独立行政法人科学技術振興機構
バイオインフォマティクス推進センター事業
追跡評価報告書

(生命情報データベースの高度化・標準化)
(第 I 期：平成 13 年度採択課題)

平成 21 年 6 月

独立行政法人科学技術振興機構

目次

I.	追跡評価の概要	2
1.	追跡評価の目的	2
2.	対象課題	2
3.	評価者	2
4.	評価方法	3
4. 1	追跡調査	3
4. 2	追跡評価	3
II.	評価結果（生命情報データベースの高度化・標準化（第 I 期））	5
1.	「2 項関係に基づくゲノムと生命システムの機能解読」	5
2.	「ゲノム生物学バックボーンデータベースの構築提供」	7
3.	「シグナルオントロジーとバイオタームバンクの開発」	9
4.	「蛋白質立体構造データベースの高度化」	11

I. 追跡評価の概要

1. 追跡評価の目的

本報告書は、独立行政法人科学技術振興機構（JST）が実施するバイオインフォマティクス推進センター事業の追跡評価結果を報告するものである。

追跡評価は、事後評価を終えて2年以上経過した課題について、副次的効果を含めて研究開発成果の発展状況や活用状況等を明らかにし、事業及び事業運営の改善等に資することを目的として、「バイオインフォマティクス推進センター事業の課題評価の方法等に関する達」（平成15年10月1日 平成15年達第74号）に基づき実施したものである。

2. 対象課題

生命情報データベースの高度化・標準化（第I期：平成13年度採択課題）（4課題）

「2項関係に基づくゲノムと生命システムの機能解読」（平成13年6月～平成18年3月）
代表研究者 金久 實（京都大学 化学研究所 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」（平成13年6月～平成18年3月）
代表研究者 菅原 秀明（国立遺伝学研究所 特任教授）

「シグナルオントロジーとバイオタームバンクの開発」（平成13年6月～平成18年3月）
代表研究者 高木 利久（東京大学大学院新領域創成科学研究科 教授）

「蛋白質立体構造データベースの高度化」（平成13年6月～平成18年3月）
代表研究者 中村 春木（大阪大学 蛋白質研究所 教授）

注：代表研究者の所属は平成21年3月現在を記載

3. 評価者

評価者は、「バイオインフォマティクス推進センター事業 追跡評価委員会」の評価委員の5名。

主査	宮田 隆	JT 生命誌研究館 顧問
委員	菊野 玲子	財団法人かずさディー・エヌ・エー研究所 主任研究員
委員	西村 いくこ	国立大学法人京都大学大学院理学研究科 教授
委員	美宅 成樹	国立大学法人名古屋大学大学院工学研究科 教授
委員	安永 照雄	国立大学法人大阪大学 微生物病研究所 教授

4. 評価方法

4. 1 追跡調査

追跡評価の対象である平成13年度に採択したデータベースの高度化・標準化課題（4課題）について追跡調査を行った。

追跡調査項目は、(1)調査対象課題の公開データ（中間評価、事後評価、終了報告書、研究実施報告書）による研究成果の把握、(2)論文引用数などのデータベース調査、(3)代表研究者への聞き取り調査による、その後の研究の発展状況や研究開発成果の波及効果、研究成果の利用状況、参画した研究員等の動向等である。

なお、代表研究者には、アンケート調査と聞き取り調査を実施した。

(1) アンケート調査

アンケート調査は、メール調査の形式で行った。

(2) 聞き取り調査

アンケート調査の結果を深めるために、有識者及び事務局が代表研究者に聞き取り調査を行った。

これらの結果を踏まえ、追跡調査報告書を作成した。

また、4課題中2課題については海外の有識者よりコメントを得た。

4. 2 追跡評価

追跡調査報告書をもとに、追跡評価委員会において以下の視点で評価を行った。

(1) 研究開発成果の発展状況や活用状況に関する評価

- ・ 構築されたデータベースの機能、開発された技術等の活用状況はどうか。
- ・ 研究開発課題終了後から現在にかけて、研究開発の継続状況、研究開発成果の発展状況等はどうか。

(2) 研究開発成果から生み出された科学技術的、社会的及び経済的な効果・効用及び波及効果に関する評価

- ・ 構築されたデータベースの機能、開発された技術等の波及効果はどうか。
- ・ 後のバイオインフォマティクス分野への影響や効果はどうか。
- ・ ライフサイエンス分野の研究の進展への貢献はどうか。
- ・ 新しい分野の人材育成の面から参加研究者の活動状況はどうか

(3) 総合評価

- ・ 総合的に判断して以下のどの評価にあてはまるか。
 - S) 優れた十分な活用がなされている。
 - A) 活用がなされている。
 - B) 活用がやや不足している。
 - C) 活用が不足している。

なお、この追跡評価結果についてはバイオインフォマティクス委員会に6月報告した。

II. 評価結果（生命情報データベースの高度化・標準化（第I期））

1. 「2項関係に基づくゲノムと生命システムの機能解読」

代表研究者 金久 實（京都大学 化学研究所 教授）

ア. 研究開発成果の発展状況や活用状況

（1）構築されたデータベースの機能、開発された技術等の活用状況

KEGG (Kyoto Encyclopedia of Genes and Genomes) は国際的に広く利用され、既に高い評価を受けている。本研究で開発された BRITE、KAAS は、KEGG に最近導入されたシステムであるが、今後大いに活用されるものと期待される。実際、既に BRITE を引用した論文が多く見られるようになってきていること、NCBI の新しいリソース Protein Cluster で利用されている。

（2）研究開発課題終了後から現在にかけて、研究開発の継続状況、研究開発成果の発展状況等

創薬・医療への応用を考慮しながら更なる機能の付加を行っている。また本データベースは NCBI などの国内外の各種データベースにもリンクされ広く利用されている。被引用件数の多い優れた論文が発表されており、BIRD 終了後の発展状況は目覚ましいものがある。KEGG は米国の NCBI、欧州の EBI、UniProt と並ぶ国際的な基盤データベースとして、汎用されている。

イ. 研究開発成果から生み出された科学技術的、社会的及び経済的な効果・効用及び波及効果

（1）構築されたデータベースの機能、開発された技術等の波及効果

KEGG は教育プログラムやトレーニングコースなどでも利用され、人材育成にも一役買っている。多くの企業ともライセンス契約を結び、世界中の大手製薬会社の多くが KEGG を利用している。このように応用研究（創薬）や教育まで幅広い分野での貢献が認められる。

（2）後のバイオインフォマティクス分野への影響や効果

階層的マッピングという、本研究課題で導入した新たな手法は、従来のパスウェイデータベースで使用されていた二項関係を超越して、本来階層的なシステムである生命をデータベースとして表現する新たな方法の開発であり、バイオインフォマティクス分野に大きな波及効果があるものと期待される。

（3）ライフサイエンス分野の研究の進展への貢献

KEGG はすでに米国の NCBI、欧州の EBI、UniProt と並ぶ国際的な基盤データベースとし

て世界で認知され、ライフサイエンス分野の基礎研究及び応用研究に広く利用されている。

(4) 新しい分野の人材育成の面から参加研究者の活動状況はどうか

本プロジェクトに参加した若手の一人が独立して准教授に昇格している。多くのメンバーが BIRD II 期の主要メンバーとして、また他研究機関のスタッフとして活躍しており、人材養成面での貢献も大きい。

ウ. 総論

総合評価

S) 優れた十分な活用がなされている。

KEGG に対する評価は非常に高い。生命科学分野のスタンダード的なデータベースの一つとなってきた。今後の KEGG の進化にも期待したい。

2. 「ゲノム生物学バックボーンデータベースの構築提供」 代表研究者 菅原 秀明（国立遺伝学研究所 特任教授）

ア. 研究開発成果の発展状況や活用状況

（1）構築されたデータベースの機能、開発された技術等の活用状況

本研究課題では、国立遺伝学研究所 DDBJ の研究者が中心となって研究を推進し、アノテーション支援（OASYS および WABI）、ゲノム関連のデータベース（GTOP）、遺伝子発現データベース（MADB および CIBEX、BSD）が開発・公開された。これらのデータベースの維持管理もよくなされ、アクセスログのデータから高い利用状況にあることが理解できる。これらのデータベースは、DDBJ 事業ならびに他プロジェクトと有機的に連携し、統合化が進んでいる。

（2）研究開発課題終了後から現在にかけて、研究開発の継続状況、研究開発成果の発展状況等

WABI は、BIRD 事業第 2 期受託事業として研究開発が継続された。KEGG ならびに PDBj との共同開発も一部行われ、機能拡張が図られている。他のデータベースもおおむね他の資金により更新が継続されている。

イ. 研究開発成果から生み出された科学技術的、社会的及び経済的な効果・効用及び波及効果

（1）構築されたデータベースの機能、開発された技術等の波及効果

OASYS の Web サービスは、PDBj、KEGG など国内の代表的なライフサイエンス統合データベースプロジェクトにおいても、統合化実現のための主要な課題の一つとして取り上げられている。また、微生物ゲノム自動アノテーションシステム（MiGAP）、Four-dimensional Ascidian Body Atlas（FABA）などへ応用され研究の発展に貢献している。

（2）後のバイオインフォマティクス分野への影響や効果

本研究でいち早く取り入れられたインターネットの新技术 Web サービスは KEGG や PDBj での採用に影響を与えた。利用しやすいデータベースの構築によって Web 情報が多くの研究者にとって身近なものとして利用されるようになった。この面での貢献度は高いと考える。

（3）ライフサイエンス分野の研究の進展への貢献

OASYS は国際アノテーションワークショップで ORF の再評価に貢献し、GTPS が予測した新規 ORF が採用されている。また遺伝子発現データベース CIBEX は、遺伝子発現データに基づいた論文を発表する際のデータベースの一つに指定されている。このように本研究課題で開発されたデータベースは、バイオ情報資源を統合的に利用できるデータベースとし

て、ライフサイエンス全体にとっての基盤情報を提供するものと期待できる。

(4) 新しい分野の人材育成の面から参加研究者の活動状況はどうか

2名が助手から教授、1名が助手から室長、1名が助手から准教授、1名がポスドクから准教授、1名がポスドクから助教に昇任あるいは任用されており、人材育成面での成果も上がっている。

ウ. 総論

総合評価

S) 優れた十分な活用がなされている。

DDBJ に欠けている部分を補いながら、大変な作業をがんばって続けているという印象が強い。

3. 「シグナルオントロジーとバイオタームバンクの開発」

代表研究者 高木 利久（東京大学大学院新領域創成科学研究科 教授）

ア. 研究開発成果の発展状況や活用状況

（1）構築されたデータベースの機能、開発された技術等の活用状況

本研究課題で開発されたデータベースは、シグナル伝達オントロジーとパスウェイデータベース INOH であり、すでに公開されていて多くのアクセスがある。当初の INOH を改良し、Web から直接利用できるようにした。この改良は利用促進を図る上で重要な進展であったと考えられる。

（2）研究開発課題終了後から現在にかけて、研究開発の継続状況、研究開発成果の発展状況等

BIRD 終了後もデータベースの高度化・標準化を視野に入れた開発研究を継続している。特にこの分野のデータベースの標準化を目指す BioPAX では標準化に向けた努力がなされている。この点は大いに評価できる。

イ. 研究開発成果から生み出された科学技術的、社会的及び経済的な効果・効用及び波及効果

（1）構築されたデータベースの機能、開発された技術等の波及効果

将来的には、ゲノムからの機能予測、細胞機能のシミュレーション、医薬品の開発に貢献するものと期待される。

（2）後のバイオインフォマティクス分野への影響や効果

パスウェイデータベースへの需要は高く、今後生物種ごとのパスウェイが整備されれば、近い将来バイオインフォマティクス研究、ライフサイエンス研究への貢献が十分期待できる。

（3）ライフサイエンス分野の研究の進展への貢献

シグナル伝達系に関する分子細胞生物学的研究は急速に進歩しており、得られる生物学的情報量も膨大になりつつある。知識データベースやパスウェイデータベースの整備は、この分野の研究を相乗的に推進することが期待される。

（4）新しい分野の人材育成の面から参加研究者の活動状況はどうか

本プロジェクトに参加した若手研究者が各分野で活躍している。

ウ. 総論

総合評価

A) 活用がなされている。

本プロジェクトが目指しているパスウェイデータベースとそのためのオントロジー開発は、現状では研究的要素が強く、世界的にも確立しているとは言い難い。従って実績は高度化・標準化課題の他の3課題と比べられない。しかし、大規模な研究経費を用いて開発したデータベースとしては、必ずしも満足いくものではないとの指摘もある。

4. 「蛋白質立体構造データベースの高度化」

代表研究者 中村 春木（大阪大学 蛋白質研究所 教授）

ア. 研究開発成果の発展状況や活用状況

(1) 構築されたデータベースの機能、開発された技術等の活用状況

国際タンパク質構造データバンクとの協力のもと、PDBj の管理と関連ソフトウェアの開発を行った。アクセス数、被リンクサイト数ともに多く、広く活用されていることが分かる。ライフサイエンス統合データベースでは基幹データベースの一つとして利用され始めている。PDB のタンパク質構造データそのものの活用は国内の主要な研究機関で活用されている。

(2) 研究開発課題終了後から現在にかけて、研究開発の継続状況、研究開発成果の発展状況等

本研究課題終了後も、JST-BIRD による生命情報データベースの高度化・標準化(第 II 期) 課題「蛋白質構造データバンクの国際的な構築と高度化 (PDBj)」に採択され、活動は着実に継続されている。米国 RCSB-PDB と欧州の PDBe と連携し、生体高分子の立体構造データベースを維持している組織として国際的な認知度も高い。PDB データベースの XML 化は日本でいち早く進められ、そのフォーマットが事実上の標準になったことは高く評価される。

イ. 研究開発成果から生み出された科学技術的、社会的及び経済的な効果・効用及び波及効果

(1) 構築されたデータベースの機能、開発された技術等の波及効果

PDB は、産学を問わず、多くの研究機関でタンパク質への薬物ドッキング・シミュレーションの基本となるデータとして活用されている。特に製薬企業では、このような研究はドラッグデザインへの応用に結びつくため、本データベースの活用度は非常に高いと思われる。教育や社会一般へのタンパク質の形に対する見識を広めた点でも功績があった。

(2) 後のバイオインフォマティクス分野への影響や効果

XML-DB を用いて PDB データを高速に検索できる仕組みを開発した。また、Web サービスを提供することによって、他のデータベースを連携させた融合的なバイオインフォマティクスの利用法が可能となった。さらに日本独自のサービスとして PDBj で開発された、分子表面データベースと機能部位の検索サービスやダイナミクスのデータベースなどを提供し、多くのバイオインフォマティクス研究者の利用に供している。これらのサイトへのアクセスも前者が年間 5 万件、後者が年間 2 万件ほどあり、バイオインフォマティクス研究に大きく貢献している。

(3) ライフサイエンス分野の研究の進展への貢献

各種二次データベースの作成・検索ツールの開発し、講習会等の普及活動を実施することにより、PDBj の認知度が高まり、広くライフサイエンスの研究者に情報を活用する機会を与えている。

(4) 新しい分野の人材育成の面から参加研究者の活動状況はどうか

ポスドクから教授が 1 名、准教授から教授が 3 名、ポスドクから准教授が 1 名、助手から准教授が 1 名と、人材育成面でも大きな成果が得られている。

ウ. 総論

総合評価

S) 優れた十分な活用がなされている。

非常に重要なデータベースの一つとして機能している。