Chapter 4. Protein Databases

Huaqin He
(College of Life Sciences, FAFU)

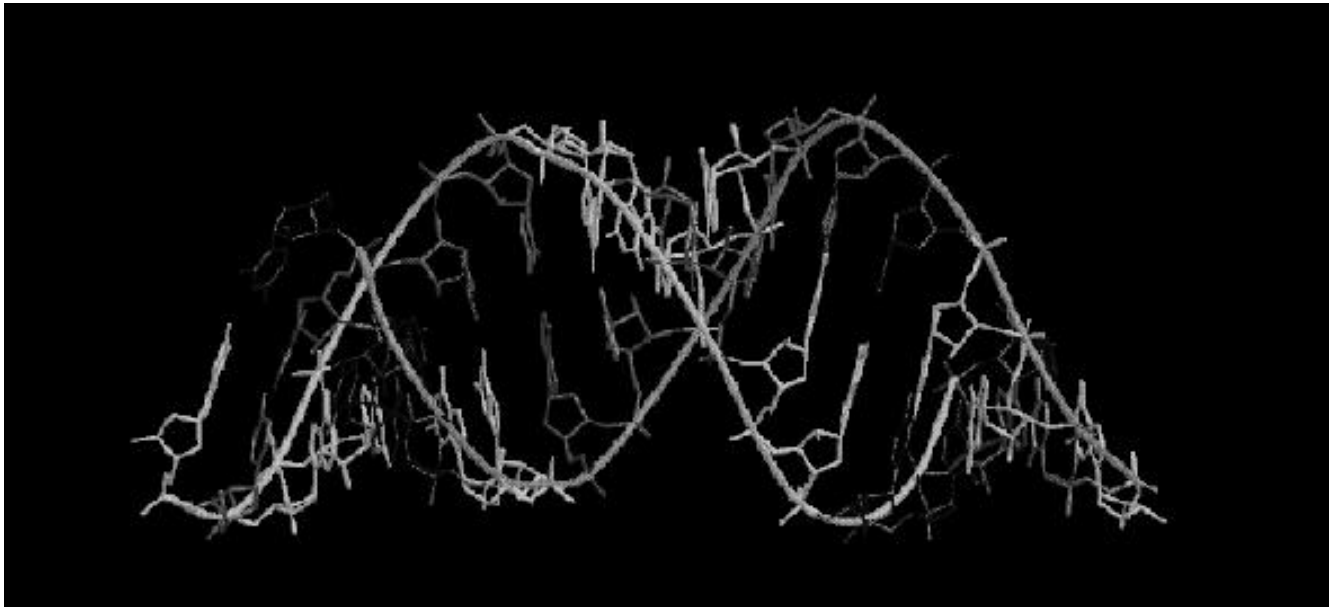
PPT slides and Message @ http://jxpt.fafu.edu.cn/meol/homepage/common/

Email: 1156743645@qq.com

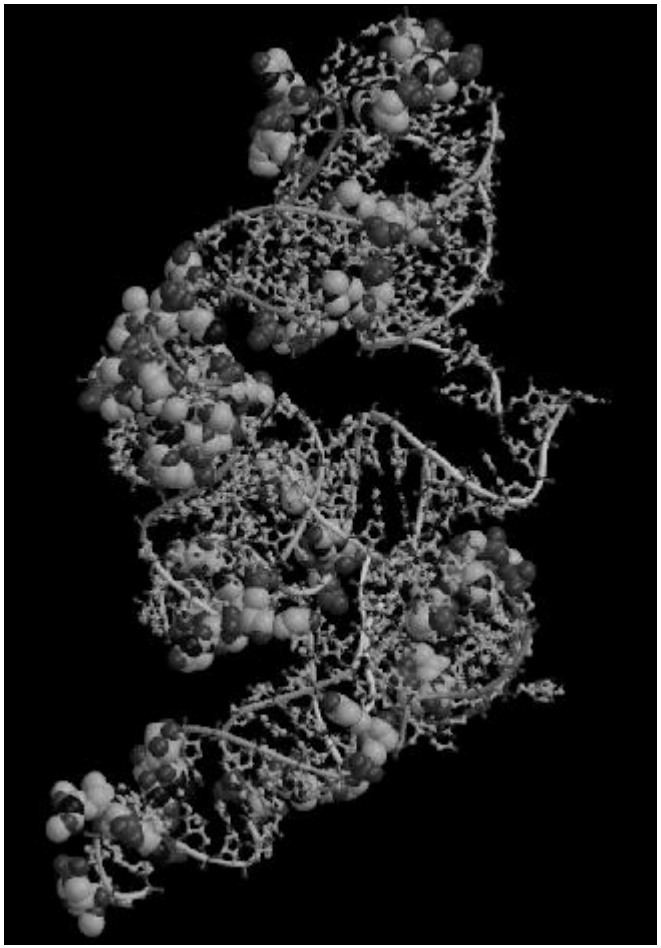
OUTLINES

1. Protein sequence databases
2. Protein motif and domain databases
3. Protein structural databases
4. Protein structural classification databases

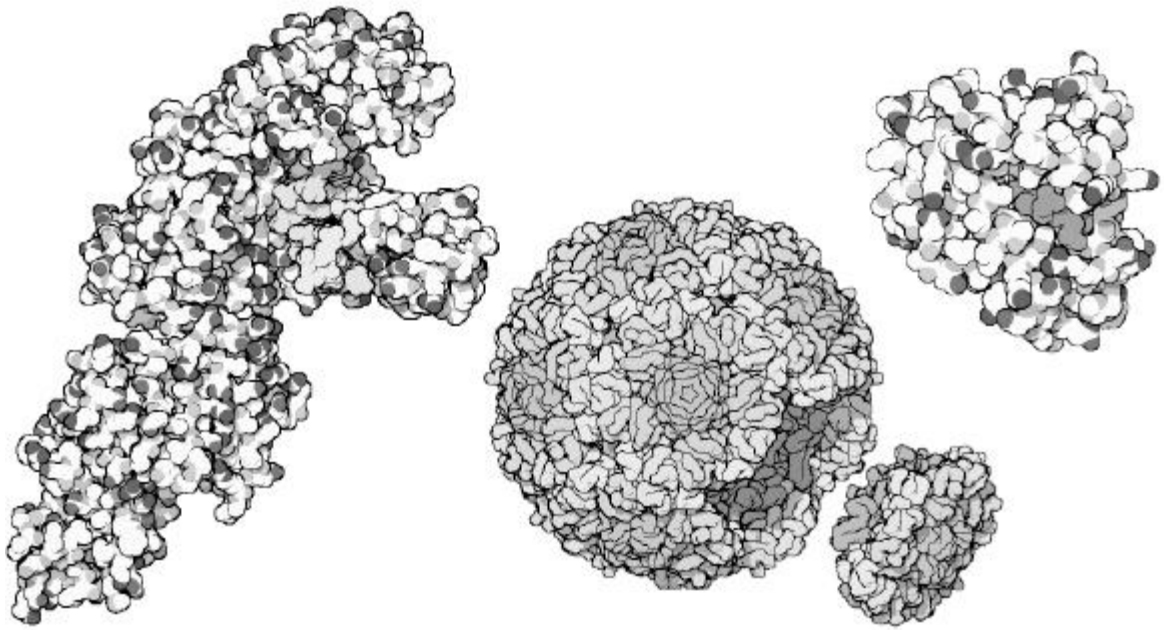
Structures---DNA



Structures---RNA

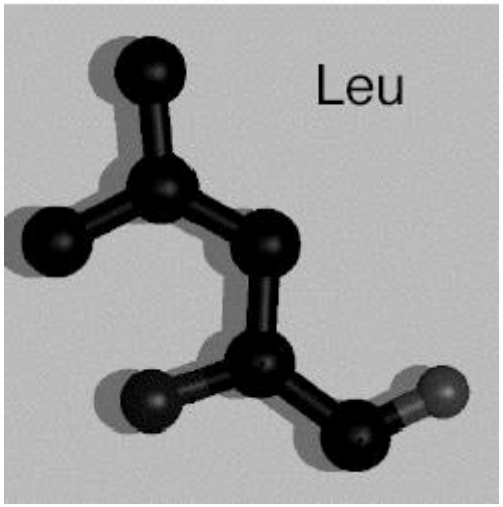


Structures---Protein



The "Average" Amino Acid

- ~ 200 residues per protein
  - -- 200 CA atoms, separated by 3.8 Å
- "average" residue: Leucine
  - -- 4 backbone atoms
  - -- 4 side chain atoms
  - -- Volume: 150!Å

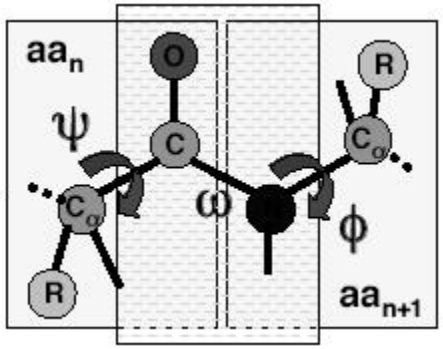


~1500 xyz triplets per protein

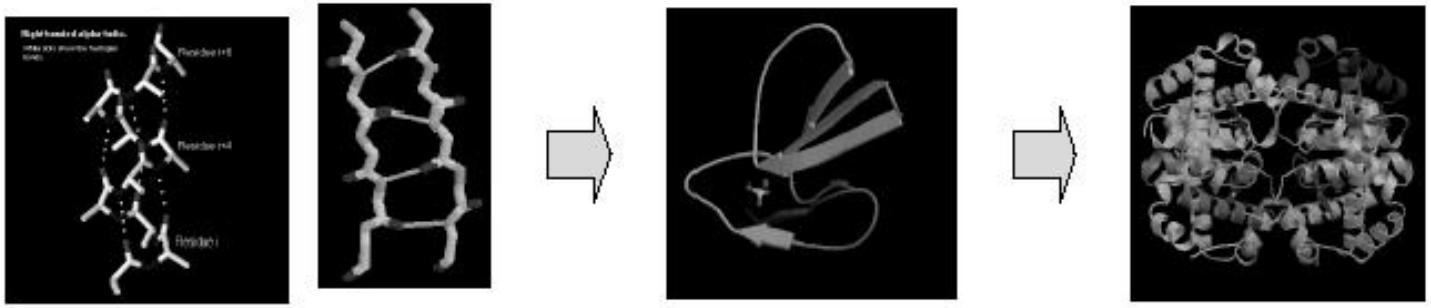
Peptide Bonds

- Primary Structure: Protein sequence
  - -- Two torsion angles: Ψ, φ
  - -- Many combinatorial possibilities of side chain interactions
- Secondary structure:
  - -- α-helix, β -strand, loops
- Tertiary structure
- Quaternary structure

Peptide Bonds



•••LPPVIHTWEAF HA GGL•••



1. Sequence Database

   PIR, SWISS-PROT

2. Motif and Domain databases:

   PROSITE, Pfam

3. Protein structure database

   PDB

4. Protein structure classification databases

   SCOP, CATH

1. Protein Sequence Databases

1.1 PIR (protein information resource) and PSD (protein sequence database)

pir.geoegetown.edu/pirwww

❶ Generated from GenBank/EMBL/DDBJ ;

❷ Sequences from publications;

❸ Submission.

1.2 SWISS-PROT/TrEMBL

www.expasy.org/swissprot

❶ Searched by SRS, ID, …

ExPASy - Swiss-Prot and TrEMBL - 夏新笔记本电脑

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 搜索 收藏夹 媒体

地址(D) http://us.expasy.org/swissprot/ 转到 链接

ExPASy Home page | Site Map | Search ExPASy | Contact us | PROSITE | Proteomics tools

Search Swiss-Prot/TrEMBL for Go Clear

Swiss-Prot
Protein knowledgebase
TrEMBL
Computer-annotated supplement to Swiss-Prot

UniProt
the universal protein resource

The UniProt Knowledgebase consists of:

- **Swiss-Prot**; a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [More details / References / Linking to Swiss-Prot / User manual / Recent changes / Disclaimer].
- **TrEMBL**; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups at SIB and at EBI.

UniProt Release 4.4 consists of:
Swiss-Prot Release 46.4 of 29-Mar-2005: 178022 entries (More statistics)
TrEMBL Release 29.4 of 29-Mar-2005: 1647645 entries (More statistics)

> ***Swiss-Prot headlines***
Adding the keyword 'Complete proteome' to fungal entries (Read more...)

Access to Swiss-Prot and TrEMBL

- SRS - Access to Swiss-Prot, TrEMBL and other databases using the Sequence Retrieval System
- Full text search in Swiss-Prot and TrEMBL

(剩下1 项) 正在下载图片 http://us.expasy.org/images/expasy_logos/sprot1.gif... Internet

开始 蛋白质序列分析 ExPASy - Swiss-P... 22:55

Swiss-Prot: P12544 - 夏新笔记本电脑

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 搜索 收藏夹 媒体

地址(D) http://au.expasy.org/cgi-bin/get-sprot-entry?P12544 转到 链接

ExPASy Home page | Site Map | Search ExPASy | Contact us | Swiss-Prot

Hosted by APAF Australia | Mirror sites: Bolivia | Brazil new | Canada | China | Korea | Switzerland | Taiwan | USA

Search Swiss-Prot/TrEMBL for GRAA_HUMAN Go Clear

Swiss-Prot: P12544

NiceProt - a user-friendly view of this Swiss-Prot entry

```
ID   GRAA_HUMAN     STANDARD;      PRT;   262 AA.
AC   P12544;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-MAY-2005 (Rel. 47, Last annotation update)
DE   Granzyme A precursor (EC 3.4.21.78) (Cytotoxic T-lymphocyte proteinase
DE   1) (Hanukkah factor) (H factor) (HF) (Granzyme 1) (CTL tryptase)
DE   (Fragmentin 1).
GN   Name=GZMA; Synonyms=CTLA3, HFSP;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=T-cell;
RX   MEDLINE=88125000; PubMed=3257574 [NCBI, ExPASy, EBI, Israel, Japan];
RA   Gershenfeld H.K., Hershberger R.J., Shows T.B., Weissman I.L.;
RT   "Cloning and chromosomal assignment of a human cDNA encoding a T cell-
RT   and natural killer cell-specific trypsin-like serine protease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:1184-1188(1988).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
```

(剩下2 项) 正在下载图片 http://au.expasy.org/images/expasy_logos/psite50x50.gif... Internet

开始 Microsoft PowerP... Swiss-Prot: P125... 22:40

ID： 序列名称和氨基酸残基数目

AC：序列编号（收录号、登录号）

DT：提交到数据库的时间及最近修改时间

DE：描述行，对蛋白质的简单说明

GN：编码蛋白质的基因名称

OS：物种来源

OC：分类学中的位置

RN：基本注释信息

CC: 按主题进行区分

Function：描述功能

PTM： 说明修饰后的翻译

Tissue specificity: 说明组织专一性

Subcellular location: 说明亚细胞定位

Similarity: 说明与该蛋白质具有相似性
或相关的某个蛋白质家族

DR：提供与其它生物信息学数据库的链接

KW：关键词

FT：特征表。包括跨膜螺旋等超二级结构单元、配体结合位点、翻译后修饰位点等。每一行都有一关键词、特征序列氨基酸残基的位置及注释信息的性质。

SQ：蛋白质序列

2. Protein Motif and Domain Databases

2.1 PROSITE (www.expasy.org/prosite/)

❶ ScanProsite:

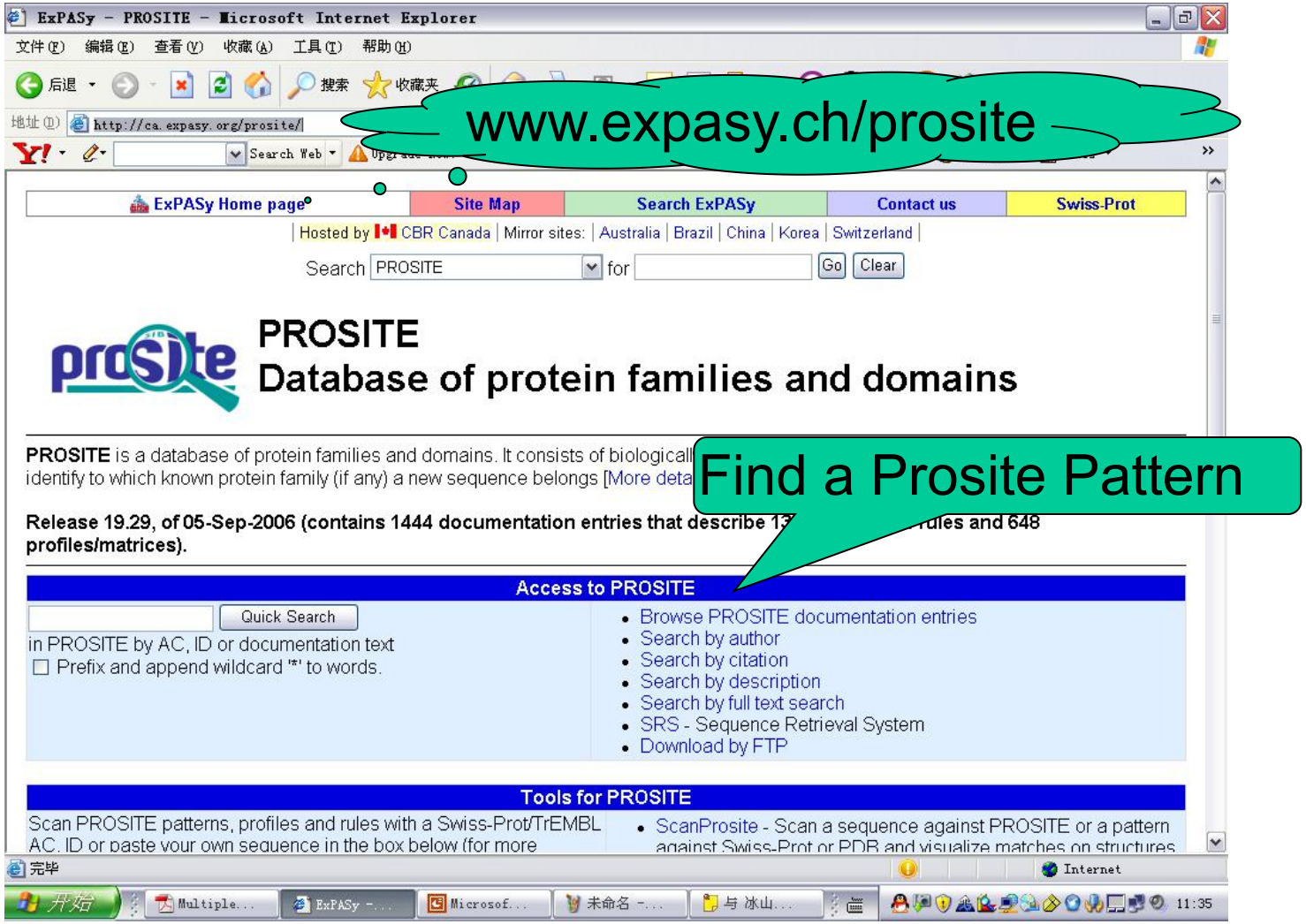
❷ MotifScan:

❸ Original curated protein family database introduced in 1989.

❹ Excellent documentation, search patterns, and position specific scoring matrices (PSSM).

❺ Hit or miss results with no statistics.

❻ Patterns are derived from consensus sequences.


ExPASy - PROSITE - Microsoft Internet Explorer
http://ca.expasy.org/prosite/
www.expasy.ch/prosite
ExPASy Home page
Site Map
Search ExPASy
Contact us
Swiss-Prot
Hosted by CBR Canada | Mirror sites: | Australia | Brazil | China | Korea | Switzerland |
Search PROSITE for Go Clear
PROSITE
Database of protein families and domains
Find a Prosite Pattern
Release 19.29, of 05-Sep-2006 (contains 1444 documentation entries that describe
profiles/matrices).
Access to PROSITE
Quick Search
in PROSITE by AC, ID or documentation text
Prefix and append wildcard '*' to words.
Browse PROSITE documentation entries
Search by author
Search by citation
Search by description
Search by full text search
SRS - Sequence Retrieval System
Download by FTP
Tools for PROSITE
Scan PROSITE patterns, profiles and rules with a Swiss-Prot/TrEMBL
ScanProsite - Scan a sequence against PROSITE or a pattern

Tools for PROSITE

Scan PROSITE patterns, profiles and rules with a Swiss-Prot/TrEMBL AC, ID or paste your own sequence in the box below (for more options, use the ScanProsite form

e.g. PTPB_HUMAN

Quick Scan | Clear

☐ Exclude patterns with a high probability of occurrence

- ScanProsite - Scan a sequence against PROSITE or a pa against Swiss-Prot or PDB and visualize matches on struc
- **raphical view and feature detection**
- Scan a sequence against the profile entries in PROSITE and Pfam
- InterProScan - Scan a sequence against all the motif datab in InterPro
- ps_scan - Perl program to scan PROSITE locally
- tools - Standalone programs to create and scan PROSIT
- Other pattern and profile search tools

Find Patterns in a Sequence

Documents

- PROSITE user manual
- PROSITE release notes
- Document describing the syntax of profiles in PROSITE
- List of abbreviations for journals cited
- List of on-line experts
- The optimal way to develop patterns

Document
n.公文，文件；证件

ScanProsite Results Viewer - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

ScanProsite Results Viewer

This view shows ScanProsite results together with ProRule-based predicted intra-domain

Hits for all PROSITE (release 19.29) motifs on sequence ENTK_ease 50.6)]:

found: 13 hits in 1 sequence

P98073 **ENTK_HUMAN** (1019 aa)
Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase) [Contains: Enteropeptidase non-catalytic heavy chain; Enteropeptidase catalytic light chain]. *Homo sapiens (Human)*

```
MGSKRGISSRHHSLSSYEIMFAALFAILVVLCAGLIAVSCLTIKESQRGAALGQSHEARATFKITS
GVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQFENGSIIVVFDLFFAQWVS
DQNVKEELIQGLEANKSSQLVTFHIDLNSVDILDKLTTTSHLATPGNVSIECLPGSSPCTDALTCI
KADLFCDGEVNCPDGSDEDNKMCATVCDGRFLLTGSSGSFQATHYPKPSETSVVCQWIIRVNQGLS
IKLSFDDFNTYYTDILDIYEGVGSSKILRASIWETNPGTIRIFSNQVTATFLIESDESDYVGFNAT
YTAFNSSELNNYEKINCNFEDGFCFWVQDLNDDNEWERIQGSTFSPFTGPNFDHTFGNASGFYIST
PTGPGGRQERVGLLSLPLDPTLEPACLSFWYHMYGENVHKLSINISNDQNMEKTVFQKEGNYGDNW
NYGGQVTLNETVKFKVAFNAFKNKILSDIALDDISLTYGICNGSLYPEPTLVPTPPPELPTDCGGPF
ELWEPNTTFSSTNFPNSYPNLAFCVWILNAQKGKNIQLHFQEFDLENINDVVEIRDGEEADSLLLA
VYTGPGPVKDVFSTTNRMTVLLITNDVLARGGFKANFTTGYHLGIPEPCKADHFQCKNGECVPLVN
LCDGHLHCEDGSDEADCVRFFNGTTNNNGLVRFRIQSIWHTACAENWTTQISNDVCQLLGLGSGNS
SKPIFSTDGGPFVKLNTAPDGHLILTPSQQCLQDSLIRLQCNHKSCGKKLAAQDITPKIVGGSNAK
EGAWPWVVGLYYGGRLLCGASLVSSDWLVSAAHCVYGRNLEPSKWTAILGLHMKSNLTSPQTVPRL
IDEIVINPHYNRRRKDNDIAMMHLEFKVNYTDYIQPICLPEENQVFPPGRNCSIAGWGTVVYQGTT
ANILQEADVPLLSNERCQQQMPEYNITENMICAGYEEGGIDSCQGDSGGPLMCQENNRWFLAGVTS
```

Prosite Scan Result Page

Scan Sequence

ScanProsite Results Viewer - Microsoft Internet Explorer

hits by profiles: [8 hits (by 6 distinct profiles) on 1 sequence]

P98073
(ENTK_HUMAN)

(1019 aa)

Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase) [Contains: Enteropeptidase non-catalytic heavy chain; En tidase catalytic light chain]. *Homo sapiens (Human)*

PS50024 **SEA** *SEA domain profile :*

52 - 169: **score** = 24.822

```
LGQSHEARATFKIT--SGVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNS
RVLQFENGSII-VVFDLFFAQWV-SDQNVKEELIQGLEANKSSQLVTFHIDLNSVDILDK
LT
```

PS50068 **LDLRA_2** *LDL-receptor class A (LDLRA) domain profile :*

183 - 222: **score** = 10.75

```
ECLPGSSPCTDaLTCIKADLFCDGEVNCPDGSDEdnKMCA
```

Predicted features:

DISULFID	184	197	By similarity	[condition: C-x*-C]
DISULFID	191	210	By similarity	[condition: C-x*-C]
DISULFID	204	221	By similarity	[condition: C-x*-C]

642 - 678: **score** = 13.3

```
PCKADHFQCKNGECVPLVNLCDGHLHCEDGSDEADCV
```

Predicted features:

DISULFID	643	655	By similarity	[condition: C-x*-C]

Graphical Overview (Applet)

ScanProsite Results Viewer - Microsoft Internet Explorer

hits by patterns: [5 hits (by 4 distinct patterns) on 1 sequence]

P98073 (ENTK_HUMAN) (1019 aa)

Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase) [Contains: Enteropeptidase non-catalytic chain; Enteropeptidase catalytic light chain]. *Homo sapiens (Human)*

PS01209 **LDLRA_1** *LDL-receptor class A (LDLRA) domain signature :*

197 - 221: CIkadlf.CDgevNCpdgsDEDnkm..C

655 - 677: CVplvnl.CDghlHCedg.SDEad...C

PS00740 **MAM_1** *MAM domain signature :*

391 - 431: GfYIstpTgpggrqervg.LlslpLdptlepaCLsFwYhmyG

PS00134 **TRYPSIN_HIS** *Serine proteases, trypsin family, histidine active site :*

821 - 826: VSAAHC

PS00135 **TRYPSIN_SER** *Serine proteases, trypsin family, serine active site :*

965 - 976: DScqGDSGGPLM

Legend:

disulfide bridge | active site | other 'ranges' | other sites

Link to PROSITE Pattern

PROSITE documentation PDOC00929 [for PROSITE entry PS01209] - Microsoft Internet Explorer

ExPASy Home page | Site Map | Search ExPASy | Contact us | PROSITE | Proteomics tools | Swiss-Prot

Hosted by CBR Canada | Mirror sites: Australia | Brazil | China | Korea | Switzerland

Search Swiss-Prot/TrEMBL for Go Clear

prosite Documentation PDOC00929

PROSITE Pattern Documentation

LDL-receptor class A (LDLRA) domain sig

Description:

Low-density lipoprotein (LDL) receptors are the major cholesterol-carrying lipoproteins of plasma. Seven successive cysteine-rich repeats of about 40 amino acids are present in the N-terminal of this multidomain membrane protein [1]. Similar domains have been found (see references in [2]) in other extracellular and membrane proteins which are listed below:

- Vertebrate very low density lipoprotein (VLDL) receptor, which binds and transports VLDL. Its extracellular domain is composed of 8 LDLRA domains, 3 EGF-like domains and 6 LDL-receptor class B domains (LDLRB).
- Vertebrate low-density lipoprotein receptor-related protein 1 (LRP1) (reviewed in [3]), which may act as a receptor for the endocytosis of extracellular ligands. LRP1 contains 31 LDLRA domains and 22 EGF-like domains.
- Vertebrate low-density lipoprotein receptor-related protein 2 (LRP2) (also known as gp330 or megalin). LRP2 contains 36 LDLRA domains and 17 EGF-like domains.
- A LRP-homolog from Caenorhabditis elegans, which contains 35 LDLRA domains and 17 EGF-like domains.
- Drosophila putative vitellogenin receptor, with 13 copies of LDLRA domains and 17 EGF-like repeats.
- Complement factor I, which is responsible for cleaving the α-chains of C4b and C3b. It consists of a FIMAC domain (Factor I/MAC proteins C6/C7), a scavenger receptor-like domain, 2 copies of LDLRA and a C-terminal serine protease domain.
- Complement components C6, C7, C8 and C9. They contain each one LDLRA domain.

3. **Protein structural database** --PDB (www.rcsb.org/pdb/)

❶ The primary public repository for macro-molecular structures

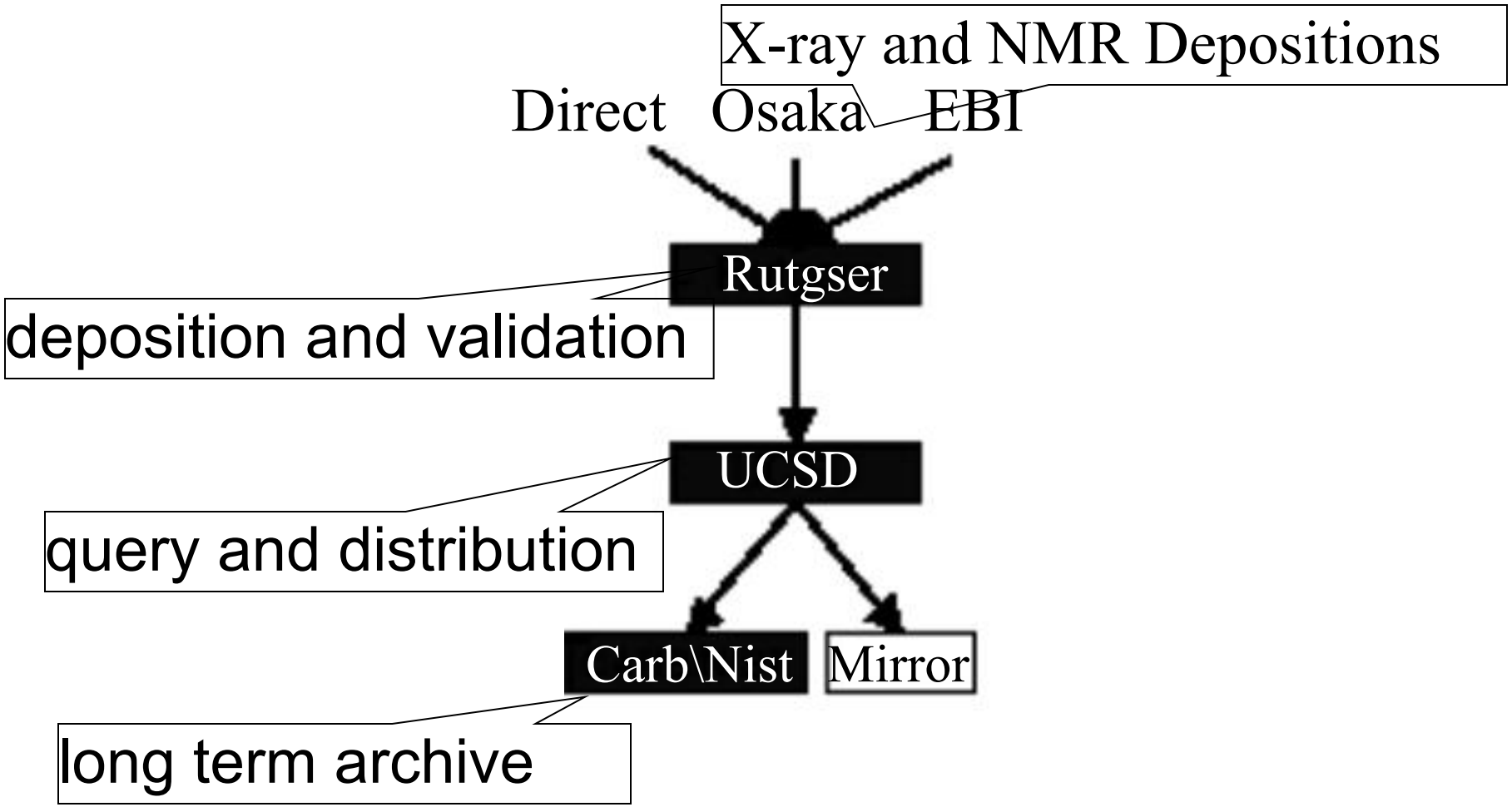
❷ Founded in 1972 (2 structures)

❸ Initially maintained at the Brookhaven National Laboratory

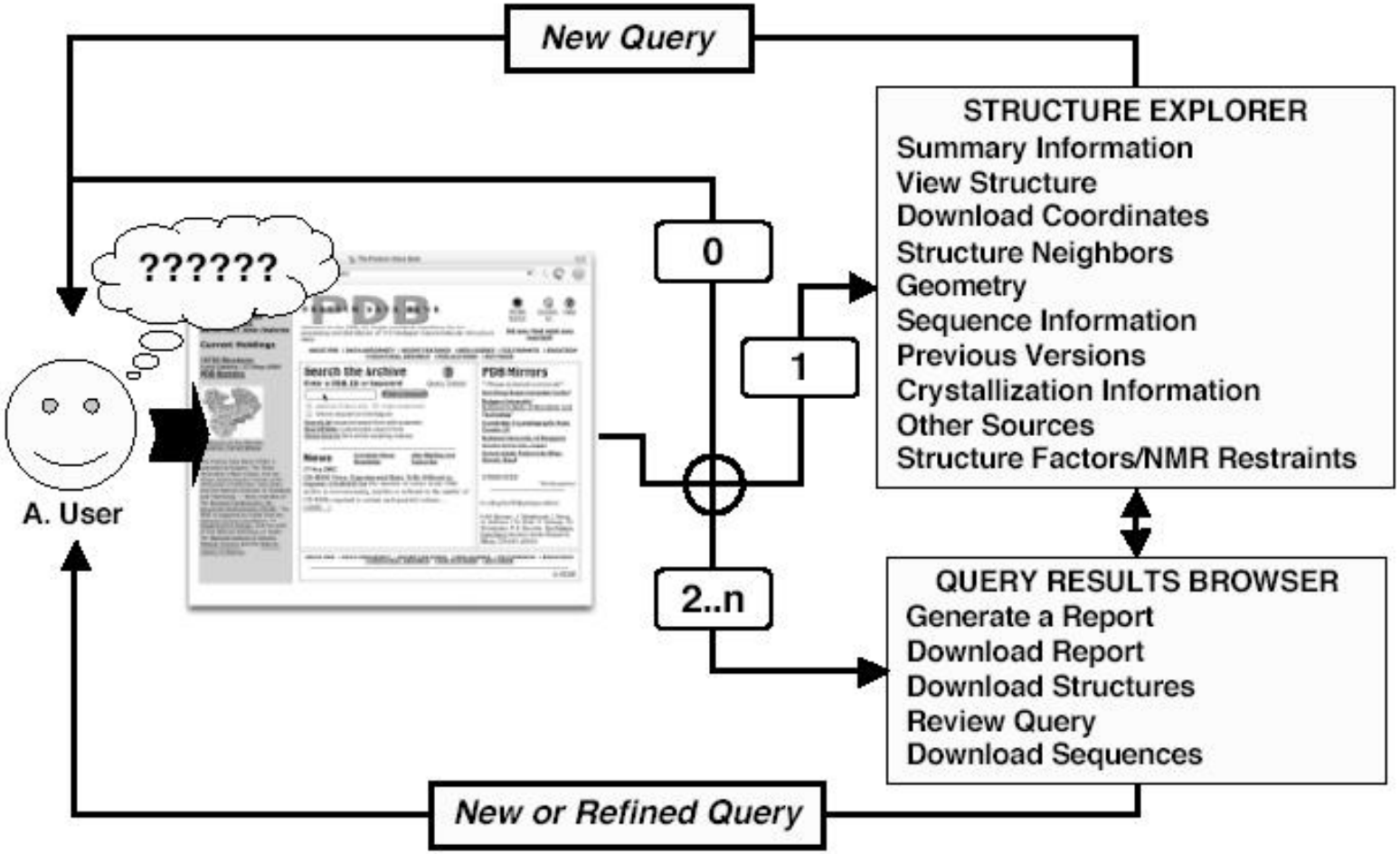
❹ Since 1999 maintained by the Research Collaboratory for Structural Bioinformatics

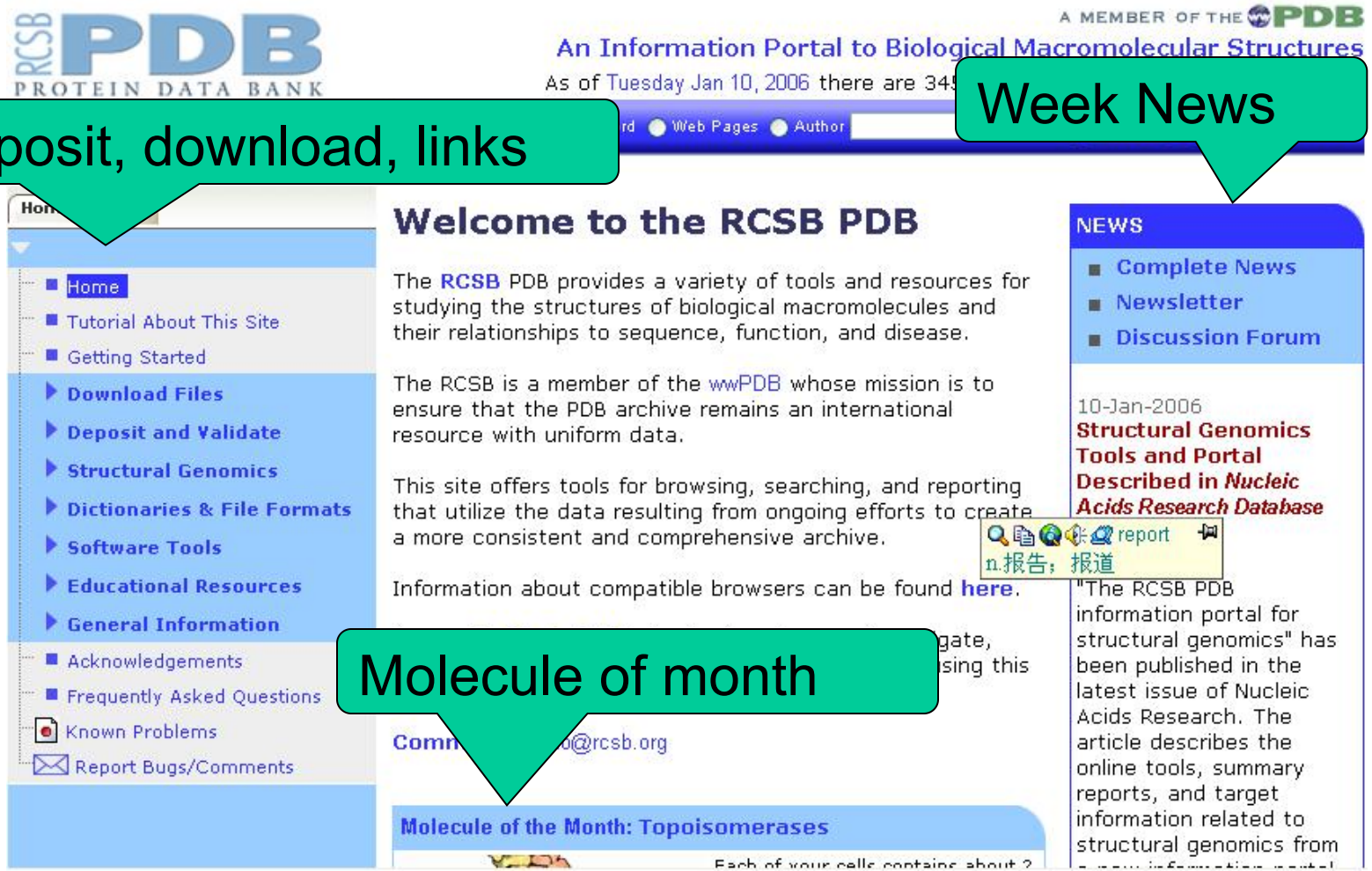
RCSB/PDB Responsibilities

- Berman group @ Rutgers U.
- Bourne group @ UCSD
- Gilliland group @ CARB/NIST
- Operates under a cooperative
  - -- agreement with NSF, NIH & DOE
  - -- Helen Berman PI


X-ray and NMR Depositions
Direct
Osaka
EBI
Rutgser
deposition and validation
UCSD
query and distribution
Carb\Nist
Mirror
long term archive

PDB Functionality Overview




RCSB PDB
PROTEIN DATA BANK
A MEMBER OF THE PDB
An Information Portal to Biological Macromolecular Structures
As of Tuesday Jan 10, 2006 there are 34
Web Pages
Author
Home
Tutorial About This Site
Getting Started
Download Files
Deposit and Validate
Structural Genomics
Dictionaries & File Formats
Software Tools
Educational Resources
General Information
Acknowledgements
Frequently Asked Questions
Known Problems
Report Bugs/Comments
Welcome to the RCSB PDB
The RCSB PDB provides a variety of tools and resources for studying the structures of biological macromolecules and their relationships to sequence, function, and disease.
The RCSB is a member of the wwPDB whose mission is to ensure that the PDB archive remains an international resource with uniform data.
This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.
Information about compatible browsers can be found here.
Molecule of the Month: Topoisomerases
NEWS
Complete News
Newsletter
Discussion Forum
10-Jan-2006
Structural Genomics Tools and Portal Described in Nucleic Acids Research Database
report
n.报告；报道
"The RCSB PDB information portal for structural genomics" has been published in the latest issue of Nucleic Acids Research. The article describes the online tools, summary reports, and target information related to structural genomics from
Deposit, download, links
Week News
Molecule of month

PDB file format

- historic format
- 5 different versions
- "header" and coordinates sections
- See

http://www.rcsb.org/pdb/cgi/explore.cgi?job=download&pdbId=1C2W&page=&pid=&opt=show&format=PDB&header=1

- Why hang on to it?
  - -- Widely used in virtual every piece of SB
  - -- software
  - -- Human readable
  - -- Moving of data to mmCIF not trivial
  - -- Moving of users to mmCIF not trivial

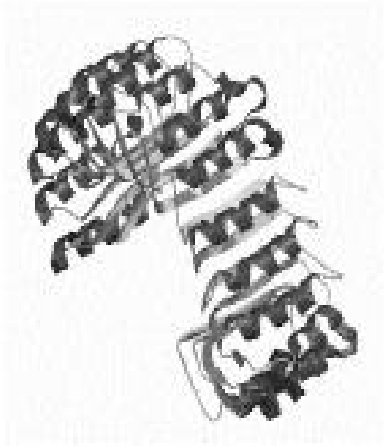
mmCIF file format (1)

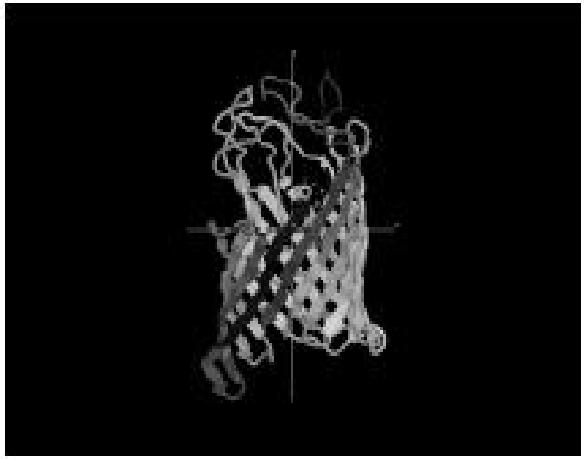
- mmCIF := macromolecular crystallographic information file
- Subset of STAR (self-defining Text Archive and Retrieval Format)
- IUCr approved standard
- WWW: http://ndbserver.rutgers.edu/NDB/mmcif

4. Protein Structural Classification Databases

- Similarities in secondary structure element assembly
- Topological units of polypeptide chains
- Regularities arise from intrinsic physical and chemical properties
- Folds are the units of protein function, structure and evolution

- Examples: propeller, horseshoe, TIM barrel…





4.1 SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/)

❶ Structure Classification of Proteins

❷ Based on evolutionary relationships

❸ Generated through visual comparison and inspection of automated structure alignments

❹ Provides links to coordinates of domains, images and sequence data

4.1.1 SCOP Hierarchy

❶ Class

-- secondary element composition

-- All α, all β, α/ β, α+β, some others

❷ Folds

-- Common core structures

-- 138, 93, 97 & 184 respectively for each class

❸ Superfamily

-- Share common structure and function

❹ Family

-- Share clear common evolutionary origin

- Fold classification most difficult step
- Differences between mixed classes:
  - α/ β
    - Principally single b-sheet with a-helices joining the individual strands
    - Two subclasses: β -sheet barrel surrounded by α-helices and planar β-sheet flanked on either side by α-helices

- α+β
  - α and β units largely separated;
  - Antiparalell strands usually joined by hairpins;
  - Small clusters of helices tightly packed against sheet.

4.2 CATH (www.biochem.ucl.ac.uk/bsm/cath_new/index.html)

❶ Class, Architecture, Topology and Homologous Superfamily

❷ Also based on evolutionary relationships

❸ Automated generation with validation of ambiguities in assignments

4.2.1 CATH Hierarchy

❶ Class

-- Determined by secondary structure composition and packing

-- mainly-α, mainly- β and α-β.

❷ Architecture

-- Description of orientation of secondary structures regardless of connectivity

-- Assigned manually

❸ Topology

-- Regards both secondary structure orientation and connectivity

❹ Homologous Superfamily

-- Evolutionary grouping based on structure, sequence and/or functional similarity

-- Proteins are clustered into sequence families at different levels of sequence identity (35%, 60%, 95%, 100%)

4.2.2 CATH Update Strategy

❶ Identify close relatives using pairwise sequence alignments;

❷ Detect distant relatives using sequence profiles and structure comparisons;

❸ Examine unclassified structures using both automatic and manual procedures to determine domain boundaries

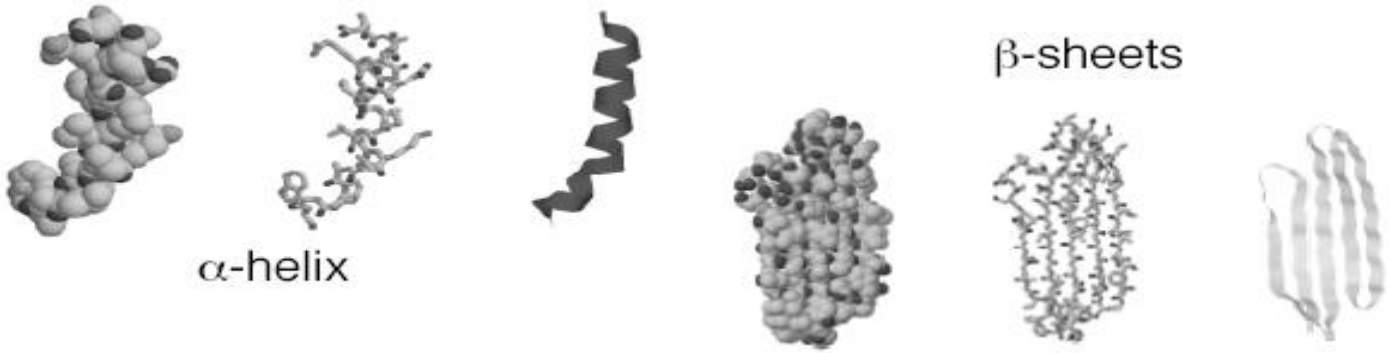
❹ Reiterate over steps 2 and 3

❺ Manual assignment to existing or new architectures

Thanks for your attentions !

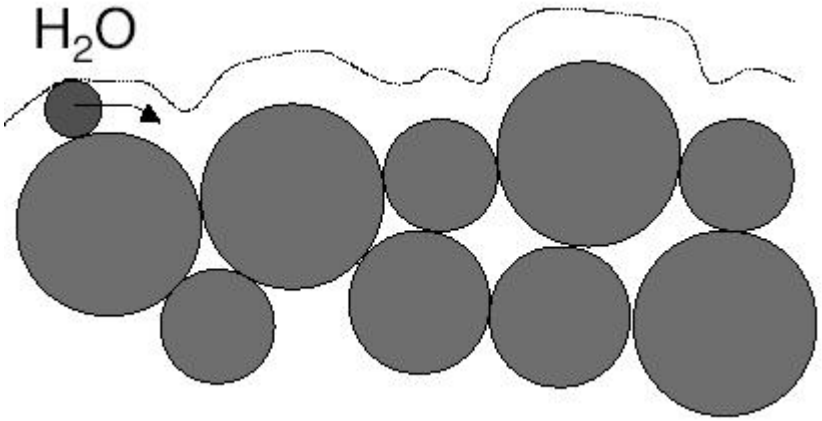
2. Basic Structure Analysis

- Secondary structure elements
- Tertiary structure motifs
- Structure domains
- B-factor, occupancy and heterogeneity
- Visualization tools



2.1 Structure Based Calculation

- Secondary structure calculation approaches
- Kabsch/Sanders
- Solvent exposure & curvature



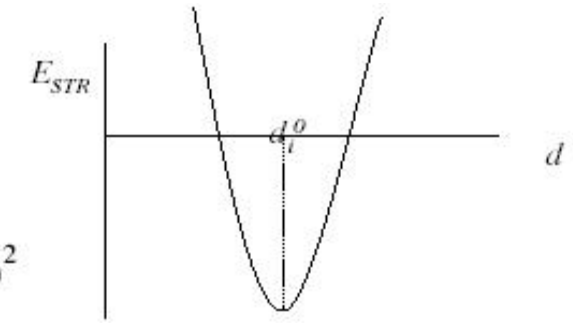
Structure Determination

- X-ray crystallography
- NMR
- Electron microscopy
- Structural genomics

Molecular Dynamics

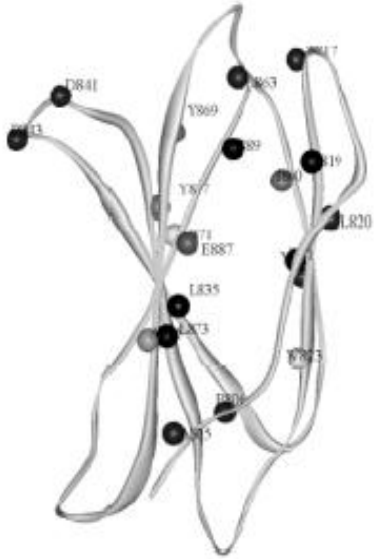
- B-factor revisited
- Large motions vs. jitterbugs
- Thermodynamic equations and energy minima
- Diffusion

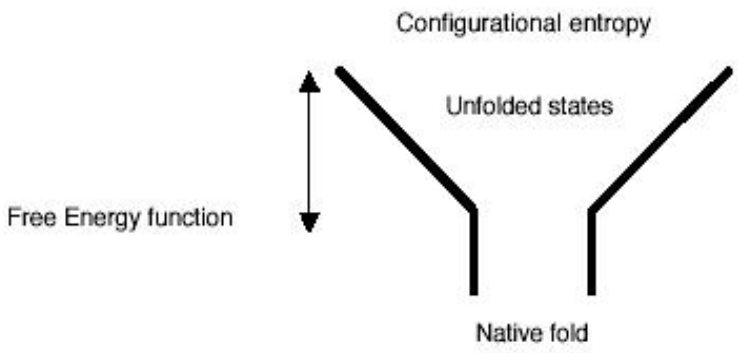
$$E_{STR} = \sum_{i=1}^{n} \frac{1}{2} k^{d_i} (d_i - d_i^0)^2$$



Protein Folding

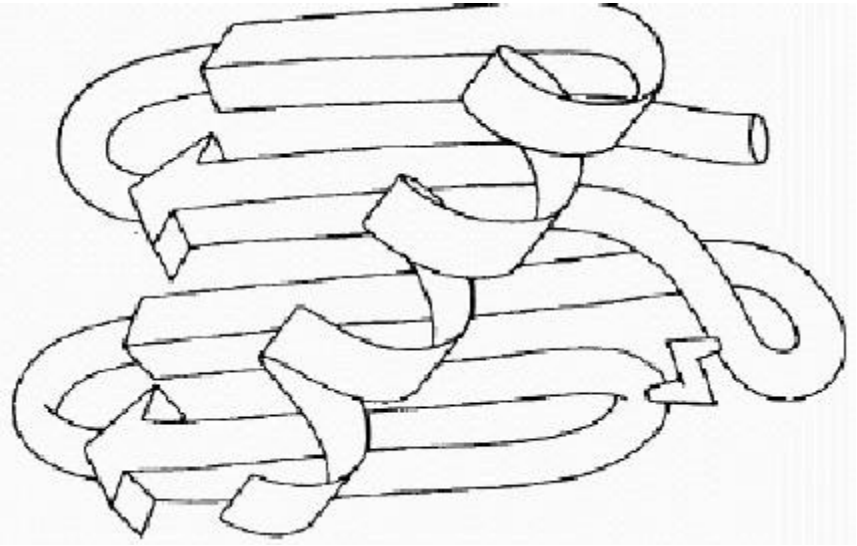
- Levinthal paradox: “proteins simply can not fold on a reasonable time scale” .
- Folding units and CKAAPs





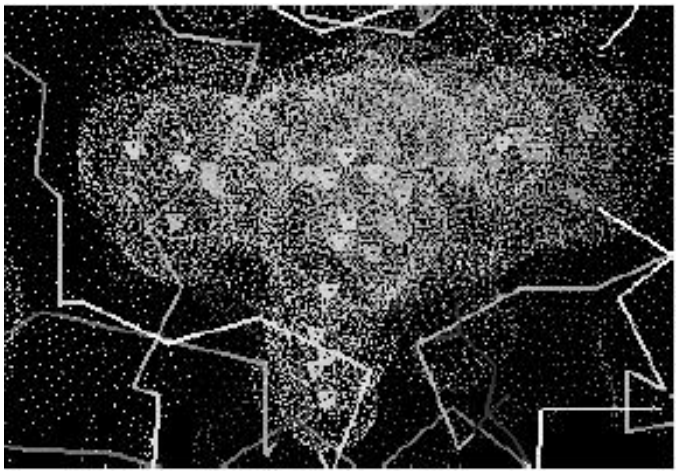
Structure Prediction

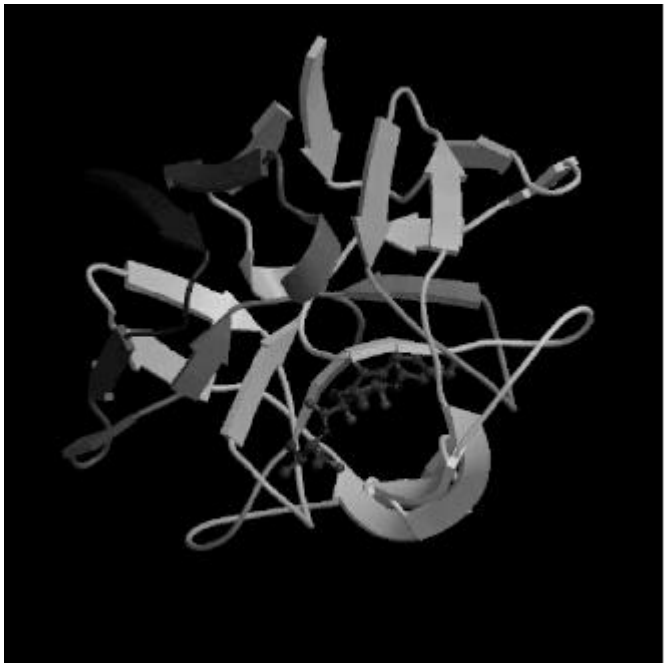
- Homology modeling
- Fold recognition
- *ab initio* prediction
- CASP



Docking

- Force fields
- Docking s/w
- Virtual screening





3. Structural Alignments

- Provide an understanding of sequence/structure and structure/function relationships
- Can help to find active sites or binding regions
- Highlight the targets of evolutionary pressure

Structural Alignment Mathematics

- Find rotation matrix *R* and translation
- vector *T* for which:
- No known deterministic algorithm
- NP hard!

$$\boldsymbol{B} = \boldsymbol{R} \times \boldsymbol{A} + \boldsymbol{T}$$

Algorithm Terminology

- NP := non-deterministic polynomial time
- "guesses" can be checked in polynomial time
- NP-hard := NP problem at least as hard or harder as all other NP problems
- "order" of algorithms = max. time needed:
  - -- e.g. O(N), O(N2), O(log(N)), O(eN) …
  - -- Want O(N) not O(Nx) or even O(xN)!!!
  - -- Polinomial time (P): aN + bN2 + cN3 + ….

Problematic Issues

- Measure used to quantify difference, i.e. a similarity score
- Non-locality of scoring function
  - -- Any three atoms can be perfectly aligned
  - -- Aligning the fourth atom requires a change to the previous alignment
  - -- Dynamic programming not applicable!
- Existence of gaps and insertions

Similarity Measure

Root Mean Square Deviation (RMSD)

$$R_{ms} = \sqrt{\sum_{i=1}^{n} \frac{(C_{Ai} - C_{Bi})^2}{n}}$$

Distance between aligned atoms

The number of aligned atoms

- Penalizes worst fitting atoms
- Contributions of individual atoms not discernable

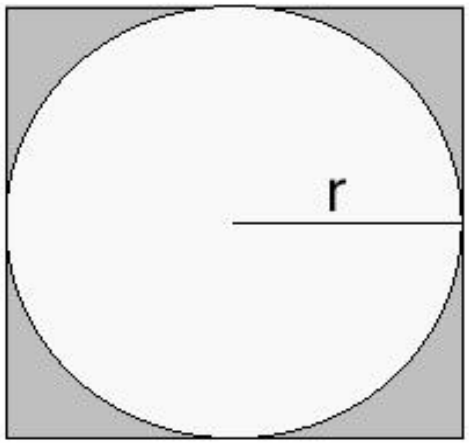
Alignment Approaches

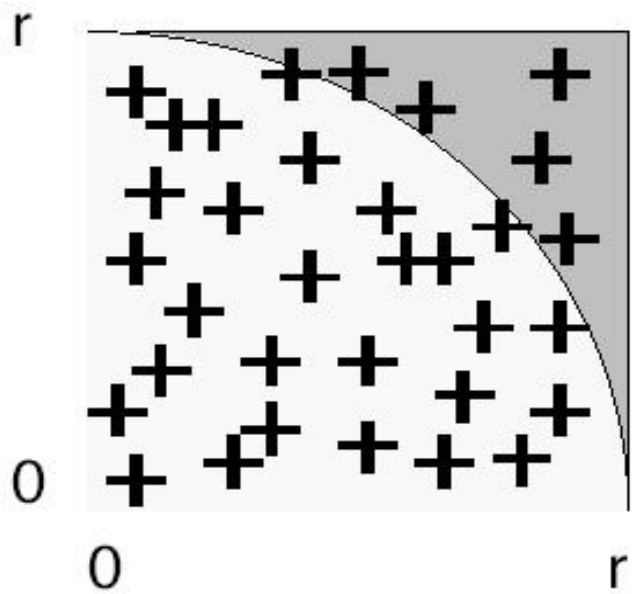
- Differences of Distance maps
  -- DALI (distance matrix alignment program)
- Contact Map overlay
- Secondary structure element (SSE) representations
  -- VAST
  -- CATH

Optimization Algorithms

- Dynamic programming (as in Smith-Waterman) [CATH]
- Monte Carlo [DALI]
- 3D clustering
- Graph theory [VAST]
- Combinatorial Extension [CE]
- Combinations

Monte Carlo





$A_S = 4r^2$

$A_C = \pi r^2$

$\pi = 4A_C/A_S$

$A_S = A_S+1;$
if (hit in circle) { $A_C = A_C + 1;$ }
$\pi = 4A_C/A_S$

Monte Carlo (2)

π = 3.141592654

1000	3.26	3.204	3.176	3.144
10000	3.1672	3.1428	3.1488	3.1448
100000	3.15764	3.14788	3.14732	3.14028
1000000	3.142848	3.142864	3.141488	3.141152
10000000	3.141352	3.142390	3.141398	3.141341
20000000	3.141586	3.142083	3.141652	3.141601
30000000	3.141718	3.141860	3.141456	3.141669
40000000	3.141715	3.141879	3.141336	3.141444
50000000	3.141698	3.141900	3.141479	3.141451
60000000	3.141806	3.141698	3.141597	3.141458
70000000	3.141974	3.141647	3.141531	3.141373
80000000	3.141938	3.141636	3.141504	3.141393
90000000	3.141868	3.141696	3.141478	3.141412
100000000	3.141822	3.141734	3.141466	3.141453

Exercise