

共通基盤技術開発 (1)

The screenshot shows a web browser window titled "SIMCOMP Help" with the URL "http://www.genome.jp/tools/simcomp/simcomp_help.html". The page features the SIMCOMP logo (two cracked eggs) and the text "SIMCOMP for chemical structure search". There are navigation tabs for "Home" and "Help". The main content area is titled "About SIMCOMP" and contains the following text:

About SIMCOMP

SIMCOMP is an efficient algorithm for comparing two chemical compounds, where the molecule is treated as a 2D graph consisting of atoms as vertices and covalent bonds as edges. It is an algorithm to solve the maximal common subgraphs of two graphs. Here, maximal common subgraphs can be found by searching for maximal cliques in the association graph, and introduced heuristics to accelerate the clique finding. (see reference)

SIMCOMP provides the atom alignments between two chemical compound graphs, then calculate the similarity of two chemical compounds by counting the number of matched atom alignments. (see reference)

For all calculations in SIMCOMP, the KEGG Atom Types are needed as a representation of atoms to detect biochemically meaningful features. The KEGG Atom Types are based on the concept of functional groups in chemistry, and 68 atom types (vertex types) have been defined for carbon, nitrogen, oxygen, and other atomic species with different environments.

The current version of GenomeNet structural search provides the availabilities for MDL/Mol file format and SMILES format. When those formats are input, they will be transformed into KCF formats internally.

KCF format and KEGG Atom Types

KCF (KEGG Chemical Function) format is a format of chemical objects like chemical compounds, chemical structures, and reactant pairs, which has been defined by KEGG. In particular case of chemical

An inset window titled "Compound Data Search Result" shows a table with columns for "No.", "SMILES", and "Structure". It lists several search results with their corresponding SMILES strings and chemical structures. Below the table is the caption "Example of search result".

類似構造検索プログラム SIMCOMP の高速化
ドッキンググラフの作成方法を改良することにより
従来より約10倍の高速化を実現した。
(7月に公開サイトに反映済み)

部分構造検索プログラム SUBCOMP の整備
二重結合の認識ミスを修正。
KEGG Atom Type による芳香環の認識。

キーワード検索やLinkDB検索に構造情報を利用するための方法は、現在検討中

共通基盤技術開発（2）

e-zyme result

http://www.genome.jp/ligand-bin/e-zyme2/result.cgi

e-zyme result

Pair 1 [KCF] [Edit alignment](#)

C00077 C00327

N1a-N1b: *-C5a:C1b-C1b

EC assignments

EC number	Weighted score	Observed freq.
3.5.1	145.5	30.0
2.3.1	54.2	20.0
6.3.2	36.2	7.0
2.3.2	14.4	3.0
6.3.1	7.6	4.0
3.4.13	7.5	
2.1.3	3.9	3.0
6.3.4	3.7	
6.3.5	3.6	
3.4.11	3.0	2.0
2.1.2	2.1	
3.5.3	1.6	
6.2.1	1.5	

酵素番号自動割り当てプログラム e-zyme の改良
データベース中の反応パターン分布に基づく
スコアリング。
検索結果表示インターフェースの改良。
(12月中を目処に公開サイトに反映予定)
IUBMB での EC 割り当てでも試験運用。

化学反応ネットワーク予測
反応パターンを利用した新規化合物の反応経路
予測システムを構築中。
解候補の爆発への対処法を検討中。
(ゲノム、パスウェイ情報を利用するなど)

統合データベース開発・運用（1）

ゲノムネット - DBGET の概要

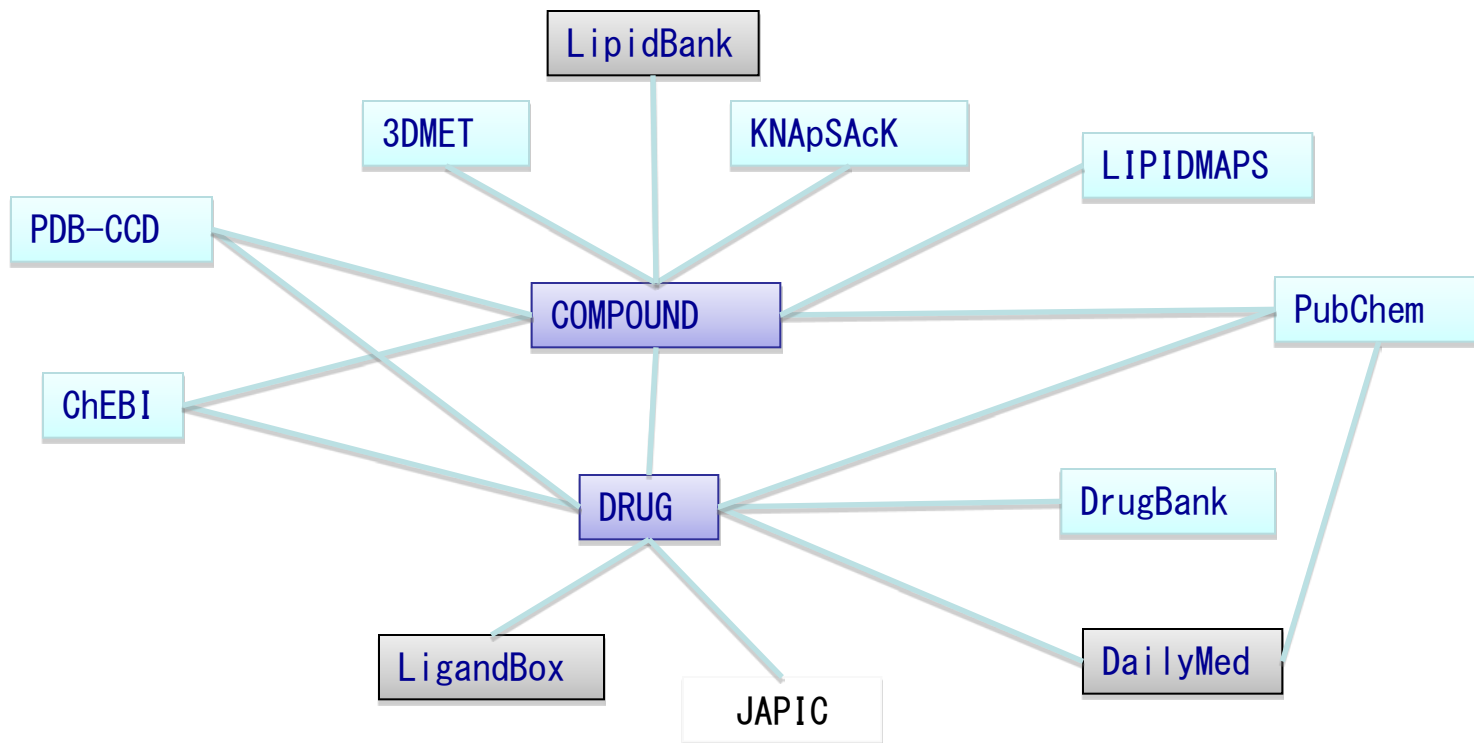
http://www.genome.jp/ja/about_dbget_ja.html

2. データベースの分類

DBGET/LinkDB システムでは多数のデータベースを統合するために、データベース利用条件の違い（ミラリング可、キーワードインデクシング可、リンクのみ）を考慮して、各データベースを以下の5つのカテゴリーに分類しています。

カテゴリー	検索コマンド			備考
	bget	bfind	blink	
1. KEGGデータベース	yes	yes	yes	ゲノムネットでミラリング
2. その他のDBGETデータベース	yes	yes	yes	
3. Web上の検索可能データベース	no	yes	yes	
4. Web上のリンクのみのデータベース	no	no	yes	各サイトのサービスを利用
5. PubMedデータベース	yes	no	yes	

- カテゴリー 1
- カテゴリー 3
- カテゴリー 4



統合データベース開発・運用（2）

ゲノムネット - DBGET の概要

http://www.genome.jp/ja/about_dbget_ja.html

2. データベースの分類

DBGET/LinkDB システムでは多数のデータベースを統合するために、データベース利用条件の違い（ミラリング可、キーワードインデクシング可、リンクのみ）を考慮して、各データベースを以下の5つのカテゴリーに分類しています。

カテゴリー	検索コマンド			備考
	bget	bfind	blink	
1. KEGGデータベース	yes	yes	yes	ゲノムネットでミラリング
2. その他のDBGETデータベース	yes	yes	yes	
3. Web上の検索可能データベース	no	yes	yes	各サイトのサービスを利用
4. Web上のリンクのみのデータベース	no	no	yes	
5. PubMedデータベース	yes	no	yes	

5. Web上の検索可能データベース (カテゴリー3)

ゲノムネットでキーワード検索のみできるデータベースは以下の通りです。

データベース	内容	元サイト
genbank		NCBI
insdc	核酸塩基配列	国立遺伝学研究所
ddbj		
embl		EBI
ncbi-gene	Entrez 遺伝子データベース	NCBI
unigene	UniGene 遺伝子データベース	NCBI
ensembl	真核ゲノムアノテーション	Ensembl
hgnc	ヒト遺伝子名	HGNC
go	遺伝子オントロジー	GO
ipi	タンパク質リファレンス配列	EBI
interpro	タンパク質ファミリー・ドメイン	EBI
omim	遺伝病	NCBI
pubchem	PubChem 化合物データベース	NCBI
chebi	ChEBI 化合物データベース	EBI
pdb-ccd	PDB リガンド辞書	PDB
lipidmaps	脂質代謝	LIPIDMAPS
knapsack	植物二次代謝産物	KNAPSAcK
3dmet	天然化合物の立体構造	3DMET
drugbank	医薬品とターゲット	DrugBank

参考文献

1. Kanehisa, M.; Linking databases and organisms: GenomeNet resources in Japan. Trends Biochem Sci. 22, 442-444 (1997). [pubmed]
2. Fujibuchi, W., Goto, S., Mizumatsu, H., Uchiyama, I., Ooiwara, A., Akiyama, Y., and Kanehisa, M.:

化合物データベースのキーワード検索機能
KEGG 以外の DB をカテゴリー3として登録
PubChem, ChEBI, PDB-CCD, LIPIDMAPS
KNAPSAcK, 3DMET
DrugBank
(7月に公開環境でサービス開始)

統合データベース開発・運用（3）

キーワード検索結果画面

DBGET Search Result: 統合データベース atp

医薬品

KEGG DRUG

D02300 [KegDraw](#) [Jmol](#)
Adenosine triphosphate disodium hydrate (JAN); Aden
5-triphosphate disodium; ATP (TN)

DRUGBANK

DB01661 [DrugBank](#)
Phosphoribosyl Atp; PRT
DB01763 [DrugBank](#)
Tatp
DB03319 [DrugBank](#)
(S)-Atpa, (S)-2-Amino-3-(3-Hydroxy-5-Tert-Butyl-Isoxaz
DB04631 [DrugBank](#)
Atpenin A5
DB04796 [DrugBank](#)
THIO-ATPA

化合物・糖鎖

KEGG COMPOUND

C00002 [KegDraw](#) [Jmol](#)
ATP; Adenosine 5'-triphosphate
C00131 [KegDraw](#) [Jmol](#)
dATP; 2'-Deoxyadenosine 5'-triphosphate; Deoxyaden
triphosphate
C02739 [KegDraw](#) [Jmol](#)
Phosphoribosyl-ATP; N1-(5-Phospho-D-ribose)-ATP; 1
C04727 [KegDraw](#) [Jmol](#)
[Propionyl-CoA:carbon-dioxide ligase (ADP-forming)]; |
forming)); Holo-[propionyl-CoA:carbon-dioxide ligase (
dioxide ligase (ADP-forming)); Holo-[propionyl-CoA-ca
CoA-carboxylase (ATP-hydrolysing)]
C04763 [KegDraw](#) [Jmol](#)
Apo-[propionyl-CoA:carbon-dioxide ligase (ADP-formir
hydrolysing)]
... » display all

PUBCHEM

10317181 [PubChem](#)

CHEBI

20855 [ChEBI](#)
ATP-sugars
20629 [ChEBI](#)
5-phosphoribosyl-ATP
18263 [ChEBI](#)
1-(5-phospho-D-ribose)-ATP
16284 [ChEBI](#)
dATP
15422 [ChEBI](#)
ATP
... » display all

PDB-CCD

ATP [PDB-CCD](#)
ADENOSINE-5'-TRIPHOSPHATE
PRT [PDB-CCD](#)
PHOSPHORIBOSYL ATP

3DMET

B01125 [3DMET](#)
ATP; Adenosine 5'-triphosphate
B01169 [3DMET](#)
dATP; 2'-Deoxyadenosine 5'-triphospha
triphosphate
B04862 [3DMET](#)
Phosphoribosyl-ATP; N1-(5-Phospho-D-ribose)-ATP; 1-(5-Phosphoribosyl)-ATP
B02104 [3DMET](#)
3'-Keto-3'-deoxy-ATP

KNAPSACK

C00001491 [KNAPsACK](#)
ATP; Adenosine triphosphate
C00007351 [KNAPsACK](#)
Phosphoribosyl-ATP

酵素反応

KEGG ENZYME

1.13.12.7

LipidBank, DailyMed は、現在のところ LinkDB のみ。

構造データに関しては LigandBox, 3DMET, PDB-CCD と COMPOUND, DRUG の関係を LinkDB で定義。

ゲノムネット医薬品データベースの副作用情報などはキーワードの抽出中（解析用）。