Digitally Assessing Protein Properties

Biochemistry Boot Camp 2022
Session #2

Nick Fitzkee
nfitzkee@chemistry.msstate.edu

Protein as Chemicals

- Molecular weight
- Chemical formula (e.g. $C_{274}H_{427}N_{69}O_{93}S_1$)
- Isoelectric point
- Sequence & Residue composition
- Solubility
- Structure
- Concentration/extinction coefficient

→ How do we access this information?

Sequence of GB3

- Primary Structure:

  NT-Met-Gln-Tyr-Lys-…-Thr-Glu-**CT**

- More convenient:

```
MQYKLVINGK TLKGETTTKA VDAETAEKAF
KQYANDNGVD GVWTYDDATK TFTVTE
```

- Can we search this (think Google)?

Website #1: Protparam

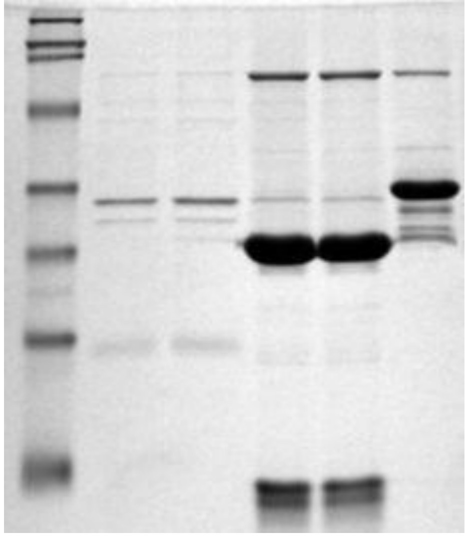
- http://web.expasy.org/protparam/

- **Input:** Protein sequence (one-letter codes)

- **Output:** Basic chemical properties
  - Molecular weight
  - Isoelectric point (pI)
  - Extinction coefficient

Molecular Weight

Polyacrylamide Gel Electrophoresis (SDS-PAGE)



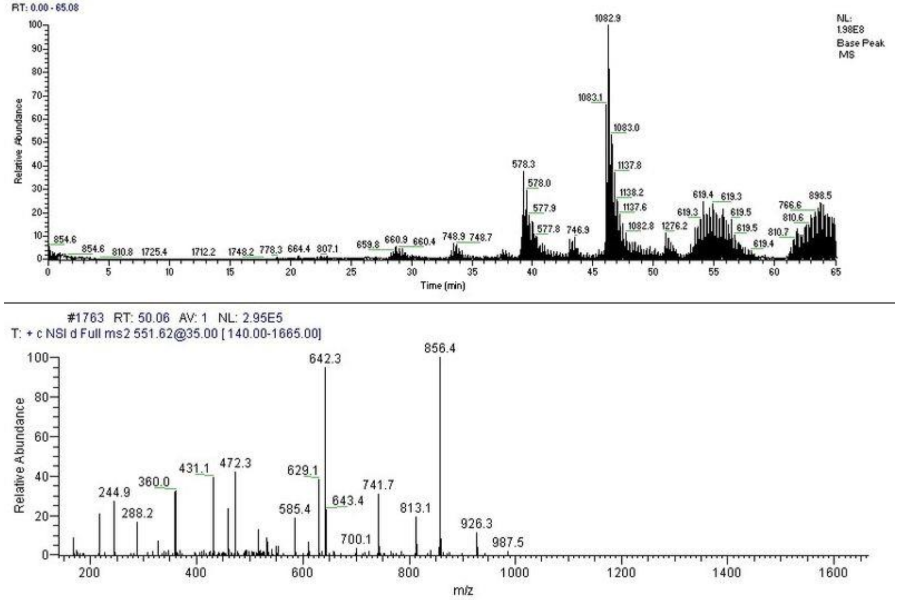
Standards

Samples
(one per lane)

High
Molecular
Weight

Low
Molecular
Weight

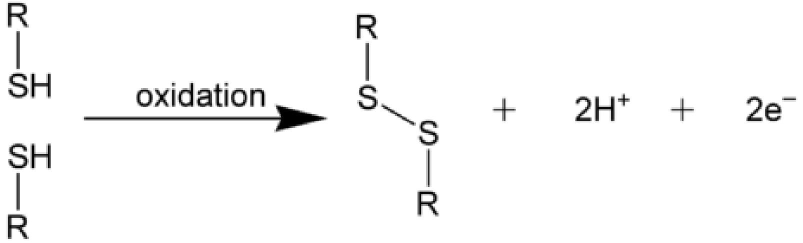
Mass Spectrometry (ESI-MS, LC-MS)



Sources: en.wikipedia.org/wiki/SDS-PAGE, en.wikipedia.org/wiki/Protein_mass_spectrometry

Residue Composition

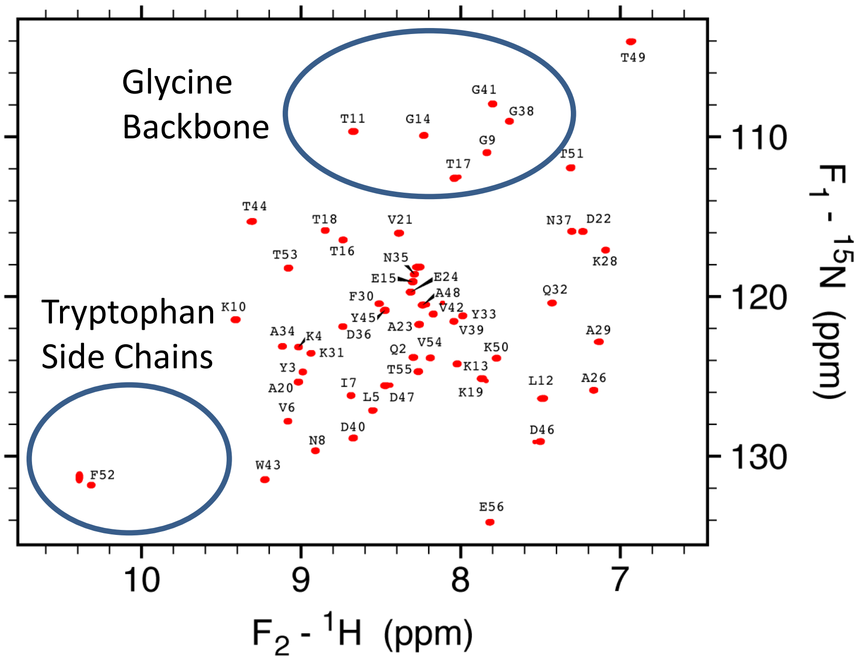
Disulfide Formation (Cysteine Content)



Reducing Agents:

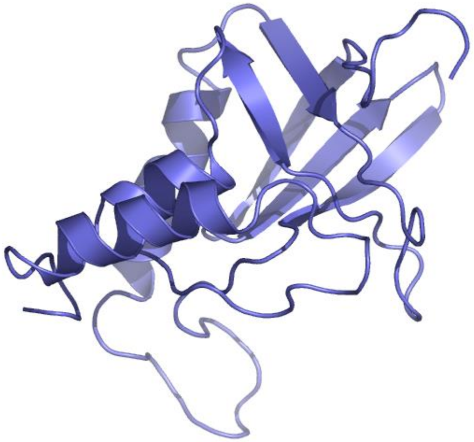
- 2-Mercaptoethanol (BME, 5-10 mM)
- Dithiothreitol (DTT, 1-5 mM)
- Tris-(2 carboxyethyl) phosphine (TCEP, < 1 mM)

Protein ^{15}N HSQC (NMR)


Glycine Backbone

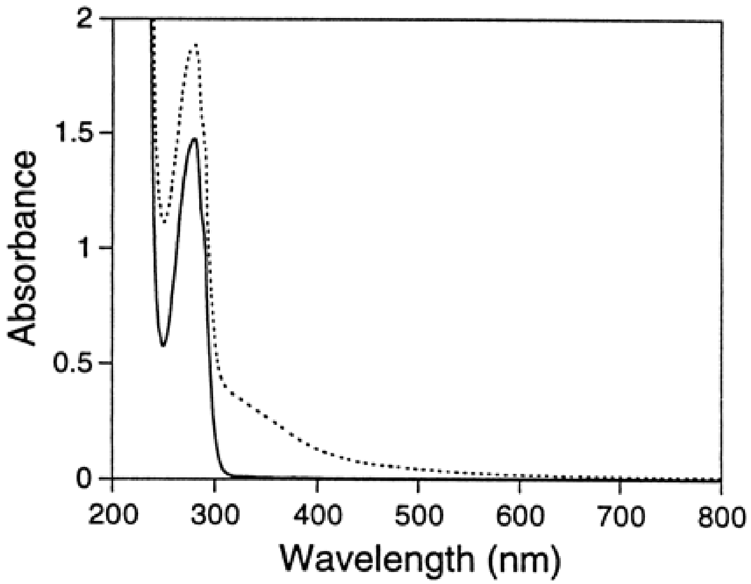
Tryptophan Side Chains

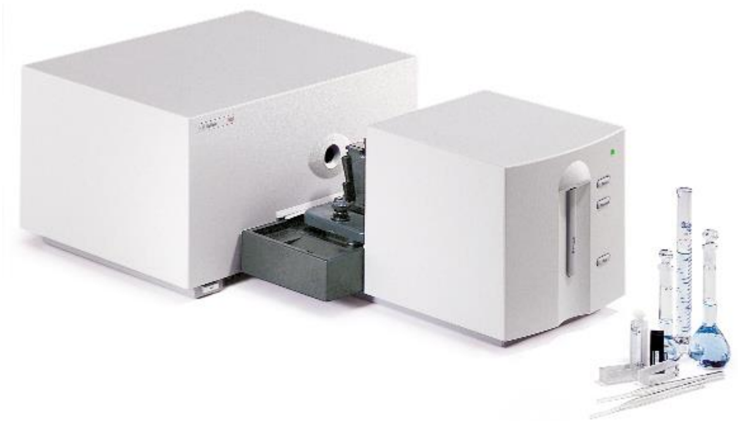
Extinction Coefficient



Tryptophan side chain absorbs light at 280 nm

More absorbance → More protein





<span></span>If we know the extinction coefficient, we can *estimate* the concentration.

Calculating Protein Concentration

(Beer's Law)

- **UV-Vis:** Absorbance at 280 nm is 0.348 in a 0.3 cm quartz cuvette
  - Most cuvettes are 1 cm

- **Protparam:** Extinction coefficient at 280 nm is 9970 M^{-1} cm^{-1}

- **Beer's Law:** $A = \epsilon C l$

Source: www.malvernstore.com

What If My Protein Doesn’t Have Trp?

- No Trp means low (no) absorbance at 280 nm

- Protein backbone has intrinsic absorbance at 205 nm
  - See Anthis, N.J. and Clore, G.M. (2013) *Protein Science.* http://www.ncbi.nlm.nih.gov/pubmed/?term=23526461
  - Website: http://nickanthis.com/tools/a205.html

- Complications:
  - Protein concentration will need to be quite low, which may introduce dilution errors
  - Many buffers absorb at 205 nm, these can overwhelm the protein signal (even when using a blank)
  - **Solution:** Careful dilution, use water as a blank if possible

Caveats: Extinction Coefficient

- Uncertainty can be as much as 10%
    - Can be worse if your technique is poor!
- Absorbance values need to be between 0.1-1.0 for highest accuracy
    - Estimate your expected A_{280} and dilute if necessary
- **Scattering of aggregates:** If the baseline is not zero at 600 nm, you are probably not getting an accurate value!
- DNA, other impurities or other compounds may artificially increase absorbance at 280 nm

Think and Discuss

The extinction coefficient can be calculated from primary structure alone. Why is this important?

Website #2: NCBI Databases

- https://www.ncbi.nlm.nih.gov/

- **Input:** Gene names, organisms, authors, etc.

- **Output:** Curated summary of research
  - Accepted DNA and protein sequences
  - Summaries of associated diseases
  - Recent research papers

NCBI Tricks #1

- Database restriction

srcdb refseq [prop]	Only search reference sequences
srcdb pdb [prop]	Only search the PDB

- Journal restriction

1998:2003 [dp]	Dates from 1998-2003
fitzkee_nc [auth]	Author name is Fitzkee, N. C.
j am chem soc [jour]	Journal name is JACS (need to know abbreviation)

NCBI Tricks #2

- Combining Terms

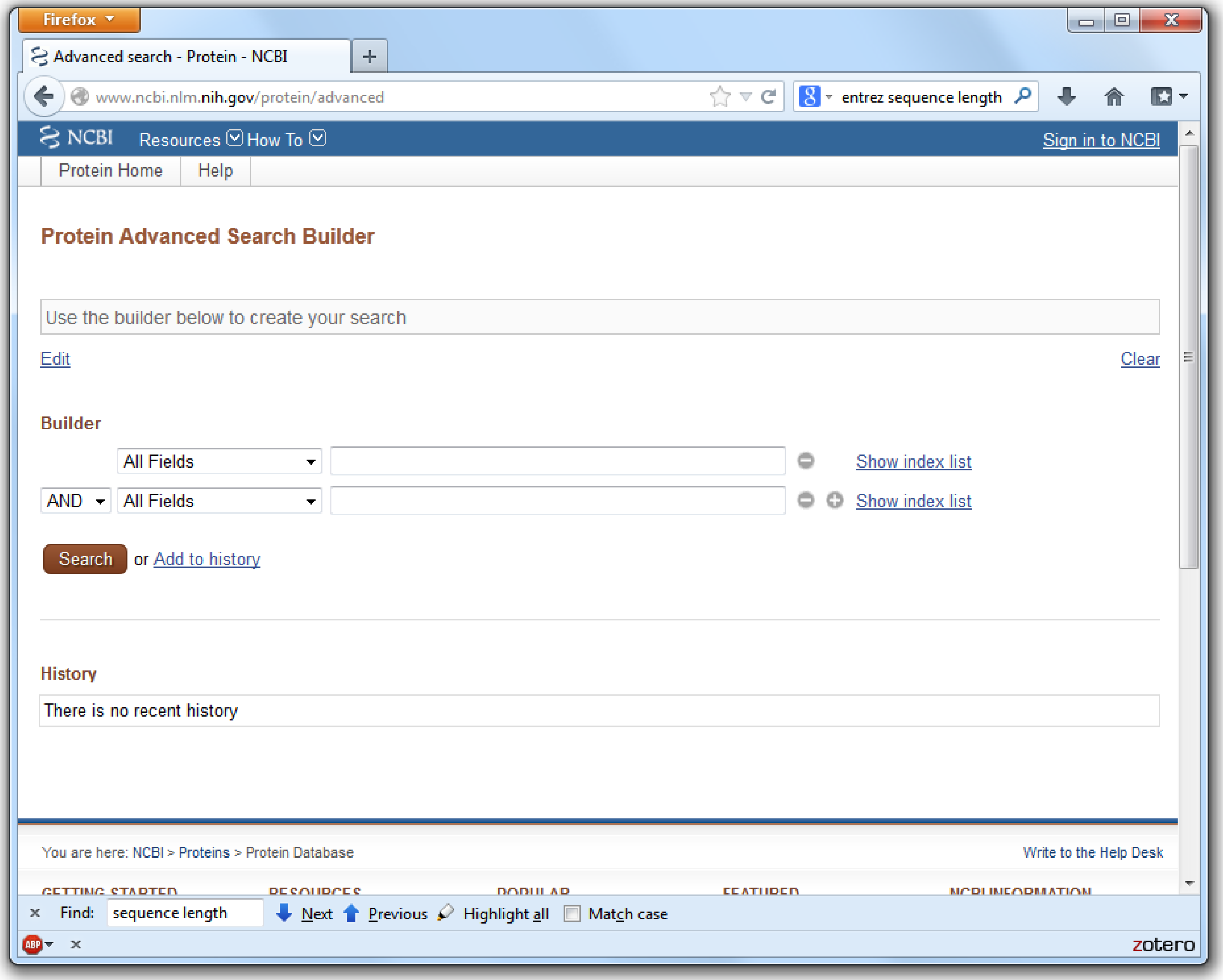
xx AND yy	Must have xx and yy
xx OR yy	Must have either xx or yy
NOT zz	Without term zz
xx AND (yy OR zz)	Complex example

- Chemical Properties

75:100 [sequence length]

3500:6000 [molecular weight]

Advanced Searches



Protein Advanced Search Builder

Use the builder below to create your search

Edit

Clear

Builder

All Fields

AND All Fields

Search or Add to history

History

There is no recent history

You are here: NCBI > Proteins > Protein Database

Practice

- What’s the sequence of your favorite protein?

- What’s the extinction coefficient of human heart fatty acid binding protein?

- What human disease is associated with phenylalanine hydroxylase?

Website #3: Protein Data Bank

- http://rcsb.org/

- **Input:** Protein name, PDB ID, authors, etc.

- **Output:** 3D coordinates of protein structures
  - Author information on methods
  - Cofactors and other information

What is a PDB file?

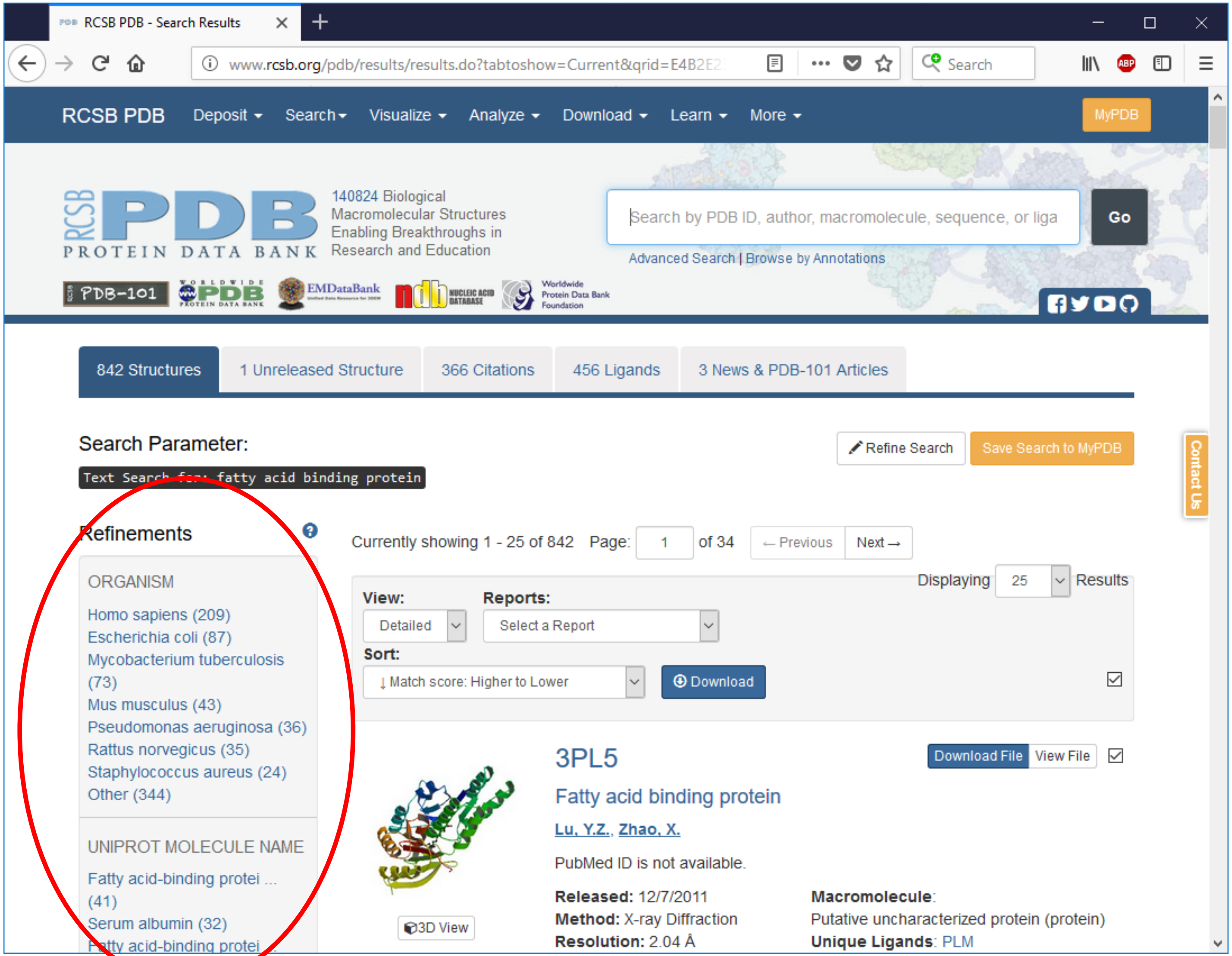
- Example: Ricin (2AAI)
- Text file contains a summary of information used in structure determination
- Most important: ATOM records contain X, Y, Z in *Ångströms* $(1 \times 10^{-10}\ \mathrm{m})$
  - Most atoms have a radius of 0.5-2 Å

Properties of PDB Files

- Experimental methodology:
  - X-Ray: Typically more precise
  - NMR: Need lots of “restraints;” sometimes hard to assess quality

- “Good” Structures (for X-Ray)
  - Low resolution (< 2Å)
  - Low R-value (< 20%)
  - Low R_{free}-value (< 25%)

Searching the PDB


RCSB PDB - Search Results
www.rcsb.org/pdb/results/results.do?tabtoshow=Current&qrid=E4B2E2
Search
RCSB PDB
Deposit
Search
Visualize
Analyze
Download
Learn
More
MyPDB
RCSB PDB PROTEIN DATA BANK
Search by PDB ID, author, macromolecule, sequence, or liga
Go
Advanced Search | Browse by Annotations
PDB-101
WORLDWIDE PDB PROTEIN DATA BANK
EMDataBank
NDB NUCLEIC ACID DATABASE
Worldwide Protein Data Bank Foundation
Refine Search
Save Search to MyPDB
Contact Us
Previous
Next
Download
Download File
View File
3D View

140824 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

842 Structures | 1 Unreleased Structure | 366 Citations | 456 Ligands | 3 News & PDB-101 Articles

Search Parameter:

`Text Search for: fatty acid binding protein`

Refinements

ORGANISM

- Homo sapiens (209)
- Escherichia coli (87)
- Mycobacterium tuberculosis (73)
- Mus musculus (43)
- Pseudomonas aeruginosa (36)
- Rattus norvegicus (35)
- Staphylococcus aureus (24)
- Other (344)

UNIPROT MOLECULE NAME

- Fatty acid-binding protei ... (41)
- Serum albumin (32)
- Fatty acid-binding protei ...

Currently showing 1 - 25 of 842 Page: 1 of 34

Displaying 25 Results

View: Detailed

Reports: Select a Report

Sort: ↓ Match score: Higher to Lower

☑

3PL5 ☑

Fatty acid binding protein

Lu, Y.Z., Zhao, X.

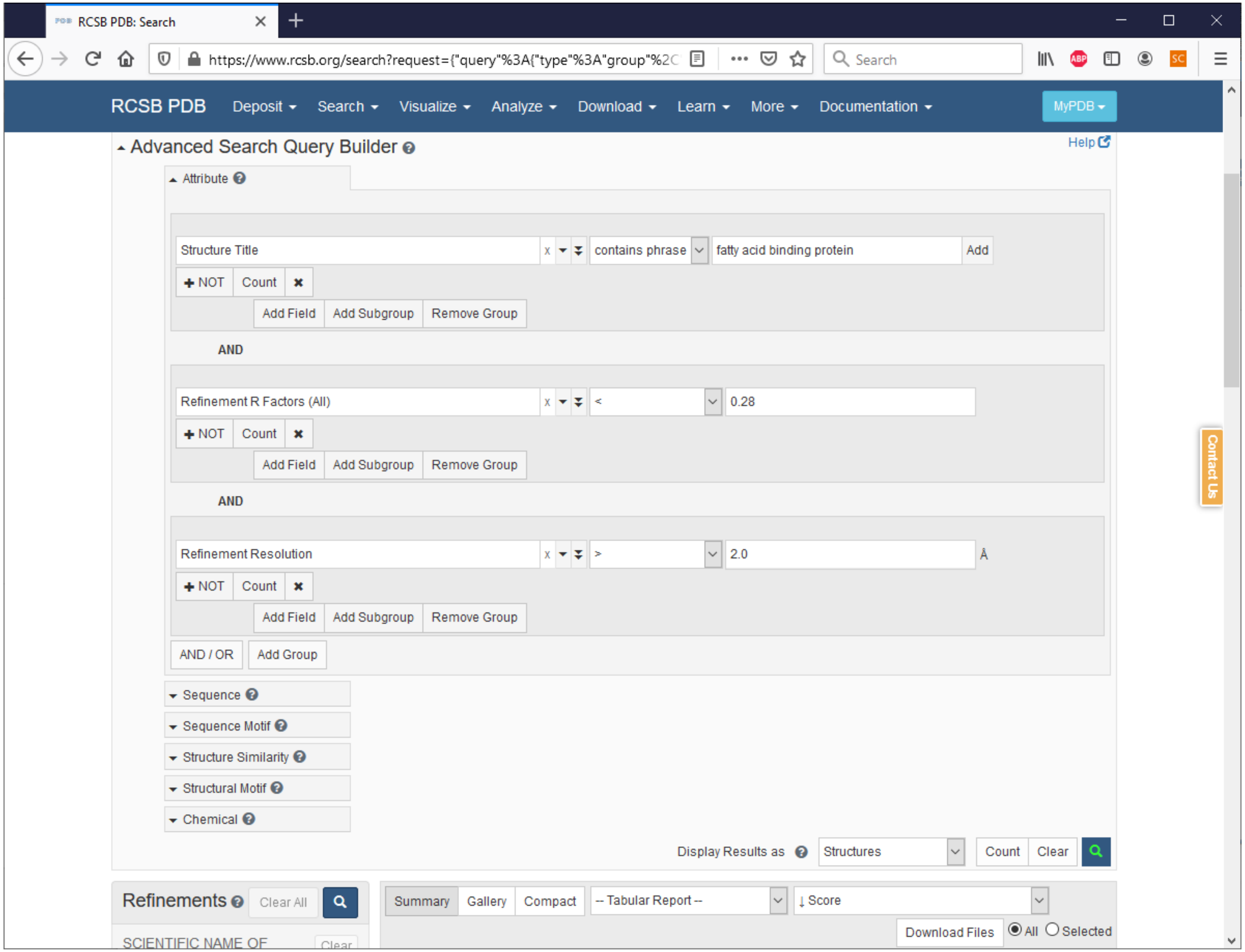
PubMed ID is not available.

Released: 12/7/2011
Method: X-ray Diffraction
Resolution: 2.04 Å

Macromolecule: Putative uncharacterized protein (protein)
Unique Ligands: PLM

Note refinements!

Advanced Searching



Advanced Search Query Builder

Attribute

Structure Title — contains phrase — fatty acid binding protein

AND

Refinement R Factors (All) — < — 0.28

AND

Refinement Resolution — > — 2.0 Å

Display Results as: Structures

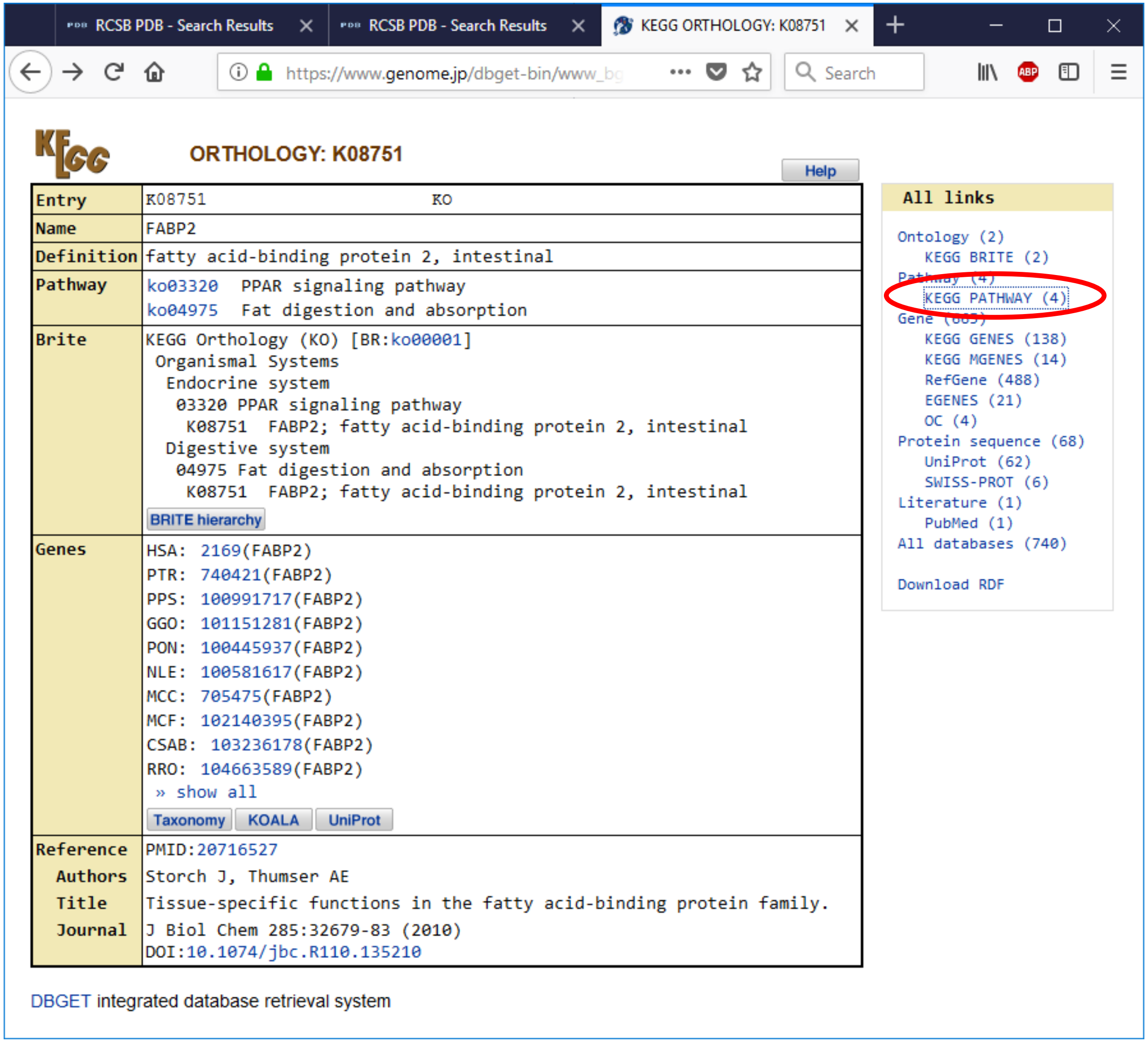
Website #4: KEGG

- http://www.genome.jp/kegg/
  (Kyoto Encyclopedia of Genes and Genomes)

- **Input:** Protein name, PDB ID, authors, etc.

- **Output:** What reactions does an enzyme catalyze?
  - Metabolic pathways
  - The "big picture"

Search Result: Intestinal FABP


RCSB PDB - Search Results
RCSB PDB - Search Results
KEGG ORTHOLOGY: K08751
https://www.genome.jp/dbget-bin/www_bg
Search
Help

ORTHOLOGY: K08751

Entry	K08751 KO
Name	FABP2
Definition	fatty acid-binding protein 2, intestinal
Pathway	ko03320 PPAR signaling pathway ko04975 Fat digestion and absorption
Brite	KEGG Orthology (KO) [BR:ko00001] Organismal Systems Endocrine system 03320 PPAR signaling pathway K08751 FABP2; fatty acid-binding protein 2, intestinal Digestive system 04975 Fat digestion and absorption K08751 FABP2; fatty acid-binding protein 2, intestinal BRITE hierarchy
Genes	HSA: 2169(FABP2) PTR: 740421(FABP2) PPS: 100991717(FABP2) GGO: 101151281(FABP2) PON: 100445937(FABP2) NLE: 100581617(FABP2) MCC: 705475(FABP2) MCF: 102140395(FABP2) CSAB: 103236178(FABP2) RRO: 104663589(FABP2) » show all Taxonomy KOALA UniProt
Reference	PMID:20716527
Authors	Storch J, Thumser AE
Title	Tissue-specific functions in the fatty acid-binding protein family.
Journal	J Biol Chem 285:32679-83 (2010) DOI:10.1074/jbc.R110.135210

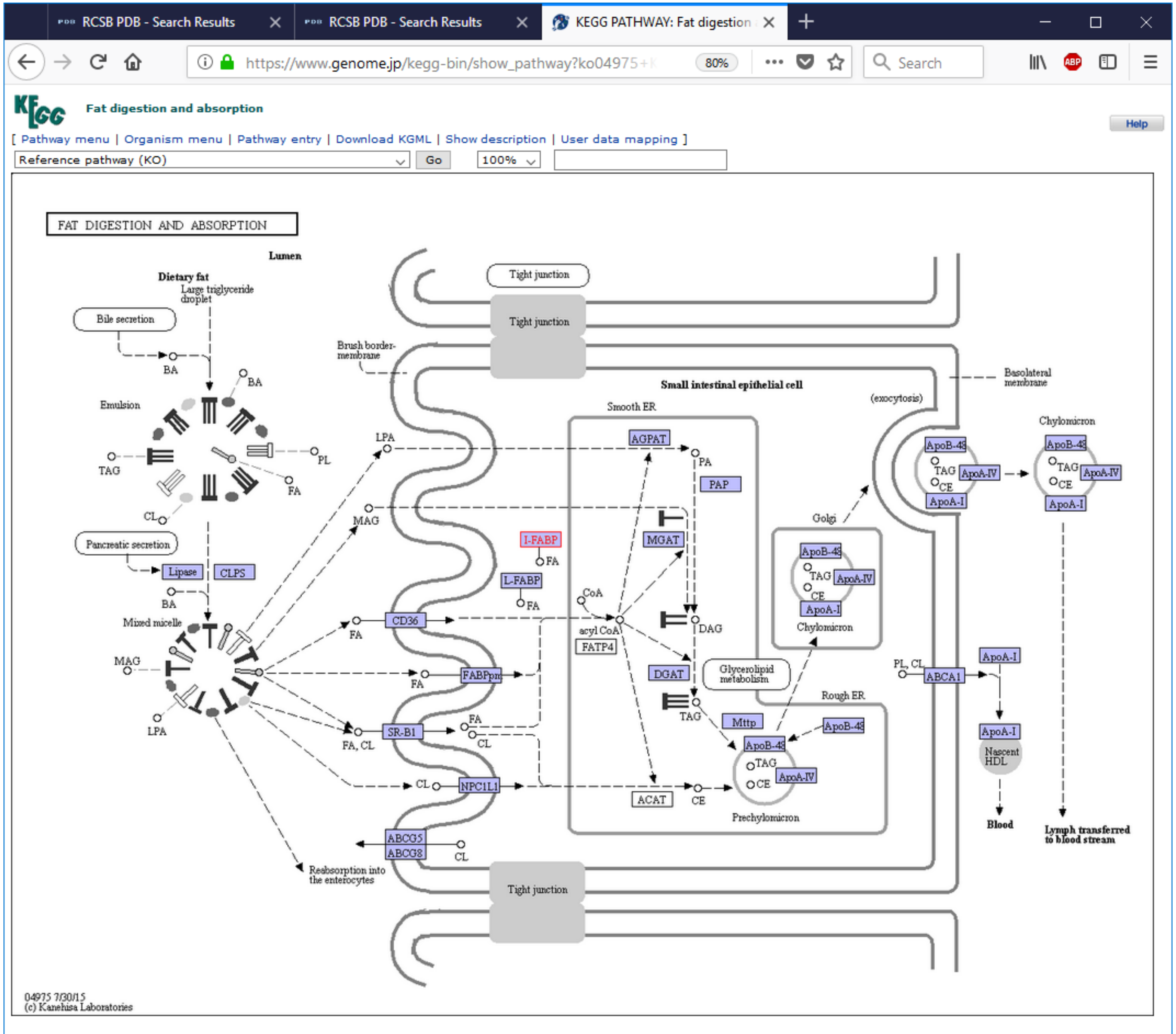
All links

Ontology (2)
KEGG BRITE (2)
Pathway (4)
KEGG PATHWAY (4)
Gene (665)
KEGG GENES (138)
KEGG MGENES (14)
RefGene (488)
EGENES (21)
OC (4)
Protein sequence (68)
UniProt (62)
SWISS-PROT (6)
Literature (1)
PubMed (1)
All databases (740)

Download RDF

DBGET integrated database retrieval system

Search Result: Fat Digestion and Absorption



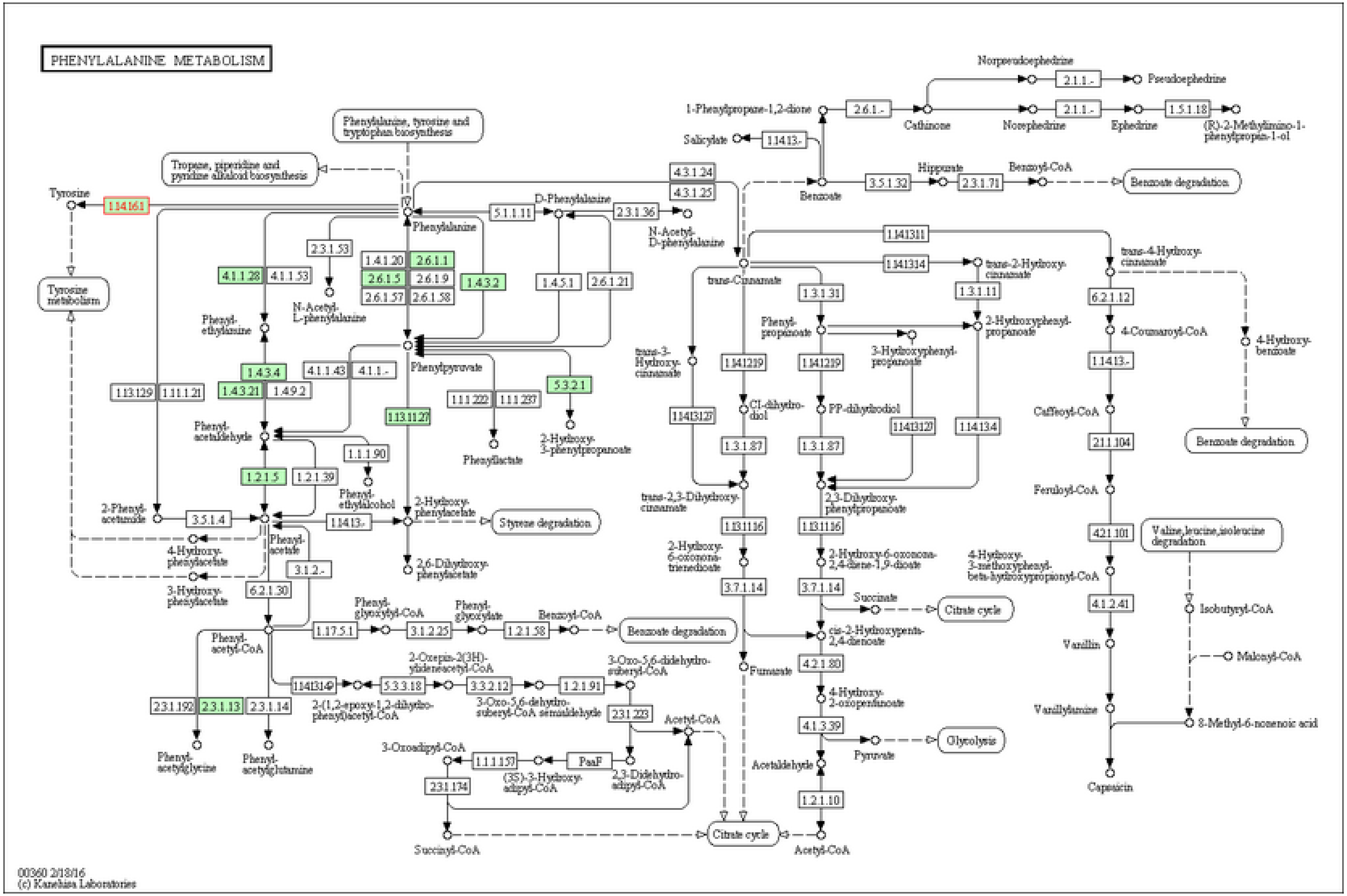
https://www.genome.jp/kegg-bin/show_pathway?ko04975+K

Fat digestion and absorption

[Pathway menu | Organism menu | Pathway entry | Download KGML | Show description | User data mapping]

Reference pathway (KO) Go 100%

Pathway for Phenylalanine Hydroxylase



Think and Discuss

What are the advantages to large, public databases of scientific information? Are there any disadvantages?

Summary

- Protein properties depend on their primary, secondary, tertiary, and quaternary structure
- Computer databases can organize huge amounts of data on biomolecular systems
- Entrez and the PDB are curated from published research worldwide