Open source and Free tools for drug development and analysis

V. Ganesh, Ph.D.
Novalead Pharma Pvt. Ltd. Pune

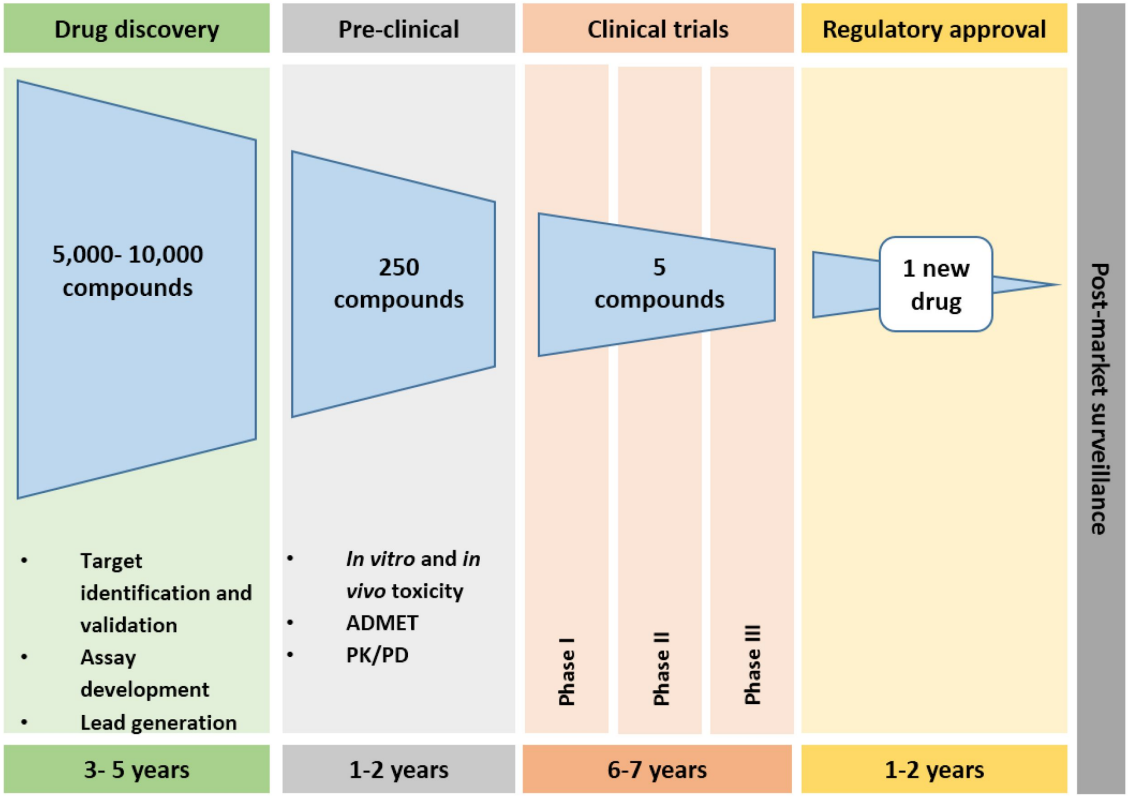
Agenda

1. Drug development life cycle
2. Algorithm Driven Drug development
3. Open source / Free tools for drug discovery and analysis

The “Funnel”

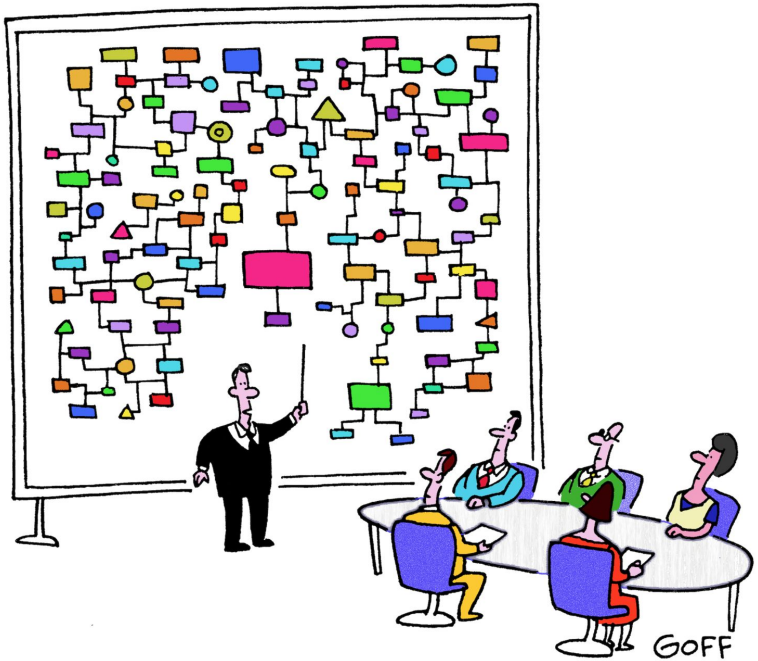
1. Multi year effort from lab to market approval
2. Initial screening may range from 1000 to a million molecule database
3. A shorter development cycle is repurposing market approved drug for a new indication.

Ref: https://www.mdpi.com/2227-7382/4/3/28/htm



Check the FDA: https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process

The need for a computer - “in-silico” drug discovery



"And that's why we need a computer."

License granting use of this cartoon to B. Villoutreix by T. Goff – USA – March 14, 2014

Ref: https://www.vls3d.com/links/51-shortlist

Drug Discovery Pipeline (in-silico)

- Database (Small Molecule / Molecule Fragments / Protein)
- Modeling Software (Geometry Optimization, Descriptors, Pharmacophore, Pocket identification and Docking, Binding and Energy calculation)
- Analysis Software (Structure Activity Relation, Toxicity, PK/PD)

Basic skills to acquire

1. File Formats (FASTA, PDB, MOL, SMILES, CML)

   (OpenBabel - https://github.com/openbabel/openbabel)

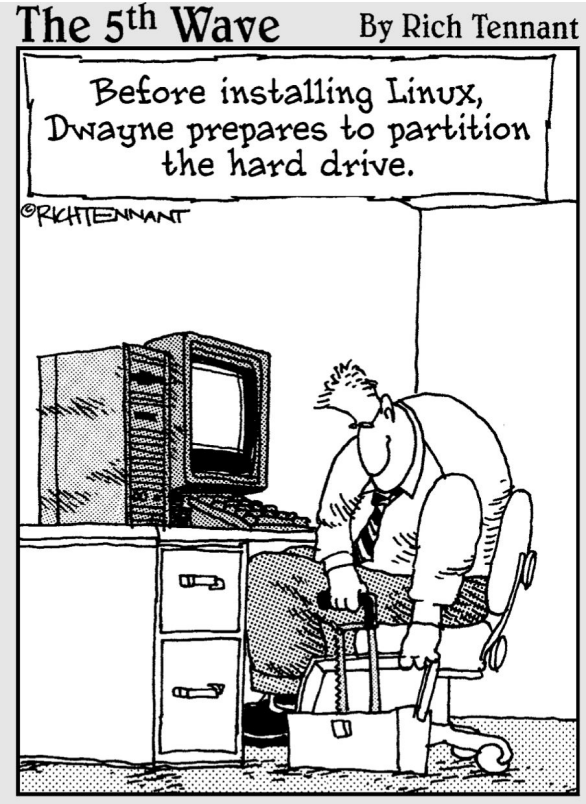
2. At Least one programming / scripting language (eg. Python, R)

   (A Byte of Python - https://python.swaroopch.com)

3. Ability to use web tools, spreadsheet and its functions

   (LibreOffice Calc is good replacement for Excel - https://www.libreoffice.org, Alternative: Google Sheets)

Basic skills to acquire

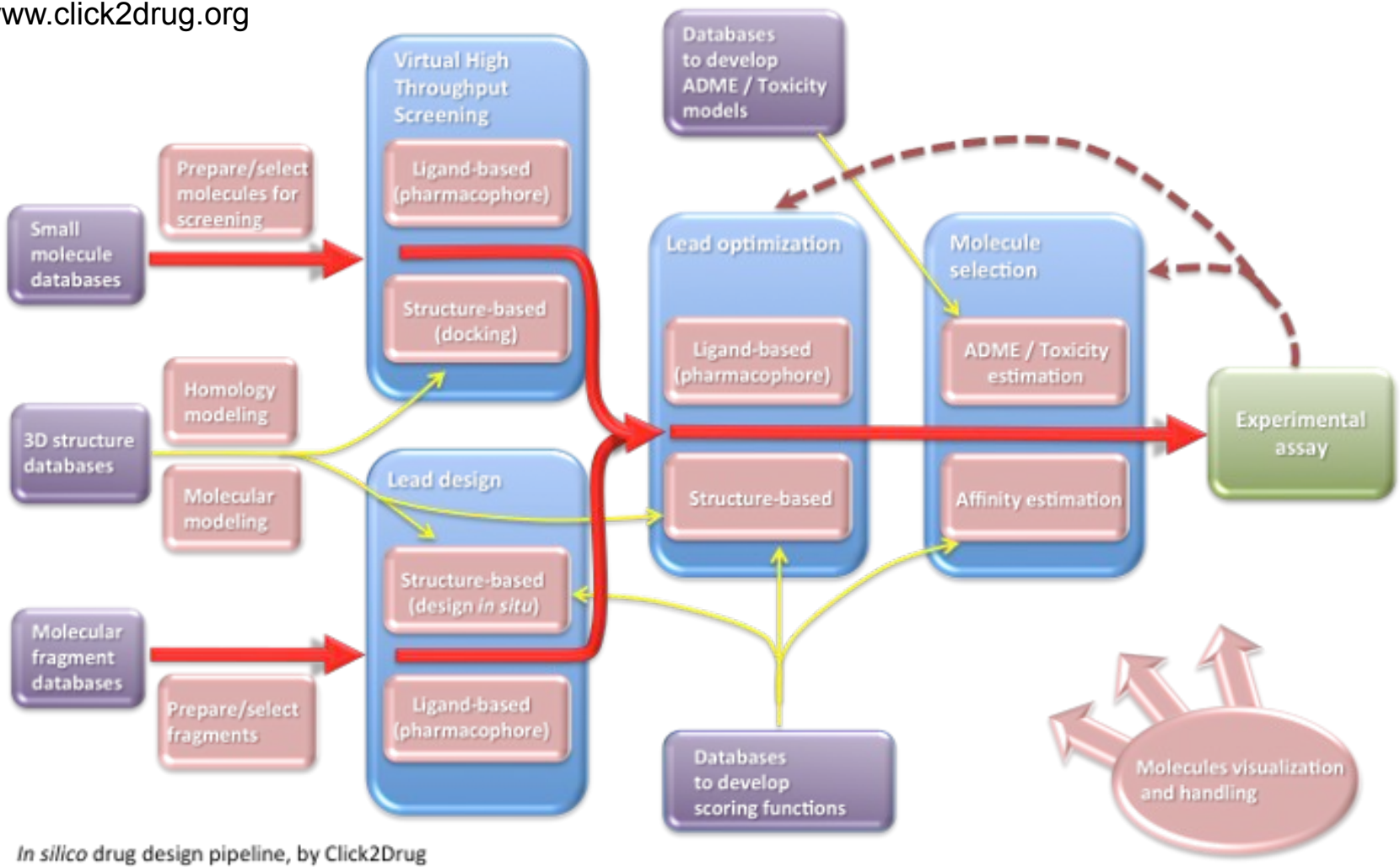


1. Good to know: Linux and Linux command line
2. Linux command line is also available on

   Windows 10 - if you are not comfortable setting

   up a different machine
3. Not that hard, but very powerful

Ref: www.click2drug.org



In silico drug design pipeline, by Click2Drug

Open source molecular modeling

Topical Perspective

Open source molecular modeling

Somayeh Pirhadi [b], Jocelyn Sunseri [a], David Ryan Koes [a]

Show more

https://doi.org/10.1016/j.jmgm.2016.07.008

Abstract

The success of molecular modeling and computational chemistry efforts are, by definition, dependent on quality software applications. Open source software development provides many advantages to users of modeling applications, not the least of which is that the software is free and completely extendable. In this review we categorize, enumerate, and describe available open source software packages for molecular modeling and computational chemistry. An updated online version of this catalog can be found at https://opensourcemolecularmodeling.github.io.

Ref: https://opensourcemolecularmodeling.github.io

Databases for in-silico drug discovery

1. Small Molecule (Zinc/15 - https://zinc.docking.org, ChEMBL - https://www.ebi.ac.uk/chembl/, PubChem - https://pubchem.ncbi.nlm.nih.gov)
2. Molecular Fragments

   https://academic.oup.com/nar/article/44/W1/W550/2499367

   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013803/
3. Proteins (Uniprot - https://www.uniprot.org, RCSB - https://www.rcsb.org)

Molecule Cleaning, Descriptor calculation...

1. CDK - https://cdk.github.io, Joelib - https://sourceforge.net/projects/joelib/,

   RDKit - https://www.rdkit.org (comprehensive tool kits)

2. Other resources: http://crdd.osdd.net/descriptors.php

3. Also check: https://webs.iiitd.edu.in/crdd/descriptors.php

Pharmacophore Identification / Search

1. PharmaGist - http://bioinfo3d.cs.tau.ac.il/PharmaGist/php.php
2. ZINCPharmer - http://zincpharmer.csb.pitt.edu/
3. Pharmer - https://sourceforge.net/projects/pharmer/

Structure based tools - Docking etc

1. fpocket - https://github.com/Discngine/fpocket
2. AutoDock vina - http://vina.scripps.edu/
3. PDBinder - http://cbm.bio.uniroma2.it/pdbinder/

QSAR

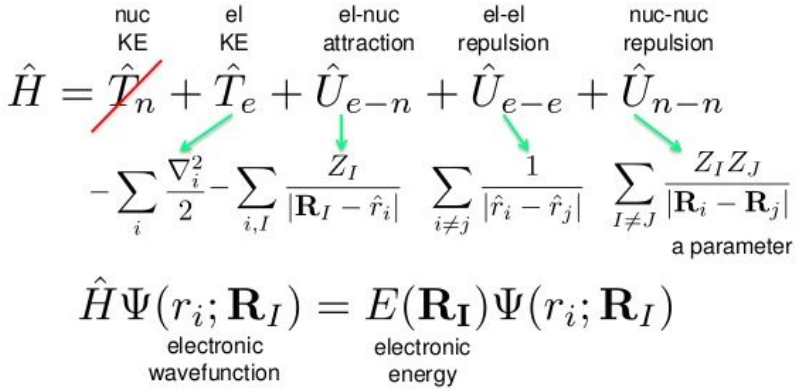
1. McQSAR - http://users.abo.fi/mivainio/mcqsar/
2. BlueDesc - http://www.ra.cs.uni-tuebingen.de/software/bluedesc/welcome_e.html
3. ODTT - https://github.com/oddt/oddt (Pipeline toolkit + molecular modeling tools)

Quantum Chemistry tools

1. Best tools: GAMESS, OCRA, NwChem, NAMD
   (Quantum Chemistry tools on Classical computers)

2. Online Submissions: https://chemcompute.org/

3. Also consider: http://www.psicode.org/



Ref: https://condensedconcepts.blogspot.com/2017/03/computational-quantum-chemistry-in.html

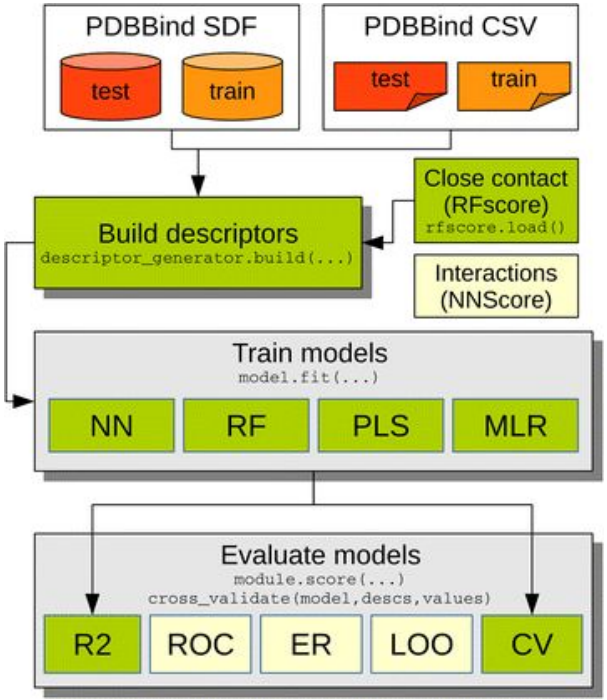
GUI Tools for submitting jobs

1. WebMO https://www.webmo.net/
2. OpenStructure https://openstructure.org/
3. Jmol http://jmol.sourceforge.net/
4. MeTA Studio https://github.com/tovganesh/metastudio

   (disclaimer: this one is written and maintained by me)

Pipeline toolkits - schedule tasks, document steps

1. ODDT - https://github.com/oddt/oddt
2. Jupyter Notebooks - https://jupyter.org
3. KNIME - https://www.knime.com, TAVERNA - https://taverna.incubator.apache.org



Ref: https://jcheminf.biomedcentral.com/articles/10.1186/s13321-015-0078-2

Systems Biology

1. Systems biology is the computational and mathematical analysis and modeling of complex biological systems
2. Open Source Computational Biology: https://github.com/opencb
3. https://github.com/biotite-dev/biotite

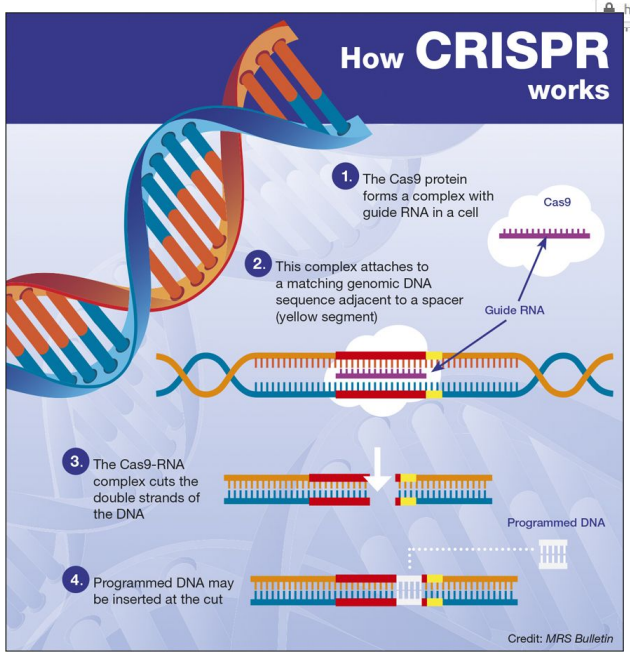
   (Biotite: a unifying open source computational biology framework in Python)

Emerging: Automated Hypothesis from Literature

1. An ML technique to generate hypothesis from existing literature using structured MESH (Medical Subject Headings) terms
2. https://www.biorxiv.org/content/10.1101/403667v1.full - Automated literature mining and hypothesis generation through a network of Medical Subject Headings
3. Novalead’s Druglead platform - open for academic collaboration

Emerging Paradigm: CRISPR

Clustered
Regularly
Interspaced
Short
Palindromic
Repeat



https://www.nature.com/articles/d41586-018-02477-1

OUTLOOK · 07 MARCH 2018

How CRISPR is transforming drug discovery

Gene editing is quietly revolutionizing the search for new drugs.

Andrew Scott

Ref: https://www.nature.com/articles/d41586-018-02477-1

Emerging Paradigm: CRISPR

1. https://github.com/crisflash/crisflash -

   off-target discovery tool for CRISPR/Cas9 guide RNAs

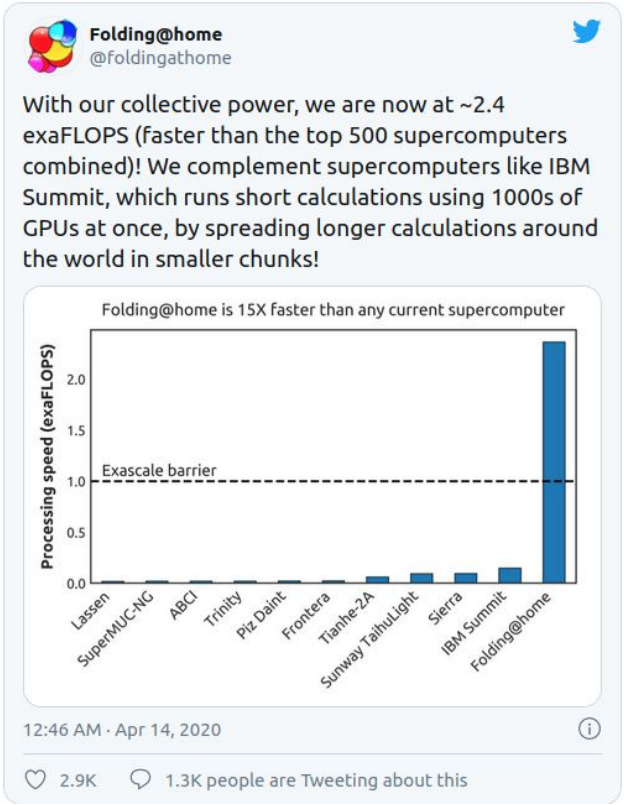
Jennifer Doudna and Emmanuelle Charpentier share the 2020 Nobel chemistry prize for their discovery of a game-changing gene-editing technique.

Ref: https://www.nature.com/articles/d41586-020-02765-9

Emerging Paradigm: Quantum Computing

1. https://dl.acm.org/doi/10.1147/JRD.2018.2888987 - Potential of quantum computing for drug discovery
2. https://cen.acs.org/business/informatics/Lets-talk-quantum-computing-drug/98/i35 - Let’s talk about quantum computing in drug discovery
3. https://science.sciencemag.org/content/369/6507/1084 - First Hartree-Fock calculation on a quantum computer

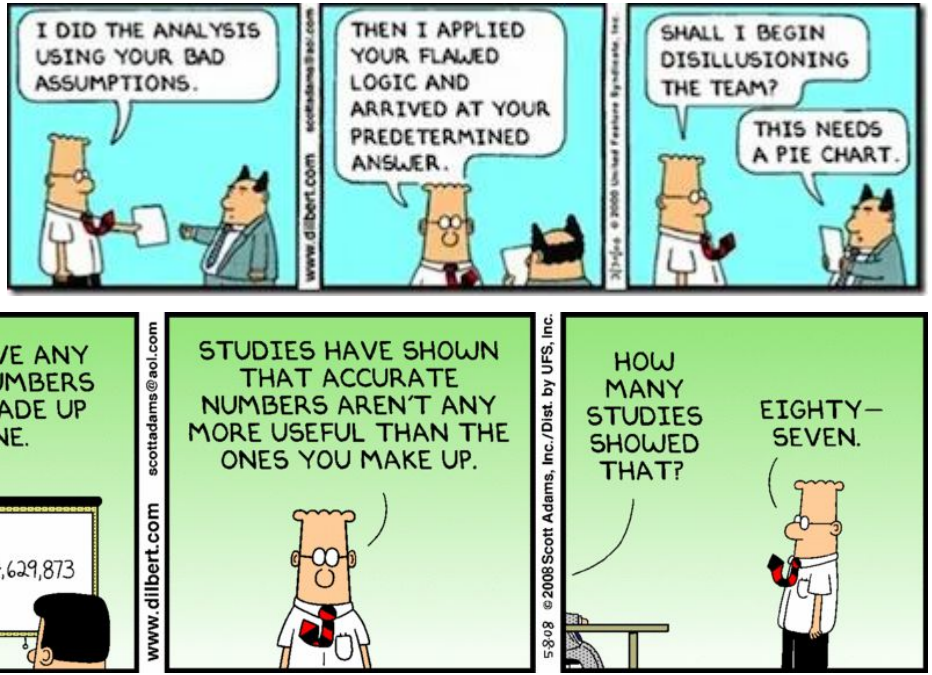
Distributed Computing Platforms



1. Computations can be expensive
2. Distributed Computing Platforms provide access to huge pool of voluntary compute resources donated by regular home computer users.
3. One such largest example is Folding@home - https://foldingathome.org/
4. Another example is: https://fold.it/ - uses gamification to solve protein folding problem

Analysis

1. Correct Data
2. Correct Reasoning



Analysis - Drug Safety and Efficacy

Analysis - Drug Safety and PK/PD

1. Toxicity analysis - TEST tool from EPA

   https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test

   https://www.nature.com/articles/d41586-018-05664-2

   (Software beats animal tests at predicting toxicity of chemicals)

2. PK/PD analysis - BioGears

   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6363067/

   https://www.biogearsengine.com/

Analysis - Drug Safety and PK/PD

3. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules http://www.swissadme.ch/

Analysis - Statistical analysis

- Tools: GNU Octave - https://www.gnu.org/software/octave/,

  R Studio - https://rstudio.com/products/rstudio/#rstudio-desktop,

  scilab - https://www.scilab.org

- Others: Python,

  Docker instances - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6042832/

  (with pre-build pipeline)

Team effort

Jupyter Notebook - https://jupyter.org/
For reproducible, sharable documentation of executable computational experiment.

Thanks

List of software tools mentioned in this presentation are available on Github: https://github.com/tovganesh/oftdd

To contribute, you can reach me at: tovganesh@acm.org