばらくすると、もうちょっと馬を増やしましょうというように、だんだん育ってくるわけです。終わりぐらいには十数匹のいい馬ができてというように、やってみながらというやり方がいいやり方になっていると思います。

勝木: CRESTというよりもさががけですね。

吉田: 何か新しいやり方、中身も形も違うものを作ったらいと思うんです。JSTの責任として結果は問わないと言わんばかりのものを作らないと。

勝木: それはそうですね。

さががけのほうは今日聞いてますます自信を深めたんですが、人を育てる方向は統合データベースには無理だから、しかしこんなに大事なものを失ってはもったいない。今日話したようなことがあるので、若者に、真に新しいものに挑戦できる場を作り続けてほしいと思います。その時のキーワードは「成果」ではなく「自由と創造」であってほしい。

吉田: 最後に檄文を書いてもいいんじゃないですか。最初と最後に檄文ありきと。

勝木: わたくし? 私の役割はもう終わりました。私の檄文は空振りになりますよ。老兵は、ただ消え去るのみです(笑)。

本日はお集まりいただき、ありがとうございました。

座談会はこれで終わります。

バイオインフォマティクスとは

バイオインフォマティクス(bioinformatics)は、コンピューターサイエンスと情報技術を生物学・医学の分野に応用した科学技術である。その第1の目的は、ゲノム情報をはじめ、現代の計測・計量技術によって得られる生物を構成する成分、構造、また機能情報を整理し、実際の生物現象と対比させ、無秩序に見える情報データの中から、生物学的な知識を発見することである。ここで言う生物現象とは、進化多様性の解析のようなマクロなものから、生体分子間の相互作用の予測といったミクロのものまでを含むが、バイオインフォマティクスは単に生物現象の解析に止まらず、現象の底にある問題の発見的手法であり、それは生物学者に新たな問題を課するものとなり、新たな学問領域を開くものと期待されている。

1980年代の終わり頃から、ゲノム研究と遺伝子研究においてバイオインフォマティクスという言葉が徐々に使われ始めた。このころ、ヒトゲノムの全塩基配列を解析するというヒトゲノムプロジェクトをはじめ、様々な生物のゲノムプロジェクトが開始された。当初、バイオインフォマティクスは、ゲノムプロジェクトという膨大なデータを産出する実験を支援する位置づけであった。しかし、ゲノムプロジェクトが進むにつれ、蓄積されたデータから規則性や法則を見出し、知識を抽出する試みがなされるようになった。

バイオインフォマティクスの開発は、遺伝子配列比較、遺伝子予測、タンパク質構造比較、タンパク質構造予測などを主な研究対象としており、ヒトの遺伝子治療研究や乾燥に強い農作物の研究開発などに役立っている。iPS細胞(induced pluripotent stem cells、人工多能性幹細胞)の作成においては、遺伝子データベースが重要な役割を果たした。生物全体を1つのシステムと捉えるシステムバイオロジーや、ゲノム情報に基づく創薬などを扱うケモインフォマティクスもバイオインフォマティクスの開発対象となってきた。近年、短時間でゲノム配列を読み取る次世代シーケンサーから産出される遺伝子配列データや、核磁気共鳴画像法(MRI)から産出される画像解析データなど、取り扱うデータ量は飛躍的に増大しており、バイオインフォマティクスの重要性がさらに高まっているとともに、スーパーコンピューターの利用が急務となってきた。

バイオインフォマティクスの開発には、分子生物学や細胞生物学などに関する知識とともに、データベースやプログラム言語などに関する知識、すなわち生物学と情報科学の両方の知識が求められる。バイオインフォマティクス推進センター事業では平成13年度(2001年度)から、「創造的な生物・情報知識融合型の研究開発」として、情報科学と生物科学との融合したアプローチにより、多彩な生物情報から生物現象の原理や法則を発見し体系化することを目指す研究開発を推進するなど、バイオインフォマティクスの開発推進に努めた。

研究開発課題紹介

生命情報データベースの高度化・標準化

第I期 研究開発課題 研究開発期間：平成13年度～平成17年度

金久 實 京都大学化学研究所バイオインフォマティクスセンター 教授
2項関係に基づくゲノムと生命システムの機能解読

菅原 秀明 国立遺伝学研究所 生命情報・DDBJ研究センター 教授
ゲノム生物学バックボーンデータベースの構築提供

高木 利久 東京大学医科学研究所 教授
シグナルオントロジーとバイオタームバンクの開発

中村 春木 大阪大学蛋白質研究所 教授
蛋白質立体構造データベースの高度化

第II期 研究開発課題 研究開発期間：平成18年度～平成22年度

稲葉 一男 筑波大学下田臨海実験センター 教授
ホヤプロテイン統合データベースの構築

金久 實 京都大学化学研究所バイオインフォマティクスセンター 教授
ゲノムと環境の統合解析による生命システムの機能解読（KEGG）

菅原 秀明 国立遺伝学研究所 生命情報・DDBJ研究センター 教授
バイオ基幹情報資源の高準化と共有化

高木 利久 東京大学大学院新領域創成科学研究科 教授
オントロジーによるパスウェイの高度化および国際標準化（INOHパスウェイデータベース）

中村 春木 大阪大学蛋白質研究所 教授
蛋白質構造データバンクの国際的な構築と高度化（PDBj）

西岡 孝明 慶應義塾大学大学院政策・メディア研究科先端生命科学研究所 教授
メタボロームMSスペクトル統合データベースの開発（Metabolome-Mass Spectral Database）

森下 真一 東京大学大学院新領域創成科学研究科 教授
マルチモーダル統合バイオDB（Multimodal BIODB）

創造的な生物・情報知識融合型の研究開発

平成13年度採択 新規研究開発課題 研究開発期間：平成13年10月～平成16年9月

伊藤 啓 岡崎国立共同研究機構 基礎生物学研究所 助手
ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築

伊藤 隆司 金沢大学がん研究所 教授
インタラクトーム解析からの生物知識獲得

大浪 修一 システムバイオロジー研究機構 理事
線虫C. elegans発生過程のシステム解析

岸野 洋久 東京大学大学院農学生命科学研究科 教授
ゲノム進化とマッピングの階層モデルと解析アルゴリズムの開発

肥後 順一 東京薬科大学生命科学部 教授
高速計算機システムによる蛋白質フォールディングの研究

森下 真一 東京大学大学院新領域創成科学研究科 助教授
遺伝子破壊株イメージ・マイニング

矢田 哲士 東京大学医科学研究所 助教授
ヒト遺伝子の転写・発現の多様性解明を目指した基盤データベースの開発

平成16年度採択 継続研究開発課題 研究開発期間：平成16年10月～平成18年9月

伊藤 啓 東京大学分子細胞生物学研究所 助教授
ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築

伊藤 隆司 東京大学大学院新領域創成科学研究科 教授
絶対定量オーミックスからの知識発見

森下 真一 東京大学大学院新領域創成科学研究科 教授
遺伝子破壊株イメージ・マイニング

矢田 哲士 京都大学大学院情報学研究科 助教授
ヒトゲノムにおける広義の遺伝子発見研究

平成17年度採択 新規研究開発課題 研究開発期間：平成17年10月～平成20年9月

内山 郁夫 自然科学研究機構 基礎生物学研究所 助手
大規模な比較ゲノム研究を展開するためのワークベンチの構築

太田 元規 東京工業大学学術国際情報センター 助教授
タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用

川戸 佳 東京大学大学院総合文化研究科 教授
脳スライス中で可視化した神経シナプスの自動解析

塩田 浩平 京都大学大学院医学研究科 教授
ヒト胚の形態発生に関する三次元データベース

白井 剛 長浜バイオ大学バイオサイエンス学部 教授
実践による超分子複合体モデリングシステムの開発

宮川 剛 京都大学大学院医学研究科 助教授
マウスを用いた脳機能表現型データベースの開発

平成18年度採択 新規研究開発課題 研究開発期間：平成18年11月～平成21年9月

石井 信 奈良先端科学技術大学院大学情報科学研究科 教授
予測技術を用いた生命システムの同定手法の開発

黒川 顕 奈良先端科学技術大学院大学情報科学研究科 助教授
メタゲノムオーソログ遺伝子統合解析システムの開発

梶原 康文 慶應義塾大学理工学部 教授
タンパク質化合物相互作用の網羅的予測手法とデータベースの開発

清水 謙多郎 東京大学大学院農学生命科学研究科 教授
高精度タンパク質間相互作用予測システムの開発

中井 謙太 東京大学医科学研究所 教授
転写制御領域の構築原理解明

平成19年度採択 新規研究開発課題 研究開発期間：平成19年10月～平成22年9月

長野 希美 産業技術総合研究所 生命情報工学研究センター 主任研究員
酵素反応分類に基づく酵素反応予測システムの開発

馳澤 盛一郎 東京大学大学院新領域創成科学研究科 教授
進化型計算と自己組織化による適応的画像分類法の開発

馬見塚 拓 京都大学化学研究所 教授
生命科学上の非構造化データの統合マイニング

水口 賢司 医薬基盤研究所 バイオインフォマティクスプロジェクト プロジェクトリーダー
ダイナミクスを考慮した膜蛋白質の構造モデリング法の開発

平成20年度採択 継続研究開発課題 研究開発期間：平成20年10月～平成22年9月

内山 郁夫 自然科学研究機構 基礎生物学研究所 助教
大規模比較ゲノムワークベンチの実践的应用に向けた研究開発

太田 元規 名古屋大学大学院情報科学研究科 教授
タンパク質の構造・機能・相互作用予測システムの開発と展開

塩田 浩平 京都大学大学院医学研究科 教授
ヒト胚の形態発生に関する三次元データベース

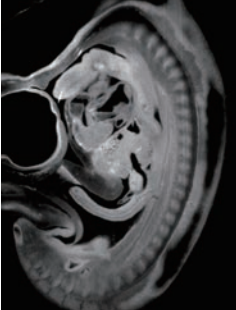
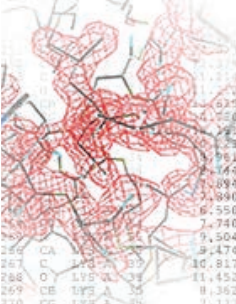
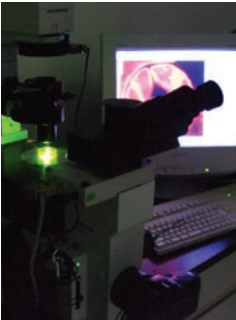
白井 剛 長浜バイオ大学バイオサイエンス学部 教授
実践による超分子ネットワークモデリングシステムの開発

平成21年度採択 継続研究開発課題 研究開発期間：平成21年10月～平成23年9月

黒川 顕 東京工業大学大学院生命理工学研究科 教授
メタゲノム統合解析システムの開発

梶原 康文 慶應義塾大学理工学部 教授
タンパク質化合物相互作用の網羅的予測手法とデータベースの開発

清水 謙多郎 東京大学大学院農学生命科学研究科 教授
高精度タンパク質間相互作用予測システムの開発





代表研究者

金 久 實

Kanehisa Minoru

京都大学化学研究所バイオインフォマティクスセンター 教授

生命情報データベースの高度化・標準化 第Ⅰ期 H13～17・第Ⅱ期 H18～22



大量データの生物学的な意味を解釈する国際標準データベース

KEGGがつなぐゲノムと社会

日常的に起こる病気は、様々な遺伝因子と環境因子が複雑に絡み合っていて、ヒトの健康や病気の容態は、生体システムの状態が安定か不安定かを反映しています。そのため、ゲノム情報（遺伝因子）とケミカル情報（環境因子、医薬品）から、生体システムと環境との相互作用を統合して解析すれば、病気の分子メカニズムを理解できると考えられます。

生命システムの情報統合データベース「KEGG (Kyoto Encyclopedia of Genes and Genomes)」は、ゲノムをはじめとした大量のデータの統合処理と生物学的な意味解釈のための標準的なリソースとして広く使用されています。KEGGをさらに高度化し、ゲノム情報と生体内・外のケミカル情報を分子ネットワークレベルで統合することで、生命システムの仕組みを理解すること、医療・創薬・環境保全に役立てることが期待されます。

研究開発開始時のKEGGは、生命システムを構成する部品の情報として遺伝子・蛋白質に関するゲノム情報 (KEGG GENES) と化合物・化学反応に関するケミカル情報 (KEGG LIGAND)、部品間の配線図情報として相互作用・反応ネットワークに関するパスウェイ情報 (KEGG PATHWAY) を統合したデータベースでした。

第Ⅰ期、第Ⅱ期を通じて、これまで生体内分子に限られていたパスウェイ情報に生体外分子とのつながりを含めて、ゲノム情報とケミカル情報との関連、生命システムと環境との相互作用に関する基盤情報を提供する一方で、ゲノムの機能解読と有効利用促進のために、細胞・個体・エコシステムといった様々な高次の知識を機能階層に取り入れ、疾患と医薬品に関する知識を集積することを目指しました。

我が国を代表する国際標準データベースに発展したKEGG

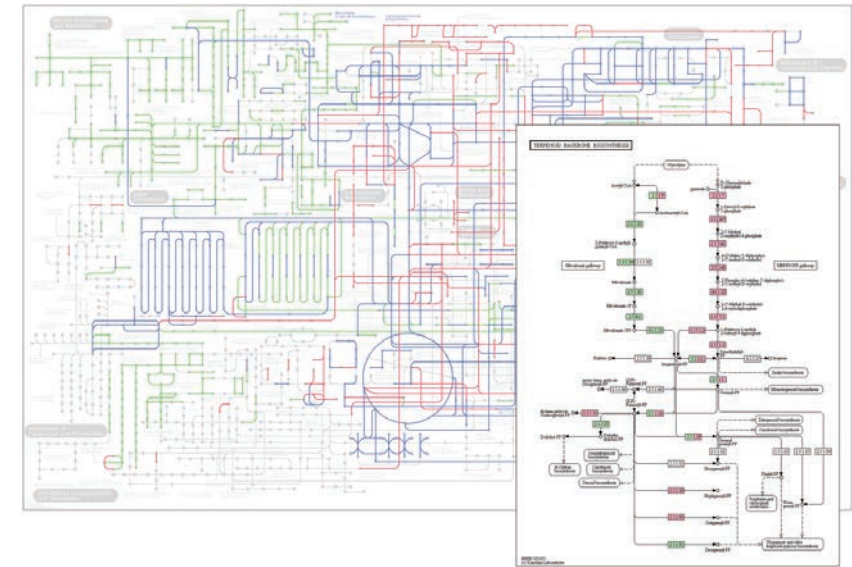
第Ⅰ期では、語彙の階層的な定義を行うGO (Gene Ontology) の考え方を取り入れたBRITEデータベースを構築し、細胞レベルの生命システムの働きを分子間相互作用ネット

ワークの「形」として表現するKEGGと、「語彙の関係」で表現するBRITEを融合しました。これにより、KEGGの階層に基づく高次機能の推論と、KEGGパスウェイでは表現できない知識のコンピュータ化を、階層テキストファイルという語彙の体系 (KO, KEGG ORTHOLOGY) で実現しました。

第Ⅱ期では、(1) KEGGの中核であるKEGG PATHWAYマップに階層的概念を導入し、従来型より上位レベルで代謝の全体像を提供するグローバルマップと、下位レベルでより細かな機能単位を定義したKEGG MODULEの開発を行いました。(2) ゲノムと生命システムをつなぐKOを着実に改良し、新たなアノテーションツールKOALAの開発を行いました。(3) ヒト疾患に関する分子レベルの知識集約では、疾患と病因遺伝子・病原体ゲノムとの関連をKEGG DISEASE疾患エントリとしてコンピュータ化しました。(4) 化学構造をベースに日米欧の医薬品を網羅的に蓄積したKEGG DRUGでは、ターゲット、薬物代謝酵素、薬物間相互作用など分子ネットワーク情報の集積に重点をおき、また薬効分類をはじめとしたKEGG BRITE機能階層の充実と日本語化を行いました。(5) ゲノムと環境との相互作用の1つの側面として、植物・微生物による二次代謝物質合成経路と、微生物による環境物質分解経路の知識を集約し、これらの物質の化学構造を蓄積しました。(6) 大量デー



KEGGの開発と運用を支えるデータベースサーバー群



ヒトゲノムと腸内細菌メタゲノムから推定される複合代謝経路

タ解析ツールとしてKEGG MapperとKEGG Atlasを開発しました。(7) KEGGとNCBIの緊密な連携体制も構築しました。

第Ⅰ期、第Ⅱ期を通じて、ゲノムをはじめとした大量データを意味解釈するためのリソースとして、我が国を代表する国際標準データベースに発展させることができました。

知的財産としてのバイオインフォマティクス

ヒトゲノム計画とともに始まったバイオインフォマティクスの研究分野は、実験技術の進歩・普及とともに、当初の大型プロジェクトを支援する役割から、個々の研究者あるいは研究グループを支援する役割へと変遷してきました。今後は、パーソナライズド医療をはじめ社会貢献の役割も広がっていくでしょう。これらでは、有用なデータベースやソフトウェアの開発・普及という支援形態が大きな役割を果たしています。一般にデータベース等には知的財産権（著作権）が存在します。これまでのバイオインフォマティクスは基礎研究を中心としたアカデミアの世界を対象としてきたため、デー

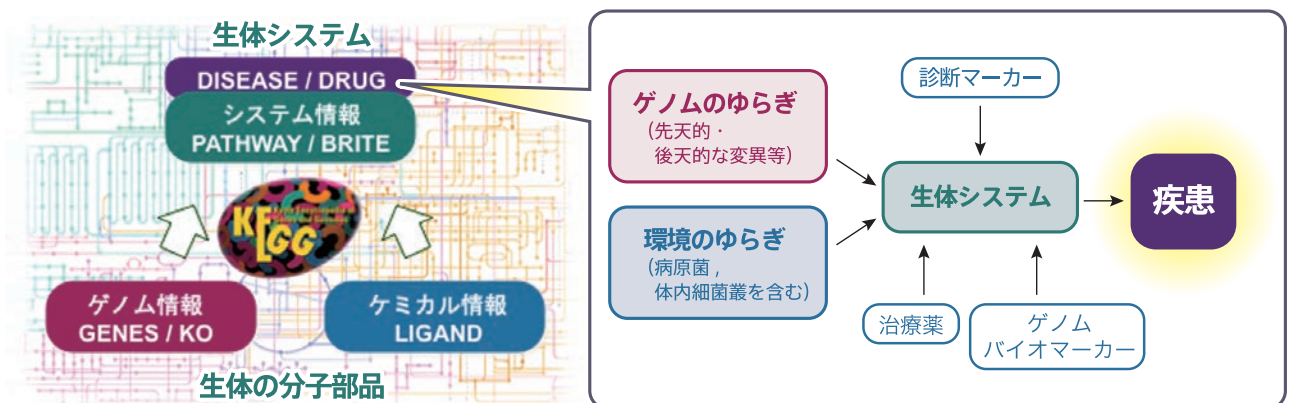
タベース等はフリーであること（無償提供すること）が重要でした。しかしながら、今後バイオインフォマティクスが社会貢献をするためには、フリーならよいという考え方だけでは対応できません。

またSome rights reservedとして少しでも権利を主張する限り（パブリックドメインでない限り）、利用許諾（ライセンス）をどのように行うか具体的な方策が必要です。我が国では、データベース等の取扱いに関する旧文部省通知、日本版バイ・ドール法、法人化後各大学での知的財産ポリシーなどの変遷がありましたが、1995年に開始したKEGGデータベースにはAll rights reservedの著作権が存在します。当初より、この知的財産の活用を促進するため、ゲノムネットを通じたフリーサービスだけでなく、企業等に対して適切な技術移転の道筋を作ることに努力してきました。両者がKEGGの国際的地位を高めることにつながり、また次のステップへの足がかりになったと考えています。

Email : kanehisa@kuicr.kyoto-u.ac.jp



疾患は生体システムのゆらぎ状態として、医薬品は生体システムへのゆらぎ物質として、KEGGデータベースに統合されている





代表研究者

菅原 秀明

Sugawara Hideaki

国立遺伝学研究所生命情報・DDBJ研究センター 特任教授

生命情報データベースの高度化・標準化 第Ⅰ期 H13～17・第Ⅱ期 H18～22



多彩なデータベースの連携を実現する情報環境は知識発見を支援する

質量ともに拡大するデータを捌くデータベースを構築

留まることのない実験技術の進歩を享受して、多様な生命現象について膨大な実験データが日々生み出され、多彩なデータベースもまた次々に構築されています。それ故に、公的データバンクには、データを網羅的かつ効率的に集約した上で信頼に足るデータベースを提供することが求められています。

そこで、三大国際DNAデータバンクの1つであるDDBJ(日本DNAデータバンク)は、本研究開発の第Ⅰ期では、ゲノム生物学のバックボーンとなるデータベース構築を目指して、それまでデータベースにおいて取り上げられることが少なかった生命現象における時間軸を視点に加えた「四次元データベース」と、データの質を高めた「高品位配列データベース」の開発に取り組みました。

第Ⅱ期では、第Ⅰ期の精神を引き継いでパイオ情報資源の品質を高める(高準化)ことに加えて、多彩なデータベースやツールを組み合わせやすく(共用化)することを目指しました。この目標は、「膨大なパイオ情報資源の仕様を正確に見通しよく整理することで、必要な資源のある場所や活用法、信頼性についての的確な判断を下すことができ、ひいては資源を最大限活用できる」という考えに基づいています。

具体的には、パイオ情報資源を分析し、実験技術と情報技術の手法と機能を一定の形式に従って記述して見通しを良くすることを試みました。並行してDDBJが提供するデータベースとツールを対象に、コンピュータプログラムから直接利用可能なサービス(Web API)を用意し、これらWeb API群を部品とする利用例を提供しました。ヒトの遺伝子頻度解析や遺伝子と疾患の相関関係等の問題解決などのワークフロー群です。

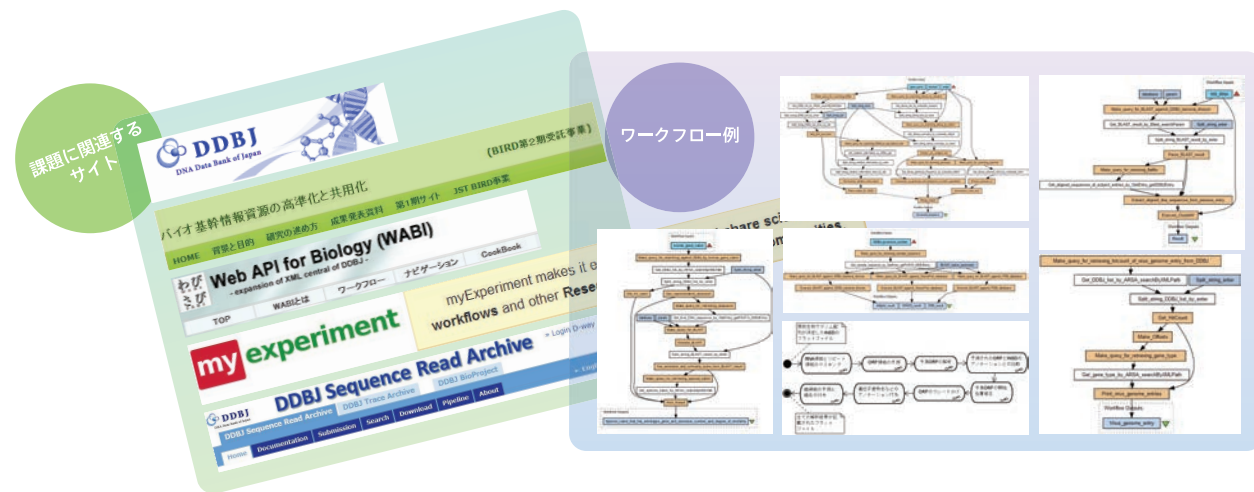
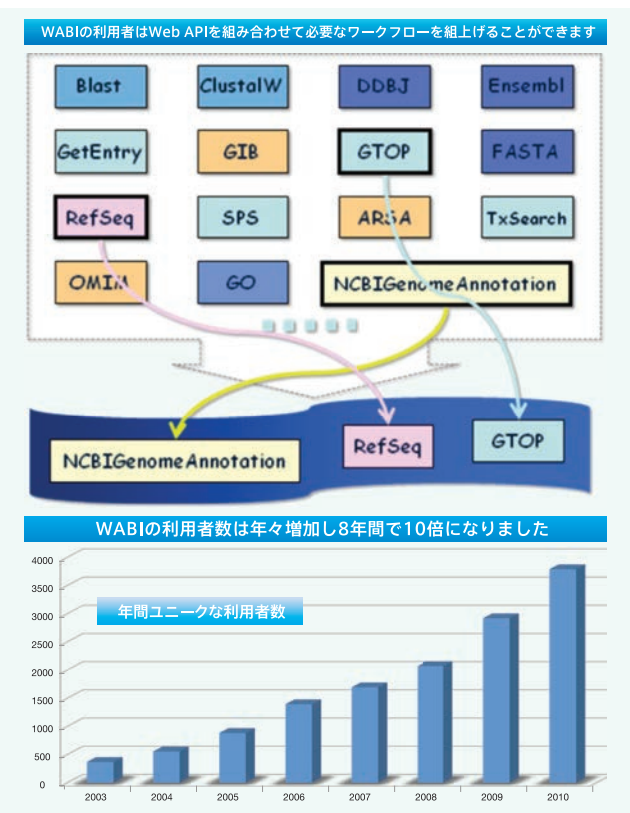
国内外のパイオ情報資源と連携しながら高度な情報環境を構築

第Ⅰ期では、四次元データベースとして、遺伝子発現情報の登録・公開、さらに時間的・空間的比較解析ができるMicroarray DatabaseとBio-simulated Databaseを開発しました。

高品位データベースとしては、国際塩基配列データベース(INSD)から対象エントリーをダウンロードし、DDBJ内外のデータバ

スからの知見を幅広く集約するOASYS(Open Annotation SYStem)と、ゲノム配列からタンパク質産物の構造を予測するGTOP(Genomes TO Protein structures and functions)を開発して、ゲノム配列アノテーションを再評価し、高品質なデータの提供を進めました。例えば、OASYSとGTOPを連携させた解析により、ORF(オープンリーディングフレーム)の確からしさを評価し、その構造やタンパク質立体構造の新たな知見を得ました。

第Ⅱ期では、代表的文献データベースPubMedに収録された論文で引用頻度が高いデータベースを対象にそのメタデータを標準形式でデータベースに蓄積することによって、様々なパイオ情報資源の透明性を高めて相互運用を促しました。さらに、論文からパイオ分野の実験手法とデータ解析手法を抽出してデータベースに蓄積して、手法の透明性を高めてその活用を促しました。



Web APIについては、DDBJの各サービスにWeb APIを装備するだけでなく、使用法を紹介するサイト「Cook Book」も構築し、さらに、Web API群をもとにWeb上でマウス操作のみでワークフローを組み立てることができるシステム「WNS(Workflow Navigation System)」を構築し、全て公開しました。その上で、こうした情報環境をWABI(Web API for Biology)と命名しました。

本研究開発課題によって、DDBJが構築している情報資源へのアクセス法が多様化し、情報技術のスキルに応じた知見取得が可能になり、また、国内外の外部資源との連携が実現しました。本課題で構築された情報環境は、今後多彩なデータベースの統合利用に貢献していくことでしょう。

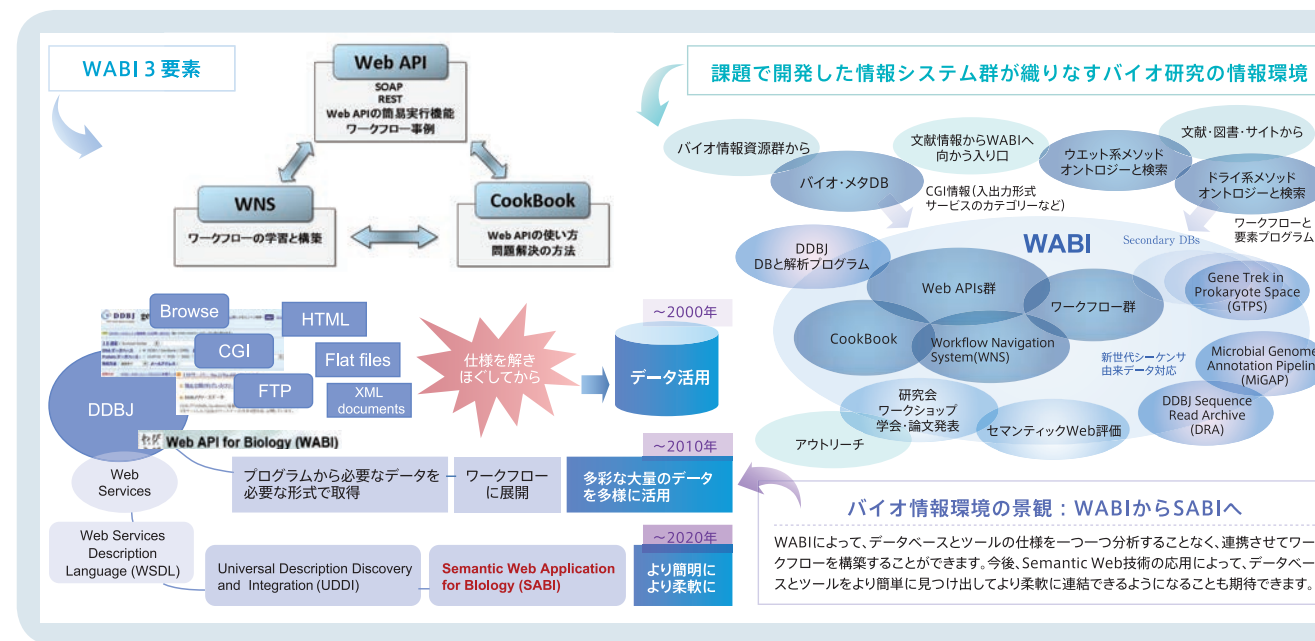
WABIからSABIへ

第Ⅰ期に取り組んだOASYSの概念と技術は、現在は、微生物ゲノムの全ORFについてその確からしさを評価したデータベースGTPS(Gene Trek in Prokaryote Space)と微生物ゲノム自動解析パイプラインMiGAP(Microbial Genome Annotation Pipeline)へ引き継がれ、幅広く利用されています。第Ⅱ期後半に、国際協力のもとに急遽着手したSRA(Sequence Read Archive)もごく短期間のうちに、配列決定が爆発的に行われる時代のバイオ研究に必須の環境となりました。公開当初はその有用性に疑問を抱かれたこともあ

たWABIも、2010年には年間4,000に迫るサイトから利用されるに至り、WABIを利用したワークフローがmyExperimentといった海外のサイトに他機関によっても登録されるようになりました。また、Web APIは広く一般のパイオ情報資源においても標準装備になっています。このように、高品質なデータベース・アーカイブと、それらの連携を簡単に実現できるWABIが、バイオ研究コミュニティの情報環境形成に貢献しつつあるとあってよいでしょう。いいかえますと、これらを今後も継続していく責任が生じていることになります。一旦サービスを開始した情報資源に「404 NOT FOUND」は許されません。

それでは、第Ⅰ～Ⅱ期の成果を維持していくだけで、バイオ分野の情報環境として十分でしょうか。十分ではありません。データベースの仕様やWABIのお作法を知らなくても、必要なデータ・情報・知識をデータベースの集合(データクラウド)から取り出せるようにできないものでしょうか。Semantic Web技術にはこれに応えられる可能性が秘められています。第Ⅱ期においてすでに、DDBJ全件をRDF(Resource Description Framework)で表現しtriplestoreに格納した上で、SPARQLでの検索を試みましたが、これからも国内外の関連活動と連携しながら、WABIに加えてSABI(Semantic Web Application for Biology)の確立に取り組んでいきます。WABI-SABIによって、パイオデータからの情報・知識抽出にとってより簡明で柔軟な情報環境が整っていくことでしょう。

Email : hsugawar@genes.nig.ac.jp





代表研究者

高 木 利 久

Takagi Toshihisa

東京大学大学院新領域創成科学研究科 教授

生命情報データベースの高度化・標準化 第Ⅰ期 H13～17・第Ⅱ期 H18～22



生物界の膨大な知識を体系化し、未来志向型データベースの開発に挑む

オントロジー開発とデータベース化による生物界の膨大な知識の体系化

あいつぐモデル生物のゲノム配列決定により、従来のひとつのモデル生物を対象とする研究から、生物界全体を対象とする研究が生まれました。このモデル生物種間の比較解析による新しい知識発見のためには、これまで蓄積された膨大なデータを生物界全体で統合し、これまで解析されたさまざまな解析技術を効果的に適用できるよう、環境を整備する必要があります。ゲノムにコードされた生命メカニズムを解読するには、シグナル伝達パスウェイの「オントロジー」構築が重要になります。

オントロジーとは、対象とする分野のデータの特徴を抽出し、概念の性質や互いの関係を明示的に定義したもので、生体内分子の複雑なネットワーク（パスウェイ）の膨大な知識の体系化に利用できる概念です。本研究開発では、パスウェイ研究におけるシグナル伝達系と生命科学用語のオントロジー開発・データベース化を進めてきました。シグナル伝達系は代謝系に比べ、研究に遅れをとっており、全世界のゲノム研究と産業界両方から、この研究の成果に期待が寄せられています。

第Ⅰ期では、シグナル伝達系の知識を集積するデータベース「INOHパスウェイデータベース(Integrating Network Objects with Hierarchies)」と、パスウェイの各構成要素をアノテーションするためのオントロジー構築に着手しました。INOHデータベースにおいては、生物種間や分子間で共通な反応を示すグループを単位としたモデルを提案し、専門化によるマニュアルキュレーションを行い、詳細にアノテーションされたデータを蓄積してきました。

第Ⅱ期では、INOHデータベースをさらに改良し、細胞のさまざまなメカニズムの知識を提供する公共の情報インフラとしての開発・提供を進めました。また、より正確な情報を簡単に検索できるよう整備し、国際標準への採用も目指しました。

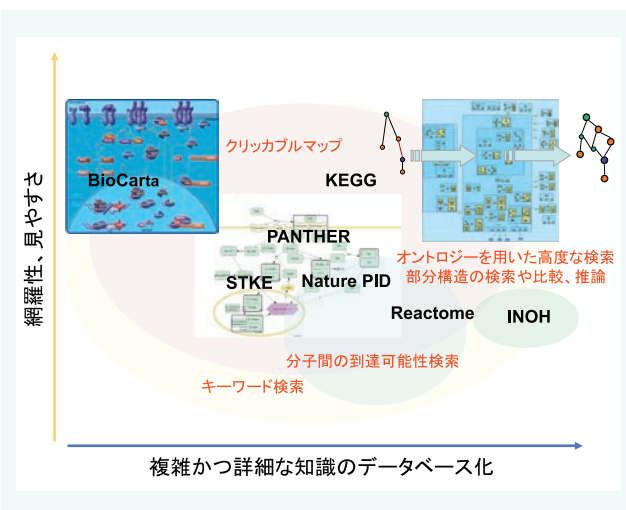
INOHが導く、次世代データベースのスタンダード

第Ⅰ期は、シグナル伝達系における、タンパク質、生命現象、相互作用、細胞局在部位の4つのオントロジーの開発に成功し、これらのオントロジーは、Ontology Viewerから一般公開しています。

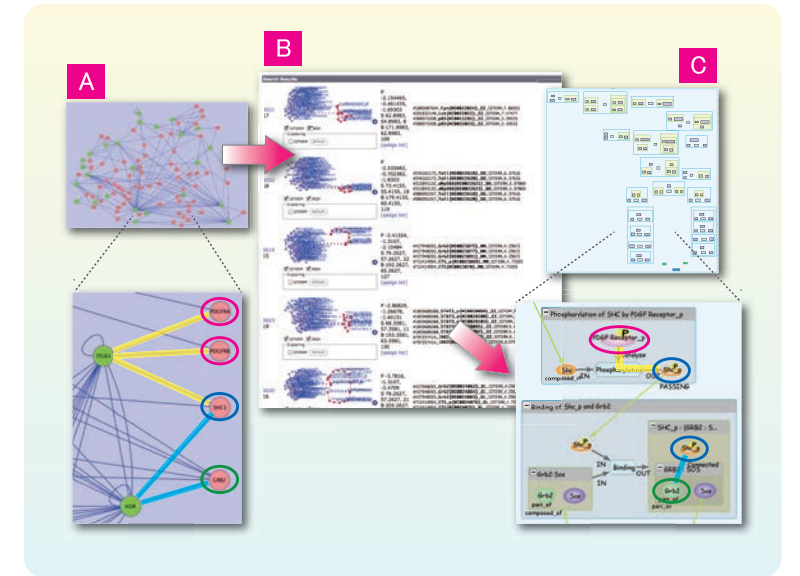
INOHデータベースでは、専門家が教科書や論文を精査してパスウェイの知識を抜き出すことで、他にはないオリジナルのコンテンツを提供でき、階層的に分類することにより、各種の検索・データ閲覧を容易にしました。それらのデータを扱う際の支援ソフトとして、検索機能を備えたパスウェイエディタ INOH Clientも開発しました。

第Ⅱ期では、INOHデータベースの知識処理技術を改良し、高精度のパスウェイデータやさまざまなオントロジーを公開しました。さらに類似パスウェイを検索・提示できる高度な推論検索システムの実用化にも成功しています。

このINOHデータベースでは、人間の利用を前提とした従来のような教科書・論文の電子化だけではなく、計算機が直接処理できるような形でデータベース化しました。これにより、パスウェイデータをプログラムから利用するための、API(Application Programming Interface)の公開も実現することができまし



世界のDBとINOHの比較



INOH推論検索システム

た。INOHデータベースは、オントロジーに基づいた新たな知識処理の枠組みを提供する、未来志向型のデータベースといえるでしょう。

データベースの国際標準化への取り組みとしては、パスウェイデータベース間でデータ交換を行う際のフォーマットである、BioPAXの一員として標準化の議論に参加し、各種仕様の提案および検証を行いました。

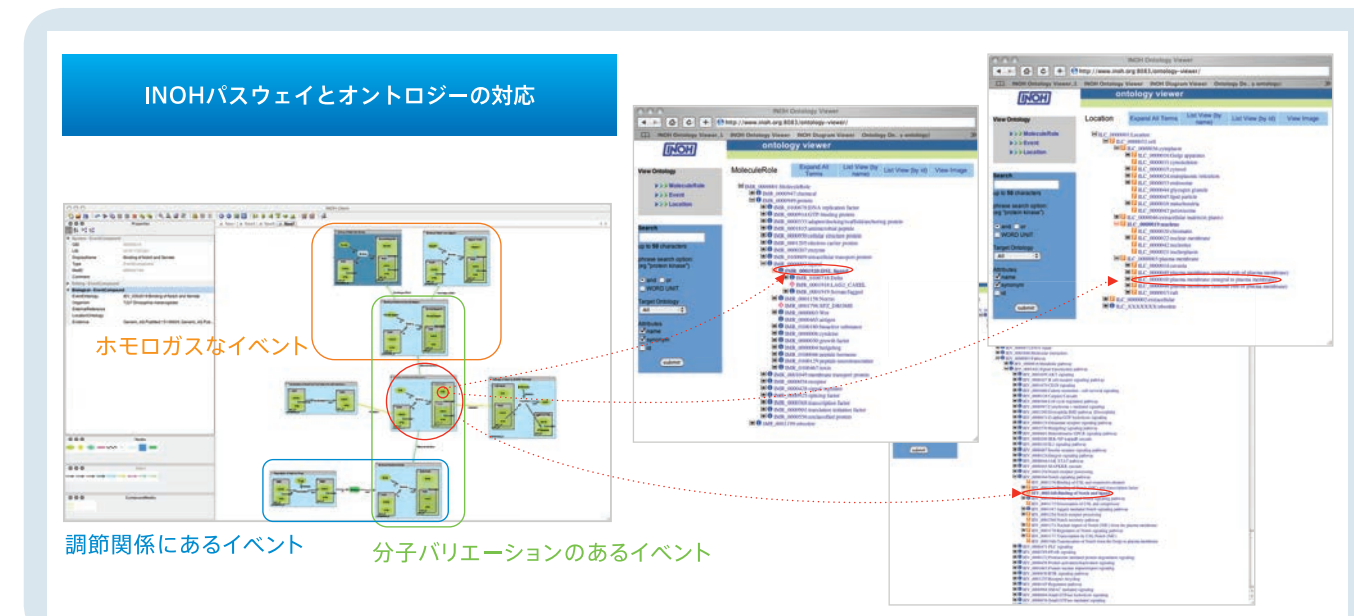
INOHデータベースは、ゲノムからの細胞の機能予測、コンピュータによる細胞機能のシミュレーション、医薬品の開発等に際して重要な情報源になることが期待されています。

生命科学の新たな展開の礎に

新型シーケンサーに代表されるように、生体の観測・計測装置の進歩には目覚ましいものがあります。それらの装置から産み出される膨大かつ多種多様なデータが生命科学に新たな展開をもたらそうとしています。そして、これらのとてつもない量のデータをどのように整理・解釈し、研究や産業に活用するかがバイオインフォマティクスにいままさに問われています。これを解決するには大きく二つの方向があります。一つは、データの生物学的意味には踏み込まず、もっと言えばデータを解釈せずにそのまま活用する方

向です。例えば、データ同士の比較や分類から予測を行うタイプです。これには新たな情報技術の開発が必要です。もう一つが本プロジェクトで目指したようなオントロジー（生物学的意味）ベースの知識の体系化を基盤にして、データの解釈、活用を図る方向です。これには情報技術はもちろんのこと、丹念な知識の編集が必要になります。どちらのアプローチも一長一短がありますが、どちらも大切であり、これからはこれら二つのアプローチがあいまって研究は進んで行くものと思われます。本プロジェクトは今回で一旦終了しますが、ここで培われた方法論、情報技術、オントロジー、データベースは、生命科学の新たな展開の礎となるものと確信しています。また、本プロジェクトに携わった研究員、技術員はそれらを担う人材として活躍してくれるものと期待しています。

Email : tt@k.u-tokyo.ac.jp



代表研究者

中 村 春 木

N a k a m u r a H a r u k i

大阪大学蛋白質研究所 教授

生命情報データベースの高度化・標準化 第Ⅰ期 H13～17・第Ⅱ期 H18～22



「形」や「動き」も検索できる！ 国際協力で作る蛋白質構造データバンク

日本発の蛋白質立体構造データベース (PDBj)

欧米・日本を中心とする国際的なゲノムプロジェクトの進展によって、多くの蛋白質立体構造が従来に比べ迅速に決定される時代を迎えています。これらのゲノム研究により得られる膨大な遺伝情報が、生命活動においてどのような機能をもつのかを解明することが、ポスト・ゲノム研究の大きな課題になっています。その課題を解き明かす「鍵」となるのが、遺伝情報を蛋白質の立体構造として理解し、その形や機能の情報を蓄積する構造生物学研究です。

世界的な構造生物学研究の進展により、種々の生物の蛋白質の立体構造が次々と明らかになっており、得られた立体構造データを登録する「蛋白質立体構造データベース」(PDB:Protein Data Bank)は、世界中で利用されていましたが、これまでは単に構造生物学研究の成果をまとめただけのデータベースにとどまっています。

そこで本研究開発では、蛋白質の立体構造とゲノム情報との結びつきを強める一方、XML等の最新情報技術を利用し、国際的な連携のもとに世界標準としての新しいデータ記述と解析ツールや二次データベースを開発して付加価値を付け加えました。そしてPDBの品質を保ちつつ、さらに大量のデータの登録・編集とその維持・管理・標準化を行い、構造生物学者だけを対象とする専門的なデータベースから、広く生命科学の研究者、産業界、さらには一般の人にも役立つデータベースに高度化することを目指しました。

将来にわたって持続可能な基盤的構造を作る

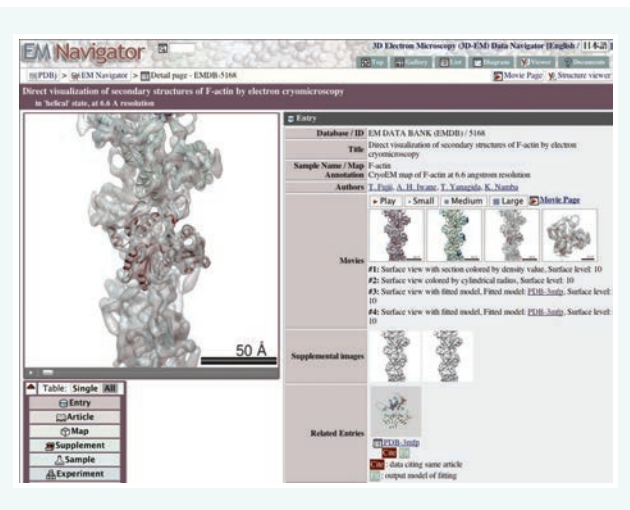
大阪大学蛋白質研究所では、2001年度に日本蛋白質構造データバンク (PDBj: PDB japan) を組織し、欧米の研究機関と共に国際蛋白質構造データバンク (wwPDB: worldwide PDB) を設立して、国際協力によるPDBの維持・管理とサービス・システムの開発を進めました。

第1期では、PDBデータの品質を将来にわたり保つだけでなく、他のデータベースとの統合化を進めるために、米国と共同して正規XML記述のPDBフォーマット(PDBML)を国際標準として

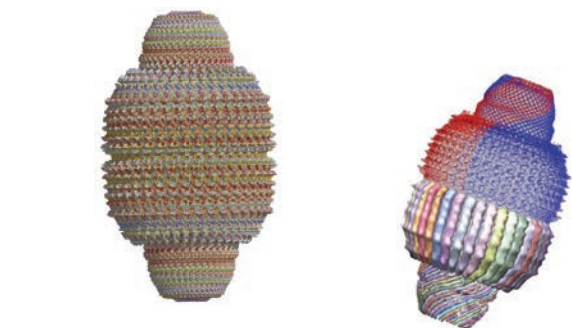
確立しました。そして、XMLのデータ拡張が容易という特長を活かして、オリジナルデータに多く欠損している分子機能や実験条件等の情報を文献から抽出して追加する作業を継続して行いました。PDBjでは、2010年末までに世界全体の約25%に達する1万6345件の構造データの検証・編集・登録処理を行いました。

第II期では、PDBjの高度化に取り組み、類似フォールドや基質結合部位、分子表面の形状など、蛋白質の「形・表面・動き」といった動的性質を検索できるアナログ検索サービス (Structure Navigator) や、進化情報を加味して配列・構造から機能を推定するサービス (SEALA)、教育的なサイトとして英語と日本語による蛋白質構造事典 (eProtS)、既知の立体構造との類似性に基づいて蛋白質の未知の立体構造を予測するサービス (Spanner) を新たに開発・公開しました。これらのサービスを利用して、タンパク質3000プロジェクトで構造が解かれた蛋白質やマイクロアレイ解析実験で発見された新たな蛋白質の機能推定や、リガンド分子認識や蛋白質間相互作用のための構造モチーフの同定等、新たな知見が生み出されています。

一方、蛋白質のNMR(核磁気共鳴)解析からは、構造情報以外にもリガンド相互作用等の機能解明に寄与する空間・時間分解能の高い情報が得られます。米国のNMRデータベース(BMRB:



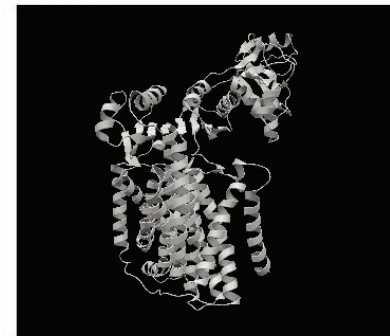
EMnaviのサービスによる、生体高分子やその複合体の
電子顕微鏡イメージと原子構造の重ね合わせの表示



巨大な蛋白質Vaultの分子表面における
静電位の表示(青が正、赤が負)

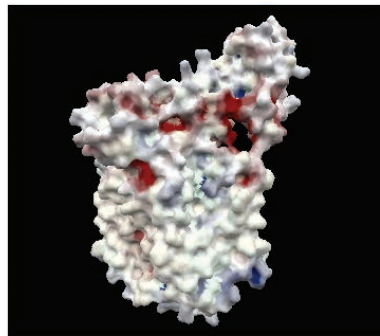


巨大な蛋白質Vaultの 原子構造

***e*F-site**

蛋白質の静電位と構造とを同時に表示するeF-siteサービス

3aqp-A



BioMagResBank)と協力して、PDBj-BMRBでは2005年からNMRデータの登録を行っており、12月末までにBMRB全体の約12%に達する760件の登録処理を実施しました。さらに、NMRによって構造解析が行われたペプチドや核酸、糖以外の低分子量生体分子構造の登録処理システム(SMSDep: Small Molecule Structure Deposition)を米国BMRBとの協力により、PDBj-BMRBで構築し運営を開始しました。

このように、PDBj およびPDBj-BMRBでは、wwPDBにおける国際連携により、コスト・パフォーマンスが高く、かつ恒久的に持続可能なデータバンク構築を行う努力を続けています。

PDBjの展開:高い品質の構造データを最新の情報技術・計算科学技術を用いて提供する

これまでのBIRDによる研究開発で構築した蛋白質立体構造データベース(PDBj)は、ライフサイエンス分野の基盤データベースであり、バイオインフォマティクスだけでなくほとんど全ての生命科学の基礎・応用両面で大きな貢献をしてきました。創業などいくつかのバイオ産業では、蛋白質立体構造データベースは必須の情報基盤となっています。

PDBjがこれまで構築・運営してきたデータベースは、国際的にも価値が高いデータベースであり、本研究開発の終了後も、国際協力によって、その登録・編集作業を引き続き円滑に進める必要があります。日本人研究者によって今後とも多くの生体高分子の立体構造データが産みだされると期待され、国内の拠点において蛋白質立体構造データバンクが長期にわたって運営されるべきです。このため、蛋白質・核酸等の立体構造解析や構造バイオインフォマティクスの専門家がいる大阪大学蛋白質研究所において、平成23年度以降も、引き続き生体高分子立体構造データの登録・編集・品質管理が行われ、データベースが構築されています。大阪大学蛋白質研究所は平成22年度から共同利用・共同研究拠点としての活動を実施していますが、これまでのPDBj活動の共同利用・共同研究に果たす役割の重要性が認められ、その拠点活動の一環としてPDBjの活動が位置づけられています。一方で、PDBjは、ライフサイエンス統合データベースセンター(NBDC)と協力し、オントロジーやセマンティック・ウェブ技術等に基づく他の生命科学データベースとの統合化も進めています。

Email : harukin@protein.osaka-u.ac.jp



PDBJ 日本蛋白質構造データバンク <http://www.pdbj.org/>

大阪大学蛋白質研究所にて実施

原子種とその座標、アミノ酸残基、実験手法、実験時の情報、実験観測データを登録し、5万6千件をデータベース化。

wwPDB (国際蛋白質構造データバンク) の一員として国際連携活動

Graphic viewer: JV
<http://www.pdbj.org/>

Molecular surface DB: eSite
<http://eSite.pdbj.org/>

DBの高度化 (付加価値を加える)
(実験・機能情報、他のDBの情報との付加、二次DBの構築と検索・解析ツール開発)

Year	Processed total at PDBJ	Processed total at wwPDB
2000	~1000	~1000
2001	~1000	~1000
2002	~1000	~1000
2003	~1000	~1000
2004	~1000	~1000
2005	~1000	~1000
2006	~1000	~1000
2007	~1000	~1000
2008	~1000	~1000
2009	~1000	~1000
2010	~1000	~1000

創造的な生物・情報知識融合型の研究開発 H13～16・H16～18
生命情報データベースの高度化・標準化 第II期 H18～22



出芽酵母からヒトへ: 遺伝子機能を解明するデータベースの構築!

モデル生物「出芽酵母」から広がるデータ・マイニング

ポストゲノム時代を迎えた現代では、様々なバイオデータが多様な形で日々蓄積されています。本研究課題ではこれらのデータを効率よく活用することを念頭に置き開発を進めました。

創造期では、モデル生物「出芽酵母」を用いて、遺伝子の網羅的な機能予測法の確立に取り組みました。従来の機能予測法は「ある遺伝子を欠失させた細胞（遺伝子破壊株）が、正常な細胞と比べてどんな変化を示すのか」を解析するものでしたが、この手法を網羅的に行うためには、ゲノム内の各遺伝子を1つ1つ欠失させた膨大な種類の細胞が必要であり、さらに機能欠損を反映した表現型を効率的に見つけ出すことも非常に困難でした。そこでゲノムにコードされる遺伝子が既に判明している出芽酵母を用いて、遺伝子破壊株の形態異常を網羅的に解析することで、遺伝子の機能予測の効率化に挑みました。その中で、遺伝子破壊株の細胞形態の変化を顕微鏡で追跡した画像情報を収集し、特徴的な状態の変化を自動的に追跡・抽出する「イメージ・マイニング」技術と、遺伝子情報との因果関係を推測する「データ・マイニング」技術の開発に取り組みました。

高度化第II期では、DNA配列、遺伝子発現量、ハイオミメージ、パスウェイ、文献等、異なるデータタイプへの問合せを容易にする「マルチモーダル統合バイオデータベース (Multimodal BIODB)」の構築に挑みました。その中で、多様な生物種のMultimodal BIODB構築を可能にするツールキット、大量データの迅速な分析を可能にする解析アルゴリズム、超並列クラスター型計算機を利用した処理手法の開発にも取り組みました。

学問の新分野を開拓する「Multimodal BIODB」

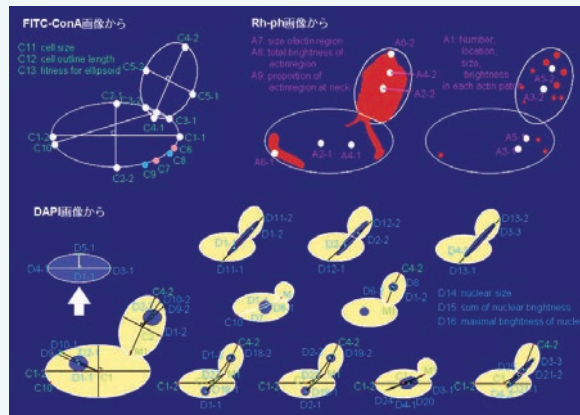
創造期では、出芽酵母の遺伝子破壊株の画像解析ソフトウェアを開発し、4780の非必須遺伝子破壊株中で約2000株の形態異常を突き止め、これまであまり重要視されていなかった非必須遺伝子の破壊によって形態異常を示す細胞が多いことを明らかにしました。また、新規の細胞周期チェックポイントを発見し、新たなシ

グナル伝達経路の存在を提示しました。

既知の遺伝子をGene Ontology (GO) により分類して細胞形態と遺伝子の関係を網羅的に調べて、形態情報から遺伝子機能を予測 (マイニング) することに成功しました。また、ヘテロ二倍体遺伝子破壊株を用いて、従来は困難だった、必須遺伝子破壊株の形態情報を収集する作業も進んでいます。

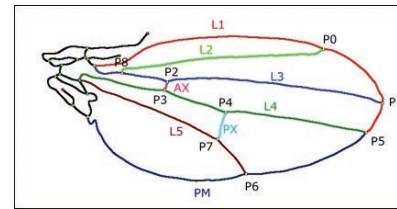
遺伝子破壊株からのイメージ・マイニングは、表現型データの網羅的収集と解析を行うフェノーム研究の典型として、国際的にも広く評価されています。この技術を応用すると、ビールやパンの発酵時の酵母の生育状況のモニタリングもできます。また、出芽酵母の遺伝子破壊株の形態を多面的にとらえたカタログを作成することで、遺伝子機能の解明だけでなく、発酵・醸造に有用な酵母の育種や、薬剤の標的遺伝子の同定・作用機構の解明にも応用できます。

高度化第Ⅱ期では、様々なデータタイプが蓄積されている出芽酵母に加えて、発生遺伝学のモデル生物であるメダカ・ショウジョウバエ、害虫研究のモデル生物であるカイコ、ヒトを対象に広げて、Multimodal BIODBの開発を行いました。短期間の間に、パーソナルな環境でゲノムブラウザを簡単に構築できる機能を備えたツールキットUTGB (University of Tokyo, Genome Browser) を公開し、超高速 DNAシーケンサーが出力するゲノ

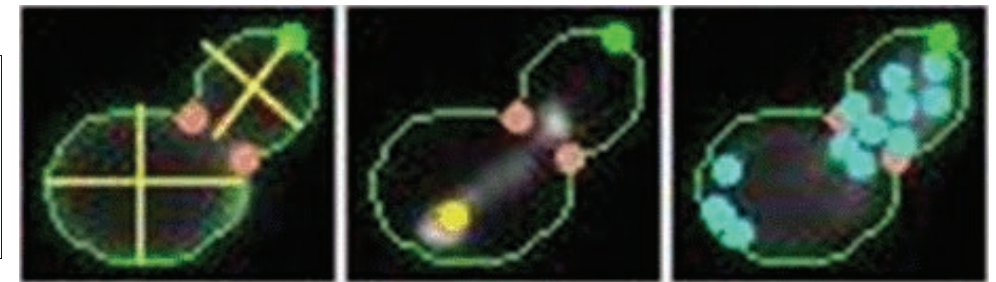


表現型の幾何学的パラメータ

表現型の変化を画像処理で正確に描出



翅脈認識の方法



出芽酵母の形態変化の認識

ムデータを分析したウェブサーバを構築しました。研究成果は7つのサーバから公開されており、毎月の平均で約1万2000の訪問者に利用されています。

また、Multimodal BIODBを用いた分析により、「クロマチン構造が遺伝的多様性に影響を与える」という結果をScienceに報告するなど、国際雑誌や国際会議で研究成果を発表しています。今後、Multimodal BIODBが学問の新分野を開拓すると期待されています。

長いひものDNAは、なぜ絡まらないのでしょうか？

Multimodal BIODB により我々のクロマチン構造に関する理解は深まり、いまはこの方向の研究をひろく展開しています。その動機を説明しましょう。ヒトDNA約30億塩基対は、仮に一直線に伸ばすことができれば、長さが2メートルにも及びます。この長いひもであるDNAは、直径10マイクロメートル程度という小さな空間である細胞核の中に押し込められています。いつもはある程度弛緩した状態にあり、細胞核内で広がっています。しかし体細胞が分裂する際には、それに先だってDNAは複製され染色体として凝縮し、2つのコピーはうまく分離されて各細胞へと分配されます。DNAは長いひもです。一般には、ひもをグシャグシャに丸めてから

伸ばすと結び目ができやすいです。そうすると、DNAが分離される際には結び目ができて絡まり、壊れやすいのでは？ にもかかわらず、うまく2つに分離されるということは、背後に何らかのうまい構造的なカラクリがある？ そう考えたくになります。

DNAという小宇宙の3次元構造は依然として未知ですが、生命の本質を理解するための重要なヒントが隠されています。たとえば、ある遺伝子群が、DNA上では遠く離れた位置でコードされているにもかかわらず、協調して働いているケースが多いのです。遠く離れているのに、なぜでしょう？ 協調する遺伝子群がDNAの3次元構造の中で近くに配置されていれば、同時に制御されるには都合が良いですね。もうひとつ例をあげます。多様な遺伝子がDNAから読みだされる量は大きく隔たりがあり、最大量と最小量のあいだには数万倍の開きがあります。これについても構造がある程度関与しているという推測があります。大量に読まれる遺伝子の周辺では、その遺伝子の開始部分と終了部分がループするようにDNAが折り畳まれていると考えると好都合です。なぜなら、遺伝子をクルクル何度も読みだすために有利な構造となるからです。このような謎を解き明かすために、Multimodal BIODB の時に研究開発したソフトも活用しています。

Email : moris@cb.k.u-tokyo.ac.jp



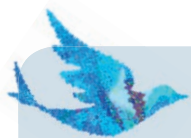
マルチモーダル統合バイオDB
(Multimodal BIODB)

目標：次世代シーケンサーが観測を可能にしたマルチモーダルなデータ（パーソナルゲノム、DNAメチル化、クロマチン構造、遺伝子発現量、転写開始点等）を、スーパーコンピューティングにより高速に問合せ、統合化した知識を表現するシステムを研究開発



多様なデータの集まり

クロマチン構造が進化と転写へ及ぼす影響を分析する視点へ変換



代表研究者

稲葉 一男

I n a b a K a z u o

筑波大学下田臨海実験センター 教授

生命情報データベースの高度化・標準化 第II期 H18～22



ホヤのタンパク質研究からヒトの進化メカニズムを解き明かす！

ヒトとホヤの意外なつながりを利用したプロジェクト

ホヤはヒトと同じ「脊索動物」に属し、脊椎動物の先祖に当たります。遺伝子構成が単純であることや、脊椎動物の先祖的器官を持つことから、ヒトを理解する上で進化系統学上、極めて重要な生物です。2002年に、ホヤの1種で世界中の海に生息するカタユウレイボヤの遺伝子配列が決定されたことから、ホヤの研究は大きく前進しました。

本研究では、ホヤの胚や幼生、成体の組織で発現するタンパク質の情報をデータベースとしてまとめ、ホヤの研究はもとより、国内外の多くのライフサイエンス研究に役立つようなプロテイン統合データベースの構築を目指しました。それにより、ヒトへの進化のメカニズムや細胞分化のメカニズムの解明や、未知の遺伝子・タンパク質の機能解明、疾患遺伝子の機構解明に貢献することを目指しました。

データベースを充実させ、国際基準を目指す

ホヤプロテイン統合データベース(CIPROサイプロ)は2006年に開発を始め、翌年11月に公開しました。ホヤの胚や幼生、成体の器官に発現したタンパク質の基本情報の提供をしています。発現場所やその量については、タンパク質をスポットとして分離する二次元電気泳動の像で見ることができます。また、他生物に存在するタンパク

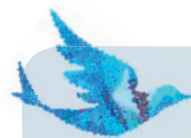
質のホヤにおける発現パターンの検索、遺伝子発現パターンとの比較、ヒト疾患に関係するタンパク質の検索、キーワード検索など、様々な情報を検索することができます。

さらに、ホヤの発生を三次元的に見ることができるデータベース「FABA」や、ホヤの様々なデータベースやソースのプラットフォームとなっている「ホヤウェブポータル」(フランス) 等の複数のデータベースとのリンクも充実しており、CIPROはホヤのタンパク質データベースの国際基準と認識されています。

理想の生命系データベースを目指して

本データベースの開発に当たっては、実験研究者がほしい情報をすばやく取り出せるような工夫をこらしました。開発者のほとんどは実験研究者です。データベースは情報の格納庫であるとともに、研究アイデアを引き出す宝庫です。必要な情報を、画像情報も交えながら効果的に提示することにより、「研究を創出」することが可能なデータベースとなり得ます。CIPROがそのようなデータベースのひな形として生命科学のフロンティア研究に貢献できたら幸いです。

Email : inaba@kurofune.shimoda.tsukuba.ac.jp



代表研究者

西岡 孝明

N i s h i o k a T a k a a k i

奈良先端科学技術大学院大学情報科学研究科 特任教授

生命情報データベースの高度化・標準化 第II期 H18～22



多様な代謝物質を分析して、生命活動の物質的基盤を理解する！

有用なデータベース群を組み合わせ、さらなる価値を引き出す

生物は、細胞内の化学反応を複雑に組み合わせた「代謝反応ネットワーク」によって多様な代謝物質を作り出し、生命活動をしています。この代謝物質を化学分析する「メタボローム解析」は、物質に基づいて生命活動を理解することによって、医療の発展や、生物を利用した物質生産を可能にする重要な研究です。

本研究開発では、メタボローム解析を進展させるべく、有用性が認識されつつもこれまで存在しなかった2次代謝物質の化学・生物学データ等を収集したデータベース(KNAPsAcK, Metabolomics.JP)と、新しい2次代謝物質の探索、発見などのメタボローム解析に特化したデータベースとして、代謝物質を高精度な質量分析(MS)で測定したマスマススペクトルを集めたデータベース(MassBank)の開発に挑みました。

また、これらのデータベースを組み合わせることで、2次代謝物質の代謝反応ネットワークを明らかにして、ゲノム情報との関連付けに役立つデータ収集と解析ツールの開発に成功しました。

低コストで維持管理できる、世界初のデータベースを構築

KNAPsAcKでは、2011年1月までに2万18種の生物種における5万48種の代謝物について、10万1500対のデータを整理し公開し

ました。さらに、Metabolomics.JP では、2次代謝物質を化学構造式に基づいて系統的に分類し、代謝反応ネットワークを推定するツールを開発しました。

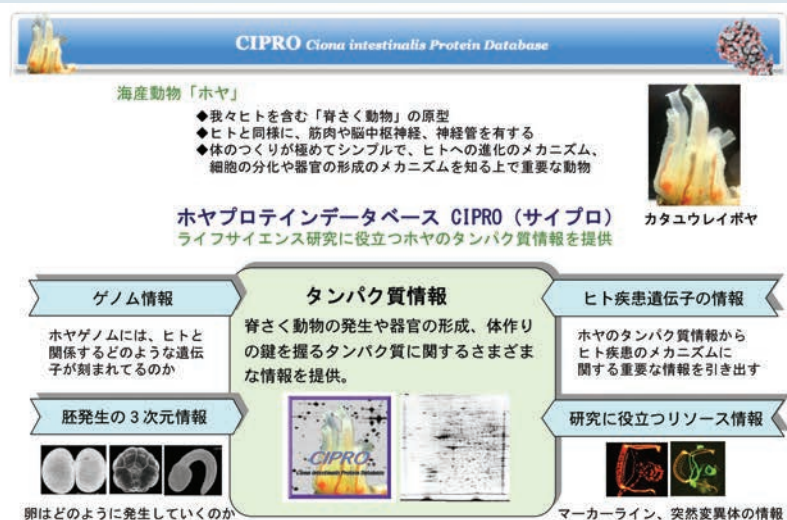
MassBankでは、データ形式を共通化することにより研究者コミュニティが低コストで維持管理できる、画期的な分散型データベースを実現しました。2010年末までに国内外19の研究グループが1万1614の化合物について測定した 3万312件のデータを、国内外の8つのサーバからWeb公開しています。

本研究で蓄積されたデータは高く評価され、特にMassBankは日本質量分析学会の公式データベースに認められています。

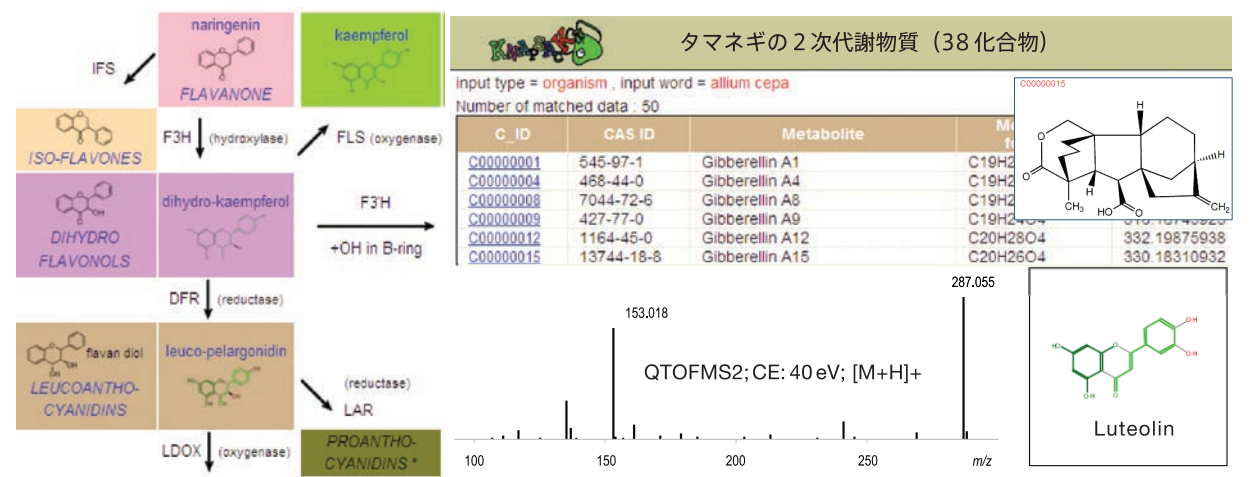
マスマススペクトルを化学構造タグとするメタボロームデータベースへ

これまでMassBankは既知物質のマスマススペクトルを公開してきました。しかし、MSで検出されたものの同定できなかったために論文に記述、報告されなかった未知の2次代謝物質が数多くあります。今後、これらのマスマススペクトルをMassBankで公開する予定です。マスマススペクトルを化学構造タグとすることによって、同じ未知の2次代謝物質を共有する生物種が明らかになります。このような知見から、新しい2次代謝物質や生合成経路の発見、生合成の進化などが解明されると期待できます。

Email : takaaki@is.naist.jp



マスマススペクトルや生物種、生合成反応のデータベースを統合して、2次代謝物質をゲノム情報とリンクする。





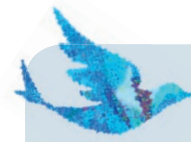
代表研究者

伊 藤 啓

I t o K e i

東京大学分子細胞生物学研究所 准教授

創造的な生物・情報知識融合型の研究開発 H13～16・H16～18



代表研究者

伊 藤 隆 司

I t o T a k a s h i

東京大学大学院理学系研究科 教授

創造的な生物・情報知識融合型の研究開発 H13～16・H16～18



2つのアプローチを組み合わせ、ショウジョウバエの脳内を解析

ボトムアップ型アプローチとトップダウン型アプローチの連携がもたらすもの

脳の研究では、単一細胞レベルから神経回路の構造や機能を調べるボトムアップ型と、コンピュータシミュレーション等によって神経回路の動作を情報科学的に推理するトップダウン型の2つのアプローチが用いられます。本研究課題ではこの2つのアプローチ手法を強力に連携させ、脳内の膨大な神経回路の構造解明を目指しました。

5年間を通し、脳構造がシンプルで脳全体を解析することが出来るキロショウジョウバエに着目し、感覚情報の入力処理部から、異種の情報を統合する連合部に至るまでの神経回路について、個々の神経回路の形態や接続状況を解析しました。この解析により、感覚情報の処理という脳機能について、「回路解析→回路動作予測→実験による検証」という総合的アプローチを実現しました。

感覚経路の脳内地図を描き出す

初期3年では、代表研究者らが作製した、様々な細胞群で任意の遺伝子を発現させられる「GAL4エンハンサートラップ系統」を用いて神経細胞の狙った部分だけを可視化する技術や、脳細胞の位置と数の定量データを抽出する技術の開発に成功しました。その結果、キロショウジョウバエの視覚、嗅覚、聴覚の3つの感覚について、感覚細胞から順にステップを追って神経回路の網羅的同定を進め、詳細

な脳内地図をデータベース化することができました。

継続2年では、前述の3つの感覚に味覚を加えた4つの神経回路の同定を進めました。結果として、従来考えられていた場所とは異なる場所に感覚神経が収斂していること、個々の感覚低次中枢とそこからつながった投射神経構造の詳細が明らかになりました。

最終的には、ショウジョウバエの脳全体の回路図を作成してコンピュータ上で動作を再現することを目指しています。

情報の統合から行動の制御へ

今後は、五感の残るひとつである体性感覚の中枢を解明してすべての感覚中枢を網羅するとともに、五つの感覚中枢からの異なる情報が実際に脳でどのように統合されるのかを明らかにしてゆきます。並行して、体の動きを制御する行動中枢から逆に情報経路をさかのぼり、行動の制御に必要な情報が作り出される経路を調べます。山の左右からトンネルを掘り進んで真ん中でぶつかるように、脳の情報の入口と出口から神経回路を組織的に調べてゆくことで、脳全体の情報の流れを明らかにします。

Email : itokei@iam.u-tokyo.ac.jp



分子間の相互作用から、生物の新たな知識を発見する！

生物システムの研究は新たなステージへ

ゲノムの解読が急速に進むにつれ、生物のシステム研究の基盤となる「生物を構成する要素(ゲノムなど生体内の分子)の同定」は大きく進歩を遂げ、次なる段階として、「要素間の関係性の解析」が求められています。本研究では定量的なデータ取得法の確立とデータの定量性を活用して新知識発見を促す方法の構築を目標に開発を行いました。

初期3年は、生体内にある分子間の相互作用を解析する「インタラクトーム解析」を進めるとともに、定量計測を支援するバイオインフォマティクス(BI)技術の開発も行うことで大規模な計測系を構築し、データの定量性を知識発見に活用する新しい手法の開発を進めました。

継続2年は、初期3年で確立した技術を駆使しながら、ゲノミクス、トランスクリプトミクス、プロテオミクス、フェノミクスなどのデータを、各オームミクスの枠を超えて相互に比較・解析を進めました。

BI技術を取り入れて解析を進化させる

初期3年では、従来行われていた蛋白質間のインタラクトーム解析だけでなく、遺伝子間、蛋白質-DNA、および蛋白質-脂質等の相互作用の解析に成功しました。また、基盤技術として開発した、遺伝子発現を絶対定量化する技術(GATC-PCR)や蛋白質相互作用を絶

対定量化する技術(PCS-MS)は様々な解析に応用でき、これまでに例のない種類のデータの創出が可能になりました。そのデータの持つ定量性は、新たな知識の発見に有用なものと予測されます。

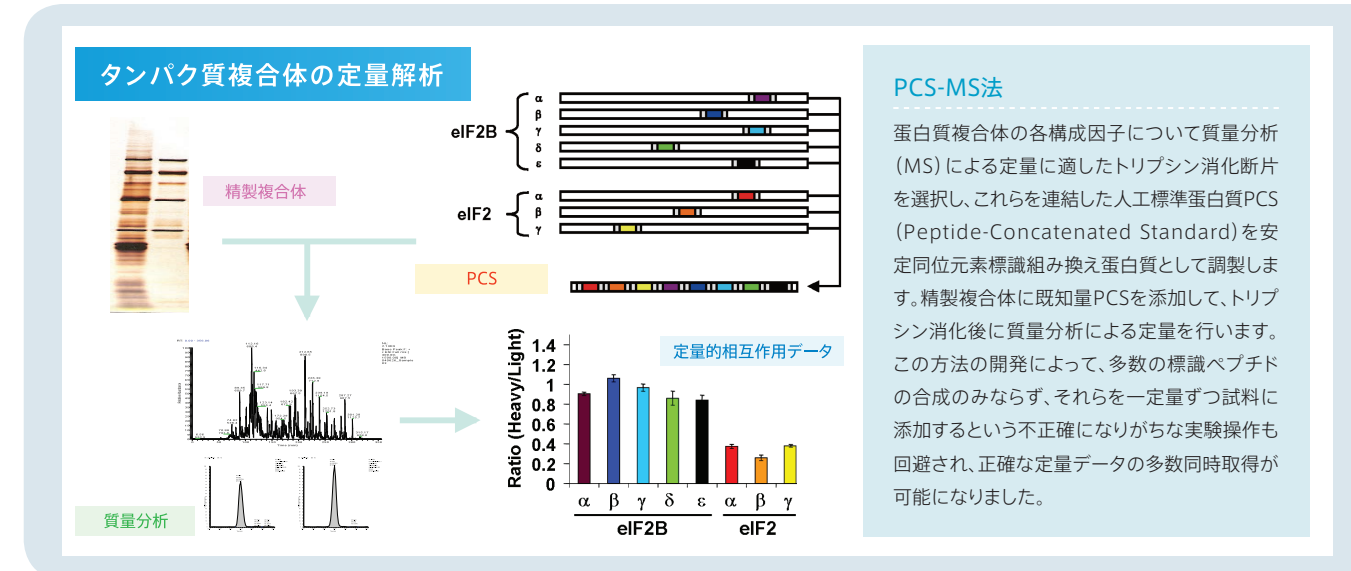
継続2年では、GATC-PCRやPCS-MSを駆使した絶対定量的解析を行うとともに、データの定量値に解釈をほどこして知識発見に役立てる手法(ART/EX)も実用化しました。これらの技術を実装したツール「EAST」は、Webでも提供しています。

本研究で開発した新たな実験手法・計測技術や、それに関連したBI技術等の独自の技術群は、新知識発見に有用な利用価値の高いものといえます。

網羅的定量データが駆動する新しい生物学へ

有名なコールド・スプリング・ハーバー研究所シンポジウムは、1934年の発足当時から一貫して定量生物学に関するシンポジウムと銘打っています。昨今の計測技術の進歩は定量データの網羅的取得を可能にしつつあり、時代は「定量オームクス」へと向かいつつあります。今も昔も大切なのはデータの定量性を活かして生物学的本質のより深い理解に至ることですから、この過程を支援するBIは今後ますますその重要性を増してゆくことでしょう。

Email : ito@bi.s.u-tokyo.ac.jp





代表研究者

大 浪 修 一

O n a m i S h u i c h i

理化学研究所 生命システム研究センター 発生動態研究チーム チームリーダー

創造的な生物・情報知識融合型の研究開発 H13～16



線虫の発生過程を解析し、生物システムの機構を明らかにする！

遺伝子機能を障害した線虫の初期胚細胞分裂パターンを測定

生物とはいわば「固有の遺伝情報により決定される、時間的空間的に動的なシステム」です。近年、ゲノム科学の発展によって蛋白質のアミノ酸配列や転写調節シグナル等、システムの部品構造に関する情報が解明されつつありますが、次の大きな挑戦は、部品が全体として作る動的システムの機構を明らかにし、性質を理解することでしょう。

システム理解のためには計算機シミュレーションの活用が効果的であることから、以前より開発していた、線虫 (*C. elegans*) の初期胚細胞分裂パターンの測定技術と細胞構造の計算機シミュレーション技術を活用し、多細胞生物の発生機構解明に挑戦することで、時間的空間的に動的な多細胞生物の発生機構を解明する新しい研究戦略の確立を目標としました。

世界に類を見ないデータを蓄積し、細胞生物学研究をリードする

本研究では、遺伝子機能と胚構造の時間的空間的な変化に関連づける客観的かつ定量的な測定データという、世界に類を見ない独創的なデータの大量蓄積に成功しました。一例として、線虫において100%の胚致死表現型を示す351遺伝子の全てを対象に遺伝子機能阻害を行い初期胚の細胞分裂パターンを測定したところ、分裂のタイミング

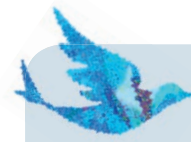
あるいは胚内の細胞の空間配置に顕著な異常が生じる等、遺伝子機能と細胞分裂パターンの関連を4500種類以上検出することに成功しました。

また、得られたデータを計算機科学的に解析する手法を開発し、胚発生の時間的空間的プログラムとそれを制御していると予想される遺伝子群の導出にも成功しました。さらに計算機シミュレーションと画像処理を利用した定量的かつ客観的な形態の測定を利用し、胚や細胞の形態変化の機構と機能を解明する新しい細胞生物学の手法/戦略の確立に成功しました。本研究で開発した手法/戦略は、今後の発生生物学、細胞生物学において中心的な位置を占めると考えています。

システムの理解から予測可能な発生生物学・細胞生物学へ

本研究で開発した、画像処理を応用した動態の定量計測と計算機シミュレーションを活用した胚や細胞のシステム解析の手法/戦略は、既に線虫以外の様々な生物の胚や細胞へ広がりはじめています。今後は、タンパク質などの生体分子の動態や細胞内／細胞間で働く力の動態等の計測と、それらのシミュレーション等との統合により、胚や細胞の「動的システム」としての理解が更に深まるでしょう。システムの理解の先には、予測可能な発生生物学、細胞生物学の実現が期待されます。

Email : sonami@riken.jp



代表研究者

岸 野 洋 久

K i s h i n o H i r o h i s a

東京大学大学院農学生命科学研究科 教授

創造的な生物・情報知識融合型の研究開発 H13～16



ゲノムを階層モデルでとらえ、生命の進化のプロセスに迫る！

進化の痕跡を探る鍵は、分子進化の速度

生物の多様化や適応進化のプロセスを知るには、それぞれの生物のゲノムに刻まれた進化の痕跡を検出する方法が有効です。本研究課題では、一般に利用されているゲノムデータベース内のゲノムの分布に階層モデルを駆使することで、そこに映し出された進化のプロセスを検出することと、ゲノム上の重要遺伝子をマッピングする手法を構築することを目的としました。

進化の痕跡は、分子進化速度の変化となって表れますが、系統ごと・遺伝子ごとの分子進化速度はデータベースを階層的にモデリングすることで、系統間・遺伝子間の相関関係等を把握でき、適切にとらえることができます。これにより、生物の多様化や適応進化の痕跡を高感度で検出することが可能になりました。また、階層モデルは重要遺伝子のマッピングにも有効であることがわかりました。

様々な角度から解析手法を構築する

上述の階層モデルを主軸としながら、遺伝子配列解析の理解をより深めるため、他にも様々な統計モデルや解析手法を開発しています。分子進化速度の変動の他にも、機能的な適応を理解するために、タンパク質進化に立体構造に関連づけたモデル、樹構造からの乖離として進化モデルの系統的な偏りを検出し、異なる系統間での収斂進化を考慮に入れて分子系統樹を頑健に推定する方法を開発しまし

た。また、分子レベルの収斂進化や組換えの検出、ウイルス集団の準種分化の推定等、モデルを超えた探索的アプローチをも積極的に開発しました。

開発された手法を適用し、哺乳動物ミトコンドリアゲノムの適応進化と寿命、植物とハエにおける分子進化速度の加速、バクテリアメタゲノムの階層構造と多様性、ウイルスの準種分化とタンパク質適度ランドスケープ、南アメリカHIVにおけるゲノム組換えの履歴、酵母における重複遺伝子の非対称性等を特徴づけることに成功しています。

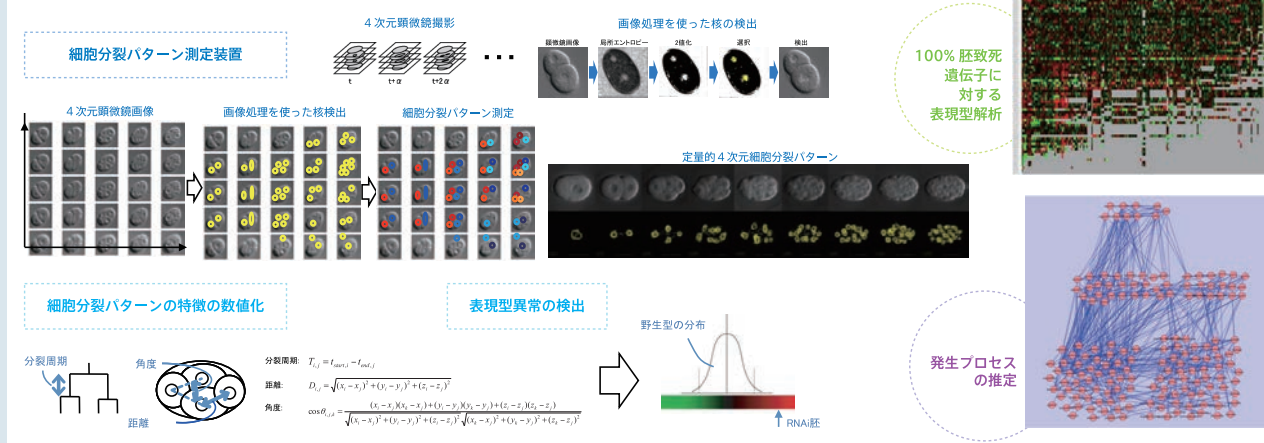
的を射る柔軟な事前情報で新たな発見の可能性を提供する

生命現象に関する知見は日進月歩で膨らんでいます。これを支えるデータベースもますます豊かになっています。その情報の本質的な部分をありのままにすくい取るために、いまや階層モデルは欠かせない手法になってきています。進化系統樹上の分子進化速度の時空間変動、タンパク質立体構造上の淘汰圧の空間集積性、遺伝子ネットワーク上の原因遺伝子の空間集積性など、的を射る柔軟な事前情報を通して、研ぎ澄まされた叡知を尽くしても漏れ落とした影にも新たな発見の可能性を提供していくでしょう。

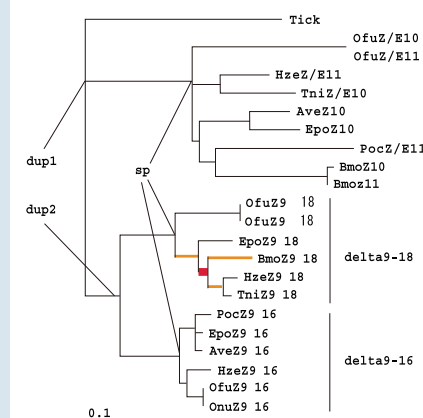
Email : kishino@lbm.ab.a.u-tokyo.ac.jp



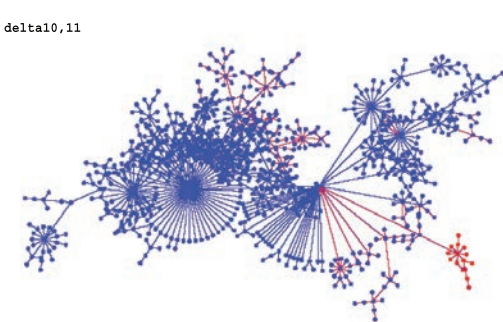
コンピュータ画像処理を使った定量的細胞分裂パターン表現型解析



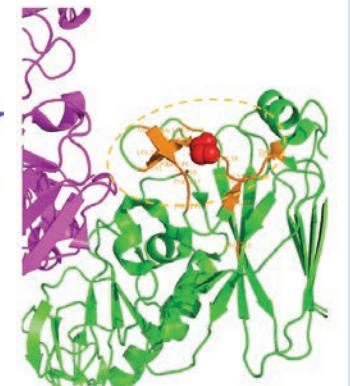
分子進化速度の時空間変動



遺伝子ネットワーク上の空間集積性



タンパク質にかかる淘汰圧の空間集積性





代表研究者

肥 後 順 一

H i g o J u n i c h i

大阪大学蛋白質研究所 附属プロテオミクス総合研究センター 客員教授

創造的な生物・情報知識融合型の研究開発 H13～16



蛋白質の立体構造を解析する、高速の計算機作りに挑む！

一個の蛋白質の観察から、複雑な立体構造を導き出す

蛋白質のアミノ酸配列からその立体構造や機能を予測することは、長らく生命科学の中心的課題でした。一方、近年の様々なゲノミクス／プロテオミクス研究から配列・配列・機能の情報が急速に蓄積されてきました。それを受けて現象論的な立体構造予測は一定の成果を上げています。しかし、配列・構造・機能間の関係について物理化学的な法則に立脚した第一原理的な理解を得ることはできていません。

そこで本研究では蛋白質の立体構造形成機構の理解を目標に、高速計算機システム開発と効率的な蛋白質立体構造探索法確立を目指しました。

第一歩として、一個の蛋白質の折れ畳み（フォールディング）を計算機で「観察」しました。立体構造探索は、分子を構成する原子間の相互作用を計算することに深くかかわっており、正確かつ高速に計算できれば効率的に複雑な立体構造を探索できるのです。

この折れ畳みの研究に、「マルチカノニカル分子動力学（McMD）」という効率的な蛋白質の立体構造探索法の技術を組み込み、立体構造予測と自由エネルギー地形の算出に挑みました。

世界トップレベルの技術をさらに高めていく

本研究開発の開始時点において、水中にある蛋白質の精密かつ徹底的な立体構造探索では、アミノ酸の化合物であるペプチド10残

基の計算が可能で、これは当時の世界記録でした。研究の終了時点では、アミノ酸約30基の構造探索が、1～2ヶ月で計算できるようになりました。また、既存の分子力場（分子や原子間に働く力の場）ではポリペプチド鎖の立体構造が再現できないことが示唆されたため、より正確な結果を求める取り組みとして新たに力場を作成し、分子力場の正確さの改善を実施しました。

従来は厳密な結果を求めると計算に膨大な時間がかかるという問題があり、1～2ヶ月で構造が算出できるというのは、世界的に見てもトップレベルにありました。そして、下で述べるように、現在では50残基を超えた蛋白質の立体構造探索が可能になっています。

立体構造予測から分子複合体形成の自由エネルギー地形へ

本研究課題を基礎にMcMD法をさらに進化させ、現在では50残基超の蛋白質の折れ畳みの自由エネルギー地形の算出が可能です。さらに、天然変性蛋白質の「折れ畳みと連動した結合」の自由エネルギー地形の算出にも成功しました（図参照）。これらの研究では、得られた様々な立体構造集団中の最大クラスター（自由エネルギー最安定状態）が天然の立体構造に対応していました。つまり、構造探索から構造・機能予測へと踏み出しました。今後、より複雑で生物学的に重要な生体分子系に挑戦します。

Email : higo@protein.osaka-u.ac.jp



代表研究者

矢 田 哲 士

Y a d a T e t s u s h i

京都大学大学院情報学研究科 准教授

創造的な生物・情報知識融合型の研究開発 H13～16・H16～18



ヒトゲノムに潜む遺伝子を全部見つけ出す！

ヒトゲノムには、まだまだ未知の遺伝子が隠れている？

ヒトゲノムプロジェクトの国際コンソーシアムが明らかにした最も衝撃的な報告は、そこに潜むタンパク質遺伝子の数が2万2287個であることでした。この数は、ショウジョウバエのゲノムに潜むタンパク質遺伝子の数とそれほど変わりがなく、ヒトの複雑さを説明するにはあまりにも少ないものでした。そこで、ヒトの複雑さを説明する鍵として、遺伝子の転写や発現の多様性を司る仕組みに注目が集まるとともに、これまでの方法では見つけ出すことができなかった遺伝子を見つけて出そうとする試みが始まりました。

この研究課題では、はじめの3年間、ヒトゲノムに潜むタンパク質遺伝子を全て見つけ出すことを目標にして、転写産物による発見法、ab initio発見法、ゲノム比較による発見法という3種類の遺伝子発見プログラムの開発に取り組みました。

それに続く2年間では、それらの遺伝子発見プログラムを駆使して、ヒトのタンパク質遺伝子の数を再評価するとともに、機能を失った遺伝子（偽遺伝子）やタンパク質をコードしていない遺伝子（RNA遺伝子）をゲノム配列から見つけ出す要素技術の開発に取り組みました。

タンパク質遺伝子の発見技術、ここに極まれり！

はじめの3年間では、新しく開発した3種類の遺伝子発見プログラムを組み合わせた遺伝子発見プロトコルを確立し、このプロトコルを用いて見つけ出したタンパク質遺伝子の情報をHAL (Human

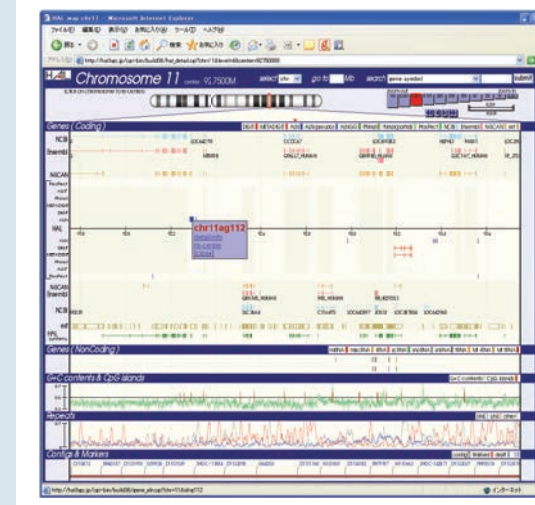
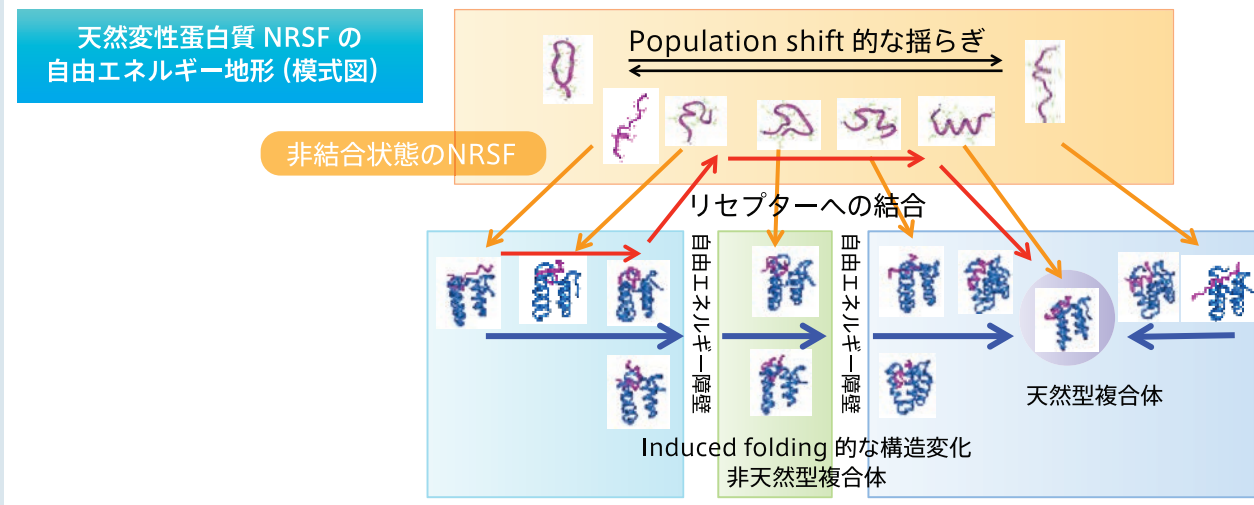
genome Annotation Library) データベースにまとめて公開しました（左図）。HALは、ヒトゲノムの配列データの更新に合わせてアップデートされ、遺伝子のゲノム上での位置や構造などに加え、外部のさまざまなデータベースへのリンクが提供されています。

それに続く2年間では、ここでの遺伝子発見プロトコルが見つけ出した未知のタンパク質遺伝子が本物であるかどうかを確認する作業を行いました。その結果、それらの少なくとも30～40%が本物の遺伝子であることが確認できました（右図）。また、偽遺伝子やRNA遺伝子をゲノム配列から見つけ出す要素技術の開発にも成功しました。とりわけ、偽遺伝子については、その形成過程に関する新しい生物学的な知見を得ることができました。

遺伝子のde novo誕生の仕組みは？

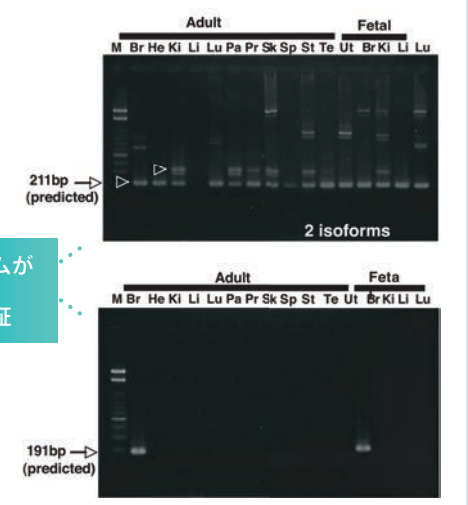
ここ数年で塩基配列を決定する革新的な技術が開発されたことにより、機能性RNA、選択的スプライシング、プロモーターなどのプロファイリングが進み、遺伝子の転写や発現の多様性の実態が明らかになりつつあります。これらのデータの蓄積によって、その分子的な仕組みが明らかになるでしょう。一方で、これらのデータは、従来の予想を上回る高い頻度で遺伝子が新規に誕生していることを示しています。しかし、その仕組みについてはほとんど分かっていません。これからの挑戦的な研究課題です。

Email : yt@i.kyoto-u.ac.jp



ヒトゲノム
アノテーション
データベースHAL

遺伝子発見プログラムが
新しく見つけ出した
遺伝子の実験的な検証





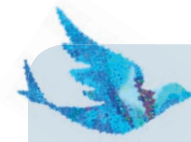
代表研究者

内 山 郁 夫

U c h i y a m a I k u o

自然科学研究機構 基礎生物学研究所 助教

創造的な生物・情報知識融合型の研究開発 H17～20・H20～22



代表研究者

太 田 元 規

O t a M o t o n o r i

名古屋大学大学院情報科学研究科 教授

創造的な生物・情報知識融合型の研究開発 H17～20・H20～22



多様な生物のゲノムを横断的に比較するための基盤システムを作る！

蓄積したゲノム情報を効果的に活用した比較ゲノム解析

シーケンス技術の進展により、近年様々な生物のゲノム配列が蓄積してきました。「比較ゲノム解析」はゲノム情報を解析する強力な手段ですが、進化的距離の異なる生物種を比較対象にすることによって、様々な知見を引き出すことが可能になります。本研究では近縁種間の比較から遠縁種間の比較まで幅広くカバーすることにより、ゲノム情報を最大限活用した比較解析を可能にするシステムの開発を目指しました。

初期3年では、微生物比較ゲノムデータベース「MBGD」を基にして、オーソログ(生物間で対応する遺伝子)の対応テーブルの作成を基軸とした、高機能な比較ゲノムワークベンチを開発するとともに、ゲノム進化の理解を促進するため、近縁ゲノム間の比較解析手法の開発等を目指しました。

継続2年では開発したワークベンチの実践的な活用に向けて、新規に解読したゲノムとその近縁種との比較解析、ならびに微生物集団を混在状態のまま解析するメタゲノム解析という2つの課題にフォーカスし、これらの研究を推進するのに有効な機能の開発を行って、オーソログ解析という共通の枠組みに載せて解析できるようにすることを目指しました。

比較ゲノムワークベンチRECOGを開発して、ゲノムの機能や進化の謎に迫る

初期3年は、多様な配列を持つ微生物数百種程度のゲノムを同時に比較することが可能な効率的なオーソログ解析手法を基軸とした、比較ゲノムワークベンチ「RECOG」を開発しました。RECOGによって、近縁ゲノム比較からすべての系統の横断的比較までをサポートするとともに、近縁種の

集合を内群、その他を外群として扱うことにより、近縁比較と遠縁比較の統合を実現しました。また、類縁度が中程度のゲノム比較を促進するため、共通祖先から受け継がれた「コア構造」の構築手法を開発しました。

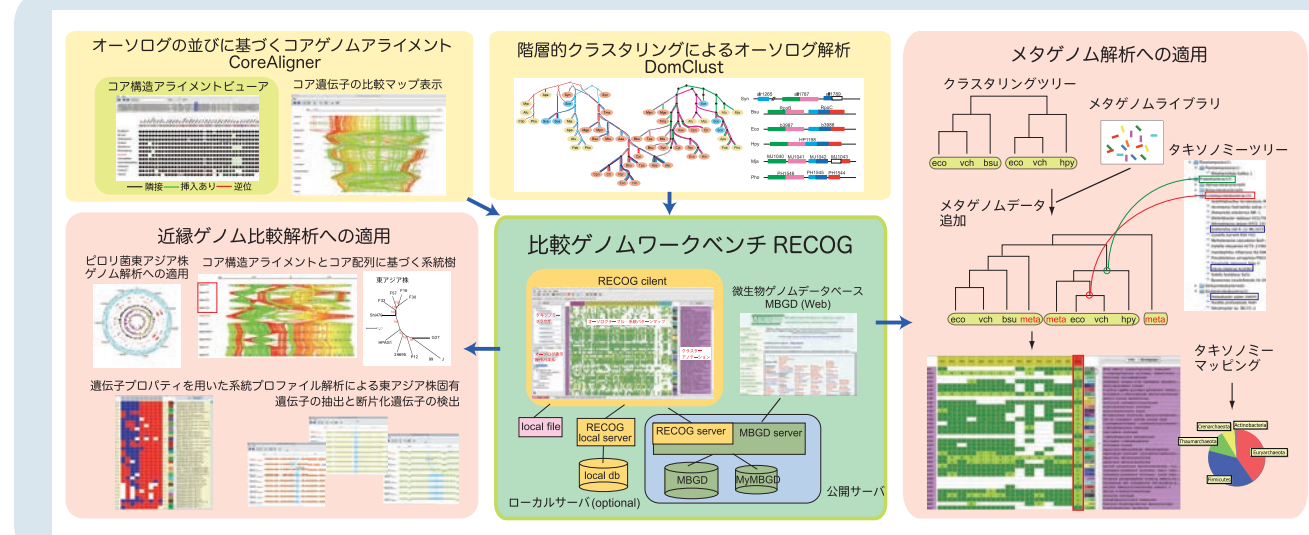
継続2年ではRECOGの改良を進め、不完全ゲノムやメタゲノムデータの取り込み機能、遺伝子プロパティと遺伝子セットに基づく比較ゲノム解析法、効率的な増分オーソロググルーピング法、ならびにオーソログ解析と連携してメタゲノムデータから由来生物種の分類群を推定する手法などの開発を行いました。

本研究で開発したRECOGシステムは、様々なゲノムプロジェクトにおいて、蓄積したゲノムデータを活用した比較ゲノム解析を進めるための、統一的な枠組みを提供します。これにより、多様な生物のゲノムの機能や進化の解明に貢献することが期待されます。

多様な生命現象の解明に向けて

シーケンス技術はさらに発展を続けており、今では自然界に存在するあらゆる生物がゲノム研究の対象となり、様々なモデル生物や類縁生物ゲノムとの比較によって、生物の多様なはたらきを調べることが可能になってきています。また、微生物ではメタゲノム解析によって、未培養菌を含めて、自然界の様々な環境における微生物集団の多様性の解明が進みつつあります。RECOGをこうした様々なゲノム研究に活用すべく、基盤となるデータベースの拡充も含めて、よりいっその研究開発を進めていきたいと考えています。

Email : uchiyama@nibb.ac.jp



遺伝子の配列情報から、分子構造や機能を予測するシステムを作る

高等生物の遺伝子配列から高次情報を抽出するのは困難

ヒト、チンパンジー、イヌ、ネズミ等様々な種のゲノム配列が解明されましたが、これら高等生物のゲノムにコードされているタンパク質の立体構造や分子機能等の高次情報を配列情報からの確に取り出す技術は、これまで確立されていませんでした。

本研究では、タンパク質のアミノ酸配列から立体構造や分子機能を予測することを目的とし、初期3年は基礎的な解析研究に加え、高等生物由来のタンパク質に適用可能な自動化されたフォールド認識法と多様な分子機能表現に対応した総合的な分子機能予測システムからなる、タンパク質の構造・機能アノテーションシステムを開発しました。

継続2年は研究対象をタンパク質複合体に拡張して、構造・機能予測法を安定稼働させるとともにヒトゲノム由来の全タンパク質配列に適用し、結果をヒトゲノム構造・機能アノテーションデータベース(SAHG)にまとめました。

蓄積した知識を整理・統合してアノテーションデータベースを構築する

初期3年では最新データに基づく基礎的な解析研究から構造・機能予測法を開発し、応用研究を広範に実施しました。それらの成果物と既存手法を組み合わせ、タンパク質の構造・機能アノテーション

システムを開発してヒト22番染色体由来の配列に適用し、成果をSAHGとしてまとめました。

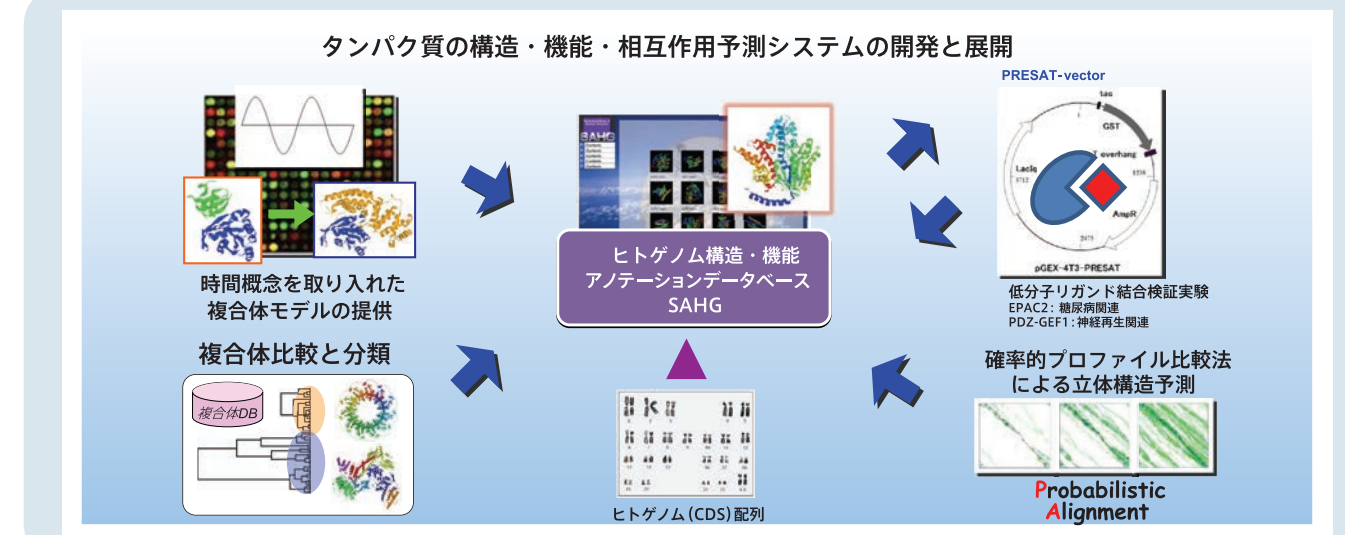
継続2年では複合体構造の比較や相互作用データの取り扱い、モデリング、タンパク質の運動の記述と比較等について広範に基礎研究を行い、成果をSAHGにて公開しました。

SAHGはタンパク質の構造、動き、分子機能、相互作用を視覚情報として提供できるので、ヒトの機能ゲノミクスを実施する際に有用な初期情報を得ることができます。さらにタンパク質ネットワークや細胞機能レベルのアノテーションに拡張することも可能であり、細胞機能の記述と理解に向けて利用価値の高いデータベースといえます。

SAHGのデータを使い倒す

SAHGの解析結果を有効に利用してもらうためには、まずは自分たちがSAHGの最初のユーザとなり、その有効性を実証していくことが大切であると思っています。SAHGにはタンパク質と低分子リガンド結合の予測結果があまねく提示されていますので、それらの中から新規な結合様式を選択し、NMRによる検証実験を行っています。また、SAHG用に開発した自動アノテーションパイプラインをヒト以外のゲノムに適用することで、高度な比較ゲノム研究を支援することも計画しています。

Email : mota@is.nagoya-u.ac.jp





代表研究者

川 戸 佳

K a w a t o S u g u r u

東京大学大学院総合文化研究科 教授

創造的な生物・情報知識融合型の研究開発 H17～20



脳内の神経シナプスを可視化！精神病治療薬の開発につなげる

記憶の貯蔵庫「スパイン」を自動的に検出するプログラムの開発

脳に存在する膨大な数の神経細胞ひとつひとつには、数十の樹状突起と数万のスパイン(=シナプス後部)と呼ばれる部分があり、記憶はこのスパインに書き込まれています。これまでの研究により、スパインはうつ病やアルツハイマー病にかかると減少し、神経栄養因子(ホルモン)を投与すると変化し、形も変化することがわかっています。そこで、スパインの数や頭部形を解析できれば、さまざまな神経疾患の原因を解明したり、治療法の開発につなげられると考えられます。従来、スパインの解析にあたっては人間の目で一つ一つ検出するしかなく、莫大な時間と労力がかかっていました。

本研究開発課題では、樹状突起とスパインの図形的な特徴を数理解算で抽出し、自動的に解析できる画期的な神経シナプス自動解析プログラム「Spiso-3D」の開発を行いました。

Spiso-3Dがもたらした新たな発見

本研究で開発した神経シナプス数理解算プログラムSpiso-3Dは、神経のスパインと樹状突起からなる構造体を輝度関数の微分を行い、画像のトポロジー的特長を抽出することで自動数理解析を行います。これまでにないアルゴリズムを用いたこのソフトウェアにより、3次元スパインの自動検出が可能になりました。Spiso-3Dを

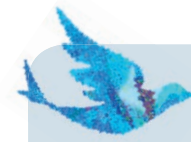
用い、様々なスパインを解析した結果、「急性的な男性・女性ホルモンの投与やストレスホルモンによってスパインが増加したり頭部が肥大化すること」「スパインの変化は、MAPK、PKA、PKCがうまく組み合わさった経路により動かされていること」「遺伝子改変マウスの神経スパインの振る舞いの異常」等の現象を新たに発見しました。ここでスパイン頭部が大きいと記憶力は高いのです。成果はCerebral Cortex(2011)に出版されました。

この成果により、現代社会の緊急の課題である、老化に伴う性ホルモン低下による更年期型痴呆症やストレス起因のうつ病等の課題に対し、その原因を解明でき、さらに治療薬の作用の新たな解析法を提供することが期待されています。

Spiso-3Dの活用で精神神経疾患のシナプス治療法開発

Spiso-3Dは研究者のみならず、製薬会社のニーズも高く、Cerebral Cortex出版後、購入希望が米国からも複数届いています。共同研究・応用研究も大きく進んでおり、老化によるスパイン密度の低下と頭部の縮小をホルモン補充療法で回復できるか解析中です。また、アルツハイマー病治療薬でスパイン密度を回復できるか解析中です。共焦点顕微鏡で実時間解析も行い、大型スパインも10分程度で大きく屈曲し消滅することが見えてきたので、薬物の作用が実時間で見えます。

Email : kawato@phys.c.u-tokyo.ac.jp



代表研究者

塩 田 浩 平

S h i o t a K o h e i

京都大学 理事・副学長

創造的な生物・情報知識融合型の研究開発 H17～20・H20～22



ヒト胚をMR顕微鏡で撮影！ゲノムワイド研究のためのデータベースを構築

ヒト胚の各発生段階の精細な画像と情報を統合的に整理する

ヒト胚の発生過程には多くの遺伝子が時期特異的・部位特異的に発現しています。こうした分子情報は断片的なものにとどまっていますが、今後は遺伝子群の相互作用や形態形成における遺伝子機能の全体像を解明するためのゲノムワイドなアプローチが必要です。そのためには胚の遺伝子発現マップを作成し、その時間的・空間的パターンを解析することが不可欠です。

本研究課題の初期3年では、ヒトの発生と先天異常解析のインフラとして、京都大学のヒト胚標本コレクションを用いて各発生段階のヒト胚をMR顕微鏡で断層撮像を行い、詳細な三次元画像データベースを構築し、ヒト胚組織における各種遺伝子の発現データをマッピングするための形態学的なリソースを整備しました。

継続2年ではそれまでの成果を発展させ、高磁場のMR装置と新規のイメージング手法「EFIC」法を用い、発生早期のヒト胚の超高解像度の画像データベースを構築しました。

世界初のヒト胚形態発生データベース

初期3年では、1200例のヒト胚標本のMR撮像を行いました。これらのMRデータを各症例の写真や臨床データとともにデータベース化し、各標本の検索表示システムを構築しました。このデータベース

は類例のない器官形成期正常ヒト胚を最高水準の解像度でとらえており、また非破壊的に撮影したため将来の遺伝子発現データと対応できるという特色があります。

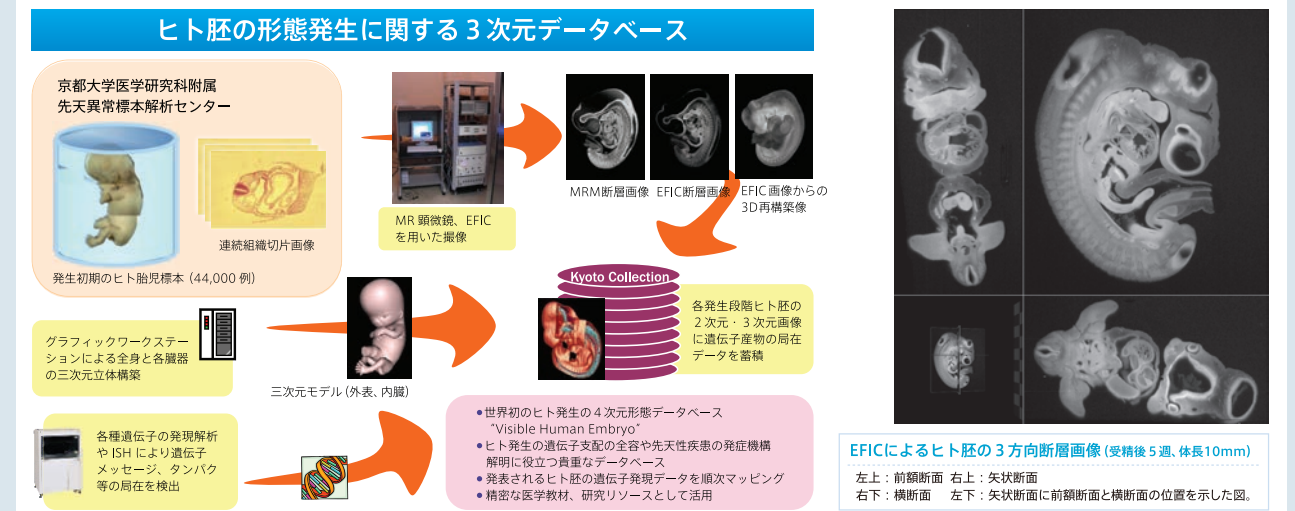
継続2年では、MRおよびEFICの画像に加え、組織標本のデジタルデータ画像をさらに高解像度で表示できるようになりました。CGを駆使したヒト胚の3次元モデルも、実際のMR計測データを使うことで極めて正確で高質なものを作成できました。

本研究は米国の「Visible Human Project」に匹敵するもので、世界初のヒト胚の形態発生データベースを構築したことから、わが国独自の知的財産を形成するという意義も大きいといえます。

ヒト発生メカニズムと先天異常解明への活用

本データベースは、器官形成期ヒト胚における遺伝子群の時間的・空間的発現パターンを解析するポストゲノム研究のための貴重な研究リソースであり、ヒト発生に関わる遺伝子の機能や遺伝子群の相互作用の解明に利用できるほか、原因不明の多いヒト先天異常の分子遺伝学的解析と発症メカニズムの解明にも資するものです。本データベースは、Webアトラスとして公開し、関連の研究利用に供します。また、医学生物学の教育、臨床における患者教育などにも利用価値が高いと考えられます。

Email : shiota@hq.kyoto-u.ac.jp





代表研究者

白 井 剛

Shirai Tsuyoshi

長浜バイオ大学バイオサイエンス学部 教授

創造的な生物・情報知識融合型の研究開発 H17～20・H20～22



生命活動の理解に必須な「超分子複合体」のモデルを作成する！

モデリングの各工程に必要なツール群を同一の仕様で開発

ノーベル物理学賞を受賞した湯川秀樹博士は「生物学は積み木細工である」と指摘しました。これはまさに慧眼であり、生物学の発展によってタンパク質・DNA・RNA等の生体分子は積み木の様に寄り集まって超分子複合体という巨大な構造を形成して働き、生命活動を支えていることがわかりました。しかし、この超分子複合体を効率よく実験的に構造解析する手段はありませんでした。この課題を計算機を使って解決するため、5年間を通して超分子複合体のモデリングツールの開発に挑みました。

初期3年はゲノムの複製や維持に必須な3R(DNA複製・修復・組み換え)超分子複合体の立体構造をX線結晶解析と電子顕微鏡で解析しながら、その過程で必要なのに欠けている超分子複合体モデリングツールを開発・整備しました。

継続2年はモデリングを分子ネットワークレベルまで拡張し、ネットワークモデリングに適したツールを開発して、複合体構造の解明を目指しました。

利用できる情報量が増えるほど、高精度のモデリングが可能に

初期3年では、未解明の複合体が多い古細菌を対象に研究を進め、ゲノムの複製や維持に必須な3R系で要となる超分子複合体

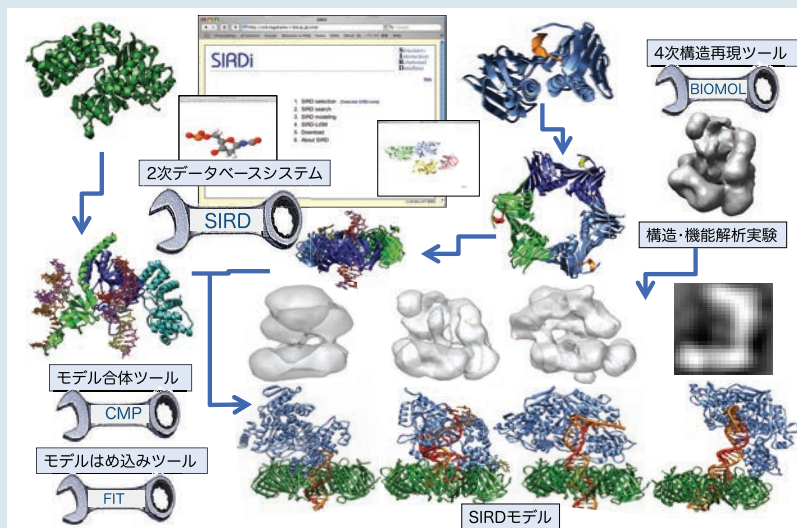
「DNAポリメラーゼ-PCNA-DNA」や「DNAリガーゼ-PCNA-DNA」の構造解析に初成功し、ツール群が作成したモデルの有効性が示されました。モデリング結果の妥当性検証のため3R複合体の機能・構造解析実験をし、PCNA複合体の解析も行いました。

継続2年ではツール群を改良してモデリングの基本過程を完遂できるシステムを構築し、平行して行った生化学・構造解析実験では、3R複合体の機能・構造の新知見を多数得ています。また、PDBを全自動で1次データベースから2次データベースへプロセスするSIRDシステム(PDB2次データベース構築検索システム)をWeb公開しました。これらは、急速に膨張し続ける生体分子情報を自動的に整理・活用する情報技術としても大きな前進です。

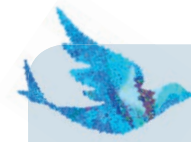
構造メタボローム2次データベースシステムSIRDmへ展開

立体構造が持つ重要な情報として、タンパク質-リガンド相互作用があります。BIRDの開発では、タンパク質だけでなく低分子リガンド構造分類にも力を注いできました。この基盤を利用して、SIRDをプロテオーム-メタボローム相互作用構造解析に適したシステム(SIRDm)へ発展させることが次の目標です。基質チャネリング機構解析、プロテオーム-エピゲノム-メタボローム相互作用解析、代謝超分子複合体モデリングなど、SIRDmの活躍できる分野は大きいと考えています。

Email : r_shirai@nagahama-i-bio.ac.jp



この研究開発により、2次データベースシステムSIRDを中心とした超分子複合体モデリングツールができました。このシステムを使って、PDBからDNA複製・修復・組み換えの要となる3R複合体のSIRDモデル(図下段。右からPCNA-ヘリカーゼ、リガーゼ、校正ポリメラーゼ、伸長ポリラーゼ-DNA複合体)を構築したところ、モデルは電子顕微鏡解析実験で得られたデータ(図中段)と良く一致することが示されました。



代表研究者

宮 川 剛

Miyakawa Tsuyoshi

藤田保健衛生大学総合医科学研究所 教授

創造的な生物・情報知識融合型の研究開発 H17～20



マウスの行動をテストで解析！脳のメカニズム解明に挑戦する

ヒトの脳を調べるには、マウスの行動解析が有効

ポストゲノム時代の現在でも、遺伝子の個体レベルでの機能はほとんど分かっていません。遺伝子機能を理解するには、全遺伝子の80%以上が発現するといわれる脳を研究することが肝要です。脳を研究するためには、そのアウトプットである「行動」を解析する手法が有効だと考えられます。

本研究課題では、ヒトの脳を統合的に理解するため、マウスの行動解析を行いました。マウスは、遺伝子の99%にヒトのホモログ(対応する遺伝子)があり、その遺伝子を自在に操作することが可能です。また、心理学的な解析をはじめとして個体レベルでの多彩な解析もできることから、ヒトの脳を理解するためのモデル動物として最適であると考えられます。

本研究では、マウスの網羅的行動テストを用いた解析から得られる脳機能表現型データをインフォーマティクスの方法を用いて解析・利用可能にすることを目的として、データ取得手法の標準化とデータベースの開発・整備・公開を行いました。

外部と積極的に連携し、解析の高度化・効率化を進める

これまでに国内外60の研究室と共同研究を行い、90系統5804匹の遺伝子改変マウスを自動化された行動テストで解析して87系統で新たな表現型を発見しました。行動テストでは164の指標を

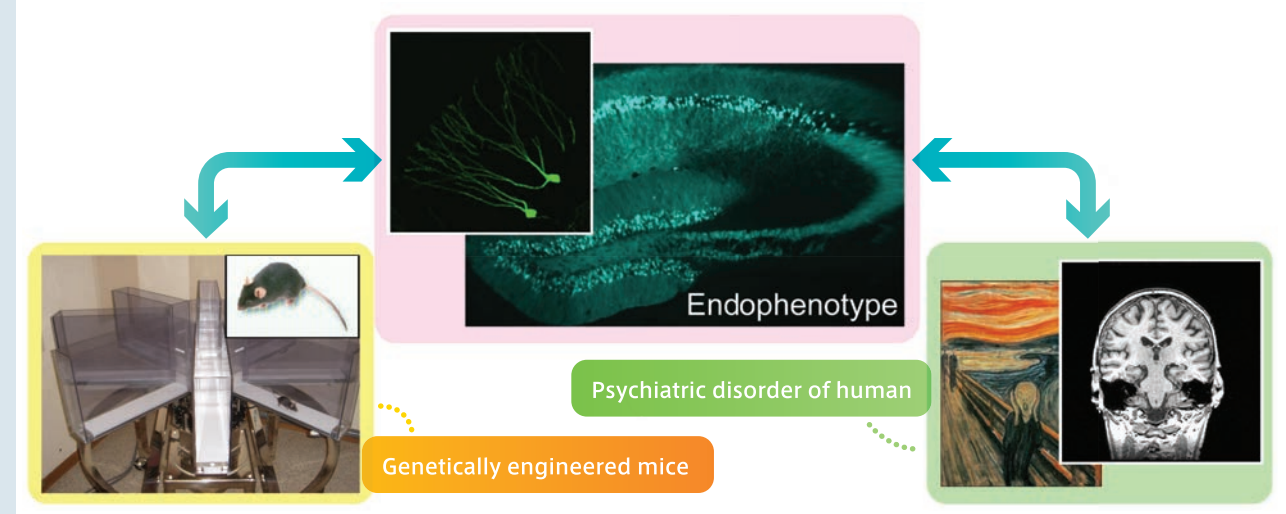
データ登録しました。これらの行動データと脳内の表現型の情報を整理・統合することで、遺伝子の機能が解明できると考えられています。開発したデータベースやプロトコル、解析サンプルのマウス等は積極的に外部へ提供し、他の研究機関でも同様の実験ができる体制を整えました。

この実績を基盤に、マウスを使った研究を起点としてヒト脳機能を総合的に研究する全国的な組織の形成が期待されます。またゲノム情報やヒト脳機能に関する他のバイオインフォーマティクスとの連携、新たな組織化が可能となり、ヒト脳機能の総理解への新たな展望も開けています。

遺伝子・脳・行動の統合的理解へ向け

今回のプロジェクトでは、多くの系統の遺伝子改変マウスの行動を解析してデータベース化しましたが、実は、これらのマウスの脳を保存・リソース化して今後の解析に使えるような状態にしています。「網羅的行動データ付きマウス脳バンク」と称しているのですが、これらの脳では遺伝子発現やタンパク発現などを調べることが可能です。今後、行動のデータベースはもちろん、これらの脳や遺伝子改変マウスそのものをあわせて活用した研究によって、遺伝子・脳・行動の関係の統合的理解が進むことが期待できます。

Email : miyakawa@fujita-hu.ac.jp





代表研究者

石 井 信

I s h i i S h i n

京都大学大学院情報学研究科 教授

創造的な生物・情報知識融合型の研究開発 H18～21



代表研究者

黒 川 顕

K u r o k a w a K e n

東京工業大学大学院生命理工学研究科 教授

創造的な生物・情報知識融合型の研究開発 H18～21・H21～23



生物の形態形成を予測できるモデルの構築に成功！

実験が難しい分野のデータ収集には、予測が有効な手段となる

生物を形作る分子は、有機的に役割分担をすることで複雑な機能を発揮しています。しかし、実験の難しさゆえに、どの分子がどのタイミングで、どの分子とどのように役割を果たすのかを知るためのデータは非常に限られていました。

そこで本研究課題では、生命システムに関して不足する情報を提示できる予測技術の開発を目指し、注目する分子のデータからその機能を予測する技術、あるいは情報抽出(コード)・解読(デコード)のための技術を開発しました。本研究の成果は実験への仮説提供、分子から機能までのプロセス同定、機能発現のためのミニマムモデルの提示等を実現します。

特に、生命機能の中でも生物の形態形成を対象としました。技術開発に際しては、主に2つの生物現象を観察しました。1つは神経細胞の形状が偏りを持つ「極性形成」、もう1つは卵細胞が分裂する過程で細胞の塊が等間隔にできる「体節形成」で、それぞれの現象においてShootin1およびHes7遺伝子がどのような役割を担っているかを予測しました。

予測結果は実験的に検証し、さらに理解を深める

生命現象を少数のパラメータで表現したミニマムモデルで表現することで、メカニズムの解釈を分かりやすく単純化し、物理的な挙動

を予測しました。

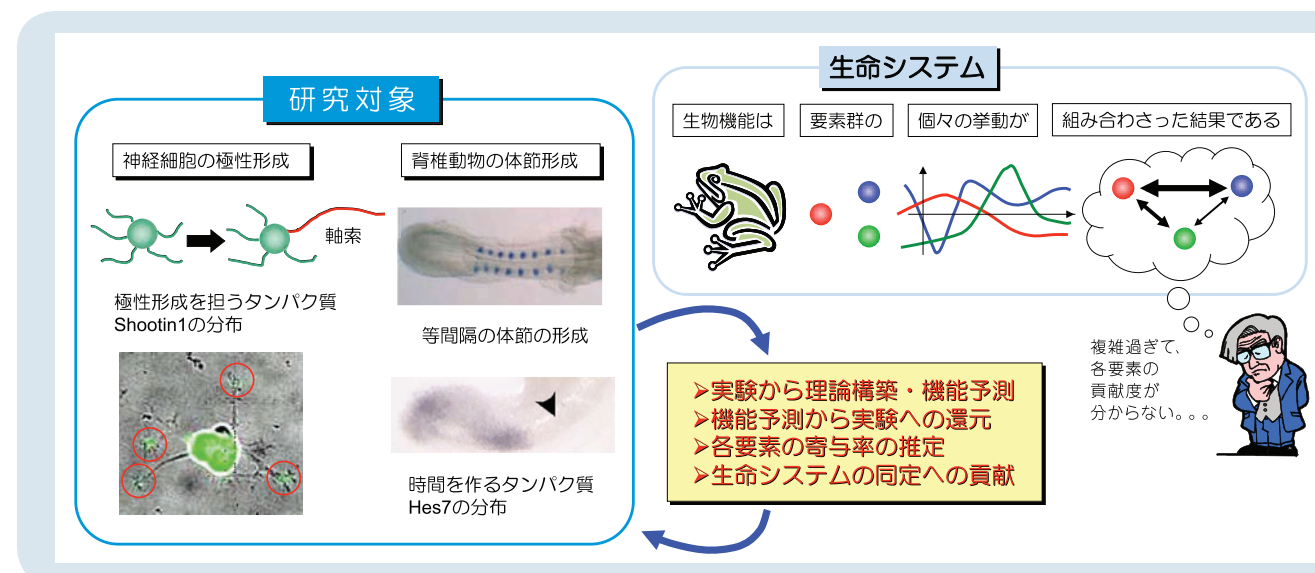
神経の極性形成では、Shootin1分子が濃くなった時に神経突起が伸びることがよく観察され、突起が縮む傾向やShootin1の濃さが増加する傾向等を解析した結果、神経の突起伸長を予測するモデルを作ることができました。このモデルを用いることで、コンピュータ上での突起伸長予測が実現されました。

体節形成では、Hes7分子がメトロノームのように時間を正確に刻むことで体節を等間隔で形成することを観察し、「時計遺伝子振動モデル」を構築しました。さらにこのモデルを用いたシミュレーションでの予測結果を、実際の発生で確認することもできました。

分子活動のライブイメージングにより生物の形態形成を予測する

これまでの研究開発によって、神経極性形成と体節形成における分子制御のミニマムモデルの構築とそれを用いた予測ができるようになりました。しかし、形態形成に関わる複数の分子の活性時系列から形態の将来を予測し、またそれら分子の関係を同定できることに至っていません。例えばライブセルイメージングの手法により、複数種の分子の動態をオンラインに観察し、形態形成の将来を予測し、それらの関係を同定するための情報学的手法の開発を今後も進めていく予定です。

Email : ishii@i.kyoto-u.ac.jp



細菌の集まりを丸ごと解析！有用な知見発見に挑む

あらゆる生態系の根幹を成す細菌叢をゲノム解析する

地球上のあらゆる環境には、多様な細菌が群集(細菌叢)を形成し生存しています。生態系の根幹を成す細菌叢の生命システムを解明するために、近年、細菌叢を丸ごとゲノム解析する「メタゲノム解析」が進んでいます。

初期3年では、メタゲノム解析で得られた大量のデータから有用な知見を発見する統合解析システムの構築を目的に、短い塩基配列断片からも正確に遺伝子領域を予測できるソフト(MGA)の開発を行いました。また、メタゲノムデータ比較のための技術開発を行い、開発した技術をヒトの各部位におけるメタゲノム解析に応用することを目指しました。

継続2年では初期3年で開発したシステムを発展させ、メタゲノム解析のワークフローを確立するとともに、細菌叢を取り巻く環境因子の要素を記述したメタデータとメタゲノムデータを統合し、メタ比較解析を可能とするシステムの構築を目指しました。

開発したシステムを、ヒト体内の細菌叢解析に応用

初期3年では、MGAで予測した遺伝子群から細菌叢の全代謝経路と細菌間の代謝相互作用を解明するため、メタゲノム遺伝子の分類群情報を付加した上で代謝マップ上にマッピングするシステムと、その結果を可視化するソフトを開発しました。これらの技術を用いて、

乳児から大人まで13名の腸内細菌叢の持つメタゲノムおよび全遺伝子情報を解析し、新たな知見を得ました。

継続2年では、ヒトのメタゲノムに特化した「Human Meta BodyMap」構築を目指し、環境因子のメタデータとメタゲノムデータを組み合わせた解析技術を開発したほか、すべての細菌が持っている16S rRNA遺伝子を利用した群集構造解析用ツール「VITCOMIC」も公開しました。

本研究で開発した技術は、メタゲノム解析だけでなく、食品加工における衛生管理、品質管理等広範な分野においての活用が期待されています。

メタゲノムデータは配列を有する表現型データ

膨大なメタゲノムデータが猛烈な勢いで公開されています。メタゲノム解析によって得られるデータは、そのほとんどが塩基配列データです。しかし、メタゲノム解析における最重要データは、膨大な塩基配列(遺伝型)そのものというよりも、それら遺伝子が存在した環境に関する情報(表現型)です。今後は、巨大化するメタゲノムデータを表現型により整理するとともに、本研究で培った解析技術、可視化技術をもさらに発展させることで、生態系の根幹を成す細菌叢の生命システムの解明に貢献していきたいと考えています。

Email : ken@bio.titech.ac.jp





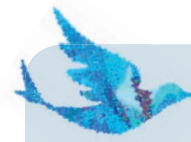
代表研究者

榊原 康文

Sakakibara Yasubumi

慶應義塾大学理工学部 教授

創造的な生物・情報知識融合型の研究開発 H18～21・H21～23



代表研究者

清水 謙多郎

Shimizu Kentaro

東京大学大学院農学生命科学研究科 教授

創造的な生物・情報知識融合型の研究開発 H18～21・H21～23



バイオインフォマティクスの予測技術を駆使して、創薬の鍵となる「リード化合物」を発見！

数千万化合物データベースからの網羅的リガンド予測

本研究では、創薬の初期ステップとなるリード化合物の探索において有用な、タンパク質と化合物間の相互作用予測を、網羅的かつ効率的に実行するための手法の開発を行いました。

初期3年では、汎用性が高く入手が容易な、アミノ酸配列データおよび化合物構造データを入力として、タンパク質化合物相互作用予測を行う手法の開発に取り組みました。予測結果をケミカルバイオロジーなどの実験手法を用いて検証・反映する「フィードバック戦略」を採用することで、予測精度のさらなる向上とデータベースの構築を行いました。

継続2年では初期3年の研究の発展的継続開発を実施するとともに、予測の機能だけを取り出して化合物探索に特化したChemical BLASTの開発と、新たにがん関連タンパク質に対するリード化合物の探索とその機能解析を目指しました。

研究の成果をがん治療薬探索につなげる

初期3年では、SVMという統計学的手法を適用して、相互作用予測システム「COPICAT」を開発しWeb公開しました。さらにフィードバック戦略を応用して、前立腺がん発生に関わるアンドロゲン受容体を阻害する新たなリード化合物を同定しました。

継続2年では、COPICATに学習モデルライブラリを組み込

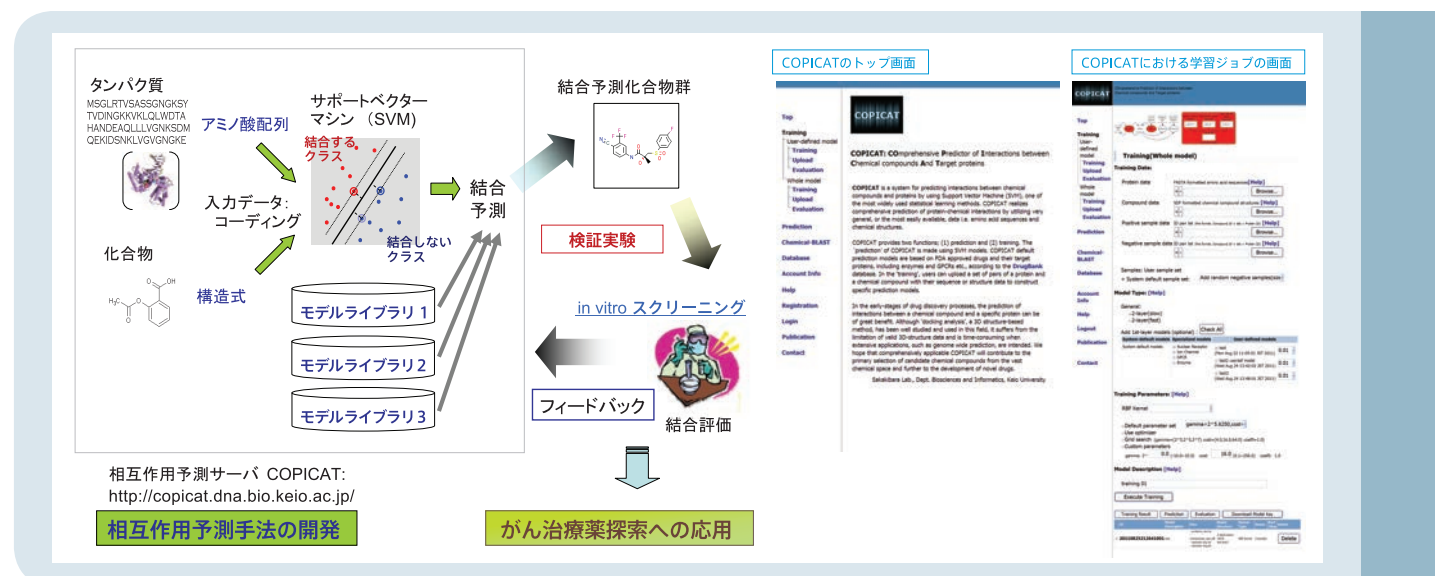
み、ユーザの学習モデル構築に対して利便性を向上させました。Chemical BLAST開発では、相互作用の情報を考慮した化合物の特徴空間を構築する手法が有効であることを確認し、PubChemデータベースからの化合物探索を可能としました。またフィードバック戦略をアポトーシス抑制タンパク質の解析に適用し、その機能を阻害する既知化合物のターゲットタンパク質であると推定される新しいタンパク質を同定しました。

本研究の成果は、創薬やがん治療薬探索のさらなる発展に寄与することが期待されています。

リード化合物探索におけるBLASTを目指して

PubChemをはじめとする化合物データベースには何千万という化合物が蓄積され、また有用微生物から天然化合物の網羅的発見が続く中、創薬におけるインシリコスクリーニングの役割はますます大きくなっていくと考えられます。本研究課題で開発した相互作用予測システムCOPICATがこの分野において、研究者がBLASTのように日常的に使用する第一のツールとしてその地位を確立し、多様性が増大するケミカルスペースにおける羅針盤となるように、システムの改良と発展を目指していきます。

Email : yasu@bio.keio.ac.jp



タンパク質の相互作用解析から、生命システムを解明する

3つのアプローチから高精度の予測システムを開発

生体内のタンパク質は、多くがタンパク質と他の分子(タンパク質、リガンド、核酸など)との相互作用で機能を発揮していることから、生命システムの解明には相互作用の解析もしくは予測が肝要です。しかし実験による解析には多大な時間とコストがかかり、一方、従来の予測法では解析を行うのに十分な精度を得ることが難しいという問題がありました。

そこで本研究課題初期3年では、相互作用部位の予測、ドッキングの予測、高精度の複合体モデリングと物理相互作用解析の3つの観点から、高速かつ高精度な相互作用予測・解析システムを開発しました。

継続2年では、より精度が高く、タンパク質間相互作用予測、相互作用部位予測、ドッキング予測、高精度相互作用解析の統合的なシステムの開発と、データベース構築を目標としました。また、多様な電子伝達系を有する酵素「芳香環水酸化ジオキシゲナーゼ (ROS)」について、特異的相互作用、電子伝達経路の解明に関する研究も行いました。

予測手法の開発から、発展的な研究開発へとステップアップ

初期3年では、アミノ酸配列情報のみを用いて2つのタンパク質の相互作用予測を行う手法を開発し、また配列情報と構造情報の両方

を用いてタンパク質相互作用部位の予測を行う手法の精度の大幅な向上を実現しました。また、タンパク質-リガンド相互作用部位予測手法の精度向上も実現しました。ドッキング予測については、独自の予測法で従来の16倍から160倍もの高速化を実現しました。

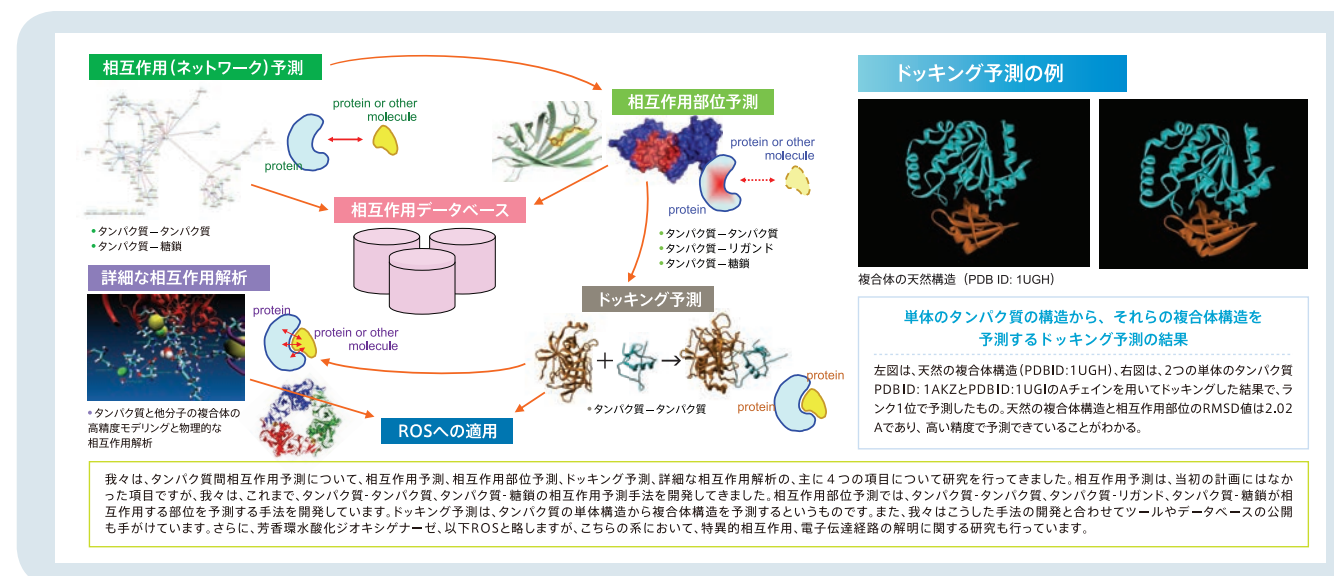
継続2年では、それまでの技術を発展させるとともに、従来より予測精度の高いタンパク質-糖鎖の結合予測手法を開発しました。また、タンパク質の4次構造を考慮したタンパク質-リガンドの相互作用データセットも構築し、データベースの構築にも着手しました。

相互作用の解析は、生命現象の解明に大きな役割を果たすものと期待されます。

総合的なタンパク質間相互作用予測・解析システムを完成させる

タンパク質とタンパク質、タンパク質と糖鎖、脂質、その他のリガンドとの相互作用予測を統合的に行えるシステムを構築し、一般に公開する予定です。実験結果をフィードバックさせ、予測精度を高めるとともに、それらをデータベース化し、知識を抽出するシステムに発展させていきます。また、必要に応じて、分子シミュレーションを駆使し、相互作用を細かく解析したり、リガンド結合に伴うタンパク質の動きを解析したりする方法を確立したいと考えています。

Email : shimizu@bi.a.u-tokyo.ac.jp





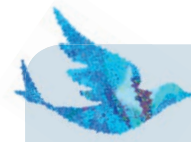
代表研究者

中 井 謙 太

N a k a i K e n t a

東京大学医科学研究所 教授

創造的な生物・情報知識融合型の研究開発 H18～21



代表研究者

長 野 希 美

N a g a n o N o z o m i

産業技術総合研究所 生命情報工学研究センター 主任研究員

創造的な生物・情報知識融合型の研究開発 H19～22



遺伝子の転写制御情報の解明に挑む！

鍵となるのは、3つの視点

ゲノム塩基配列に書き込まれた遺伝情報、中でも転写制御情報の解読は非常に困難であり、現代分子生物学の最大の課題の1つといえます。本研究は、単細胞生物からヒトにわたる様々な生物種のゲノムにおいて、遺伝子の転写制御情報の原理を解明すること、そこから発展して人工のプロモーター（遺伝子の転写を始める領域）設計の指針になる理論モデルを作ること为目标として進められました。

転写制御情報の原理を解明するには、転写制御領域のデザイン原理を理解する必要があります。そこで3つの視点（プロモーター比較に基づく進化的視点、共発現遺伝子制御領域のモデル化による数理的視点、プロモーター活性の網羅的データ解析に基づく工学的視点）から研究を行い、デザイン原理をより直接的に探求し、最終的にはこれらの研究を総合して、普遍的な原理を発見することを目指しました。

「プロモーター」から読み解く転写制御

進化的視点では、マウスとヒトの選択的プロモーターの進化的保存度を網羅的に比較して、多くの選択的プロモーターの保存度は低いという結果を我々の転写開始点データベース（DBTSS）の解析から見出しました。

数理的視点では、ホヤ等の組織特異的発現のプロモーターをモデ

ル化し、そのモデルを使って筋肉特異的に発現する未知の遺伝子をゲノム中から探索しました。また685個の出芽酵母のプロモーターを解析し、転写因子間の新たな相互作用を発見しました。

工学的視点では、ヒトの培養細胞を用いて転写開始点の頻度を測ることでプロモーターの活性を実測し、配列からの予測値との間に十分強い相関を示すことができました。

これらの研究過程で生み出されたデータベースやウェブツール等を積極的に一般公開するとともに、19報の論文も発表しました。

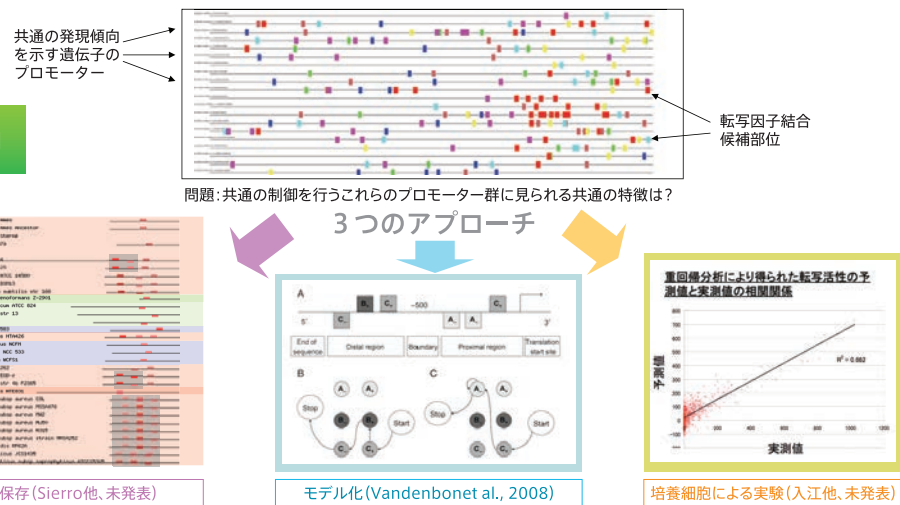
新たな展開を示す転写制御領域の理論研究

本プロジェクトで我々が目指した転写制御領域の構築原理の研究は、プロジェクト終了後も新たな展開を示して発展中です。すなわち、次世代シーケンサーを用いたChIP-seq法などの発展により、転写因子の網羅的な結合状態や、ヒストンの修飾状態などがわかるようになり、当時よりも理解が大きく前進しています。今後、我々の研究は、未知のRNA遺伝子の探索や合成生物学の発展、ひいては再生医療における細胞状態の変化の理解などにも寄与して行くことでしょう。

Email : knakai@ims.u-tokyo.ac.jp



転写制御領域の構築原理の解明



酵素の機能分類を自動化する！「酵素反応予測システム」

立体構造から酵素機能を予測する

酵素とは、生物の細胞内で様々な化学反応を助ける蛋白質です。酵素は5000種類以上あるとされ、創薬等の産業応用において重要であり、これを有効活用するためには、機能別に分類しておく必要があります。機能分類には蛋白質の立体構造が重要ですが、酵素のID番号として世界的に普及しているEC番号では、立体構造が考慮されておらず、その機能を正確に把握できません。

このEC番号の問題点を踏まえ、これまでに酵素データベース「EzCatDB」を開発して、酵素の活性部位を登録し、独自の酵素反応階層分類（RLCP分類）を行ってきました。

本研究では、構造ゲノミクスで出されている膨大な蛋白質の立体構造のデータから機能予測するために、酵素の活性部位の立体構造と触媒機能の相関関係から、酵素機能を予測して自動的に分類する「酵素反応予測システム」の開発を行いました。

類似反応予測システム「EzMetAct」の開発

蛋白質の局所構造比較によって活性部位を予測する際、従来はRMSDという原子間距離を単純に平均した尺度が用いられていました。ここに機械学習技術を導入して、高精度の活性部位予測を試みました。その結果、局所構造比較による活性部位予測に用いられる2種類の解析（シングル・テンプレート予測、マルチプル・テンプレート

予測)のためのアルゴリズムを開発しました。

実際の立体構造を使った予測実験では、2種類の解析のどちらの場合も、開発したアルゴリズムにより、従来法を有意に上回る性能向上となる結果を得ました。この予測アルゴリズムを基に、「酵素反応予測システム」のプロトタイプを開発しました。

具体的には、EzCatDBに登録されている活性部位の情報をを用いて、蛋白質の立体構造データをもとに、類似反応予測システム「EzMetAct」を開発しました。これはクエリとなる立体構造データに対して局所構造比較を行い、データベース中の活性部位テンプレートとの一致度を統計的なスコアを用いて、Web上に表示するシステムです。

酵素反応予測、蛋白質機能予測への大規模展開

現在、EzMetActが対応している活性部位テンプレートの数は、200種類程度なので対応できる酵素の種類は限定的ですが、今後、非蛋白質原子なども活性部位に導入できるシステムに改良するなど、活性部位の種類を増やすことによって、多様な蛋白質機能予測へ展開したいと考えています。特に、近年の構造ゲノミクス研究で、立体構造が解かれても機能が未知の蛋白質のデータは数多くあるので、本システムを展開する事ができれば、より大規模な機能予測を行う事が出来ると考えられます。

Email : n.nagano@aist.go.jp



酵素反応予測システム

蛋白質の酵素反応の予測は機能既知の酵素の反応部位を切り取ったテンプレートと比較することで行う。本研究開発では、計算機に「学習」とよばれる知的処理をさせて比較の高度化を行う。

EzCatDB
RLCP分類

機能既知酵素

酵素反応部位テンプレート

予測

機能既知酵素

テンプレートと学習結果

学習

テンプレート

学習前

学習後

学習前

学習後

学習前

学習後

学習前

学習後



代表研究者

馳 澤 盛 一 郎

Hasezawa Seiichiro

東京大学大学院新領域創成科学研究科 教授

創造的な生物・情報知識融合型の研究開発 H19～22



「バイオメディカルイメージング技術」の高度化を目指す！

多種多様な画像に対応できる適応的画像分類法

生命現象を画像情報として捉える「バイオメディカルイメージング技術」は、微細な分子構造から地球規模の生態系に至る幅広いスケールをカバーする、生命科学・医学における中心的な解析手段です。近年、可視化技術や撮像法の発達にともない、画像データの多様化・大規模化が進み、従来の解析法では処理しきれなくなっています。また、技術の進展によって極めて多彩な画像が生み出され、従来の画像分類アルゴリズムによるアプローチは困難になっています。

本研究開発では、多くの画像の種類や解析の目的に柔軟に対応できる、汎用性の高い画像分類や画像評価のシステムを構築することを目指して、新たなアルゴリズムに基づく適応的画像分類ソフトウェアに関する研究開発を行いました。

「進化型計算」と「自己組織化」が医療の発展を導く

生物進化のメカニズムを模倣してデータ構造を最適化していく「進化型計算」と、煩雑なデータを自動で秩序立てて整理する「自己組織化」を組み合わせ、そこに各分野の専門家の知識や判断基準を取り入れて最適な画像の評価基準（特徴）を探し出す「部分教師付学習アルゴリズム」の開発に成功しました。

このアルゴリズムを適用して光学顕微鏡画像から細胞内部の動きを解析したところ、ライブイメージング画像から研究者と同等の精度

で、細胞周期等の自動的判定ができました。

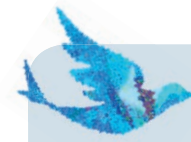
また、より複雑な高分解能MR画像データの診断でも画像分類法の有用性を示しました。さらにがん組織の画像診断に応用するため、マウス移植モデルの開発と画像の評価基準の検討を行いました。評価基準の測定に関する個別のプログラムはフリーの画像解析ソフトウェアImageJ用のプラグインの形態でWeb公開しています。

今回開発したアルゴリズムは、光学顕微鏡やMRI等の画像機器に搭載することにより、医薬品開発や医療画像の読影の省力化・高速化に貢献できると期待されています。

特徴量の適応性の向上とイテレイティブクラスタリングの高速化

今後は新たに採択されたJST先端計測の課題として、(1)特徴量の適応性の向上と(2)イテレイティブクラスタリングの高速化に取り組みます。(1)については、bag of visual words 法を用いて、特徴量候補を創出することで未知の画像に対しても十分に性質を記述できるようにします。(2)については、特徴マップ生成器として新たにEvolving Treeを導入し、クラスタリング速度を改善します。また、収束性能の向上をParticle Swarm Optimization (PSO)の導入により実現します。さらに、(3)企業との協働により本技術のイメージング機器への組込み、製品化を目指します。

Email : hasezawa@k.u-tokyo.ac.jp



代表研究者

馬 見 塚 拓

Mamitsuka Hiroshi

京都大学化学研究所 教授

創造的な生物・情報知識融合型の研究開発 H19～22



「非構造化データ」から、効率的に知識発見を行う！

「非構造化データ」から体系的に知識を発見する

生命科学におけるビッグサイエンスの潮流と実験技術の進歩により、多様なデータが大量に蓄積されつつあります。cDNAマイクロアレイのように表（実験条件×遺伝子）形式の「構造化データ」に留まらず、遺伝子配列のような「文字列」、低分子化合物の構造式のような「グラフ」や代謝パスウェイやタンパク質相互作用等の「ネットワーク」といった「非構造化データ」も含まれます。これら様々な「非構造化データ」から効率的かつ自動的に知識発見（マイニング）する手法は未発達で確立が切望されています。

本研究では、生命科学で得られた大量データのうち「非構造化データ」に着目し、データを統合しつつ効率的にマイニングする技術の構築を目指しました。

マイニング技術が拓く、バイオインフォマティクスの未来

一般にマイニング手法は、データに頻出する部分パターンを抽出する「頻出パターンマイニング」、データをグループ分けする「クラスタリング」、およびグループ化済みデータからグループ分け規則を抽出する「分類」の3つに分けられ、それぞれにおいて非構造化データ及びデータ統合の2点で、効率的な新規知識発見手法を開発しました。例えば、cDNAマイクロアレイデータ（構造化データ）と既知の遺伝子ネットワーク（非構造化データ）を統合し、遺伝子の「クラスタリング」また

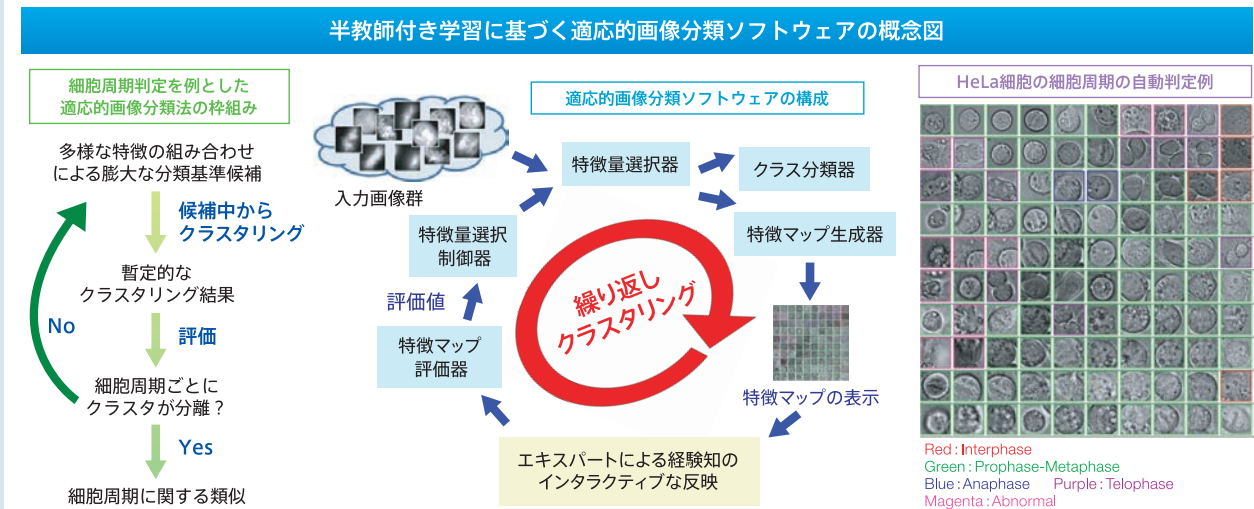
はマイクロアレイデータの「分類」を行う効率的かつ高精度なマイニング手法を確立しました。また、薬剤とターゲット（ともに非構造化データ）のペアデータから、薬剤・ターゲット結合のキーとなる部分構造ペアを効率的に発見する「頻出パターンマイニング」手法を開発しました。これにより薬効に重要な部分化学構造とターゲット部分の組み合わせが自動的に抽出され創薬効率化の重要な手がかりとなります。

上記の例を含め、生命科学の非構造化データから効率的にマイニングする様々な手法を確立しバイオインフォマティクスの発展に貢献しています。

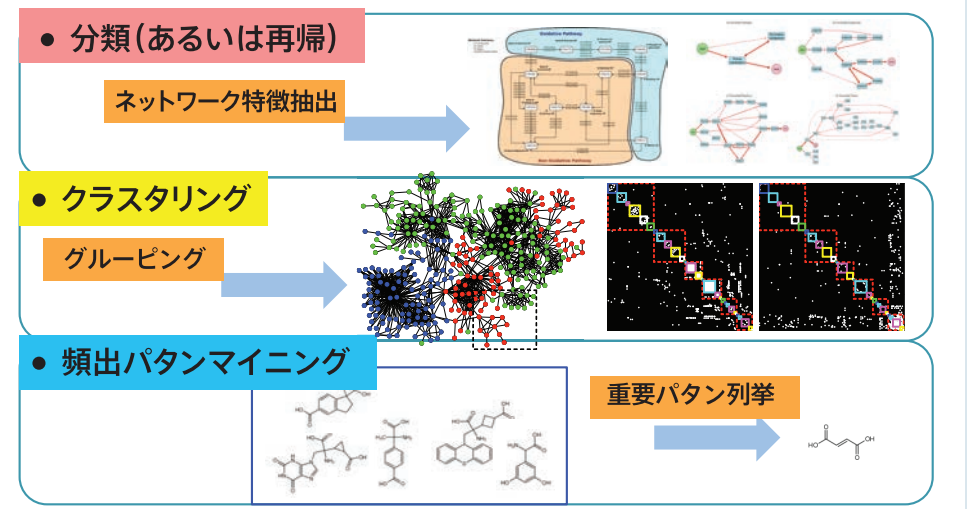
非構造化データのマイニングからシステムズバイオロジーへ

本研究の主眼は、遺伝子クラスタリングと化合物の頻出部分グラフマイニングでしたが、本研究を通じ（遺伝子等）ネットワーク上のマイニング手法を数多く構築し、ネットワーク上の動的な振る舞いを把握する、体系的な手法群を確立出来ました。一方、生命科学では、生体分子個々ではなく、分子同士が相互作用し生命活動を進めるネットワーク（システムズ）バイオロジーの考え方が浸透しつつあります。ネットワークマイニング手法をさらに深耕させシステムズバイオロジーから生命科学の進展へ貢献します。

Email : mami@kuicr.kyoto-u.ac.jp



非構造化データ
(グラフ)
マイニング





代表研究者

水口 賢司

Mizuguchi Kenji

医薬基盤研究所バイオインフォマティクスプロジェクト プロジェクトリーダー

創造的な生物・情報知識融合型の研究開発 H19～22



最新技術を融合して、膜蛋白質の立体構造を突き止める！

膜蛋白質の立体構造を高精度で予測する

膜蛋白質は、細胞内外での物質のやり取りや情報伝達、エネルギー合成等の重要な役割を担っており、多くの既知医薬品のターゲットになっています。近年、X線結晶構造解析によって、膜蛋白質の立体構造が次々に決定されていますが、膜蛋白質の立体構造解析には、「結晶構造の分解能が低い」「膜蛋白質の構造変化が不明である」「膜蛋白質複合体の立体構造が未解明である」「脂質分子との相互作用が明らかでない」等の様々な困難があり、多くの膜蛋白質について、より詳細な分子機能を解明するための妨げになっています。

そこで本研究では、実験的研究の弱点を補うべく、バイオインフォマティクスの手法と物理化学に基づく分子シミュレーションの最新技術を融合し、膜蛋白質の立体構造を、その膜蛋白質がどのように動くかも含めて高精度で予測する構造モデリング法の開発を目指しました。

膜蛋白質の機能および構造解析のための基盤的な技術

膜蛋白質のアミノ酸配列から、立体構造が変化を起こしやすい部位を予測する方法を開発しました。また、膜蛋白質の分子シミュレーションを実行するために必要となる、基本的な技術を複数開発しました。さらに、構造変化を網羅的に抽出したデータベース (MultiCoMP) を開発しました。

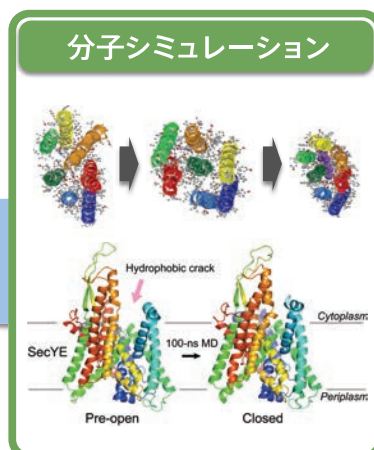
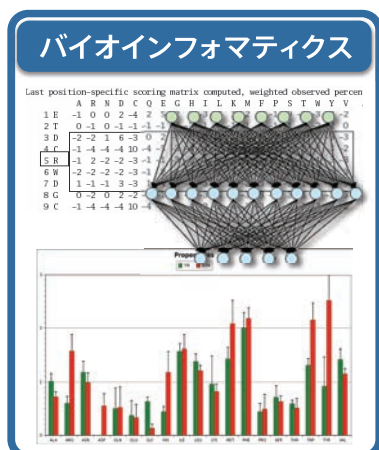
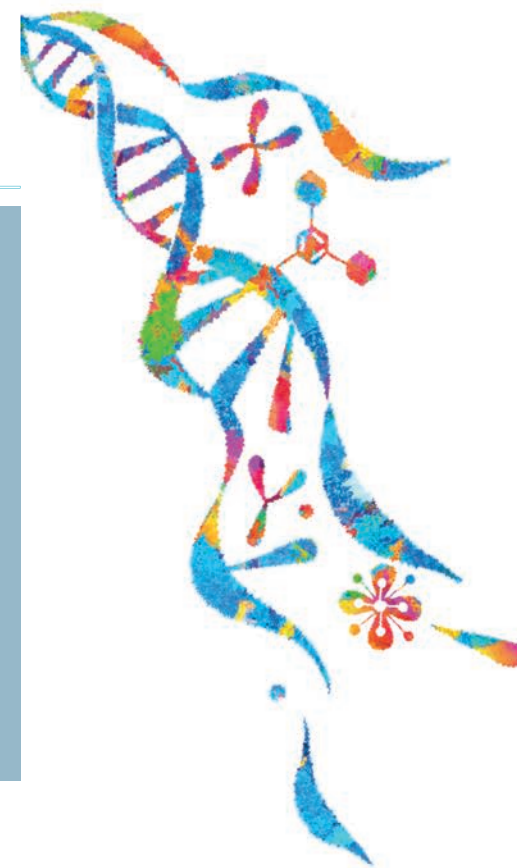
これらの技術の具体的な応用例として、従来の技術では難しかった、蛋白質膜透過装置であるSec複合体のモデリング解析を実施し、クロスリンクなどの生化学実験を用いて、この膜蛋白質複合体の構造変化に脂質分子の側方運動が関与する新しいメカニズムを発見しました。

今回開発した技術は、創薬ターゲットとなる膜蛋白質複合体の構造モデリングに応用でき、将来的には、阻害剤やアロステリック医薬の開発等の具体的なプロジェクトへの展開が期待されています。

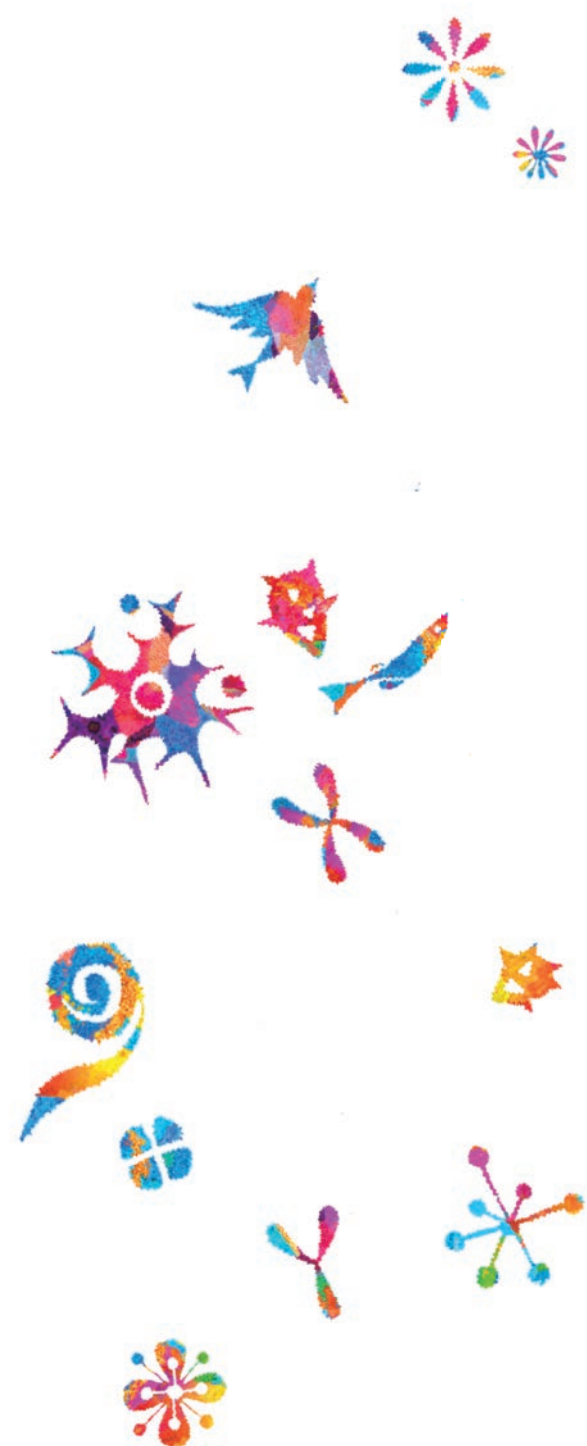
構造モデリングを創薬ターゲットに応用する

本課題で開発したバイオインフォマティクスおよび分子シミュレーションに関する基盤技術は、アカデミア、産業界を問わず、膜蛋白質に関わる多くの研究課題 (創薬ターゲット探索、分子モデリングや化合物ドッキング等を含む) に活用されることが期待されます。また、実験技術の進展によるデータ蓄積速度の加速に伴って、開発したデータベースが拡充され、膜蛋白質の構造や機能に興味を持つ幅広い分野の研究者に対し、基本的なリソースになると期待されます。

Email : kenji@nibio.go.jp



膜蛋白質の構造・機能の解析のための要素技術 → 革新的創薬へ！



参考資料

データベース、ツール一覧

主要なデータベースに関するデータ

ゲノムリテラシー講座

BIRD 年表

データベース、 ツール一覧

1. 研究機関が提供するデータベース

■ BMRB (BioMagResBank) 提供機関:大阪大学蛋白質研究所 URL: http://bmrb.protein.osaka-u.ac.jp/ 内容:生体高分子のNMRデータのデータベース
■ BSD (Bio Simulated Database) 提供機関:国立遺伝学研究所 URL : http://bsd.genes.nig.ac.jp/bsd_web/ 内容:マイクロアレイ、EST、SAGE データを扱う遺伝子発現解析システム
■ BUDDY-system (タンパク質・リガンド結合状態／非結合状態へデータベース) 提供機関: 東京大学 URL : http://www.bi.a.u-tokyo.ac.jp/services/buddy/current/ 内容: タンパク質の単体構造からリガンド結合状態の構造を結合しているリガンドとともに網羅的に検索可能
■ CIBEX (Center for Information Biology gene EXpression database) 提供機関:国立遺伝学研究所 URL : http://cibex.nig.ac.jp/index.jsp 内容: マイクロアレイデータを中心とする遺伝子発現データベース
■ CIPRO (Ciona intestinalis Protein Database : ホヤプロテイン統合データベース) 提供機関: 筑波大学 URL : http://cipro.ibio.jp/ 内容: ホヤのタンパク質に関する様々な情報を提供
■ COPICAT (COverprehensive Predictor of Interactions between Chemical compounds And Target proteins : タンパク質化合物相互作用予測システム) 提供機関: 慶應義塾大学 URL : http://copicat.dna.bio.keio.ac.jp/ 内容: タンパク質アミノ酸配列の部分配列の頻度情報および化合物構造における部分パスの頻度情報を組み合わせて多次元ベクトル化し、SVMを用いてそれらの相互作用を網羅的に予測するシステム
■ DBTBS 提供機関: 東京大学 URL : http://dbtbs.hgc.jp/ 内容: 枯草菌を中心とした転写情報のデータベース
■ DBTGR (DataBase of Tunicate Gene Regulation) 提供機関: 東京大学 URL : http://dbtgr.hgc.jp/ 内容: ホヤ等の被囊類の転写因子とプロモーター情報のデータベース
■ DBTSS (DataBase of Transcriptional Start Sites) 提供機関: 東京大学 URL : http://dbtss.hgc.jp/ 内容: 各種生物の転写開始点と上流制御領域のデータベース
■ DDBJ (DNA Data Bank of Japan) 提供機関:国立遺伝学研究所 URL: http://www.ddbj.nig.ac.jp/index-j.html 内容:全世界の研究者が実験によって決定したDNA、またはRNAの塩基配列データを、日欧米で定めたデータ構築規範に従い収集・編集し、フラットファイル形式で提供
■ eF-site (electrostatic surface of Functional-site) 提供機関:東京大学医科学研究所 URL: http://ef-site.hgc.jp/eF-site/index.jsp 内容: タンパク質の分子表面に関するデータベースで、静電的ポテンシャルと疎水的性質を活性部位のコノリー表面と合わせて表示し、分子認識機構の分析に有用な情報を提供
■ eProtS (Encyclopedia of Protein Structures) 提供機関: 大阪大学蛋白質研究所 URL : http://www.pdbj.org/eprints/index_ja.cgi? 内容: 生物学的に重要なタンパク質を選び、その立体構造を表示するとともに、タンパク質の構造と機能についてわかりやすく解説
■ GPI-anchored protein prediction (GPI アンカー型タンパク質予測システム) 提供機関: 東京大学 URL : http://bolero.bi.a.u-tokyo.ac.jp:8201/GPI-Predictor/ 内容: 与えられたタンパク質が翻訳後脂質修飾の一種であるGPI アンカー型であるかどうかを予測するプログラム。 タンパク質のアミノ酸配列のみからSVMを使って予測する
■ GTOP (Genomes TO Protein structures and functions) 提供機関: 国立遺伝学研究所 URL : http://spock.genes.nig.ac.jp/~genome/gtop-j.html 内容: ゲノムにコードされる全タンパク質の配列データを解析した結果をまとめたデータベース
■ INOH (Integrating Network Objects with Hierarchies) 提供機関: 東京大学／産業技術総合研究所 URL : http://www.inoh.org/ 内容: 文献から人の手で抽出したパスウェイ情報を格納したデータベース。シグナル伝達する複数の生体分子間の関係のような高次機能情報をオントロジーを使ってデータベース化している
■ KEGG (Kyoto Encyclopedia of Genes and Genomes) 提供機関: 京都大学化学研究所 URL : http://www.genome.jp/kegg/ 内容: システム情報、ゲム情報、ケミカル情報に大別された16の主要データベースから構成される生命システム情報統合データベース
■ KEGG-BRITE 提供機関: 京都大学化学研究所 URL : http://www.genome.jp/kegg/brite.html 内容: 生体系の様々な側面に関する知識を階層的に分類したデータベース
■ Lectin prediction (糖鎖結合タンパク質予測システム) 提供機関: 東京大学 URL : http://bolero.bi.a.u-tokyo.ac.jp:8201/Lectin-Predictor/ 内容: 与えられたタンパク質が糖鎖と結合するかどうかを予測するプログラム。タンパク質のアミノ酸配列のみからSVMを使って予測する
■ MassBank 提供機関: 慶應義塾大学 URL : http://www.massbank.jp/ 内容: 代謝物質を高分解能質量分析によって測定したマススペクトルを収集したデータベース
■ Melina II 提供機関: 東京大学 URL : http://melina2.hgc.jp/public/index.html 内容: 頻出モチーフ抽出支援ツールの改良版
■ PDBj (Protein Data Bank Japan) 提供機関: 大阪大学蛋白質研究所 URL : http://www.pdbj.org/index_j.html 内容: 米国RCSB、BMRB、および欧州PDBeと協力して、生体高分子の立体構造データベースを国際的に統一化されたPDBアーカイブとして運営するとともに、様々な解析ツールを提供している
■ Promode 提供機関: 早稲田大学 URL : http://cube.socs.waseda.ac.jp/pages/jsp/ja/index.jsp 内容: 基準振動解析によって得られたタンパク質立体構造の動的構造に関するデータを集めたデータベース
■ SCMD (Saccharomyces Cerevisiae Morphological Database) 提供機関: 東京大学 URL: http://scmd.gi.k.u-tokyo.ac.jp/datamine/ 内容: 出芽酵母の突然変異体の顕微鏡写真を集めたデータベース

2. JST が提供するサービス

■ GBIF JAPAN (Global Biodiversity Information Facility in JAPAN : 地球規模生物多様性情報機構) URL : http://bio.tokyo.jst.go.jp/GBIF/gbif/japanese/index.html 内容: 全世界の生物多様性に関するデータを各国で分散的に収集し、ネットワークを通じて利用することを目的とした国際的な科学プロジェクト
■ GenLibi (Gene Linker to bibliography : 遺伝子文献情報連携システム) URL : http://gene.jst.go.jp/cgi-bin/about.cgi 内容: JST 文献情報から得られる遺伝子機能や関連する疾患／表現型を提供するためのシステム
■ HOWDY (Human Organized Whole genome Database : ヒトゲノム情報統合データベース) URL : http://howdy.biosciencedbc.jp/HOWDY/top_ja.pl 内容: ヒトのゲノム情報および遺伝子について、複数のデータベースを一度に検索できるシステムを提供
■ JSNP (Japanese Single nucleotide polymorphisms : 日本人一塩基多型データベース) URL : http://snp.ims.u-tokyo.ac.jp/index.html 内容: 東京大学医科学研究所とJSTが協力して提供する、日本人の一塩基多型データのデータベース
■ MdeR (The metadata element repository in life sciences : ライフサイエンス分野のメタデータ要素レポジトリ) URL : http://mder.biosciencedbc.jp/ 内容: メタデータ要素の検索、検索結果の比較、収録メタデータのメタデータ要素の一覧表示を行うことができる
■ Web リソースポータルサイト for バイオインフォマティクス URL : http://tools.lifesciencedb.jp/cgi-bin/WebResourcePortal/WebResourcePortal.cgi 内容: バイオインフォマティクス解析ツールと解析ワークフローのポータルサイト
■ WINGpro (データベースポータルサイト) URL : http://wingpro.biosciencedbc.jp/dbpwiki/index.php/ 内容: ライフサイエンス分野の400あまりのデータベースについての情報を収録
■ ゲノムリテラシー講座 URL : http://www-bird.jst.go.jp/jinza/literacy/ 内容: 種々のデータベースやバイオインフォマティクス関連技術の利用法、バイオインフォマティクスの研究動向等について講座を開催。平成21年度以降の講座はストリーミング配信を実施
■ ゲノム解析ツールリンク集 URL : http://stga.biosciencedbc.jp/cgi-bin/index.cgi 内容: 分子生物学に関わるデータ解析ツールを集めたリンク集

3. BIRD からの提供を終了したサービス

■ ALIS (Advanced Life-science Information Systems) 内容: DNA塩基配列などのテキスト情報に生物学的な知識を付加したデータベース、およびそれを応用したツールを開発し、情報基盤の整備を行った
■ ECGB (真核生物比較ゲノムブラウザ) 内容: ヒトやチンパンジー、マウス、ラットなどの生物のDNA配列を比較し、生物種の間で保存されているDNA配列を提供
■ Ensembl ミニコース 内容: Ensembl (ヒトゲノム版) の使い方をGDNF遺伝子の探索を通して解説
■ HAL (Human genome Annotation Library; Jabion Genome Viewer から提供) 内容: ヒトゲノムに潜む遺伝子の位置と構造、その機能などの注釈情報を提供
■ HGS (Human Genome Sequencing) 内容: ヒトの7染色体(Ch1, 3, 6, 8, 9, 21, 22)を対象として行ったプロジェクト。シーケンシングデータを利用できる
■ NCBI ミニコース日本語版 内容: NCBI mini-courses は米国立生物工学情報センター (NCBI) が公開している解析ツールやデータベースの実践的な使用方法を紹介している教育サイト (http://www.ncbi.nlm.nih.gov/Class/minicourses/) で公開されている。本サイトは、NCBI mini-courses を日本語に翻訳し、JST が独自に補足説明を加えたものである
■ SILA (Structural Initial data Library of Amino acid residues) 内容: アミノ酸残基の構造データライブラリで、立体構造未知のアミノ酸配列を持つタンパク質の立体構造予測に利用される
■ UCSC ゲノムブラウザミニコース 内容: UCSC ゲノムブラウザ (ヒトゲノム版) の使い方をGDNF遺伝子の探索を通して解説
■ ゲノム解析ツール 内容: 原核生物の遺伝子構造予測ソフト、DNA配列からのモチーフ抽出ソフト、ホモロジー検索ソフト、DNA配列比較ソフトなどを開発して公開した
■ 国際 HapMap 計画 内容: ヒトの病気や薬に対する反応性に関係する遺伝子を発見するためのプロジェクト「国際 HapMap 計画」のデータ (SNP とハプロタイプ) を提供
■ 多型情報ネットワーク 内容: JSNP、JG-SNP、mtSNP、dbProP 等、遺伝子の個人差に関する情報を集めたデータベースをネットワーク化

主要なデータベースに関するデータ

BIRDで提供している主要なデータベースについて、アクセス数、データベース登録数等のデータをまとめました。

<対象データベース>

■ KEGG (Kyoto Encyclopedia of Genes and Genomes)

<http://www.genome.jp/kegg/>

システム情報、ゲノム情報、ケミカル情報に大別された16の主要データベースから構成される生命システム情報統合データベース

■ PDBj (Protein Data Bank Japan)

http://www.pdbj.org/index_j.html

米国RCSB、BMRB、および欧州PDBeと協力して、生体高分子の立体構造データベースを国際的に統一化されたPDBアーカイブとして運営するとともに、様々な解析ツールを提供

■ DDBJ (DNA Data Bank of Japan)

<http://www.ddbj.nig.ac.jp/index-j.html>

全世界の研究者が実験によって決定したDNA、またはRNA の塩基配列データを、日欧米で定めたデータ構築規範に従い収集・編集し、フラットファイル形式で提供

■ JSNP (Japanese Single Nucleotide Polymorphisms)

<http://snp.ims.u-tokyo.ac.jp/index.html>

東京大学医科学研究所とJSTが協力して提供する、日本人の一塩基多型データのデータベース

■ HOWDY (Human Organized Whole genome Database)

http://howdy.biosciencedbc.jp/HOWDY/top_ja.pl

ヒトのゲノム情報および遺伝子について、複数のデータベースを一度に検索できるシステムを提供

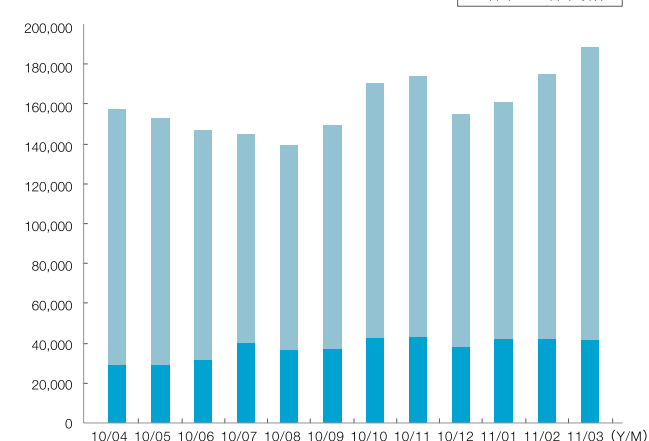
■ WINGpro (データベースポータルサイト)

<http://wingpro.biosciencedbc.jp/dbpwiki/index.php/>

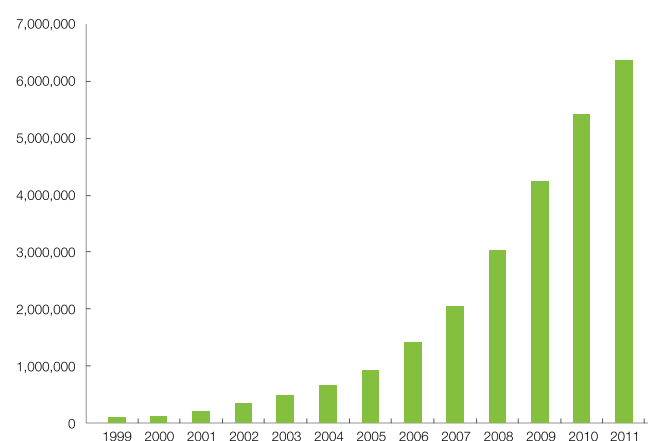
ライフサイエンス分野の400あまりのデータベースについての情報を収録

KEGG:Kyoto Encyclopedia of Genes and Genomes

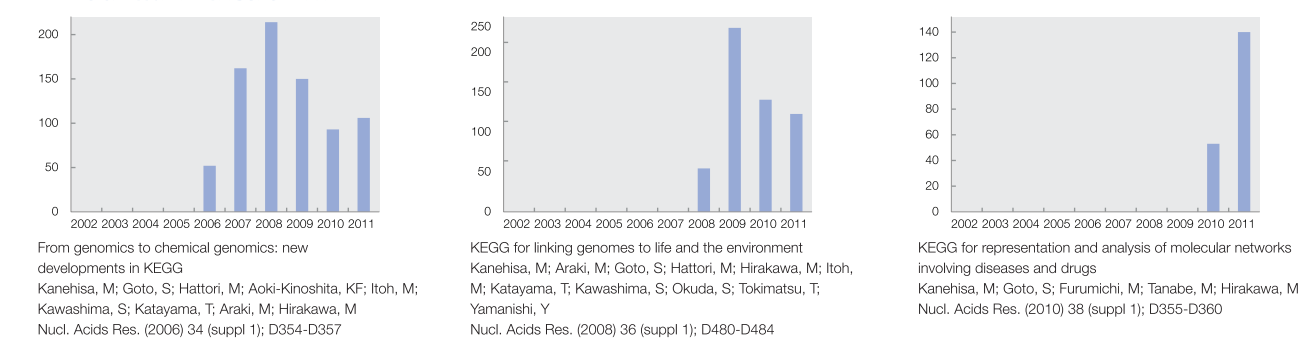
KEGGアクセス数(ユニークIP)



KEGGデータ登録数(GENESのみ)

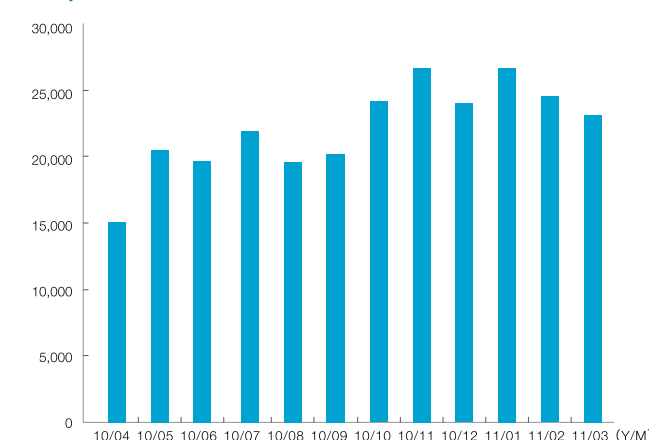


KEGG関連論文の被引用回数

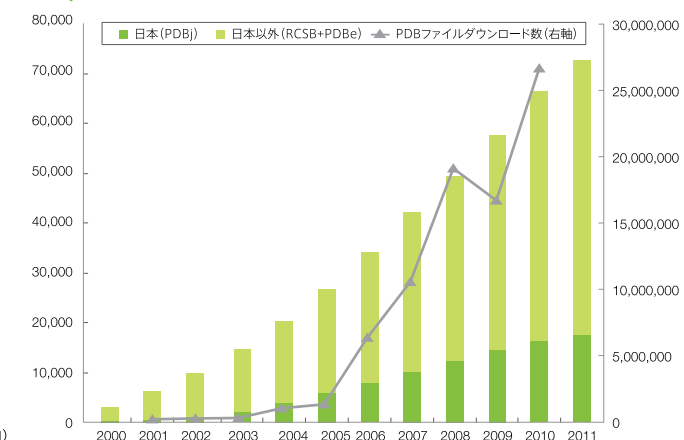


PDBj:Protein Data Bank Japan

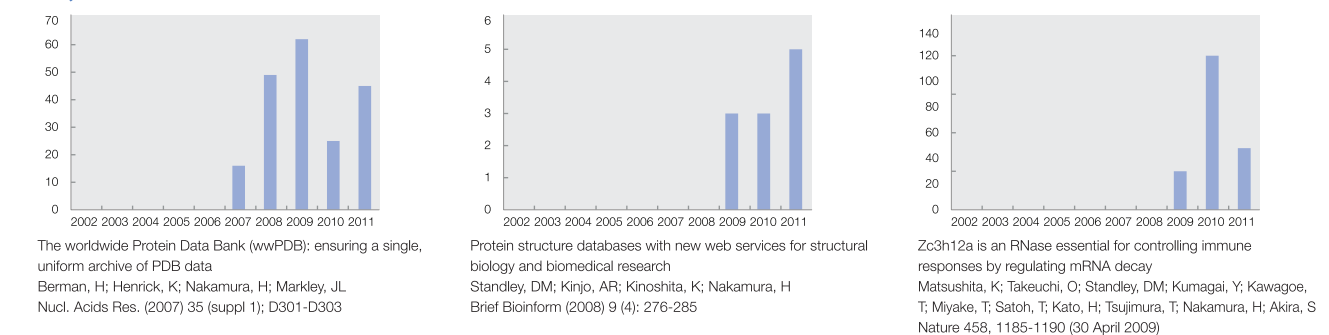
PDBjアクセス数(ユニークIP)



PDBjデータ登録数、PDBファイルダウンロード数

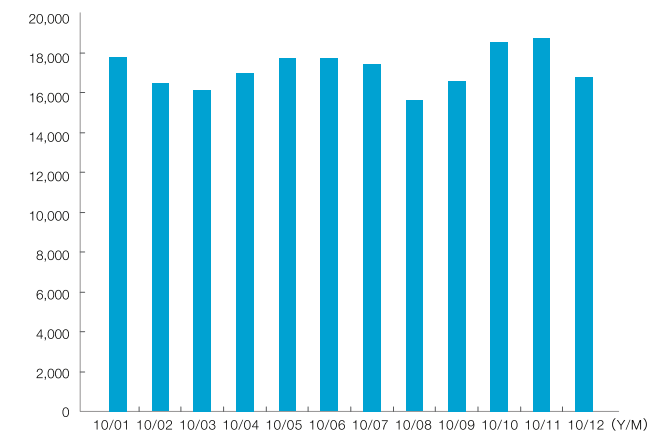


PDBj関連論文の被引用回数

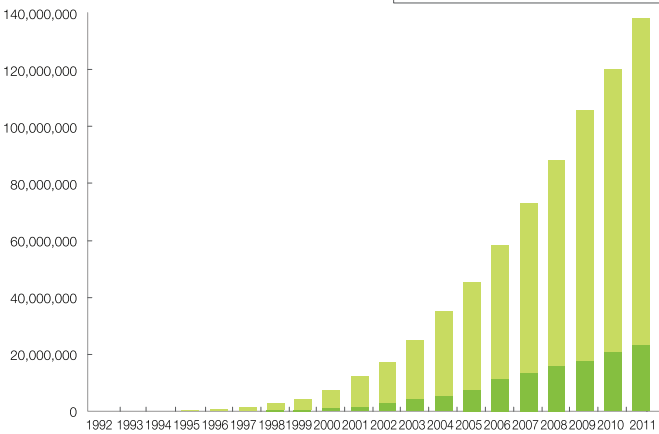


DDBJ:DNA Data Bank of Japan

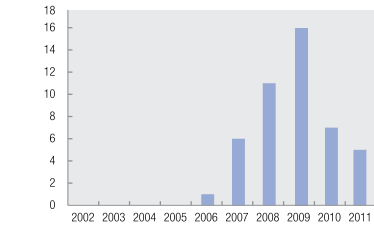
DDBJアクセス数(ユニークIP)



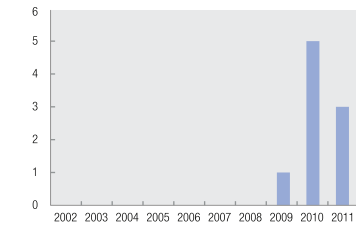
DDBJデータ登録数



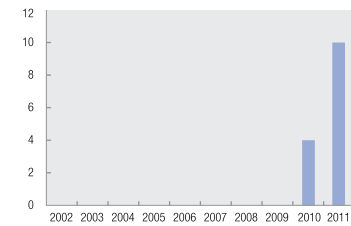
DDBJ関連論文の被引用回数



Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples
Abe, T; Sugawara, H; Kinouchi, M; Kanaya, S; Ikemura, T
DNA Res (2006) 12 (5): 281-290



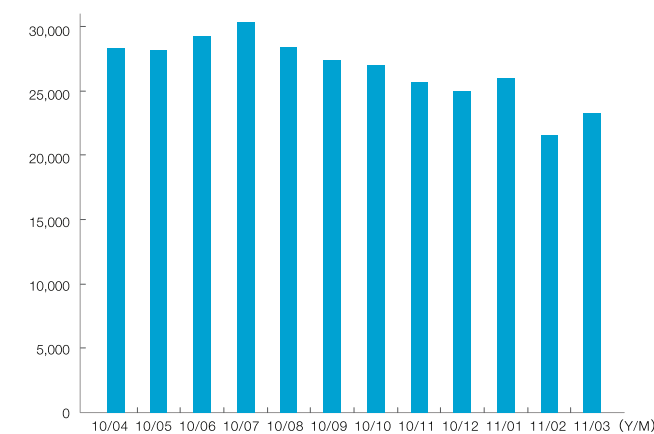
The GTP database in 2009: updated content and novel features to expand and deepen insights into protein structures and functions
Fukuchi, S; Homma, K; Sakamoto, S; Sugawara, H; Tateno, Y; Gojobori, T; Nishikawa, K
Nucl. Acids Res. (2009) 37 (suppl 1): D333-D337



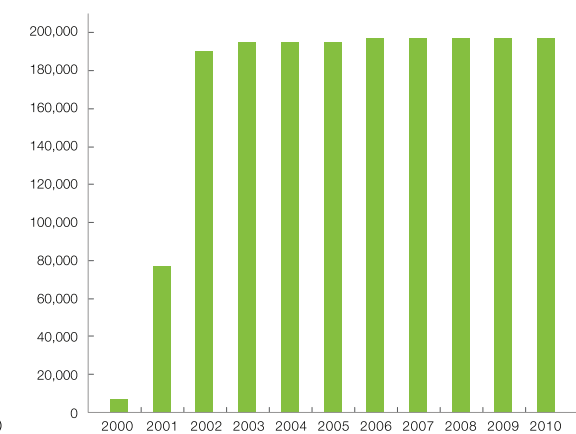
Archiving next generation sequencing data
Shumway, M; Cochrane, G; Sugawara, H Shumway, Martin; Cochrane, Guy; Sugawara, Hideaki
Nucl. Acids Res. (2010) 38 (suppl 1): D870-D471

JSNP:Japanese Single Nucleotide Polymorphisms

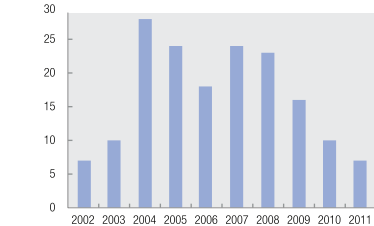
JSNPアクセス数(ユニークIP)



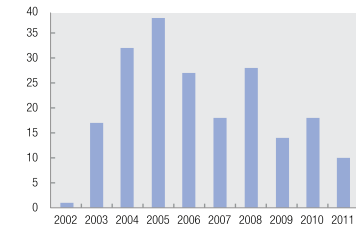
JSNPデータ登録数



JSNP関連論文の被引用回数



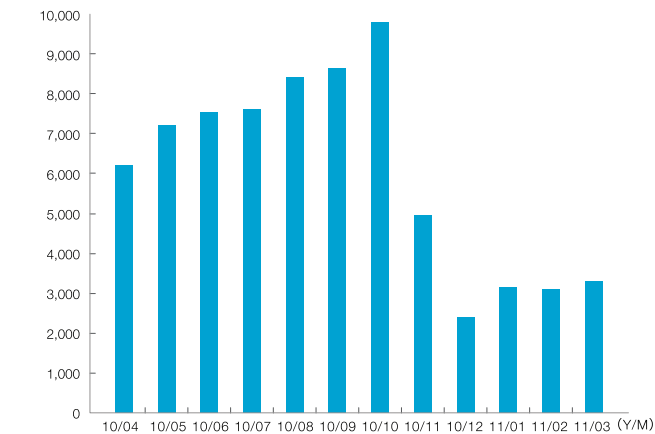
JSNP: a database of common gene variations in the Japanese population
Hirakawa, M; Tanaka, T; Hashimoto, Y; Kuroda, M; Takagi, T; Nakamura, Y
Nucl. Acids Res. (2002) 30 (1): 158-162



Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190562 genetic variations in the human genome
Haga, H; Yamada, R; Ohnishi, Y; Nakamura, Y; Tanaka, T
Journal of Human Genetics (2002) 47, 605-610

HOWDY:Human Organized Whole genome Database

HOWDYアクセス数(ユニークIP)



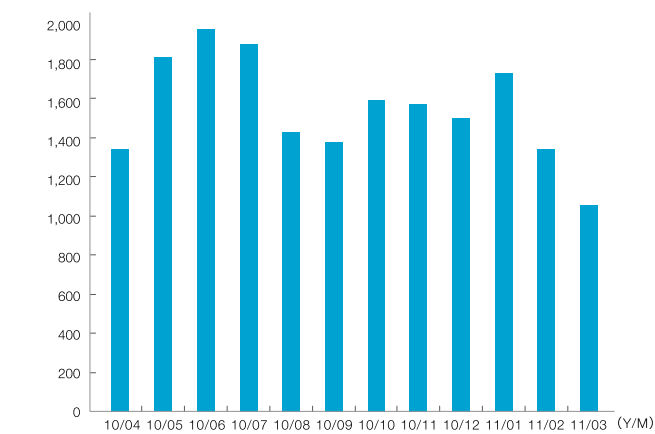
HOWDYデータ登録数

	Data Class	Entries
Gene	GDB (Data)	-
	HGNC	35,004
	EntrezGene	166,945
Protein	UniProtKB/Swiss-Prot	20,235
	ReSeq (NP)	31,304
	RefSeq (XP)	3,816
SNPs	dbSNPs	30,443,446
	JSNPs	167,195
Marker	UniSTS	303,294
	GDB (Amplimer)	-
Nucleic Acid Sequence		675,128
Refseq (NM)		31,304
Refseq (XM)		3,816
Unigene		122,727
OMIM		21,394
Contig		11,495
ContigMap		26
ALIS-HGS		46
e-PCR		790
GDB (Map)		-

2011年9月

WINGpro

WINGproアクセス数(ユニークIP)



WINGproの登録データベース

タイプ	件数
オントロジー	4
データバンク	22
プログラム	87
プロジェクト	155
ポータル	19
リソース	31
解析サービス	5
解析プログラム	2
辞典	46
知識モデル	27
注釈	10
目録	23

2011年9月



ゲノムリテラシー講座

講座名 遺伝子機能予測			
■「ホモロジー検索」			
■「モチーフ検索」			
■「ホモロジー以外の機能予測」(局在予測、膜タンパク質解析、立体構造予測)			
講座名 cDNA 解析・EST 解析			
■ゲノムマッピング			
■EST クラスタリング、配列解析			
講座名 多型解析と疾患遺伝子同定			
■連鎖解析「罹患同胞対連鎖解析および連鎖不平衡解析の実際」			
■「QTL 解析」			
■多型(SNP、マイクロサテライト)解析「ゲノムワイド遺伝的相関解析の実際」			
講座名 アレイ・インフォマティクス (遺伝子発現解析)			
■アレイ設計・作成のための情報処理			
■発現プロファイル解析 薬理ゲノミクス			
■発現情報の診断への応用			
講座名 バイオインフォマティクス技術と、バイオインフォマティクスを用いた研究成果の特許による保護について			
■バイオインフォマティクス特許			
■バイオインフォマティクスを用いた研究成果の特許による保護			
講座名 バイオインフォマティクス入門			
■「perl 入門とデータベース検索サービス構築法入門」			
講座名 ゲノムアノテーション			
■「ヒトゲノム配列の読解術：遺伝子の発見とその周辺」			
講座名 オーソログ同定とパスウェイ解析			
■オーソログ同定			
■パスウェイ解析			
■「オーソログ同定」と「パスウェイ解析」			
講座名 生命のシステム的理解 (シミュレーション)			
■システム同定の現在 (概日時計の分子機構の解明)			
■シグナル伝達経路のシステム解析			
講座名 バイオインフォマティクス入門2			
■データベース・Webツールの利用方法			
講座名 テキストからの情報抽出			
■テキストからの情報抽出技術の必要性と開発動向の紹介			
■具体的な抽出技術の解説			
講座名 比較ゲノム解析			
■比較ゲノム解析			
■比較解析の基礎と実習			
講座名 ゲノム進化・マッピング階層モデルについて			
■ゲノム進化・マッピング階層モデルについて			
講座名 インタクトームおよび定量オーミックスデータからの知識発見入門			
■インタクトーム解析と定量オーミックスの最前線			
■タンパク質相互作用予測と分子シミュレーション			
■インタクトームデータ/定量オーミックスデータ解析ツールの紹介			
講座名 高速計算機システムによる蛋白質フォールディング			
■蛋白質フォールディングの分子シミュレーション			
■蛋白質のフォールディング・シミュレーションの実際			
講座名 データベースを利用した蛋白質の立体構造予測			
■蛋白質の立体構造予測			
■全ゲノム立体構造予測データベース GTOP：機能と使い方			
■Web サーバを使った一次元構造予測：二次構造と埋もれ度の予測			
講座名 ゲノムアノテーションーデータベースとツールの活用ー			
■ゲノムデータ活用の基礎知識			
■KEGG DAS と KEGG API で繋がるゲノムアノテーションとパスウェイ			
講座名 ゲノムから得られるすべてのタンパク質を理解するには…			
■ゲノムから得られるすべてのタンパク質を理解するには…			
■SOSUI 群 (アミノ酸配列情報解析ツール)			
講座名 プロテオミクスにおける IT 技術			
■プロテオミクスにおける情報技術の原理と応用			
■プロテオームデータの更なる活用 - ゲノム配列へのマッピング -			
■プロテオミクスにおけるデータベース検索と情報処理			

講座名 RNA干渉を用いた効果的な遺伝子発現抑制技術			
■RNA干渉を用いた効果的な遺伝子発現抑制技術			
講座名 遺伝子からみた生物の系統樹			
■分子系統学講義			
■分子系統樹の作成			
講座名 KEGG におけるケモインフォマティクスへの取り組み			
■KEGG の化学情報データベース			
■ケミカルゲノミクスのためのデータ検索と処理			
講座名 プロテインインフォマティクスによるタンパク質の機能推定			
■プロテインインフォマティクスによるタンパク質の機能推定			
講座名 E-Cell による細胞シミュレーション			
■バイオシミュレーション講義			
■E-Cell モデリング入門			
講座名 UniProt/Swiss-Prot データベース			
■タンパク質情報の宝の山 Swiss-Prot をどのように攻略するか			
■Swiss-Prot の賢い使い方			
講座名 タンパク質立体構造データベース PDB と BMRB			
■Protein Data Bank と PDBj			
■BMRB データベースと BMRB データを利用した応用研究の一例			
■PDBj のウェブサービスの利用方法			
講座名 生物系研究のための XML・SOAP 技術			
■バイオデータベースの構築と利用			
■XML の取り扱いと SOAP サービスの利用			
講座名 トランスクリプトームからの代謝経路の解析			
■トランスクリプトームからの代謝経路解析の実際			
講座名 脳型データベースの推論機能を使ったボジショナルキャンディートクローニング入門			
■文献を疑似ニューロンとするデータベースの推論機構と遺伝学への応用			
■理研GSCのウェブサイトを使った候補遺伝子選択演習			
講座名 決定的ハプロタイプの概要とそのデータベース D-HaploDB の利用法			
■決定的ハプロタイプ構造の概要と構造の意義			
■D-HaploDB の利用法・活用法			
講座名 バイオインフォマティクスを活用した新規ドメインの発見・解析技術			
■バイオインフォマティクスを活用した新規ドメインの発見・解析技術			
講座名 遺伝子からみた生物の系統樹			
■分子系統学講義			
■分子系統樹の作成			
講座名 細菌の比較ゲノム解析			
■細菌の比較ゲノム解析からメタゲノム解析へ			
■統合ソフトウェアをもちいた細菌の比較ゲノム解析の初歩			
講座名 ゲノムネットと KEGG から始めるバイオデータベース検索			
■バイオデータベース入門			
■バイオデータベースの利用法			
講座名 バイオインフォマティクスに基づいた遺伝子改変マウスの表現型解析と精神疾患の研究			
■バイオインフォマティクスの手法を活用した遺伝子改変マウスの表現型解析と精神疾患の研究			
講座名 データベースと構造生物学			
■PDB からの超分子構造のモデリング：非対称単位から4次構造モデルを作るときに知っていることと便利なこと			
講座名 タンパク質構造予測とデータベース実習			
■タンパク質構造予測の基礎			
■構造予測実習			
講座名 ケモインフォマティクスを利用したリード化合物の探索入門			
■ケモインフォマティクスを利用したリード化合物の探索入門			
講座名 転写制御領域の配列解析入門			
■転写制御領域の配列解析入門			
■共発現遺伝子の上流配列解析実習			
講座名 バイオインフォマティクスの基礎：情報学一般			
■バイオインフォマティクス技術者認定試験：情報学分野の解説			
■バイオインフォマティクス技術者認定試験：情報学分野の演習			
講座名 バイオインフォマティクスの基礎：生命科学一般			
■生命体の成り立ちと情報の流れ			
■ヒト遺伝子データベースの活用法			
講座名 バイオインフォマティクスの基礎：配列解析、タンパク質立体構造解析			
■バイオインフォマティクス基礎講座：配列解析、タンパク質立体構造解析			
■配列解析と立体構造解析の基礎演習			
講座名 バイオインフォマティクスの基礎：分子生物学データベース			
■分子生物学データベースの利用法及び分子系統解析の解説			
■分子系統解析の基礎演習			
講座名 バイオインフォマティクスの基礎：パスウェイ解析・システム生物学			
■パスウェイ解析・システム生物学入門			
■パスウェイ解析・システム生物学演習			
講座名 バイオインフォマティクスの基礎：トランスクリプトーム解析・プロテオーム解析			
■トランスクリプトーム解析・プロテオーム解析入門			
■トランスクリプトーム解析基礎演習			

BIRD 年表		
JSTの動き		国内の動き
平成3年 (1991)	科学技術振興調整費による受託研究「染色体遺伝子地図データベース作製技術の開発」(平成3～5年度)	文部省の創成的基礎研究「ヒトゲノム解析研究」および重点領域研究「ゲノム解析に伴う大量知識情報処理の研究」発足
平成4年 (1992)	- 理化学研究所からの受託調査研究「ゲノム情報獲得手法の研究開発」 - ゲノムデータベース (GDB) の試験的導入	
平成5年 (1993)	- 理化学研究所からの受託調査研究「ゲノムデータベース (GDB) の開発・導入」(平成5～7年度) - 科学技術庁／NIH 協力協定締結 - ジョンズ・ホプキンス大学との間で GDB 公式ノード設置に関する協定締結 - GDBnet 設置 (米国 NSFnet と接続)	
平成6年 (1994)	- GDB 公式ノードとして運用開始 (2 月) - 科学技術振興調整費による受託研究「ヒト遺伝子マッピング情報の収集・整理」 (平成6～7年度)	
平成7年 (1995)	- 一般会計出資金による「高機能基盤生体データベース開発事業」開始 - ALIS (Advanced Life-Science Information Systems) プロジェクト (10月1日) - ヒトゲノムシーケンシングの実施 (平成7～11 年度) - シーケンシング・データベースシステムの開発 (平成7 年度～) - GDB 国際ノードとして提供 (10月1日)	科学技術基本法を制定 (11 月)
平成8年 (1996)	- 科学技術振興事業団 (JST) 設立 (10月1日) - ヒトゲノムシーケンシングデータの公開開始 - ヒトゲノム4M (400 万塩基) のシーケンシングの完了を目指し配列データの公開開始 (11月5日) - ヒトゲノム統合データベースシステム (HOWDY) の開発 (平成8 年度～)	- 第1 期科学技術基本計画 (平成8～12 年度) 閣議決定 (7 月) - 文部省の重点領域研究「ゲノムサイエンスヒトゲノム解析に基づくバイオサイエンスの新展開」発足
平成9年 (1997)	- ゲノム解析ツールリンク集公開開始 - インターネット上で提供されている分子生物学に関わるデータ解析ツールをカテゴリ別に分類し、簡単な解説を付けて提供	
平成10年 (1998)	- 遺伝子領域推定機能、高速ホモロジー検索システム (2 月) - 試験的提供開始 (以後、順次機能拡充)	
平成11年 (1999)		- 産業活力再生特別措置法 (日本版バйдール条項) 制定(8 月) - ミレニアム・プロジェクト (新しい千年紀プロジェクト) について内閣総理大臣決定 (12 月 19 日)
平成12年 (2000)	- 平成11年度補正予算より「多型情報データベース開発事業」開始 - ミレニアム・プロジェクトの一環として、東京大学医科学研究所ヒトゲノム解析センターと JST との共同プロジェクトとして「ヒトゲノム標準多型データベースの開発」実施 (平成 11-13 年度 データ取得) - ヒトゲノム標準多型データベース (JSNP データベース) 公開開始 - ミレニアム・プロジェクトの成果を公開するデータベース。東京大学医科学研究所の SNP (一塩基多型) 探索の成果を中心に公開 (7 月 7 日) - ヒトゲノム統合データベース (HOWDY) 公開開始 - ヒトゲノムに関する 14 の公共データベースを情報源とし、まとめて検索可能にするデータベース (7 月 7 日)	- 日本とイギリスを中心にヒト染色体 22 番染色体のシーケンシングが完成 - 日本が担当した 21 番染色体のシーケンシングが完成 6 番染色体も日本が分担して完成 - 産業技術力強化法制定 (4 月) - 報告書「ゲノム情報科学における我が国の戦略について」科学技術会議ライフサイエンス部会ゲノム科学委員会 (11 月 17 日)
平成13年 (2001)	- 平成 12 年度補正予算より「バイオインフォマティクス推進センター事業」開始 - 地球規模生物多様性機構 (GBIF) の技術専門委員会の事務局設置 - 生物多様性に関するデータを集積し、全世界的に利用することを目的とする国際協力による科学プロジェクト (3 月 1 日)	- 中央省庁再編、文部科学省発足 (1 月) - 総合科学技術会議発足 (1 月) - 第2 期科学技術基本計画 (平成13～17 年度) 閣議決定 (3 月)

JSTの動き		国内の動き
平成13年 (2001)	- バイオインフォマティクス推進センター設置 - 急速に進展している先端生物科学と膨大な遺伝子情報等を解析するための情報科学との融合によるバイオインフォマティクスの推進を目的として設置 (4 月 1 日) - バイオインフォマティクス委員会設置 - BIRD ホームページを公開 (6 月 20 日) - 生命情報データベース高度化・標準化第1 期 4 課題採択 (6 月 20 日) - 創造的な生物・情報知識融合型の研究開発 平成13 年度7 課題採択 (9 月 28 日) - ゲノム情報科学サーバシステム設置 (12 月) - 医学生物学データベース案内 (WING) 公開 - 国内外に多数ある医学・生物関連データベースの情報を提供	68 の国立試験研究機関の独立行政法人化 (4 月) - 知的財産基本法を制定 (11 月)
平成14年 (2002)	バイオインフォマティクス相談開始 (11 月 26 日)	- タンパク 3000 プロジェクト (平成 14～18 年度) - 第1 期ナショナルバイオリソースプロジェクト (平成14～18 年度) 「バイオテクノロジー戦略大綱」BT 戦略会議 (12 月 16 日)
平成15年 (2003)	- 独立行政法人科学技術振興機構 (JST) 設立 (10 月 1 日) - 多型情報ネットワーク - ミレニアムプロジェクトの一環として、JSNPをはじめとする、JST のデータベース化支援事業で支援された遺伝子の個人差に関する情報を集めたデータベースをネットワーク化して提供 (10 月 14 日)	第1 期個人の遺伝情報に応じた医療の実現プロジェクト (平成 15～19 年度)
平成16年 (2004)	- 国際 HapMap 計画データベースのミラーサイト開設 - ヒトゲノム上の多型情報を医学応用していくために不可欠であるハプロタイプ地図作成に関する国際的な枠組みである「国際 HapMap 計画」の成果データベースミラーサイトを開設 (5 月 25 日) - 真核生物比較ゲノムブラウザ公開 - 比較ゲノムブラウザにて、ヒトやチンパンジー、マウス、ラットなどの真核生物の比較ゲノムク解析の結果を提供。新規遺伝子の発見や、ゲノム進化についての新しい知見の蓄積を支援 (12 月 24 日)	国立大学を法人化 (4 月)
平成17年 (2005)	- 第1 回研究開発成果報告会開催 (1 月 28 日) - 創造的な生物・情報知識融合型の研究開発 平成17 年度6 課題採択 (8 月 24 日)	- 経産省ゲノム情報統合プロジェクト (平成 17 年～19 年度) - 内閣府連携施策群「ライフサイエンス分野のデータベースの統合化に関する調査研究」(平成 17～19 年度)
平成18年 (2006)	- 生命情報データベース高度化・標準化第1 1 期 7 課題採択 (2 月 2 日) - 第2 回研究開発成果報告会開催 (3 月 13 日) - NCBI ミニコース日本語版を公開 (9 月 8 日) - 第3 回研究開発成果報告会開催 (11 月 2 日) - 創造的な生物・情報知識融合型の研究開発 平成18 年度5 課題採択 (11 月 14 日)	- 第3 期科学技術基本計画 (平成 18～22 年度) 閣議決定 (3 月) - 報告書「我が国におけるライフサイエンス分野のデータベース整備戦略のあり方について」科学技術・学術審議会研究計画・評価分科会ライフサイエンス委員会データベース整備戦略作業部会 (5 月 17 日) - 統合データベースプロジェクト (平成 18～22 年度) - 新産業創出のための農林水産生物ゲノム情報統合データベースの構築 (平成 18～22 年度)
平成19年 (2007)	創造的な生物・情報知識融合型の研究開発 平成19 年度4 課題採択 (9 月 27 日)	- 第2 期ナショナルバイオリソースプロジェクト (平成 19～23 年度) - ターゲットタンパク研究プログラム (平成 19～23 年度)
平成20年 (2008)	第4 回研究開発成果報告会開催 (11 月 4 日)	第2 期個人の遺伝情報に応じた医療の実現プロジェクト (平成 20～24 年度)
平成21年 (2009)	第5 回研究開発成果報告会開催 (11 月 17 日)	- 報告書「ライフサイエンスデータベースの統合・維持・運用の在り方」科学技術・学術審議会研究計画・評価分科会ライフサイエンス委員会ライフサイエンス情報基盤整備作業部会 (1 月 13 日) 「統合データベースタスクフォース報告書」総合科学技術会議ライフサイエンス PT 統合データベースタスクフォース (5 月 27 日)
平成22年 (2010)	- 多型情報ネットワーク検索サービス、解析ライブラリー、多型情報データベースリンク集、ゲノム解析ツールの Genome Scenario 提供終了 (3 月 29 日) - 第6 回研究開発成果報告会開催 (11 月 10 日)	
平成23年 (2011)	- 第7 回研究開発成果報告会開催 (3 月 8 日) - 一部サービスの終了 - HGS、ECGB、多型情報ネットワーク、SILA、Alis、ゲノム解析ツール、HAL、NCBI ミニコース日本語版、Ensembl ミニコース、UCSC ゲノムブラウザミニコース (3 月 31 日) - バイオサイエンスデータベースセンター (NBDC) 設立 (4 月 1 日) - バイオサイエンスデータベースセンターからのサービスに移行 - 遺伝子文献情報連携システム (GenLibi)、HOWDY、WINGpro、ライフサイエンス分野のメタデータ要素レポジトリ (MDeR)、ゲノム解析ツールリンク集、Web リソースポータルサイト for バイオインフォマティクス (4 月 1 日) - 第8 回研究開発成果報告会開催 (11 月 14 日)	第4 期科学技術基本計画 (平成 23～27 年度) 閣議決定 (8 月)
平成24年 (2012)	「バイオインフォマティクス推進センター事業」終了 (3 月)	



バイオインフォマティクス推進センター事業 成果集 2011年11月

編集・発行

独立行政法人科学技術振興機構 バイオサイエンスデータベースセンター

〒102-0081 東京都千代田区四番町 5 番地 3

TEL: 03-5214-8491

FAX: 03-5214-8470

URL: <http://biosciencedbc.jp/>



独立行政法人
科学技術振興機構

