FA602 Biophysical aspects of structural biology

Adapted for on-line course = essential knowledge/skills

Spring 2020, L. Trantirek (lukas.trantirek@ceitec.muni.cz)

1. Literature searching – general, yet essential, knowledge

- **Scientific discoveries are disseminated in a form of PUBLICATION (text or metadata)**
- **Each publication is given unique DIGITAL OBJECT IDENTIFIER (DOI)**
- **DOI is a permanent way to identify an online document.**
  This identification is not related to its current location. *Example*: doi: 10.1074/jbc.RA120.012914

If you want to find scientific text/metadata then …

web search engines that indexes the full scietific text or metadata

Public/free of charge	**Public/requiring subscription**
• MEDLINE/PUBMED	Web of Science
• Google scholar	Scopus

NOTE: Various indexes do overlap, but they are not necessarily the same!!!

INPUT:

1] DOI
2] name of the author of the publication
3] keywords (example: DNA, CD spectroscopy, transcription, …)

OUTPUT:

list of publications satisfying your criteria
for each publication you will get full reference and abstract – brief text describing the work
And usually a link to electronic location of the full text/metadata

- **PUBMED** – mostly natural sciences and medicine, does not allow cross-referencing, but gives you indication of related relevant publication/database objects, etc.
- **Google scholar** – everything (non selective about the source), it may list even your bachelor thesis, allows cross-referencing
- **Web of Science** - used by our government, official scientiometry, does not generally include books/book series, low quality journals are not indexed, broad scope from art & humanities to medicine and physics, allows cross-referencing
- **Scopus** – very similar to Web of Science, allows cross-referencing

Let us practise

- PUBMED; Google scholar; Web of Science, Scopus

… switch your web browser on

However!

PUBMED; Google scholar; Web of Science, or Scopus search give you only the reference (publication info and link to publisher web page) and abstract.

What to do if you need full text?

You follow the link and go to publisher web page and you hope that
A] the text is free of charge (publishers tends to open older articles for public (free) use.
B] text was published in so-called Open Access (for you it means that it is free to read)
C] that your institution has subscription to the journal (in this case you can downloaded for free)

Otherwise you are expected to pay (typically ~ 30 USD) for access to the paper

ALTERNATIVELY you can use SCI HUB - **ethically problematic!!! It is a website that provides free access to millions of research papers and books, without regard to copyright, by bypassing publishers' paywalls.**

Let us practise

How to get to the full text?

… switch your web browser on

Useful knowledge: Placing references into your text. People are most frequently using EndNote (paid) and Zotero (free).

Optional homework:

1) **Install Zotero to your computer.**
2) **Search PUBMED for publications authored by Trantirek between 2000-2010.**
3) **Record selected publications into Zotero library.**
4) **Open new WORD document and complete the sentence:**
   "In between 2000-2010, Trantirek published X research papers (insert the publication from Zotero library, e.g. as [1-X]"
5) **What you should get is "References" - list of publications with all details**
   (author list, journal name, volume, year of publication, title, DOI, …)

You will use this knowledge when you are writing, bachelor/master/PhD thesis or scientific paper.

3. Search for basic information – essential knowledge for this course

How to obtain proteins' primary structure (sequence) & how to annotate proteins' basic functional elements?

INPUT: protein name

OUTPUT: protein sequence & annotations of functionally important parts of the protein structure

Basic TOOL: PUBMED Central – Proteins & PUBMED Central – Resources – Domains & Structures - Conserved Domain Search

Primary sequence Annotation

Let us practise

e.g., human intestinal cell kinase (ICK): Obtain hICK primary sequence and identify residues responsible for ATP binding

3. Structural data search – essential knowledge for this course

Let us focus on high-resolution data on biomolecules (DNA, RNA, proteins and their complexes) from X-ray, NMR, and cryo-EM

… this is what structural biology is mostly about*

PROTEIN DATABASE (PDB) – primary source

- **curates and annotates all biomolecular structural (3D) data according to agreed upon standards**
- **Each item is associated with unique identifier, PDB ID (e.g., 1QWB)**
- **Structural data are accessible in PDB format (sort of standard/reference format in the field)**
- **The database is freely accessible**
- **The database provide number of tools for structural, statistical, bioinformatics analysis**

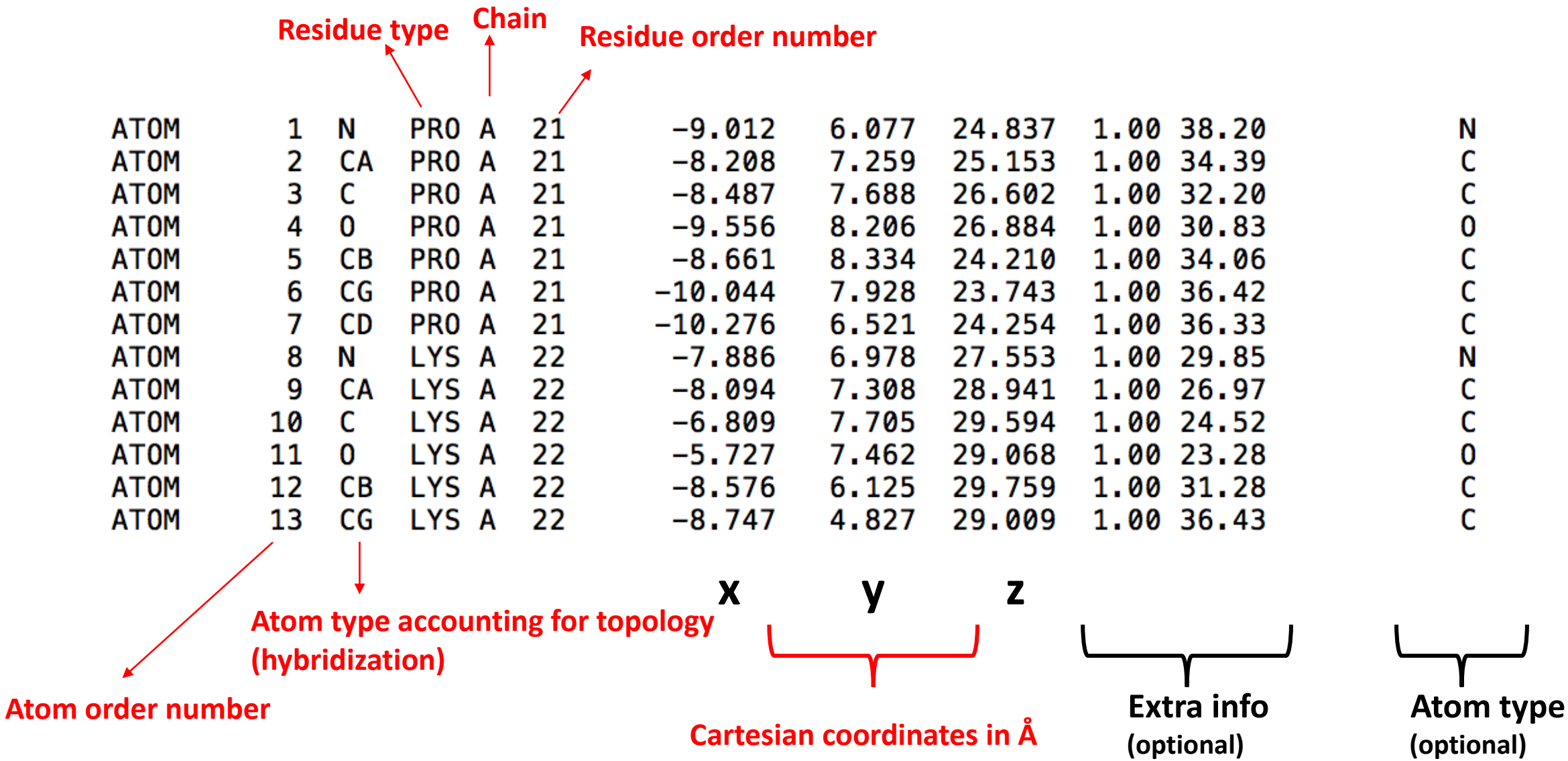
www.pdb.org

Nucleic Acids Database (NDB) – focuses on nucleic acids and their complexes, objects identified with NDB ID (which is in most cases identical with PDB ID); PDB include all information in NDB (not vice versa); NDB, however, has specialized tools to analyse NA structures.

*technically speaking, the term also involves other methods (MS, FRET, CD/IR/RAMAN spectr., chem. probing as well as modelling)
* next to NA & proteins – also (poly)-saccharides and lipids

PDB format – plain text format = you can display it in any text editor

However, only as a text. To visualize 3D structure you need specialized software



PDB format – contains also other fields than those marked by ATOM

```
HEADER    EXTRACELLULAR MATRIX                      22-JAN-98   1A3I
TITLE     X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE    2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA    X-RAY DIFFRACTION
AUTHOR    R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR   2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000        0.00000
...
SEQRES   1 A    9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES   1 B    6  PRO PRO GLY PRO PRO GLY
SEQRES   1 C    6  PRO PRO GLY PRO PRO GLY
...
ATOM      1  N   PRO A   1       8.316  21.206  21.530  1.00 17.44           N
ATOM      2  CA  PRO A   1       7.608  20.729  20.336  1.00 17.44           C
ATOM      3  C   PRO A   1       8.487  20.707  19.092  1.00 17.44           C
ATOM      4  O   PRO A   1       9.466  21.457  19.005  1.00 17.44           O
ATOM      5  CB  PRO A   1       6.460  21.723  20.211  1.00 22.26           C
...
HETATM  130  C   ACY   401       3.682  22.541  11.236  1.00 21.19           C
HETATM  131  O   ACY   401       2.807  23.097  10.553  1.00 21.19           O
HETATM  132  OXT ACY   401       4.306  23.101  12.291  1.00 21.19           O
...
```

Source: Wikipedia; cf. Wikipedia “PDB format” for detail description

Let us practise

PDB database

… switch your web browser on and go to www.pdb.org

4. Visualization of 3D structures – essential knowledge for this course

How to visualize of 3D structures (data in PDB file)?

You will need a special software.

We will learn how to use UCSF CHIMERA

(… cause, it is a freeware, it is intuitive, and allows you to do almost anything you might need)

Self-study

1) Download & install UCSF CHIMERA to your computer (https://www.cgl.ucsf.edu/chimera/)
2) Learn how to handle UCSF CHIMERA (longest video has ~ 5 min)

A] https://www.youtube.com/watch?v=hQxKYSUdiD8
B] https://www.youtube.com/watch?v=eLxhKc7Ljjk
C] https://www.youtube.com/watch?v=HRPVmRD5e1U
D] https://www.youtube.com/watch?v=oThN3LG8LQU

Note: For those interested – you might find a lot more videos on youtube on use of CHIMERA (making molecular movies, making mutant models, docking, etc). **A]-D] these are essential basics, which you will need later (exam)**

Homework: Using CHIMERA, map heparin binding site on the 3D structure of human FGF2