3DinSight : 生体分子統合データベース

URL : <http://www.rtc.riken.go.jp/jouhou/3dinsight/3dinsight.html>

理化学研究所筑波研究所 皿井明倫, 安 虹江*, 北島浩司

(*現在の所属 : The Scripps Research Institute, USA)

3DinSight is an integrated database and search tool for the structure, property and function of biomolecules. 3DinSight consists of an integrated database, a set of search programs, a WWW interface and visualization tools. It has been open to public through the Internet. 3DinSight would help researchers to get insight into the relationship among the structure, property and function of biomolecules.

database / biomolecules / structure / function / property

1. はじめに

インターネットの普及に伴ってさまざまなデータベースに手軽にアクセスできるようになったが, その反面, インターネット上に分散する情報を効率よく検索することが難しくなっている. 異なるデータベースエントリー間にリンクが張られているだけでは, 断片的な情報しか得られず, 検索の効率と品質は低い. 効率的な検索には, それらが統合された環境が必要である. 我々は, 分子のレベルで重要な配列, 構造, 機能と物性に関するさまざまな情報を1つのデータベースに統合し, 検索や可視化のインターフェイスを通して, これらの情報の間の関係を解析するシステムを作り, インターネット上に公開しているので, ここでその利用法を紹介する.

2. データベースの内容

タンパク質の立体構造情報はPDB, 配列データはPIR, SWISSPROTなど, 機能情報はPROSITE, Protein Mutant Database (PMD) など, 物性情報としてはアミノ酸の物理化学性質, 独自に文献から収集したタンパク質の熱力学データ¹⁾, タンパク質-核酸相互作用データ²⁾などを統合して関係型データベースのテーブルに入れている. テーブル間はエントリーIDなどによって関連付けられ, データベース検索言語によって柔軟かつ効率的な検索に対応できる. データ間の関連づけと

しては単にリンクを張るだけではなく, より具体的情報までデータベース化している. たとえば, 機能モチーフとタンパク質配列および構造との関係について, あるタンパク質はどういうモチーフをどこにもつか, 逆に, あるモチーフがどういったタンパク質のどこにあるかを, 前もってデータベースに統合している. これにより, 断片的な事象からタンパク質の構造と機能の相関などの全体像を洞察することが可能になる.

3. 3DinSight の利用法

上記のアドレスを入力すると, 初期画面が表示される. 画面の左欄に, Quick Search (簡単なキーワード検索), Advanced Search, What's New (最新改訂情報), More about 3DinSight (より詳細な説明とヘルプ)などの項目がある. Start search from では, 後述のようにさまざまな検索インターフェースから検索を開始することができる.

Advanced Searchはメインの検索インターフェイス (Fig.1) である. 検索画面では, 複数の項目が指定された場合, 項目間には論理積 (AND) を取る. なお, 項目の詳細な説明を参照するには, 各項目の後ろにあるヘルプマーク (?) をクリックする. 画面の Database Options では, 検索目的に従って検索対象を non-redundant な構造サブセット, DNA, DNA-タンパク質複合体などに絞ることができる. の Display Options では, 立

3DinSight : Integrated Database for Structure, Property and Function of Biomolecules

Akinori SARAI, Jianghong AN* and Kouji KITAJIMA

The Institute of Physical and Chemical Research, Tsukuba Institute

*Present Address : The Scripps Research Institute, USA

体構造を検索、表示する際、特定の機能情報が付随するものだけに絞って検索し、構造にマッピングするかどうかを指定できる。の入力部分では、Entry (PDBのエントリーコードを指定)、Keywords (2つまで指定でき、AND、あるいはORの関係性を指定できる; また、ワイルドカード*で任意の文字列、?で任意1文字を指定できる)、Authors、Resolution (解像度)などで検索できる。Sequence patternでは、正規表現による部分配列の検索ができる。たとえば、A(S,T)2DはAに続いてSかTが2回あってからDが来るパターンである。ASSD、ASTD、ATTD、ATSDがその例である。検索範囲は特定の2次構造 (HELIX, SHEET, TURN)、あるいは配列全体のいずれかを指定する。で結果リストの表示したい範囲を指定する。Defaultは1~30になっている。ヒットが多い場合、結果を1回で表示すると遅いので、これを調整することにより、ヒットリストの任意の部分が見られる。のStartボタンをクリックすると検索が開始される。

Fig.1 Main search interface of 3DinSight. After selecting database options, display options, filling searching condition and display range (30 is default), click start button.

Fig.1では、例として「解像度3 Å以内のProtein-DNA complex 構造データセットから、アミノ酸Pで始まる Helix を含み、リガンドを結合している」エントリーを検索しており、結果の一部がFig.2に示されている。この画面には対応するPDBコードとともに、分子名、著者、どのChainにどのようなリガンド (化学式と名前) が結合しているか、Helixのアミノ酸配列と位置、解像度、構造に含まれるChain, Helix, Sheet, Turnの数がリスト表示される。この内容はDisplay Optionにより多少異なる。ここでPDBコードのボタンをクリックすると、Fig.3のような画面になる。ここから、リガンドがハイライトされた構造をRasMolやVRMLで表示したり、アミノ酸の性質 (疎水性など) のプロファイルを構造情報や機能情報とともに2次元のグラフで表示したり (AMINO ACID ANALYSISからリンク) できるほか、SCOPや配列データ、文献データへもリンクしている。また、そのタンパク質に対応する熱力学データや相互作用データがあれば、それらの詳しいデータにリンクできる。この例では、タンパク質-DNA複合体に関する詳細な情報 (PROTEIN-NUCLEIC ACID COMPLEXからリンク) と、複合体中のアミノ酸と塩基の特異相互作用の情報 (BASE-AMINO ACID INTERACTIONからリンク) を調べることができる。

以上がメインの検索の手順であるが、ほかのいろいろなインターフェイスからも検索をスタートできる。

- (1) PIR, SWISS-PROTなどの配列データのキーワード検索
- (2) PROSITE モチーフのキーワード検索
- (3) PMD をキーワード、変異種類などから検索
- (4) 疾病の名前や関与するタンパク質の変異情報から検索
- (5) 結合リガンド情報をキーワード、分子名、分子式、距離などから検索

Search Condition										
Selected Database Protein-DNA Complex Subset										
Display Linking Ligand information										
Limited list displayed to: start 1 end 30										
Pattern <P in HELIX										
Resolution 3 <= A										
ENTRY	HEADER	AUTHOR	CH	FORMULA	NAME	HELIX	POS	PAT	P	RES C H S T
1a1v	COMPLEX (RNA HELICASE/DNA)	J.L.KIM, K.A.MORGENTERN, J.P.GRIFFITH, M.D.DWYER, J.A.THOMSON, M.A.MURCKO, C.LIN, P.R.CARON	A	C5 H11 N1 O3 S2	S-(THIOETHYLHYDROXY)CYSTINE	PITKYIMTC	614	1	2.200000	2 16 5 0
			A	C5 H11 N1 O3 S2	S-(THIOETHYLHYDROXY)CYSTINE	PAETTVRLRAYMN	506	1	2.200000	2 16 5 0
			A	C5 H11 N1 O3 S2	S-(THIOETHYLHYDROXY)CYSTINE	PYLVAYGATVCARA	558	1	2.200000	2 16 5 0
1a74	COMPLEX (HOMING ENDO NUCLEASE/DNA)	M.S.JURICA, K.E.FLICK, R.J.MONNAT JUNIOR, B.L.STODDARD	A	ZN1 2+	ZINC ION	PLH	108	1	2.200000	4 8 6 0
			B	ZN1 2+	ZINC ION	PLH	108	1	2.200000	4 8 6 0

Fig.2 Search result. Searching conditions are shown at upper left. Only a part of the list is shown here.

pd1a1v.ent [display help]

Rasmol
 RASMOL ☐ Original BACKBONES
☒ Modified

Version
 VRML ☐ 1.0 BACKBONES (Line)
☒ 2.0

compress
☒ none
☐ gzip

FULL DATA
[Amino Acid Analysis](#)
[Protein-Nucleic Acid Complex](#)
[Base-Amino Acid Interactions](#)
[Cross-Link to SCOP database of protein family 1a1v](#)

for identity more than %

```

HEADER      COMPLEX (RNA HELICASE/DNA)                17-DEC-97   1A1V
TITLE       HEPATITIS C VIRUS NS3 HELICASE DOMAIN COMPLEXED WITH
TITLE       2 SINGLE STRANDED DNA
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: HELICASE;
COMPND      3 CHAIN: A;
COMPND      4 SYNONYM: NS3 DOMAIN;
COMPND      5 ENGINEERED: YES;
COMPND      6 MUTATION: N-TERMINAL MET, K221Q, A277G, S301L, S332P,
COMPND      7 S410A, G530E, R582W, AND A 10 RESIDUE (GSGSHHHHH)
COMPND      8 HISTIDINE TAG ATTACHED TO THE C-TERMINUS;
COMPND      9 BIOLOGICAL_UNIT: MONOMER;
COMPND      10 MOL_ID: 2;
COMPND      11 MOLECULE: 5'-D(*UP*UP*UP*UP*UP*UP*U)-3';
COMPND      12 CHAIN: B;
COMPND      13 ENGINEERED: YES;
COMPND      14 OTHER_DETAILS: SINGLE STRANDED DNA
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HEPATITIS C VIRUS;
SOURCE      3 ORGANISM_COMMON: HCV;
SOURCE      4 STRAIN: H;
SOURCE      5 GENE: NS3;
SOURCE      6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      7 EXPRESSION_SYSTEM_STRAIN: JM109;
  
```

Fig.3 A result page corresponding to a PDB structure.

- (6) タンパク質と変異体の熱力学データから検索
- (7) タンパク質-核酸複合体の場合, 相互作用実験データ, 複合体情報から検索, また, 塩基とアミノ酸の特異的な相互作用をキーワード, 原子種類, 距離情報などから検索
- (8) アミノ酸 / 塩基配列のホモロジー検索: FASTA, SSEARCHおよびBLASTにより, PDB中の構造データに対応する配列を検索しオンラインで結果を見ることができる。

なお, 機能部位をマップした構造を表示するため RasMol のプログラムを改良しており, これらの構造表示のためには, あらかじめ改良した RasMol をダウンロードしてインストールしておく必要がある。詳しい説明は More about 3DinSight にある。

3DinSight の目的は, 生体分子の構造, 機能と物性の情報を統合することにより, 個別の情報だけでは得ることが困難な知識を新たに抽出することにある。たとえば, タンパク質のいろいろな構造部位に起きるさまざまなアミノ酸変異が安定性などの熱力学量や他分子

との相互作用にどのような影響を与えるかを系統的に調べることにより, タンパク質の構造安定性のメカニズムや分子認識の特異性に関する洞察を得ることが可能になる。また, こうしたアミノ酸変異がひいてはどのような疾病に結びつくのかを調べることもできる。最近では情報の統合化がいろいろ試みられるようになってきたが, 3DinSight の特徴は, 情報の分子レベルでの統合に特化し, 自前の熱力学データや相互作用データなどを統合しているところにあると思う。

データベースのより詳しい解説については文献3と4を参照されたい。また, ご意見を jouhou@rtc.riken.go.jp までお寄せいただきたい。

文 献

- 1) Prabakaran, P. *et al.* (2001) *Bioinformatics* **17**, 1027.
- 2) Gromiha, M. M. *et al.* (2002) *Nucleic Acids Res.* **30**, 301.
- 3) An, J. *et al.* (1998) *Bioinformatics* **14**, 188.
- 4) An, J. and Sarai, A. (1999) *Trends in Glycoscience and Glycotechnology* **11**, 429.

皿井明倫

皿井明倫 (さらい あきのり)

理化学研究所筑波研究所 副主任研究員

名古屋大学理学研究科博士課程修了(物理学専攻; 理学博士)後, Illinois 大学生物物理学科客員研究員, 米国 National Institute of Health 客員研究員などを経て現職.

研究内容: 生体分子とその相互作用の研究. また, これらの研究に必要なデータベースの開発を行っている.
連絡先: 〒305-0074 つくば市高野台 3-1-1

E-mail: sarai@rtc.riken.go.jp

安 江虹 (あん じあんぼん)

Scripps Research Institute, Postdoctoral researcher

筑波大学工学研究科修了(工学博士)後, 理化学研究所を経て現職.

研究内容: 生体分子情報の統合データベースの開発とそれを利用した情報解析.

連絡先: Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037, U. S. A.

E-mail: jianghon@scripps.edu

北島浩司 (きたじま こうじ)

アドバンステクノロジーインスティテュート(株) 研究員

明星大学情報学研究科修士課程修了

研究内容: データベース開発

連絡先: 〒305-0074 つくば市高野台 3-1-1 理化学研究所筑波研究所

E-mail: kkitajim@rtc.riken.go.jp