Core biochemistry toolbox: software, databases, etc. to get familiar with

Bri Bibel (the bumbling biochemist); updated 8/5/26

The exact programs and tools you will need will depend on your project, but these are some core ones most biochemists will need (this is an abbreviated version of my longer list)

Databases & related sites to get comfortable using:

- **UniProt**: database with tons of info about "every" protein & links to tools (BLAST, Clustal alignment, etc.) & other databases with specialized info
- **PDB** (Protein Data Bank): structural models (and now predicted structural models) of proteins and complexes w/corresponding data, validation reports, structural alignment tools, etc.
- **EMDB** (Electron Microscopy Data Bank): structural models, maps, and validation info for structures solved using cryo-electron microscopy (cryo-EM)
- **AlphaFold server:** for predicting protein structures and **AlphaFold database:** for viewing already-predicted protein structures
- **Expasy ProtParam:** biochemical stats for proteins (pI, extinction coefficient, #s of each amino acid, etc.)
- **Clustal Omega (ClustalW):** for aligning multiple protein sequences
- **Human Protein Atlas** & **EMBL-EBI Expression Atlas**: information about where genes are expressed (in what organisms, tissue types, intracellular locations, etc.)
- **BLAST**– lets you search for nucleic acid sequences, amino acid sequences, perform sequence alignments, etc.
- **NCBI Nucleotide** (which contains GenBank, RefSeq, etc.) & Ensembl (good for genome level): nucleic acid sequence data
- **PubMed** & **Google Scholar**: help you find scientific papers and receive alerts for new ones you may be interested in
- **PubMed central:** hosts full free versions of scientific papers
- **NCBI Bookshelf**: Free books, book chapters, and other documents
- **bioRxiv**: main preprint server (for papers that haven't been peer-reviewed)
- **GitHub:** a code repository where you can download open-source software programs, scripts associated with research papers, etc.
- **Expasy Swiss Bioinformatics Resource Portal:** lots of bioinformatics tools
- **BioNumbers: The Database of Useful Biological Numbers:** it looks for published stats on your query (concentrations, rates, sizes, etc.)
- **Google**: get good at Googling strategically and critically!

Software programs to choose & learn to use:

- A **reference manager** (Zotero, Mendeley, Endnote, Pages, Readcube etc.) with associated browser and word processor extensions to quickly save, organize, and cite papers
- Some sort of **plasmid viewer**: Benchling (free), ApE (free), Snapgene (viewer's free but limited features), CLC Main Workbench, DNAStar, etc. for molecular cloning
- **Vector graphics program** (e.g., Adobe Illustrator or free alternative InkScape)
- **Molecular visualization software (**especially **PyMol** and/or **ChimeraX)**: for interactively viewing (and making figures of) structures of macromolecules (proteins, DNA, RNA, complexes, etc.) which you can fetch from the PDB

- **Sequence alignment viewer** such as **JalViewJS (web version) or JalView desktop version:** for interactively viewing multiple sequence alignments (MSAs) of proteins and nucleic acids
- An **e-notebook** (OneNote, Benchling, etc.)
- **Slideshow software** (PowerPoint, Keynote, Google Slides, and/or Canva)
- **Spreadsheet software** (Excel, Google Sheets, etc.)
- **GraphPad Prism:** this isn't free, but if you have access to it, it's a great tool for graphing, modeling, and statistical analysis, of numerical data
- Also helpful - an **RSS reader** (Newsblur, Feedly, The Old Reader, Inoreader, etc.) to stay up to date on new articles

Other random things

- **Coding**
    - I recommend learning basic Python which you can work with in
        - **Jupyter Lab/Jupyter notebook**
        - **Google Colabs**
    - And if you need to learn/use R, check out **RStudio**
- **LinkedIn:** for networking, especially post-conferences and/or when looking for job opportunities
- **ORCiD:** get a free unique identifier that you can use to associate yourself with publications, sign in to various sites and set up a profile page
- **Clipboard manager:** to store things you copy so you can paste them even after you copied another thing
- **unpaywall browser extension**: see if free versions of pay-walled articles are available when you visit an article's page
- **PubPeer browser extension**: see if articles you visit or articles a visited article cites have been retracted or have had questions raised about them (usually, but not always, bad)
- **Cloud storage desktop drive (Box Drive, Google Drive, etc.)** for backing up and quickly accessing cloud files just as if they were on your hard drive

Links:

Proteins

- **UniProt** https://www.uniprot.org/
    - More about it here: https://bit.ly/uniprotprotparam & YouTube: https://youtu.be/f75f6QCe1gA
- **Expasy ProtParam** https://web.expasy.org/protparam/
    - More about it here: https://bit.ly/uniprotprotparam & YouTube: https://youtu.be/taVHOa7XnHg
- **Human Protein Atlas** https://www.proteinatlas.org/
- **EMBL-EBI Expression Atlas** https://www.ebi.ac.uk/gxa/home
- **Clustal Omega (ClustalW)** (also accessible through UniProt) https://www.ebi.ac.uk/jdispatcher/msa/clustalo
- **JalView/JalView JS** https://www.jalview.org/jalview-js/JalviewJS/

Macromolecular structures

- **PDB (Protein DataBank)** https://www.rcsb.org/ (US-maintained site), https://www.ebi.ac.uk/pdbe/ (European-maintained site)
    - More about it here: http://bit.ly/thepdb & https://bit.ly/pdbstructures (focus on x-ray crystallography structures) & on YouTube: https://youtu.be/klB4T0X-sVU & https://youtube.com/playlist?list=PLU1mvOopp1Ik&si=LLBBOZkQfvOX-YIe

- **EMDB (Electron Microscopy Data Bank)** https://www.ebi.ac.uk/emdb/
    - More: http://bit.ly/cryoemxray & https://youtu.be/G57aBLsBgOk
- **AlphaFold Protein Structure Database (AFDB)** https://alphafold.ebi.ac.uk/
- **AlphaFold** https://alphafoldserver.com/
    - More about it here: https://thebumblingbiochemist.com/365-days-of-science/interpreting-alphafold-predictions/
- **Molecular structure viewers**
    - **PyMOL** https://pymol.org/
        - More here: https://bit.ly/pymolintro & https://youtu.be/CeTTmm5DI_k & https://www.youtube.com/playlist?list=PLUWsCDtjESrG-46VDvPZfYPQyyXhZbGCN
    - **ChimeraX** https://www.cgl.ucsf.edu/chimerax/
    - Note: less powerful, browser-based alternatives include iCN3D and Mol*
        - **iCN3D** https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html
        - **Mol*** https://molstar.org/

Nucleic acids

- **BLAST** https://blast.ncbi.nlm.nih.gov/Blast.cgi
    - More about it here: http://bit.ly/sequencetermstools
- **NCBI Nucleotide** (which contains GenBank, RefSeq, etc.) https://www.ncbi.nlm.nih.gov/nucleotide/
- **Ensembl** https://uswest.ensembl.org/index.html
- **JalView/JalView JS** https://www.jalview.org/jalview-js/JalviewJS/
- **Plasmid viewers**
    - **Benchling (free)** https://www.benchling.com/
    - **ApE (A plasmid Editor)** totally FREE! https://jorgensen.biology.utah.edu/wayned/ape/
    - **SnapGene** viewer is free but has limited features https://www.snapgene.com/
    - **CLC Main Workbench** https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-main-workbench/
    - **DNASTAR** https://www.dnastar.com/software/lasergene/molecular-biology/

Papers & searching

- **PubMed** https://pubmed.ncbi.nlm.nih.gov/
- **Google Scholar** https://scholar.google.com
- **PubMed Central** https://pmc.ncbi.nlm.nih.gov/
- **NCBI Bookshelf** https://www.ncbi.nlm.nih.gov/books
- **bioRxiv** https://www.biorxiv.org/
- **Guide to strategic Googling: Blog:** https://bit.ly/scientificgoogling

Reference managers

- **Zotero** https://www.zotero.org/
- **Mendeley** https://www.mendeley.com/reference-management/reference-manager
- **Endnote** https://endnote.com/
- **ReadCube Papers** https://www.digital-science.com/product/readcube/
- More on reference managers: http://bit.ly/referencemanagementtools; YouTube: https://youtu.be/p2G0LZThsfA

Vector graphics programs

- **Adobe Illustrator**
  - More here: https://bit.ly/adobe_illustrator_fundamentals & https://www.youtube.com/playlist?list=PLUWsCDtjESrHU5vdztBYlQ4hK3iiccSnH
- **Inkscape** FREE https://inkscape.org/
- **bioicons** isn't a vector graphics program, but it has free science vector graphics you can use in those programs! (or elsewhere) https://bioicons.com/
- More about vector graphics here: http://bit.ly/makinggraphics

RSS Viewers

- **NewsBlur** https://newsblur.com/
- **Feedly** https://feedly.com/
- **The Old Reader** https://theoldreader.com/
- **Inoreader** https://www.inoreader.com/
- More about them here: https://bit.ly/science_alerts_RSS_feeds

Programming stuff

- **Jupyter lab** https://jupyterlab.readthedocs.io/en/stable/
  - More here: https://bit.ly/jupyter_tips
- **RStudio** https://posit.co/
- **GitHub** https://github.com/

Other random stuff

- **BioNumbers: The Database of Useful Biological Numbers** https://bionumbers.hms.harvard.edu/search.aspx
- **Expasy Swiss Bioinformatics Resource Portal** https://www.expasy.org/
- **GraphPad Prism** https://www.graphpad.com/
- **ORCiD** https://info.orcid.org/
  - More here: https://thebumblingbiochemist.com/365-days-of-science/orcidids/ & https://youtu.be/C6o7k7ktyXc
- **LinkedIn** https://linkedin.com
- **Clipboard manager**
  - I use one called CopyClip
  - More here: https://youtu.be/3g1CEXyxFFY
- **unpaywall browser extension** https://unpaywall.org/
- **PubPeer browser extension** https://pubpeer.com/

Resources page with a downloadable version of this, as well as a guide to my resources and others: https://thebumblingbiochemist.com/lets-talk-science/resources/

Page with resources I've found helpful**:** https://bit.ly/bumblingbiochemisttoolbox

More comprehensive toolbox of databases: https://bit.ly/databases_guide

Tools for studying proteins: https://thebumblingbiochemist.com/365-days-of-science/proteintools/

Structural biology resources: https://bit.ly/structural_biology