

ライフサイエンス分野の統合データベース整備事業

ライフサイエンス知識の階層化・統合化事業

21年度 研究成果報告書

平成22年3月

国立大学法人京都大学 化学研究所

五斗進

本報告書は、文部科学省の科学技術試験研究委託事業による委託業務として、京都大学が実施した、平成21年度の「ライフサイエンス知識の階層化・統合化事業」の成果を取りまとめたものです。

目次

1. 委託業務の目的	4
2. 平成21年度の実施内容	4
2. 1 実施計画	4
2. 2 実施内容（成果）	5
(1) 共通基盤技術開発	5
(2) 統合データベース開発・運用	9
(3) プロジェクトの総合的推進	10
2. 3 成果の外部への発表	11
2. 4 活動	12
2. 5 実施体制	12
2. 6 整備実績一覧	13

1. 委託業務の目的

本計画は現在すでに世界有数のバイオ情報サービスとなっているゲノムネットを京都大学の事業と位置づけ、化学研究所バイオインフォマティクスセンターにおいて分子情報を中心とした統合データベースを構築する。革新的なウェブ技術とKEGGにおいて人手で構築された知識の体系を融合して、平成22年までにライフサイエンス分野における世界最高水準の知的情報基盤を確立する。

2. 平成21年度の実施内容

2. 1 実施計画

(1) 共通基盤技術開発

統合データベースを構築する基盤技術はこれまでのKEGGプロジェクトですでに確立しているので、本計画では化合物をはじめとする分子データを扱う統合データベースを利用するための技術開発が中心となる。

①知識処理技術開発

平成20年度に引き続き化学構造や化学反応に関する解析技術を開発し、ソフトウェア化を行う。とくに化学研究所バイオインフォマティクスセンターの研究成果をもとに、以下の開発を行い、実用的なソフトウェアをゲノムネットサービスの一部として順次公開する。

- ・ 光学異性体のチェック機能改良などによる化学構造検索システムの高機能化。
- ・ 酵素番号の自動割り当てシステムの改良と遺伝子情報へのリンクによる高機能化。
- ・ 化学反応ネットワーク予測のための反応経路検索システムの整備。

また、医薬品やターゲットに関連した化学構造、反応の解析技術も開発する。

②ウェブ技術開発

化学情報データベースに対するキーワード検索システムの高機能化を継続する。具体的には、以下の2点を中心に開発する。

- ・ 同義語や階層分類情報を考慮した検索。
- ・ LinkDBの等価エン트리間の情報を考慮した統合検索。

(2) 統合データベース開発・運用

本計画ではライフサイエンスの観点から、医薬品・化学情報データベースの開発と運用を行う。また、統合化の基本となるLinkDBの開発・運用を継続して行う。

①医薬品データベース開発・運用

ゲノムネット医薬品データベースとして公開しているJAPIC添付文書情報の更新とKEGG DRUGとのリンク付けなど毎月の更新作業を継続して行う。またライフサ

イエンス辞書の同義語を用いた検索を実装して公開する。

②化学情報データベース開発・運用

各種化合物データベースの統合を引き続き進めるとともに、キーワード検索と構造検索を統合した検索を開発して公開する。また、食品・環境物質といった生体システムと関連の深い物質・病原性との関連物質に加え、化学反応の知識を統合し、ゲノムネット化合物データベースを発展させた化学情報データベースとして提供する。

③LinkDB開発・運用

データベース間のリンク情報に関するLinkDB検索システムの高速化、高機能化を継続して行う。高機能化に関しては、キーワードや類似化合物情報を用いた検索を開発する。また、医薬品・化合物データベース間のリンクの新規定義・更新作業も継続して行う。

(3) プロジェクトの総合的推進

分担機関である京都大学は、中核機関である情報・システム機構の全体戦略に従い連携して本事業を推進する。本プロジェクトの成果は直ちにゲノムネットサービス (<http://www.genome.jp/>) に反映し、利用者の意見を収集して今後の展開に資するとともに、中核機関での横断検索システムでの検索対象となるようにする。

2. 2 実施内容 (成果)

本統合データベースプロジェクトはゲノムネット (<http://www.genome.jp/>) をKEGGと分離して開発・運用するために提案し実施してきた。KEGGは現時点ではゲノムネットの主要サービス (<http://www.genome.jp/kegg/>) であるが、京都大学と東京大学の金久研究室が別予算で構築しており、KEGG独自のウェブサイト (<http://www.kegg.jp/>) も存在する。本計画ではゲノムネットを京都大学の事業と位置づけ、DBGET/LinkDBシステムを中心に統合化を行うものである。KEGGは統合化の対象データベースの中心であり、また化学情報解析ツールの一部はこれまでKEGGの中で開発されていたものを引き継いで本計画で開発している。平成21年度は当初計画で掲げた化学構造・化学反応解析ソフトウェアの高機能化、化合物データベースの開発、DBGET/LinkDBの高機能化を推進し、順次公開した(図1)。平成21年度の実施内容は以下の通りである。

(1) 共通基盤技術開発

①知識処理技術開発

平成20年度に引き続き、化合物の化学構造と化学反応に関する解析技術を開発し、ソフトウェア化を行った。具体的には、平成20年度に実施した調査に基づき化合物類似構造検索ツールSIMCOMPと部分構造検索ツールSUBCOMPの光学異性体のチェック機能を改善したプログラムを実装し、公開環境へと反映した。これにより、cis, transを正確に判定できるようになった。また、光学異性体の類似スコアを見直し、光学異性体の違いをより正確に反映できるようにするとともに、構造アライメントの精度向上を実現した(図2)。さらに、SIMCOMP/SUBCOMPのパッケージ化を進めており、平成22年度の早い段階に、FTPでダウンロードできる

ようにする予定である。

GenomeNet 統合データベース検索システム

Search for

ゲノムネット
ゲノムネットとは
お知らせ
謝辞

統合データベース
統合DBの概要
DBGETの概要
リリース情報
データベース増加図

医薬品データベース
利用法

KEGG
varDB
研究支援データベース

計算ツール
その他のツール

フィードバック

DBGET/LinkDB: ゲノムネット統合データベース検索システム

DBGET は世界中に存在する分子生物学データベースのウェブを対象とした統合データベースシステムです。上のサーチボックスを含め、ゲノムネットや KEGG のバックボーンシステムとして利用されています。分子生物学データのウェブは、各データベースのエントリー（ページ）をノードとし、エントリー間の参照情報をエッジ（リンク）とした巨大なグラフです。各データベースエントリーはデータベース名とエントリー名（またはアクセッション番号）のペアで指定され、これは一般には対応するページの URL に変換することができます。このような名前空間を考え、名前同士のつながりを蓄積したのが LinkDB データベースです。

DBGET サーチ
LinkDB サーチ
英文ドキュメント
How to use DBGET
URLs for making DBGET queries
リリース情報
データベースリリース情報（日々更新）
主要データベースの増加図（1982年より）

2007年4月より文部科学省統合データベースプロジェクトの支援を受け、日本語支援環境の整備を行います。LinkDB の拡張と新たな検索システムの開発を行っています。また、ゲノムネット化合物データベースとして、化合物に関する様々なデータベースの統合化を進めています。これまでに以下のデータベースが DBGET/LinkDB システムに組み込まれています。

ゲノムネット化合物データベース

区分	データベース	内容
化学構造	PubChem	化合物全般
	ChEBI	化合物オントロジー
	KEGG COMPOUND	代謝化合物、生体外化合物
	LIPIDMAPS	脂質
	LipidBank	脂質
	KNASack	植物二次代謝化合物
	KEGG DRUG	医薬品
	DrugBank	医薬品
	KEGG GLYCAN	糖鎖
	PDBChem	低分子立体構造
立体構造	3DMET	代謝化合物立体構造モデル
	LigandBox	医薬品立体構造モデル
化学反応	KEGG REACTION	生体内化学反応
	KEGG RPAIR	反応ペア

ゲノムネット医薬品データベース

区分	データベース	内容
添付文書	JAPIC 医療用医薬品	処方箋添付文書
	JAPIC 一般用医薬品	大衆薬添付文書

Last updated: February 22, 2010
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

図1. ゲノムネット統合データベース検索システムの日本語ページ。ゲノムネットでサポートしている化合物データベースをまとめて、ゲノムネット化合物データベースとして参照できるようにしている。平成21年度はゲノムネット医薬品データベースへの入口を明示的に表示するようにして、ゲノムネット統合データベース中での位置づけを明確にした。また、検索ボックスで「医薬品」を選択し、日本語でキーワードを入力すると、KEGG DRUGの日本語版以外にJAPICの添付文書情報を同時に検索できるようにした。平成19年度から引き続き開発している化学構造・化学反応解析ソフトウェアへのリンクは左側の赤枠（計算ツール）で示している。

Page : 1 Go of 22 Items : 1 - 20 of 437 Top Previous Next Bottom

Top 10 Clear Select operation (2) View

No	Entry	Structure	Name
1	C00124	 Similarity Score : 1.00	D-Galactose
2	C00737	 Similarity Score : 1.00	D-Aldose D-Aldose1 D-Aldose2
3	C00738	 Similarity Score : 1.00	D-ribose Ribose
4	C00962	 Similarity Score : 1.00	beta-D-Galactose
5	C00031	 Similarity Score : 0.98	D-Glucose Grape sugar Dextrose Glucose
6	C00021	 Similarity Score : 0.98	beta-D-Glucose

図 2. 化合物類似構造検索ツールSIMCOMPの検索結果画面の例。D-Galactose (C00124) をクエリ構造として検索するとキラル情報が反映され、D-Glucose (C00031) はスコアが低くなるようになっている。

酵素番号の自動割り当てシステムの改良については、入力された反応が複数の基質や生成物を持つ場合に、自動割り当てに必要な基質と生成物の 1 対 1 のペアを抽出する方法を実装しテストした。遺伝子情報へのリンク機能も含め引き続きテストを行い、平成 22 年度に公開する予定である。

化学反応ネットワーク予測ツールに関しては、始点と終点（または始点のみ）の化合物構造を入力し、その間の反応経路を予測するシステムPathPredを公開した

（図 3 : <http://www.genome.ad.jp/tools/pathpred/>）。PathPredでは、環境物質の分解経路で使われている反応から抽出した変換ルールを適用することによって、環境物質分解経路を予測できるようにしている。また、医薬品やターゲットに関連した解析の第一歩として、漢方薬などの成分として重要な植物二次代謝産物の整合性経路で使われている反応から抽出した変換ルールも用意して経路予測に応用できるようにした。

②ウェブ技術開発

平成 20 年度に引き続き、化学情報データベース、医薬品データベースに対するキーワード検索システムの高機能化を行った。具体的にはライフサイエンス辞書を用いた同義語検索を実装し、医薬品データベースに適用した。また、LinkDBのリンク情報を各エントリーの検索結果に直接表示し、関連情報を見やすくした。さらに、LinkDBの等価エントリー間の情報を利用し、各種化合物・医薬品データベースのキーワード検索結果から、対応するKEGGのエントリーを直接取得できるようにした（図 4）。

PathPred: Pathway Prediction server

PathPred is a web-based server to predict plausible enzyme-catalyzed reaction pathways from a query compound using the information of RDM patterns and chemical structure alignments of substrate-product pairs. This server provides plausible reactions and transformed compounds, and displays all predicted reaction pathways in tree-shaped graph.

Reference pathway:

- ☒ Xenobiotics Biodegradation (Biodegradation of initial compound)
- ☐ Biosynthesis of Secondary Metabolites (Biosynthesis of initial compound)

Enter initial compound: (in one of the four forms)

MOL File Name: (ファイルを選択) ファイルが選択されています

MOL File Text:

SMILES: (Ex) Clc1cc(Cl)c(Cl)cc1Cl

KEGG Compound ID: (Ex) C00594

Options:

E-mail Address:

Simucomp Threshold: (0.0 - 0.5)

Prediction cycle: cycles (>= 0)

Pathway Prediction Result

Path List: <show all path>

Predicted Pathway: [Top] [Help]

図3. 化合物の構造を入力して、その化合物が代謝される経路を計算するツールPathPredの画面。構造入力画面ではリファレンスとなる反応パターンが実際に使われている経路を選択できるようにしており、現在は微生物による分解系と植物の二次代謝産物合成系に特徴的な反応パターンを選ぶことができる(赤枠で囲った部分)。右の図はテトラクロロベンゼンをクエリ構造として計算した結果と、その中から一番上の経路を表示したもの。ベンゼン環が開裂してコハク酸(C00042)まで分解されていく様子が分かる。ここで、CXではじまる番号がついている化合物(テトラクロロベンゼンから7つ)はデータベースに登録されていないもので、中間代謝物として新たに予測された構造である。

GenomeNet

Search: Chemical substance - Search term: GDP-L-fucose

Database: Chemical substance - Search term: GDP-L-fucose

KEGG COMPOUND

C00325 [KeggDraw](#) [Jmol](#)

GDP-L-fucose; GDP-beta-L-fucose

KEGG GLYCAN

G10615 [KeggDraw](#)

GDP-L-fucose; (GDP-L-fucyl)

PUBCHEM

24895171 [PubChem](#)

6-Deoxy-beta-D-galactopyranosylguanosine 5'-diphosphate; G4401; SIGMA; GDP-Fuc; GDP-fucose; Guanosine 5'-diphosphate-beta-L-fucose sodium salt

26756442 [PubChem](#)

GDP-beta-L-fucose; Disodium Salt; Guanosine 5'-diphosphate-beta-L-fucose, 2Na

2782 [PubChem](#)

2-amino-9-[4-hydroxy-5-[[hydroxy(hydroxy(3,4,5-trihydroxy-6-methyl-tetrahydropyran-2-yl)oxy)phosphoryl]oxy-phosphoryl]oxymethyl]tetrahydrofuran-2-yl-1,9-dihydropurine-6-one; GDP-L-fucose; GUANOSINE-5'-PHOSPHATE; FUCOSE; guanosine 5'-(trihydroxy-2'-phosphoryl)-beta-L-fucose; GDP-L-fucose; (GDP-beta-L-fucose); [CPO-C00325]

ChEBI

17009 [ChEBI](#) [KEGG-link](#)

GDP-L-fucose; [CPO-C00325]

28530 [ChEBI](#) [KEGG-link](#)

GDP-4-epi-beta-L-fucose; [CPO-C00369]

13332 [ChEBI](#)

GDP-beta-L-fucose

3DMET

B04668 [3DMET](#) [KEGG-link](#)

GDP-L-fucose; GDP-beta-L-fucose; [CPO-C00325]

KNAPSACK

C00007245 [KNAPSACK](#) [KEGG-link](#)

GDP-L-fucose; [CPO-C00325]

PubChem, ChEBI, 3DMET, KNAPSACK 等の検索結果から対応する KEGG へのリンク

KEGG

COMPOUND: C00325

Entry: C00325

Name: GDP-L-fucose; GDP-beta-L-fucose

Formula: C14H25N9O15P2

Mass: 589.0822

Structure:

Remark: Same as C10615

Reaction: R01951 R03519 R04081 R04553 R04646 R05104 R05692 R07040

Pathway: PATH: ko0051 Fructose and mannose metabolism; PATH: ko00520 Amino sugar and nucleotide sugar metabolism; PATH: ko01109 Metabolic pathways

Enzyme: 1.1.1.271 2.4.1.65 2.4.1.69 2.4.1.152

Other DB: PubChem: 24895171; ChEBI: 17009; KNAPSACK: C00007245; 3DMET: B04668

KEGG data: [Show](#)

All links

- [Pathway \(352\)](#)
- [KEGG PATHWAY \(352\)](#)
- [KEGG COMPOUND \(1\)](#)
- [Chemical substance \(6\)](#)
- [KEGG GLYCAN \(2\)](#)
- [PubChem \(1\)](#)
- [ChEBI \(1\)](#)
- [Jmol \(1\)](#)
- [KNAPSACK \(1\)](#)
- [PubChem \(1\)](#)
- [KEGG reaction \(24\)](#)
- [KEGG ENZYME \(7\)](#)
- [KEGG REACTION \(8\)](#)
- [KEGG ENZYME \(9\)](#)
- [All databases \(3092\)](#)

図4. 各種化合物データベースへのキーワード検索結果から、対応するKEGGのエントリーへの直接リンク(左)と、KEGGのエントリーに埋め込まれたLinkDB検索結果(右)。LinkDBにはKEGGのエントリーには直接書かれていないリンク(例ではMASSBANKへのリンク)も含めて一覧することができる。

(2) 統合データベース開発・運用

①医薬品データベース開発・運用

平成20年度に引き続き、JAPIC添付文書情報の更新とKEGG DRUG、PubMed、J-STAGEとのリンクづけを、毎月の更新に伴い継続的に行っている。また、ライフサイエンス辞書を利用した同義語・類義語検索についても平成20年度に実装したプロトタイプを公開した。また、副作用情報の調査も行い、ドイツのグループが提供しているSIDERをLinkDBに取り込むことで検索できるようにした(図5)。

②化学情報データベース開発・運用

平成20年度に統合検索を実現した7つのデータベースとKEGGデータベース間のリンクを、引き続き更新するとともに、新たにMassBank、日化辞、HMDB、JCGGDB、SIDERの5つのデータベースについてKEGG DRUG、COMPOUND、GLYCANとのリンク情報を定義してLinkDBでの検索を実現した(図5 All links参照)。キーワード検索と構造検索の統合についてはJAPICデータとKEGG DRUGの対応関係を利用した構造検索の方式を検討し、医薬品データベース検索への適用を予定している。また、食品・環境物質などの生体システムと関連の深い物質として、植物の二次代謝産物、バクテリアによって分解される環境物質に着目し、これらの生合成経路や分解経路で使われる反応と反応パターンの知識を整理し、知識処理技術開発における化学反応ネットワーク予測へと応用した(図3参照)。

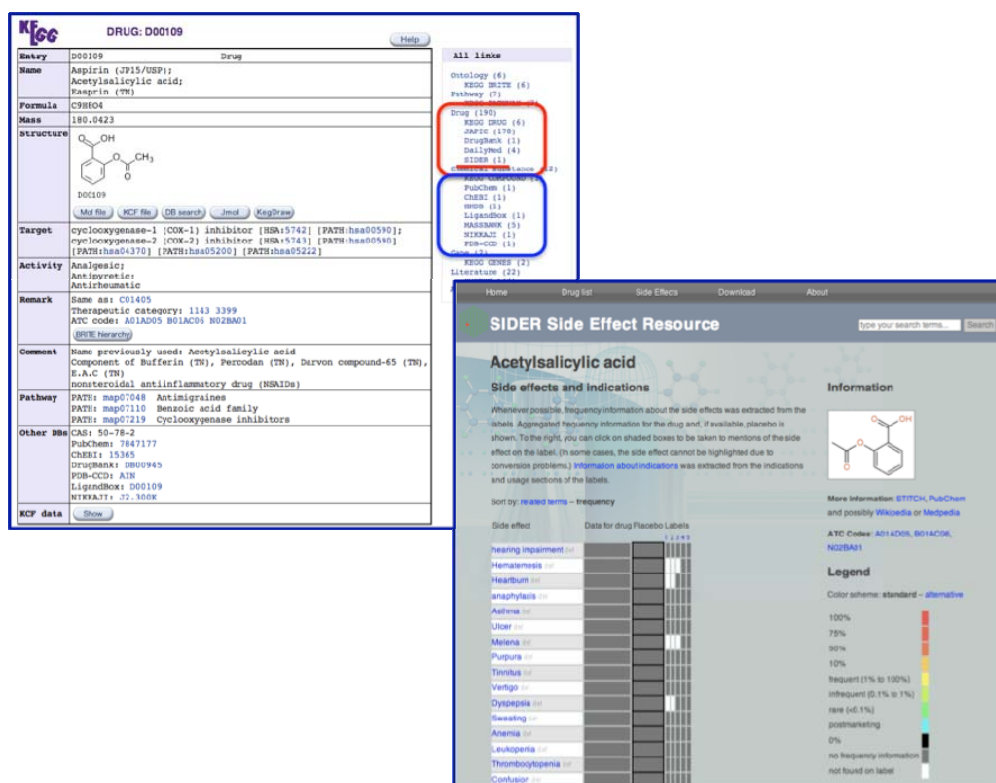


図5. 医薬品データベースKEGG DRUGから副作用情報データベースSIDERへのリンク。SIDERに副作用情報がある医薬品についてはKEGG DRUGのエントリーとの対応をLinkDBに登録しており、All linksでリンクを表示している(赤線部分)。右側は対応するSIDERのエントリー。副作用に関連しては医薬品添付文書データベースであるJAPICやDailyMedへのリンクも同様に表示している(赤枠部分)。このDRUGエントリーは平成21年度に新たに追加したHMDB、MassBank、日化辞(NIKKAJI)へのリンクも含んでいる(青枠部分)。

③LinkDB開発・運用

ゲノムネット化合物データベースとして提供する5つを新たに統合検索できるように LinkDBおよびSEQNEWを改良した。また、データの追加、変更、更新作業は順調に進んでいる。検索機能の高機能化については、ウェブ技術開発においてキーワード検索から直接KEGGのエントリを取得できるようにしたので、そちらを利用するようにした(図4参照)。また、ユーザ定義のリンク情報を使った検索機能をコマンドラインレベルで実装した。統合データベースプロジェクト内のデータベース間のリンクについて、ProNIT、ProTherm、長浜バイオ大のデータベースとKEGG GENESとの対応関係を定義し、LinkDBに組み込んだ(図6)。

The screenshot displays the LinkDB interface for the entry 'Thermococcus kodakarensis: TK2290'. The interface is divided into several sections: 'Gene name', 'Definition', 'Pathway', 'Class', 'SSDB', 'Motif', 'Structure', 'Position', 'AA seq', and 'NT seq'. The 'Definition' field is highlighted with a red box, and a red arrow points from it to the 'All links' section. The 'All links' section lists various databases and their corresponding IDs. The 'Structure' field shows a 3D protein structure. The 'Position' field shows the gene location on the chromosome. The 'AA seq' field shows the amino acid sequence. The 'NT seq' field shows the nucleotide sequence. The 'All links' section lists various databases and their corresponding IDs. The 'Structure' field shows a 3D protein structure. The 'Position' field shows the gene location on the chromosome. The 'AA seq' field shows the amino acid sequence. The 'NT seq' field shows the nucleotide sequence.

図6. 遺伝子データベースKEGG GENESから長浜バイオ大学で開発中の持続可能型社会への貢献遺伝子データベースと九州工業大学で開発中の熱力学パラメーターデータベースへのリンク。

(3) プロジェクトの総合的推進

中核機関の全体戦略に従い、化合物・医薬品を中心とする統合データベースの開発を推進している。また、平成21年度までの成果はすべてゲノムネット統合データベース、化学情報データベースとして公開しており、中核機関の横断検索のインデクシング対象となっている。システム開発に関する本プロジェクトの成果も速やかにゲノムネットサービスに反映し、公開している。参考までにゲノムネット医薬品データベースへのアクセス件数を表1に示す。また、平成22年2月にはゲノムネットデータベース利用講習会を開催した。

表1. ゲノムネット医薬品データベース (<http://www.genome.jp/kusuri/>) のアクセス数

	アクセス数	ユニーク IP数		アクセス数	ユニーク IP数
2008年4月	297,048	11,129	2009年4月	952,436	22,346
2008年5月	306,584	13,225	2009年5月	862,742	22,745
2008年6月	372,326	13,650	2009年6月	1,006,818	22,639
2008年7月	565,126	16,481	2009年7月	967,659	27,196
2008年8月	528,769	15,656	2009年8月	790,645	23,732
2008年9月	607,950	16,593	2009年9月	979,522	23,017
2008年10月	775,339	19,494	2009年10月	776,307	23,913
2008年11月	759,223	20,161	2009年11月	793,245	21,813
2008年12月	862,632	21,206	2009年12月	701,599	19,253
2009年1月	952,436	22,346	2010年1月	730,095	21,539
2009年2月	862,742	22,745	2010年2月	729,103	20,744
2009年3月	1,006,818	22,639	2010年3月	888,723	20,752

2. 3 成果の外部への発表

1. 学会等における口頭・ポスター発表				
発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
ゲノムネット化合物データベースの最新状況、ポスター発表	時松敏明、小寺正明、守屋勇樹、中川善一、服部正泰、五斗進、金久貴	第51回日本植物生理学会年会	2010.3.21	国内
Reclassification of Oxidoreductase Subclasses Based on the Relationships of Reductants and Oxidants, ポスター発表	Kotera, M., Tokimatsu, T., Nakagawa, Z., Moriya, Y., Hattori, M., Goto, S., and Kanehisa, M.	GIW2009	2009.12.15	国内
Structural Clustering of Plant Secondary Metabolite to Estimate Compound Classes Reflecting their Biosynthetic Pathway, ポスター発表	Tokimatsu, T., Nakagawa, Z., Kotera, M., Goto, S., Kanaya, S., and Kanehisa, M.	GIW2009	2009.12.14	国内
植物研究のための代謝化合物統合データベース、口頭発表	時松敏明、寺島悦子、木下広美、五斗進、金久貴	BMB2009	2009.12.12	国内
ゲノムネット化合物・医薬品データベースと関連ツールの開発、ポスター発表	五斗進、服部正泰、時松敏明、小寺正明、守屋勇樹、中川善一、金久貴	BMB2009	2009.12.10	国内
酵素反応式からの基質/生成物ペアの予測、ポスター発表	小寺正明、服部正泰、時松敏明、中川善一、五斗進、金久貴	BMB2009	2009.12.10	国内
E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs, 口頭発表	Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M.	ISMB2009	2009.6.30	国外
ゲノムネット化合物・医薬品データベース、ポスター発表	五斗進、服部正泰、時松敏明、小寺正明、守屋勇樹、中川善一、金久貴	統合データベースプロジェクト シンポジウム 2009	2009.6.12	国内
2. 学会誌・雑誌等における論文掲載				
掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌等名）	発表した時期	国内・外の別
varDB: a database of antigenic variant sequences — current status and future prospects	Diez, D., Hayes, N., Joannin, N., Normark, J., Kanehisa, M., Wahlgren, M., Wheelock, C.E., and Goto, S.	Acta Trop	2010	国外
Tools for investigating mechanisms of antigenic variation: new extensions to varDB	Hayes, C.N., Diez, D., Joannin, N., Kanehisa, M., Wahlgren, M., Wheelock, C.E., and Goto, S.	Genome Informatics	2009	国外

2. 4 活動

平成22年2月4日、5日

ゲノムネットデータベース講習会

発表者：五斗進他、開催場所：東京大学

概要：PCを用いた実習形式での講習会。ホームページ上で一般から30名の参加者を募った。大学、公的機関の研究所、企業から幅広く集まった。

2. 5 実施体制

研 究 項 目	担当機関等	研究担当者
1. 共通基盤技術開発 (1) 知識処理技術開発 (2) ウェブ開発技術	京都大学化学研究所 京都大学化学研究所 京都大学化学研究所	小寺正明 守屋勇樹 ○金久實
2. 統合データベース開発・運用 (1) 医薬品・化合物データベース開発 ・運用 (2) 統合データベース開発	京都大学化学研究所 京都大学大学院薬学研究科 京都大学化学研究所 京都大学化学研究所 京都大学化学研究所 京都大学化学研究所 京都大学化学研究所	服部正泰 金子周司 時松敏明 Nelson Hayes Diego Diez Ruiz 中川善一 ◎○五斗進
3. プロジェクトの総合的推進	京都大学化学研究所	◎○五斗進

注1. ◎:課題代表者、○:サブテーマ代表者

2. 6 整備実績一覧

整備実績一覧【代表機関名：京都大学】

(1) 保有データ情報
特になし

(2) データ（又はDB）の連結、統合化整備（※試験的、限定的公開済みのものも含む。）

通 番	データ（又はDB）の名称 ※URLがあれば記述	公開／ 未公開	概要（データの種別（生物種）・数量（kB等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述） ※ 公開している場合は、開始年月、利用状況（平均利用者数、アクセス数、ダウンロード数等の数値的指標で記述） ※ 必要に応じて画面コピー等の図表添付可
1	ゲノムネット医薬品データベース http://www.genome.jp/kusuri/	公開	研究の最先端と医療の現場さらには一般社会をつなぐ日本語の医薬品統合データベース。JAPIC 医薬品添付文書情報（医療用医薬品 13,514 件、一般用医薬品 12,313 件、平成 22 年 3 月現在）を検索可能。KEGG DRUG の構造情報やターゲット情報と統合している。また、PubMed や J-STAGE などの文献データベースへのリンクも付加している。公開は平成 19 年度よりしているが、JAPIC データの更新に伴い毎月の更新を行っている。また、平成 21 年度はゲノムネット統合データベースからキーワード検索できるようにした。利用状況については成果報告書の 2.2 を参照のこと。
2	ゲノムネット化合物データベース http://www.genome.jp/ja/gn_dbget_ja.html	公開	平成 20 年度までに「DBGET/LinkDB: ゲノムネット統合データベース検索システム」を化合物・化学反応データベースの統合化に応用して整備してきたデータベース。平成 21 年度は HMDB、MassBank、日化辞、SIDER、JCGGDB を LinkDB 対象データベースとして組み込んだ。

(3) DB基盤システム、ツール等開発成果物の整備（※試験的、限定的公開済みのものも含む。）

通 番	DB基盤システム、ツール等の名称	公開／ 未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述） ※ プログラムプロダクトに限らず、データ形式共通化、標準化のための仕様書、共通規約等のドキュメントについてもリリースしているものは対象とする。 （リリース済みドキュメントは参考として目次一覧、抜粋を添付） ※ 必要に応じて画面コピー等の図表添付可
3	SIMCOMP http://www.genome.jp/tools/simcomp	公開	類似化合物検索システム。グラフ比較に基づいた精度の高い類似度計算を実現している。平成 21 年度は平成 20 年度に実現した光学異性体の認識をより正確にできるようにし、類似スコアにも反映するように改良した。平成 22 年度中にはソースコードも含めて FTP でダウンロードできるようにする予定。
4	SUBCOMP http://www.genome.jp/tools/subcomp	公開	化合物部分構造検索システム。ビットストリングを用いた高速な部分構造検索システム。平成 20 年度に調査した結果に基づき、平成 21 年度は光学異性体の違いを認識するための改良、および、クエリを含む構造の検索を実装して公開した。平成 22 年度中にはソースコードも含めて FTP でダウンロードできるようにする予定。
5	E-zyme http://www.genome.jp/tools/e-zyme	公開	化学構造変化に基づく反応予測システム。基質と生成物を与えると、その間の反応パターンを抽出し、EC 番号との対応づけなどを行う。平成 20 年度中にデータベースを充実化し精度の向上を実現したが、EC 番号を自動的に割り当てることのできない反応があるため、平成 21 年度はそれらの反応を分類し、解析してきた。平成 22 年度中にその結果に基づいて改良する予定。
6	PathPred http://www.genome.jp/tools/pathpred/	公開	化学構造変化に基づく反応経路予測システム。基質（と生成物）を与えると、その基質を分解する経路や、生成物を合成するための経路を予測する。バクテリアが環境物質を分解する経路に特徴的な反応パターン（化学構造変化）や植物の二次代謝産物生合成経路に特徴的な反応パターンを用いて経路探索の効率化を図っている。平成 22 年 1 月 12 日に公開した。

(4) その他の成果物（(2)、(3)に該当しないもの）
特になし