

A Python package for Structural Alignment

Sommersemester 2020 - Projekt D

Jaime Rodríguez-Guerra, Andrea Volkamer
2020.01.31

1 About us

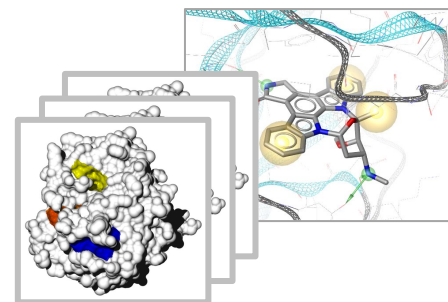
About us, AG Volkamer

Structure
-based

Structural bioinformatics

Using protein structure information^{2,3}

- Pharmacophore perception⁴
- Active site comparison
- (Off)target prediction⁵

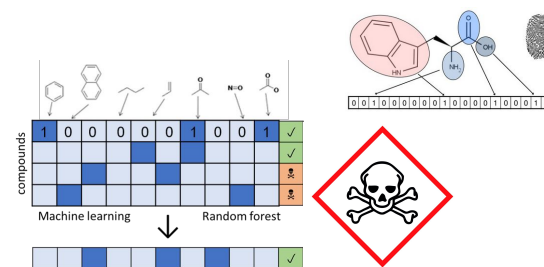


Ligand-
based

Cheminformatics

Toxicity or activity prediction¹

- KnowTox project
- Cytotoxicity prediction
- Automated ML pipeline



Applied
to kinases

Kinase-centric computational drug development⁶

- 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

The map illustrates the Charité hospital area in Berlin, highlighting its proximity to the Spree river and major public transport hubs. Key streets shown include Alexanderufer, Margarete-Steffin-Str., Kapelle-Ufer, Unterbaumstr., Reinhardtstr., Schumannstr., and the main thoroughfares Invalidenstr. and Hannoversche Str. Public transport routes are color-coded: yellow for bus lines (e.g., 120, 123, 142, 147, 245, M41, M85, N20, N40, TXL), blue for tram lines (M5, M8, M10), and orange for taxi services. The map also shows the locations of the U-Bahn stations U55 and U6, and the S-Bahn stations S3, S5, S7, and S9. An inset photo provides a visual reference of the Charité hospital building, a large, modern structure with a glass facade. The map is oriented with North at the top, and the Spree river flows from the bottom left towards the top right.

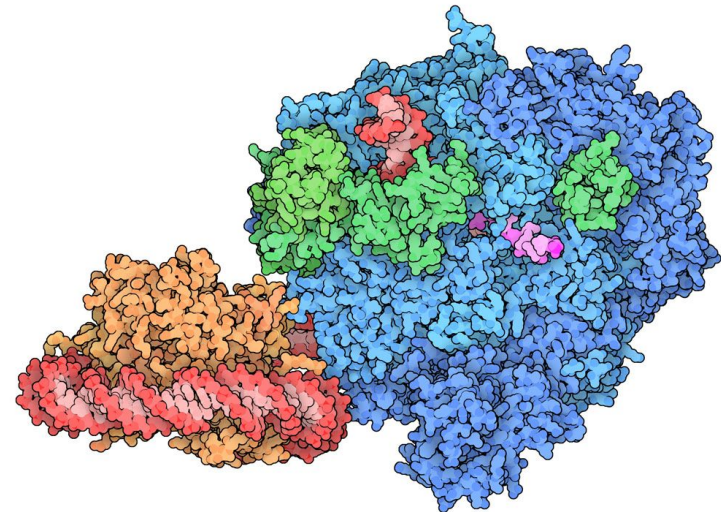
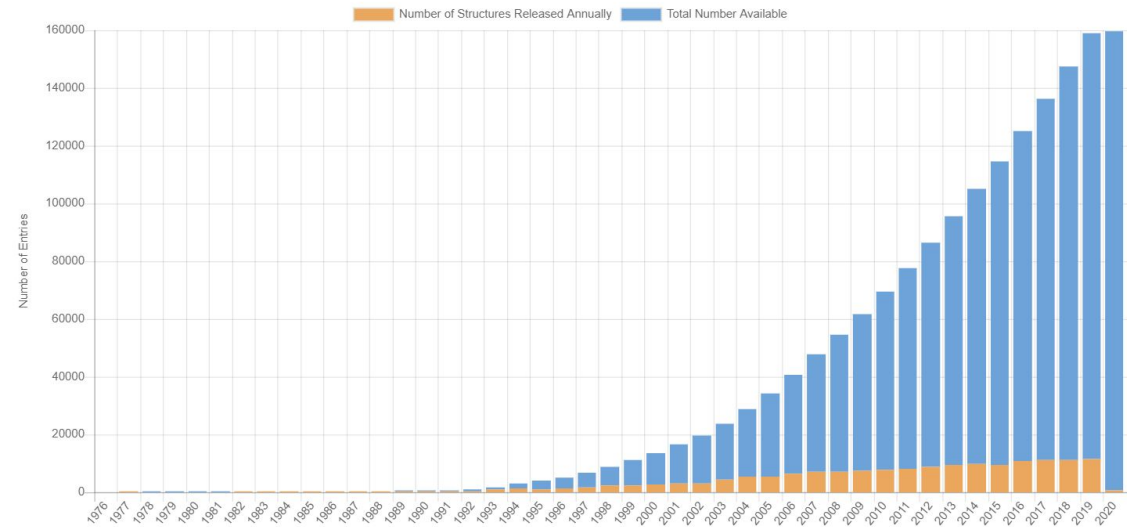
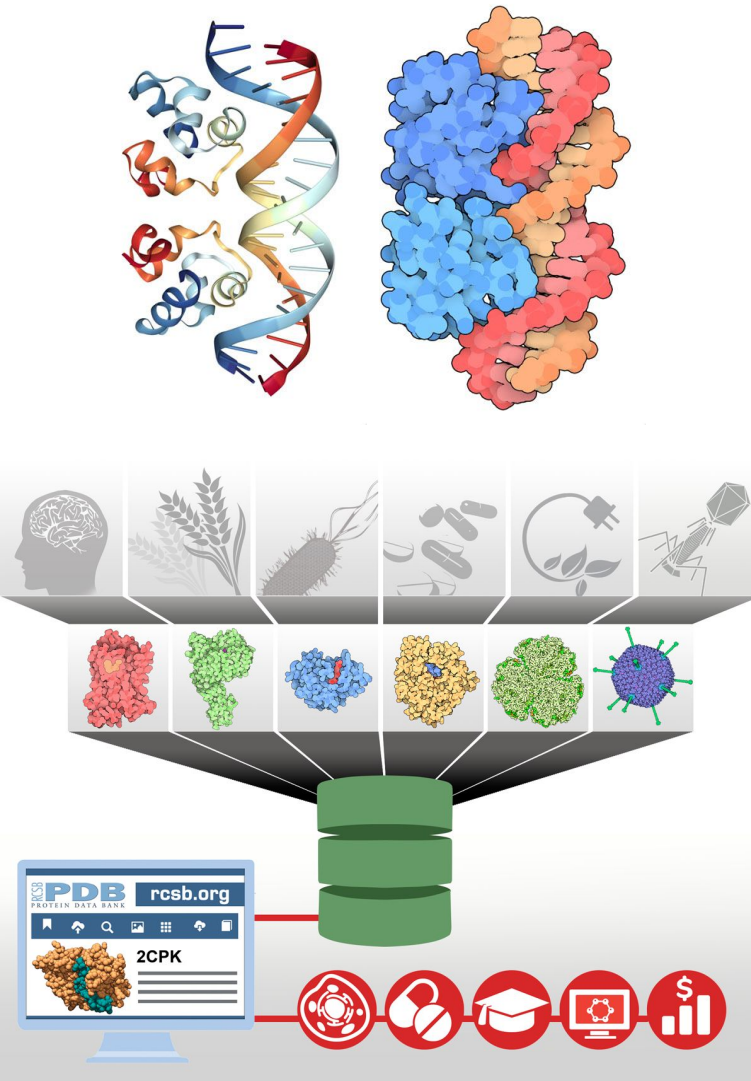


2 The project

Structural alignment & superposition

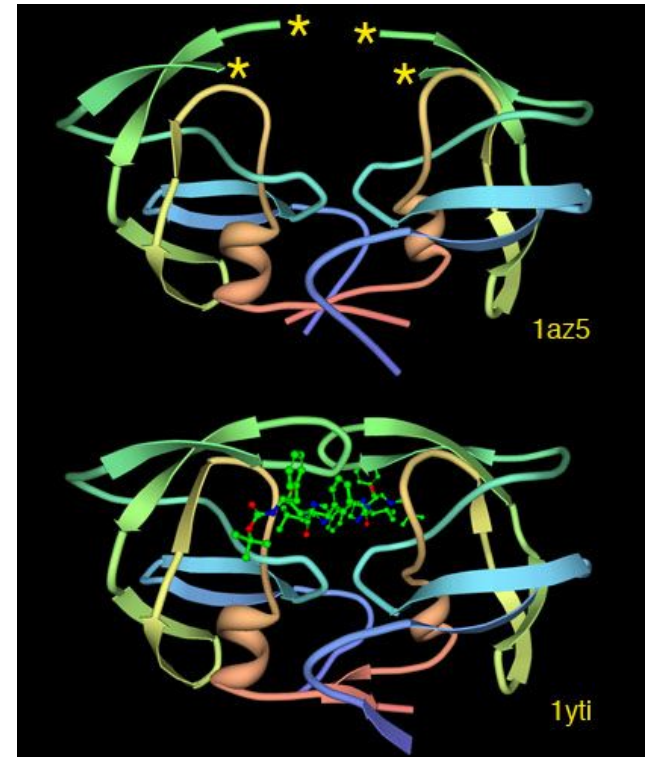
Project D

Structural databases: PDB



Structures can have issues

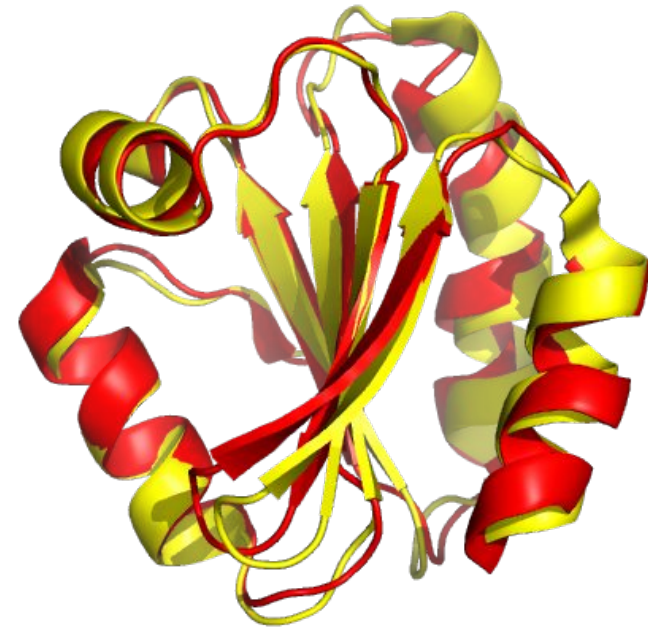
- Flexible regions are difficult to resolve
 - Missing loops or chains
 - Intrinsically disordered regions
- Not all proteins have been crystallized
 - But we can model the missing ones thanks to homology!
- **Homology modeling**
an uncharacterized protein can be modeled by mimicking the structure of 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

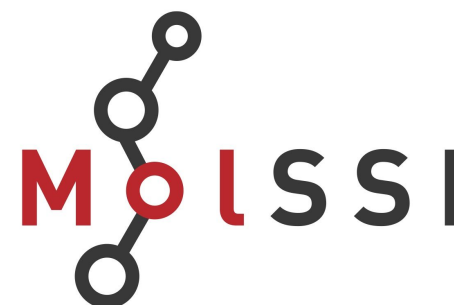
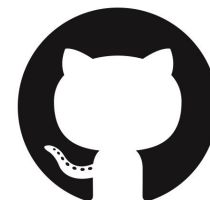
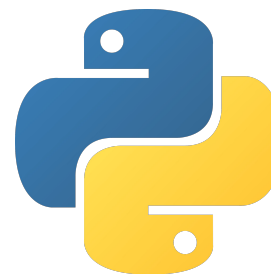
- Similar structures can be superposed for their analysis
- Useful for model comparison or homology preparation
- There are several strategies; one of them:
 - 1. Align their sequences if possible
 - 2. Compute the residue pairings
 - 3. Minimize the residue-residue distance
- Most researchers will use 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!
- You will join a free 2-day workshop:
 - From scripts to packages
 - Unit testing
 - Continuous Integration
 - Documentation
- Remote collaboration as a 5-person team
- Regular meetings for updates & QA
- Final presentation of results: the library and its packages



Tentative dates

2020 FEBRUARY						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

2020 MARCH						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

2020 APRIL						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

- **Feb 12th & 13th:** Workshop
- **Mar 3rd:** Project starts
- **Apr 28th:** Final presentation

Questions

Thanks!

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