A Python package for Structural Alignment

Sommersemester 2021 - Projekt D

Jaime Rodríguez-Guerra, Andrea Volkamer
2021.01.11

1 About us

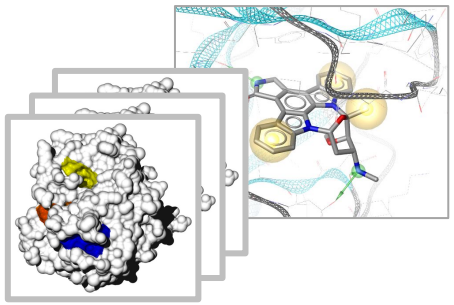
About us, AG Volkamer

Structure -based

Structural bioinformatics

Using protein structure information[2,3]

- Pharmacophore perception[4]
- Active site comparison
- (Off)target prediction[5]

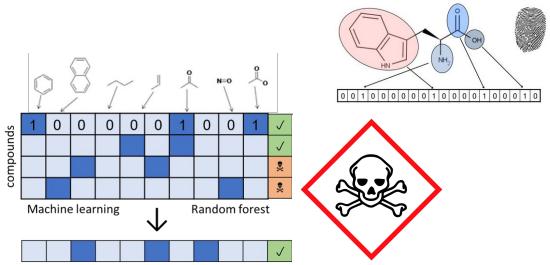


Ligand-based

Cheminformatics

Toxicity or activity prediction[1]

- KnowTox project
- Cytotoxicity prediction
- Automated ML pipeline



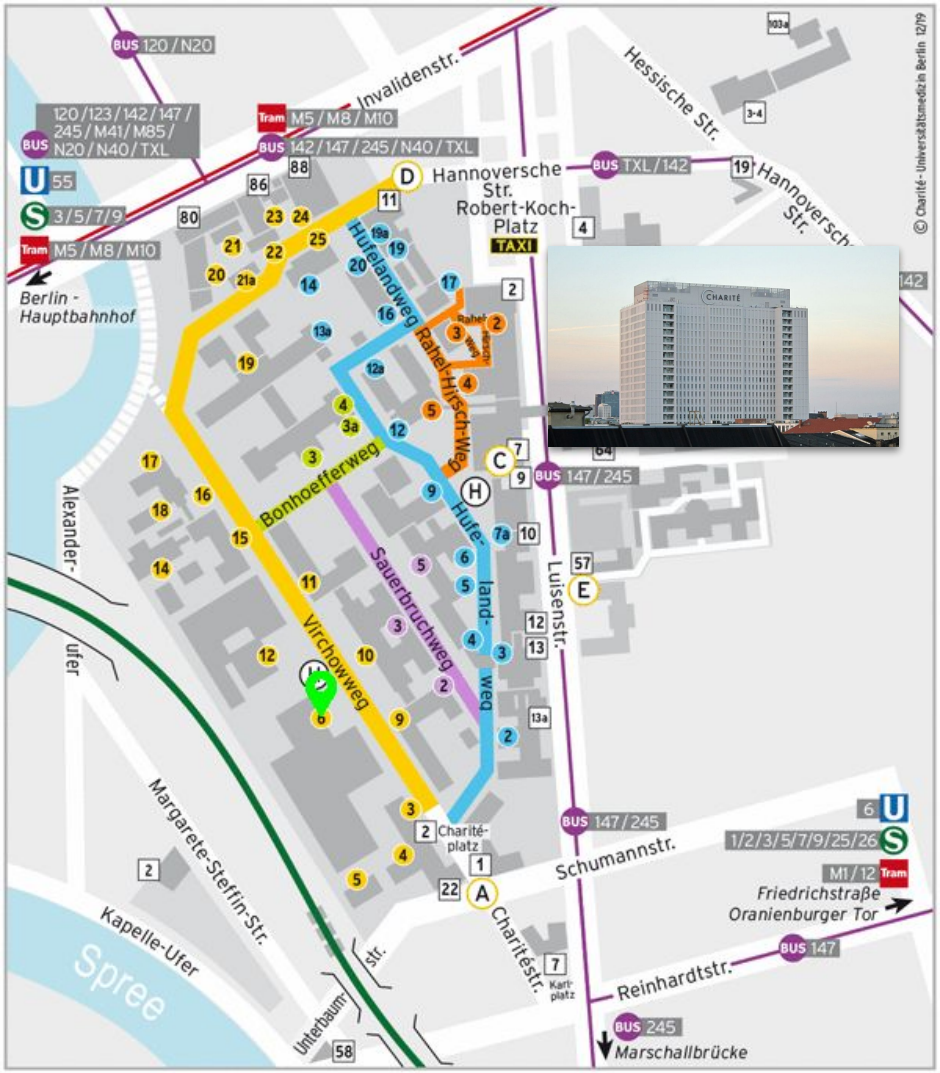
Applied to kinases

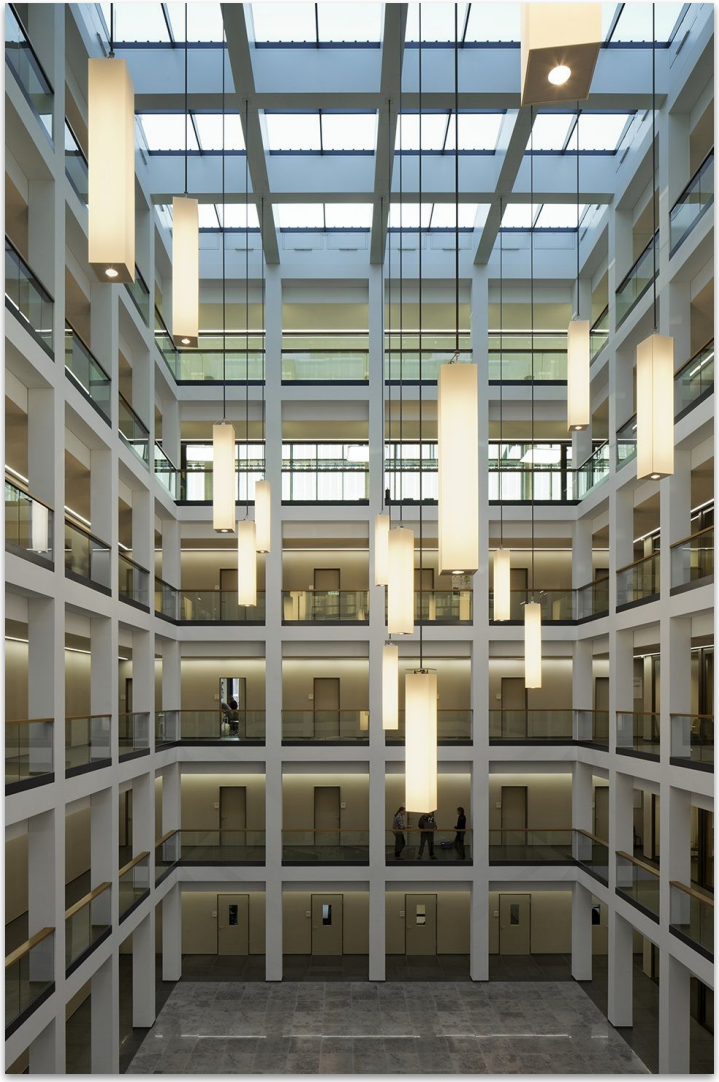
Kinase-centric computational drug development[6]

- Kinase comparison and (off)targets
- Novel fragment-based kinase inhibitor design
- Kinome-wide scalable predictions of kinase polypharmacology

[1] Lang, Volkamer, *et al., ALTEX* , ***2018,*** 35(1): 126-128
[3] Fährrolfes, *et al.*, NAR, **2018**, 45(W1): W337-W343
[3] Volkamer, *et al.,* Cheminformatics (Wiley), **2018**
[4] Mortier, Dhakal, Volkamer, *Molecules*, **2018**, 23(8), E1959
[5] Sydow, *et al.*, *JCIM*, **2019**, epub
[6] Kooistra, Volkamer, *ARMC V.50, Elsevier*, **2017**, 153-192

Location @ CCO, Charité Campus Mitte



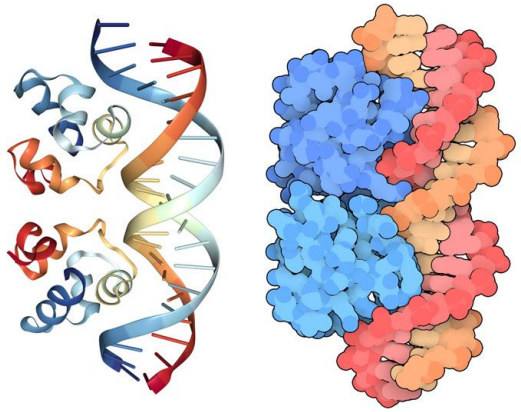


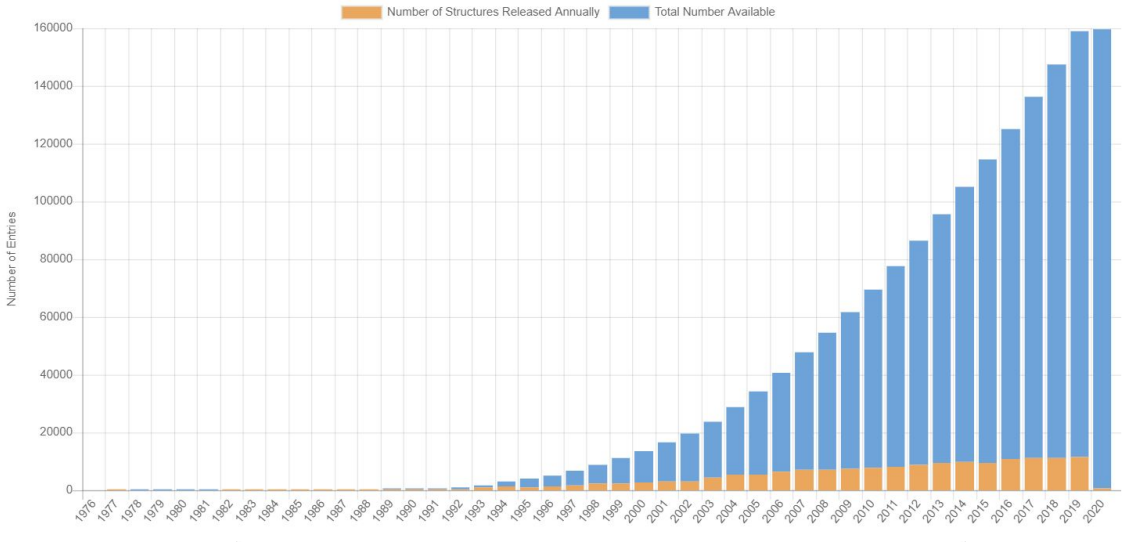
2 The project

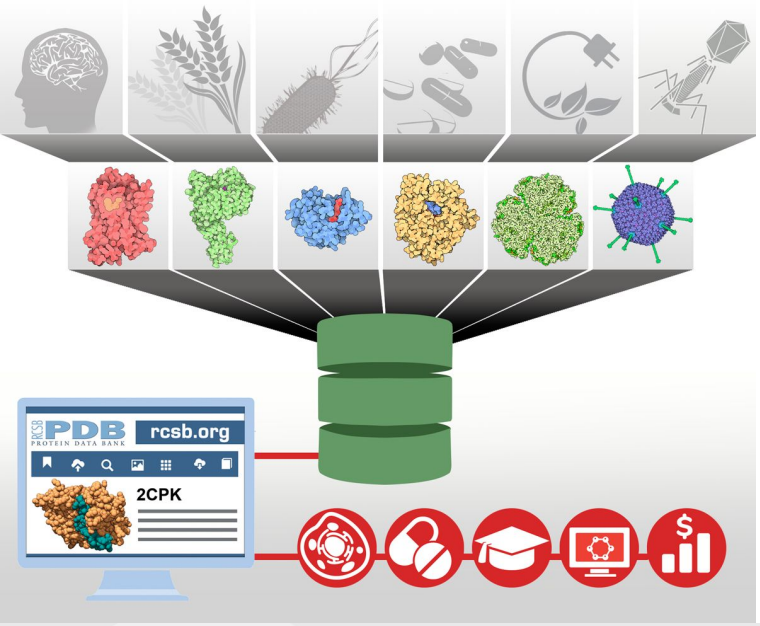
Structural alignment & superposition

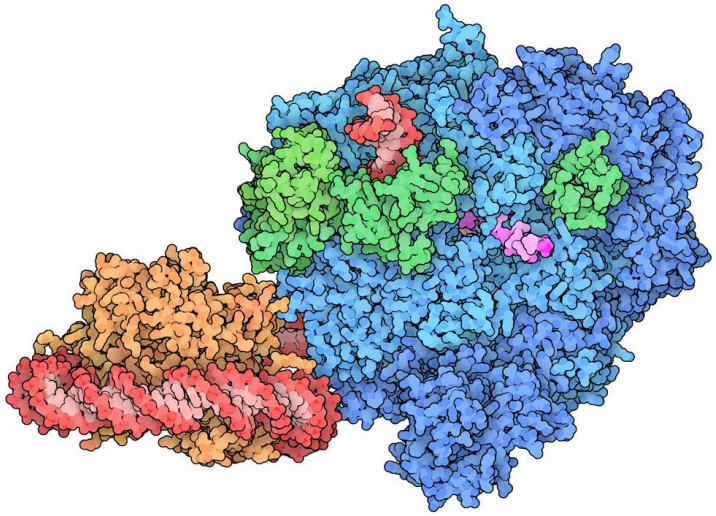
Project D

Structural databases: PDB



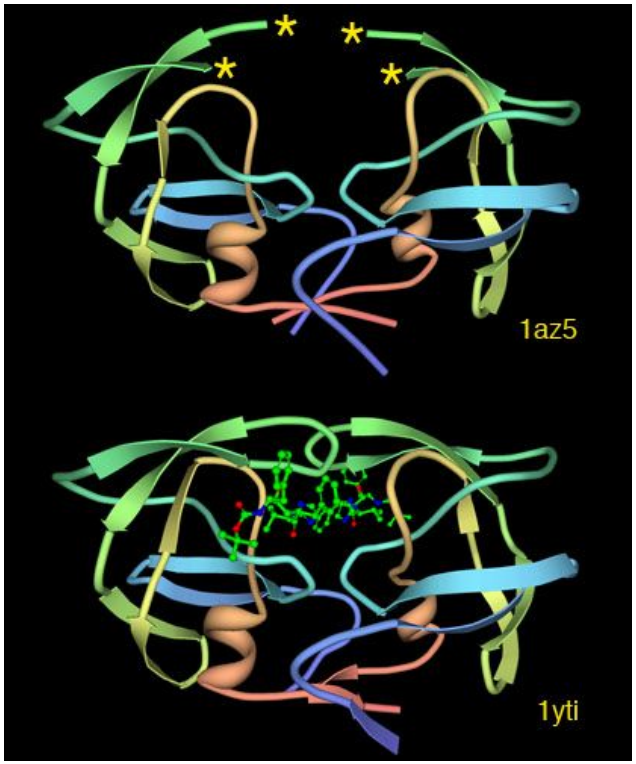






Structures can have issues

- Flexible regions are difficult to resolve
  - Missing loops or chains
  - Intrinsically disordered regions
- Not all proteins have been crystallized
  - But we can model the missing ones thanks to homology!
- **Homology modeling**
  an uncharacterized protein can be modeled by mimicking the structure of homologous proteins (we assume they are structurally similar because their sequences are closely related)



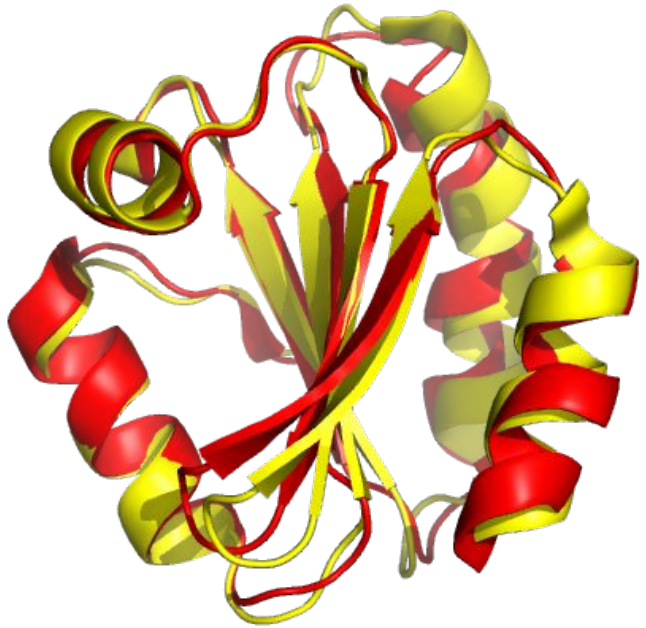
The structure of SIV protease solved without its active site (PDB entry 1az5) had two loops that were too flexible to be seen in the experiment (shown with stars in the upper picture). When the protein was crystallized with inhibitors, however, the loops adopted a stable structure that may be seen (PDB entry 1yti). -https://pdb101.rcsb.org/

Introducing structural alignment

- Similar structures can be superposed for their analysis

- Useful for model comparison or homology preparation

- There are several strategies; one of them:
  - 1. Align their sequences if possible
  - 2. Compute the residue pairings
  - 3. Minimize the residue-residue distance



- Most researchers will use graphical interfaces.
- What if we want to perform alignment in our program?
  - There is no Python standalone package for this!
  - We would need to install a huge package (not easily distributable) *just to use a single function*.

3 How we will work

How we will work

- Requirements:
  - Comfortable with **Python**
  - Familiar with Git and **GitHub**
  - Excited about structural bioinformatics & **best practices** in software development!
    - Modular **libraries** are better than one-off scripts
    - **Unit testing** helps discover bugs
    - **Continuous Integration** makes sure everything works
    - **Documentation** helps people get started

- Remote collaboration as a 4-person team

- Regular meetings for updates & QA
- Final presentation of results: the library and its packages

Current state and future goals

- Started last year with students like you!
- We built Superposer, now a part of OpenCADD
- Fully functional for simple cases, but we want to make it **better**

- Goals:
  - Get familiar with structural alignment and current codebase
  - Fix issues with alignment-specific edge cases, such as:
    - Mismatched residue numberings
    - Structural gaps
    - Locally-driven alignment (e.g. prioritize a binding site)
    - Structural equivalency of different types of residues
  - Benchmark the tool against known and unknown alignments

Tentative 8-week plan

- **Feb 15th or 22nd:** Project starts
- **Apr 21st or 28th**: Final presentation
- Unavailable

Questions ?

Please contact us at:

jaime.rodriguez@charite.de · andrea.volkamer@charite.de
www.volkamerlab.org