



MICHIGAN STATE UNIVERSITY



Structure-based Drug Design

Translational Bioinformatics Workshop

Jing Xing

Aug 14, 2020

Agenda

- Protein targets & drug design
- 3D structures of protein pockets and small molecules
- Molecular docking
- Target fishing

Key Proteins as Disease Targets



Review Article | [Free Access](#)

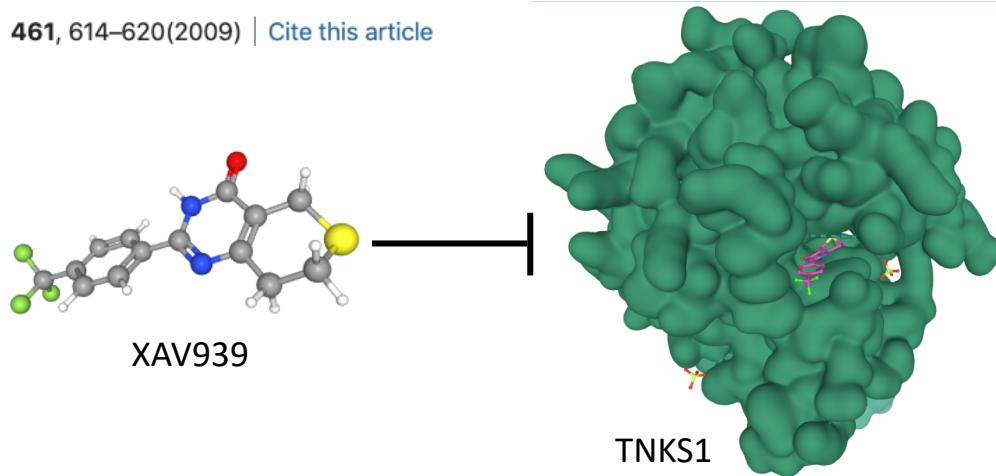
PARP inhibitors in breast cancer: Bringing synthetic lethality to the bedside

Anita A. Turk MD, Kari B. Wisinski MD

Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling

Shih-Min A. Huang, Yuji M. Mishina, [...] Feng Cong

Nature 461, 614–620(2009) | [Cite this article](#)



Review > Expert Opin Drug Discov. 2016 Sep;11(9):907-16.

doi: 10.1080/17460441.2016.1201057. Epub 2016 Jun 23.

The discovery of vemurafenib for the treatment of BRAF-mutated metastatic melanoma

Alex Kim ¹, Mark S Cohen ¹

> Cardiovasc Drugs Ther. 2018 Apr;32(2):135–145. doi: 10.1007/s10557-018-6778-x.

Combined SGLT2 and DPP4 Inhibition Reduces the Activation of the Nlrp3/ASC Inflammasome and Attenuates the Development of Diabetic Nephropathy in Mice with Type 2 Diabetes

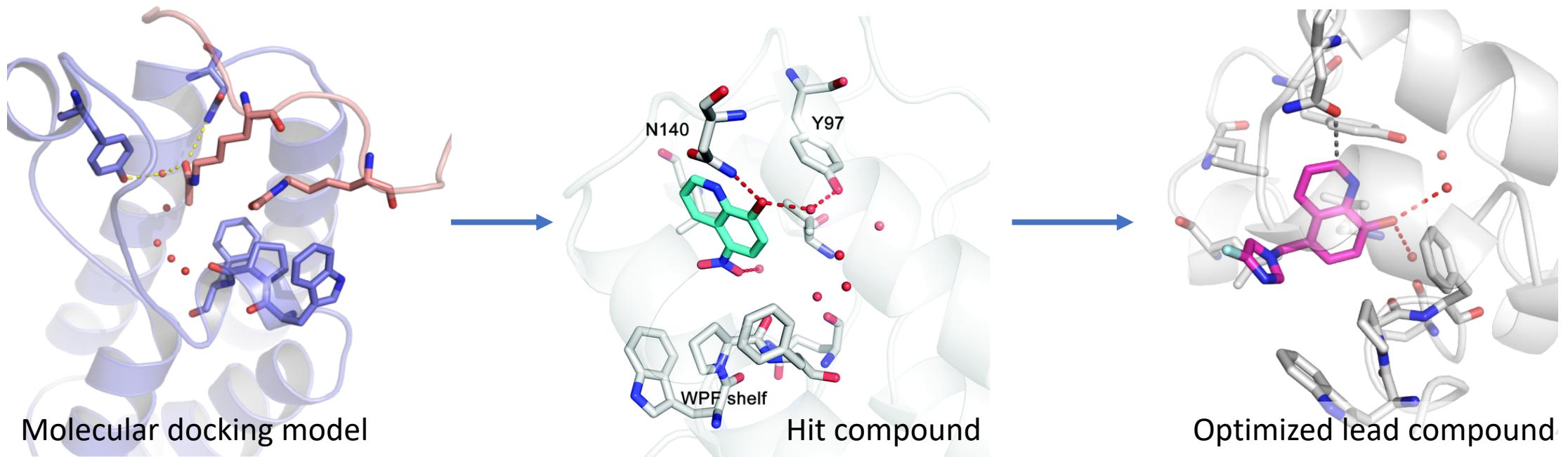
Yochai Birnbaum ¹, Mandeep Bajaj ², Hsiu-Chiung Yang ³, Yumei Ye ⁴

Biological mechanism studies suggested key proteins to intervene specific disease development. Some of the proteins are “**targetable**” by small molecules. Target-based drug discovery.

Structure-based Drug Design

BRD4, a drug target in oncology, e.g. leukemia

1. Obtain the 3D structure of BRD4
2. Develop a docking model to screen compound library against BRD4 pocket
3. Bench validation of the top scored compounds
4. Analyze binding mode & structure-activity relationships (SAR)
5. Based on the receptor pocket structure, optimize the hit compounds for better activity / selectivity



3D Structures of Macromolecules

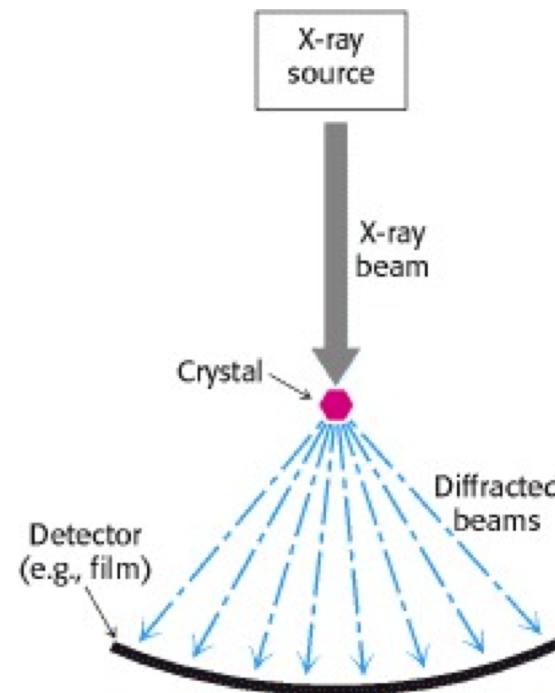
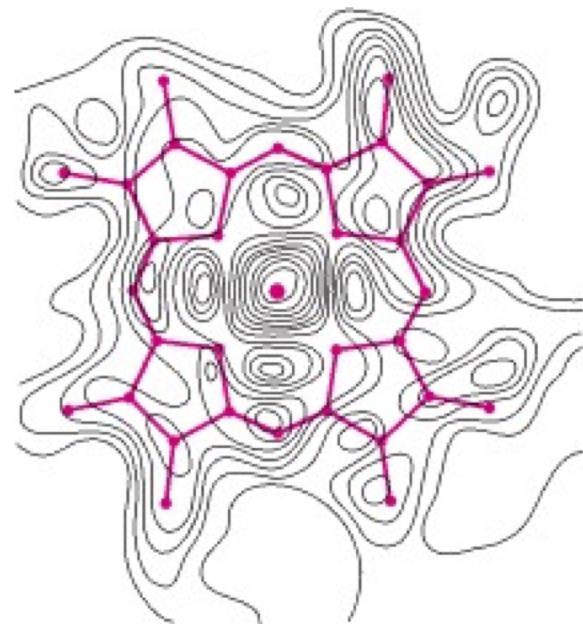
How are they determined?

X-ray crystallography:

Resolution $1\sim 2 \text{ \AA}$.

Crystals of proteins.

Electron density maps derived from X-ray reflections.



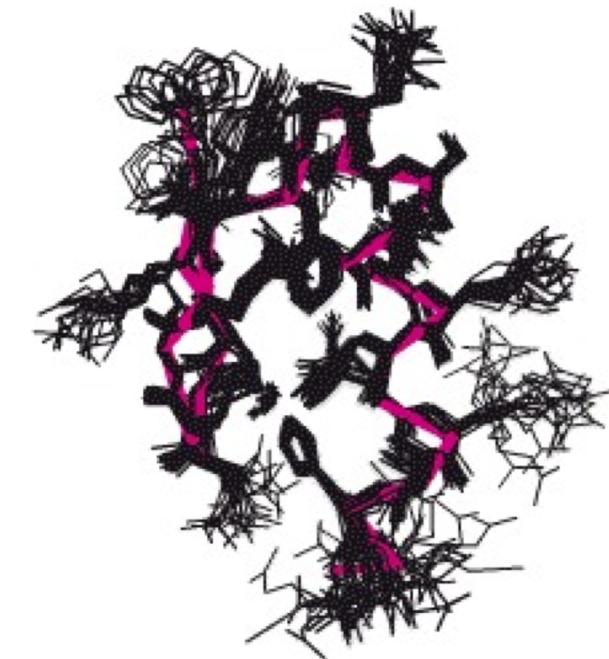
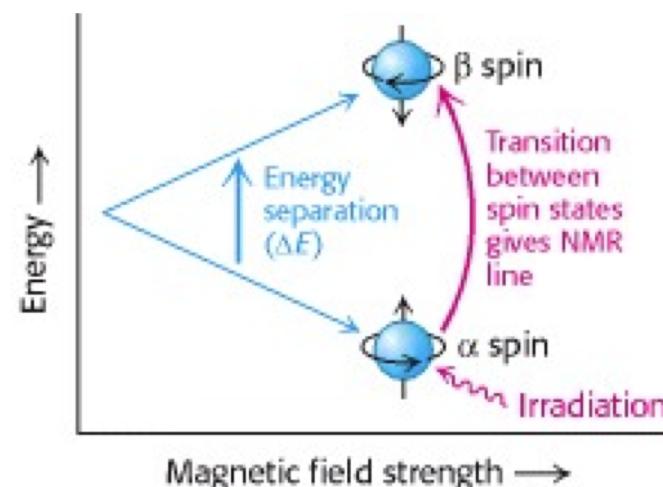
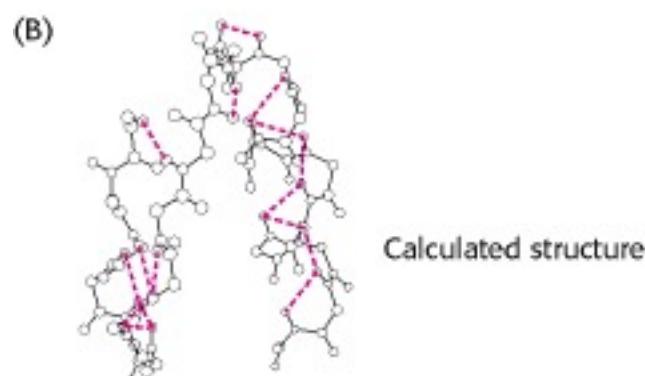
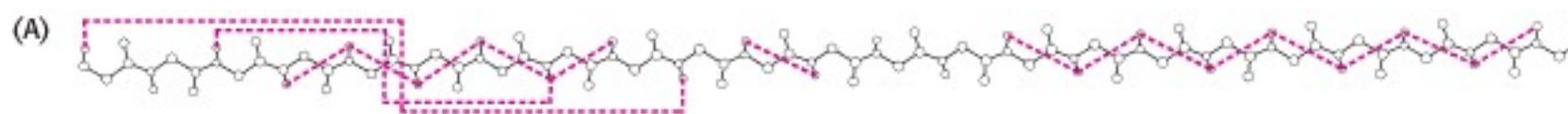
3D Structures of Macromolecules

How are they determined?

Nuclear magnetic resonance (NMR) spectroscopy:

In solution, small proteins < 15 kD.

When extra magnetic field is applied, atomic spin changes. The nuclear Overhauser effect (NOE) identifies pairs of protons that are in close proximity

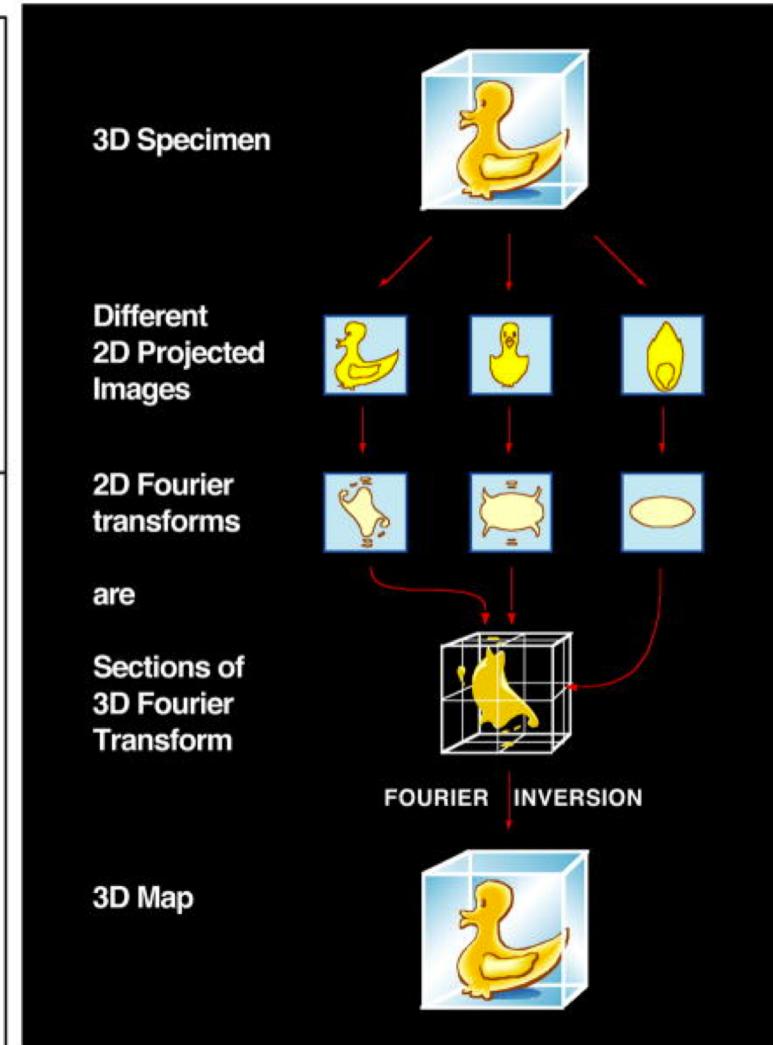
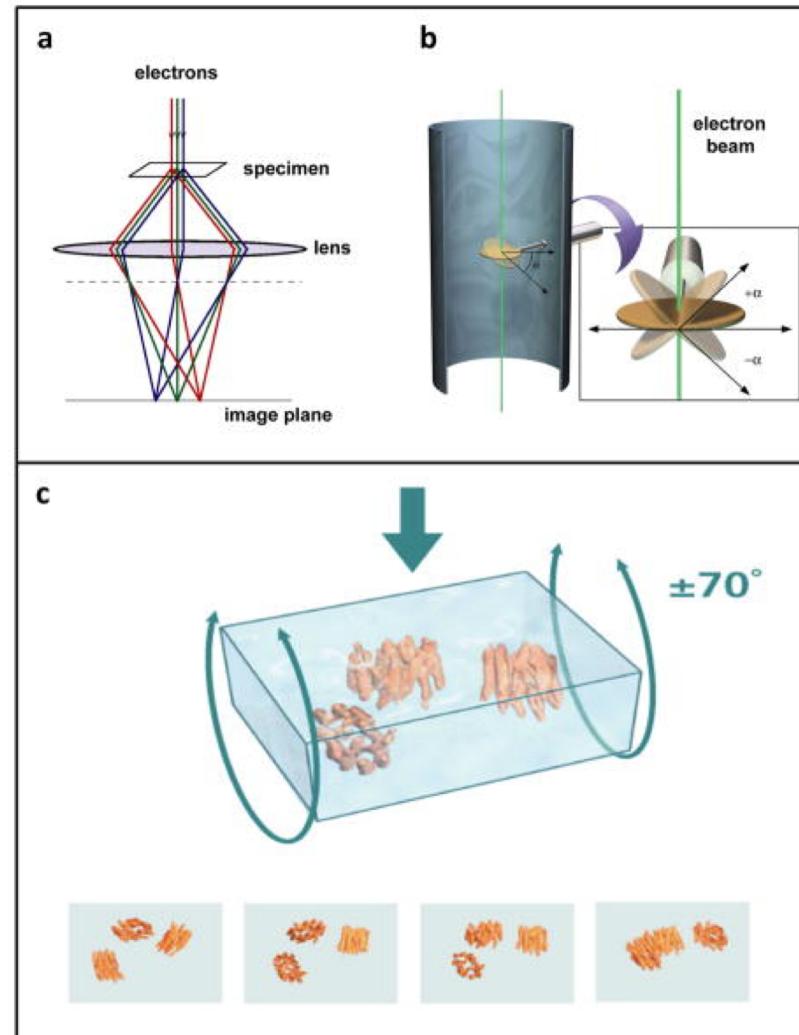


3D Structures of Macromolecules

How are they determined?

Cryo-electron microscopy (Cryo-EM):

Does NOT require crystallization.
Works in solution for larger proteins than NMR can do.
Image radiation-sensitive specimens in a transmission electron microscope under cryogenic conditions.



3D Structures of Macromolecules

Public deposit: Protein Data Bank (PDB) <https://www.rcsb.org/>

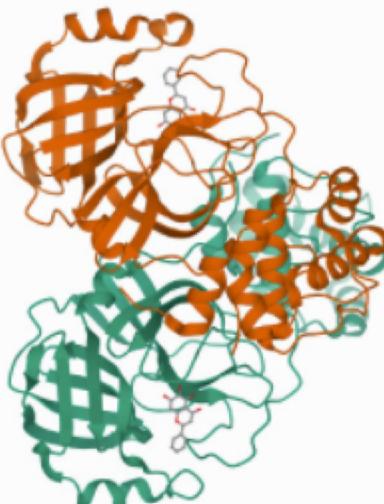
The screenshot shows the top navigation bar of the RCSB PDB website. It features the RCSB PDB logo, a search bar with placeholder text "Enter search term(s)", and a magnifying glass icon. Below the search bar are links for "Advanced Search" and "Browse Annotations". To the left of the search bar, there is text about the database containing 167,132 biological macromolecular structures. The right side of the header displays a 3D molecular structure.

The screenshot shows the main content area of the RCSB PDB website. On the left, a sidebar lists navigation options: Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area features a section titled "A Structural View of Biology" which describes the archive of 3D shapes of proteins, nucleic acids, and complex assemblies. Below this is a banner for "COVID-19 CORONAVIRUS Resources" featuring a 3D model of a coronavirus. To the right, a section titled "August Molecule of the Month" shows a 3D molecular structure of the "Phytosulfokine Receptor".

3D Structures of Macromolecules

Public deposit: Protein Data Bank (PDB) <https://www.rcsb.org/>

Biological Assembly 1  



PDB ID 6M2N 

Display Files  Download Files 

SARS-CoV-2 3CL protease (3CL pro) in complex with a novel inhibitor

DOI: [10.2211/pdb6M2N/pdb](https://doi.org/10.2211/pdb6M2N/pdb)

Classification: **VIRAL PROTEIN**

Organism(s): Severe acute respiratory syndrome coronavirus 2

Expression System: Escherichia coli BL21(DE3)

Mutation(s): No 

Deposited: 2020-02-28 Released: 2020-04-15

Deposition Author(s): Su, H.X., Zhao, W.F., Li, M.J., Xie, H., Xu, Y.C.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.20 Å 

R-Value Free: 0.254
R-Value Work: 0.225
R-Value Observed: 0.227

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.253
Clashscore		6
Ramachandran outliers		0.1%
Sidechain outliers		2.3%
RSRZ outliers		6.9%


3D View: Structure | Electron Density |
Ligand Interaction

Global Symmetry: Cyclic - C2  (3D View)
Global Stoichiometry: Homo 2-mer - A2 



6m2n.pdb ▾

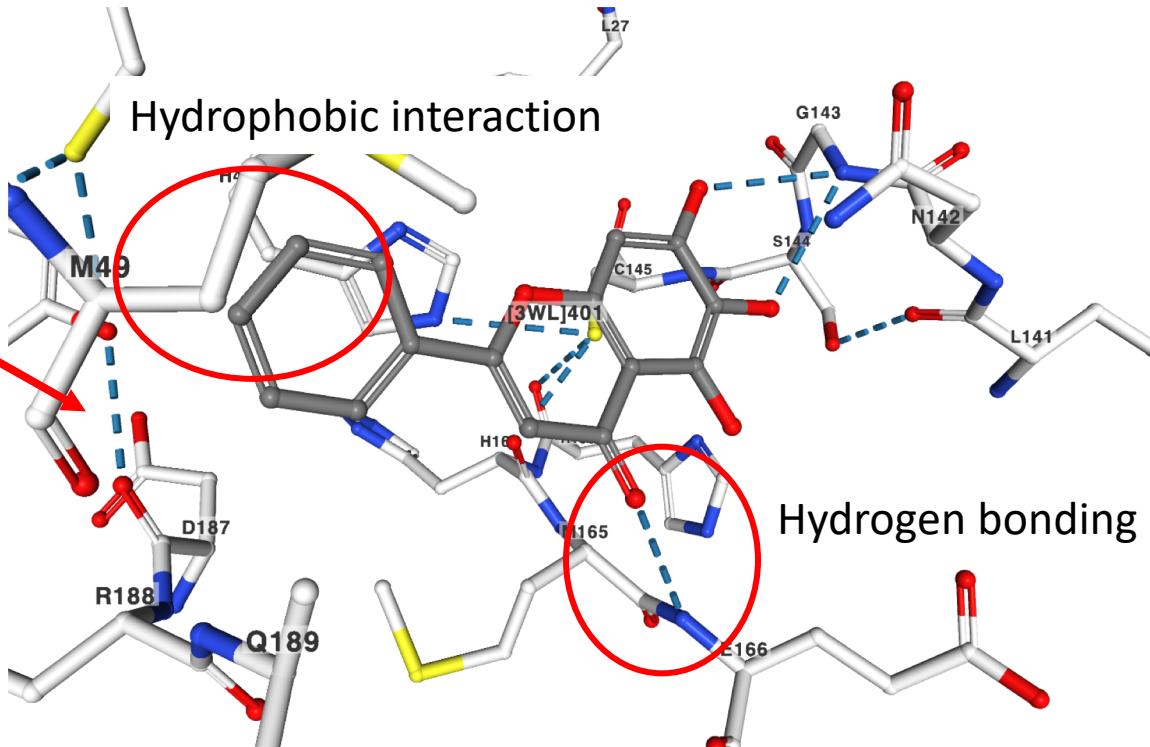
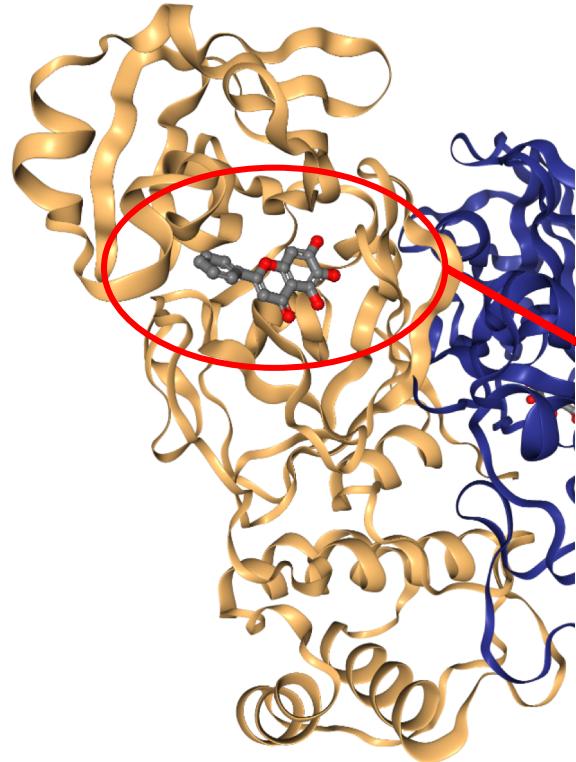
HEADER VIRAL PROTEIN 28-FEB-20 6M2N
TITLE SARS-COV-2 3CL PROTEASE (3CL PRO) IN COMPLEX WITH A NOVEL INHIBITOR
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: SARS-COV-2 3CL PROTEASE;
COMPND 3 CHAIN: A, B, C, D;
COMPND 4 ENGINEERED: YES
SOURCE MOL_ID: 1;
SOURCE 2 ORGANISM_SCIENTIFIC: SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS
SOURCE 3 2;
SOURCE 4 ORGANISM_TAXID: 2697049;
SOURCE 5 EXPRESSION_SYSTEM: ESCHERICHIA COLI BL21(DE3);
SOURCE 6 EXPRESSION_SYSTEM_TAXID: 469008;
SOURCE 7 EXPRESSION_SYSTEM_STRAIN: BL21(DE3)
KEYWDS SARS-COV-2, 3CL PRO, VIRAL PROTEIN
EXPDTA X-RAY DIFFRACTION
AUTHOR H.X.SU,W.F.ZHAO,M.J.LI,H.XIE,Y.C.XU



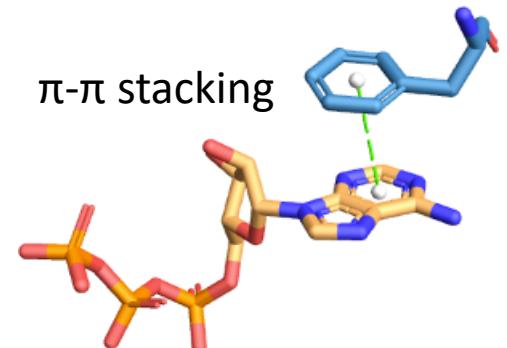
6m2n.pdb

ATOM	1	N	SER	A	1	-34.855	-24.069	48.481	1.00	39.91	N
ATOM	2	CA	SER	A	1	-33.448	-24.317	48.189	1.00	41.89	C
ATOM	3	C	SER	A	1	-33.060	-25.734	48.595	1.00	42.86	C
ATOM	4	O	SER	A	1	-33.830	-26.434	49.252	1.00	38.59	O
ATOM	5	CB	SER	A	1	-32.561	-23.293	48.901	1.00	44.70	C
ATOM	6	OG	SER	A	1	-31.192	-23.520	48.621	1.00	54.15	O
ATOM	7	N	GLY	A	2	-31.859	-26.144	48.208	1.00	43.10	N
ATOM	8	CA	GLY	A	2	-31.430	-27.515	48.372	1.00	36.69	C
ATOM	9	C	GLY	A	2	-31.807	-28.375	47.180	1.00	39.45	C
ATOM	10	O	GLY	A	2	-32.666	-28.033	46.369	1.00	34.88	O
ATOM	11	N	PHE	A	3	-31.134	-29.515	47.072	1.00	37.20	N
ATOM	12	CA	PHE	A	3	-31.438	-30.464	46.006	1.00	31.30	C

Protein-Ligand Binding



Red: oxygen
Blue: nitrogen
Yellow: sulfur



Non-covalent molecular interactions:

Hydrogen bond, Water bridge, $\pi-\pi$ stacking, Hydrophobic interaction, Ionic interaction...

https://en.wikipedia.org/wiki/Non-covalent_interaction

Molecular Docking

Determination of a protein-ligand binding mode could be time-consuming and labor intensive.

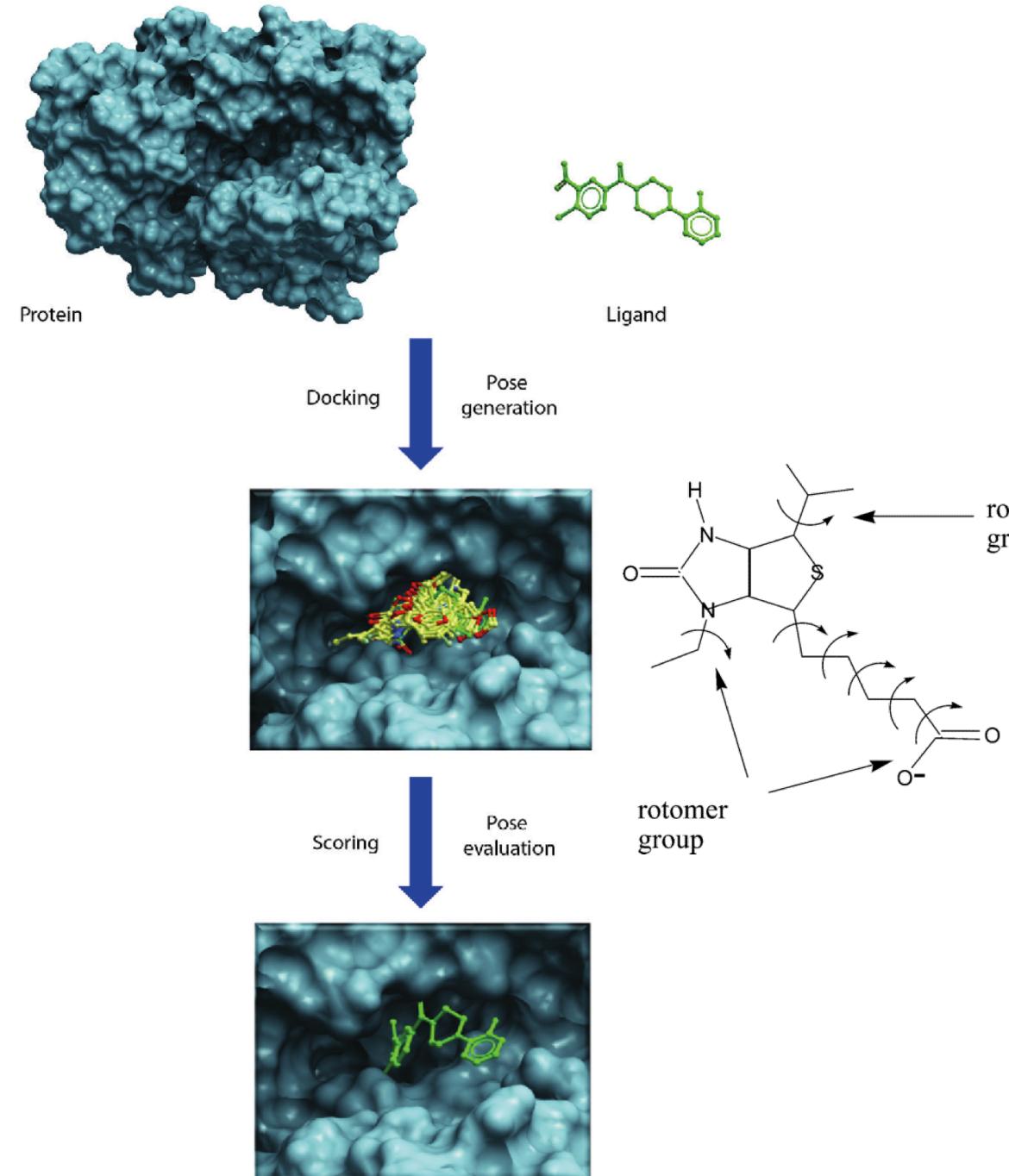
Molecular docking predicts the preferred orientation of a small molecule forming a stable complex with a protein and their binding affinity.

Application:

1. Explain the structural-activity relationship (SAR) to provide guidance for medicinal chemical modifications.
2. Conduct large-scale virtual screening to discover novel active compounds.

Steps:

1. Specify the binding pocket on the protein (receptor)
2. Sample the orientation (pose) of the small molecule (ligand)
3. Score each pose
4. Output the best fit



Scoring Function

Example: Glide score v2.5

Empirically trained from >200 pdb samples with known binding free energies.

$$\Delta G_{\text{bind}} = C_{\text{lipo-lipo}} \sum f(r_{lr}) +$$
$$C_{\text{hbond-neut-neut}} \sum g(\Delta r) h(\Delta \alpha) +$$
$$g_1(\Delta r) = \begin{cases} 1 & \text{if } \Delta r \leq 0.25 \text{ \AA} \\ 1 - (\Delta r - 0.25)/0.4 & \text{if } 0.25 \text{ \AA} < \Delta r \leq 0.65 \text{ \AA} \\ 0 & \text{if } \Delta r > 0.65 \text{ \AA} \end{cases}$$
$$C_{\text{hbond-neut-charged}} \sum g(\Delta r) h(\Delta \alpha) +$$
$$C_{\text{hbond-charged-charged}} \sum g(\Delta r) h(\Delta \alpha) +$$
$$C_{\text{max-metal-ion}} \sum f(r_{lm}) + C_{\text{rotb}} H_{\text{rotb}} +$$
$$C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} +$$
$$C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms}$$

Software Programs for Docking

https://en.wikipedia.org/wiki/List_of_protein-ligand_docking_software

Free softwares



AutoDock



Gold

Commercial suites

SCHRÖDINGER.
Glide



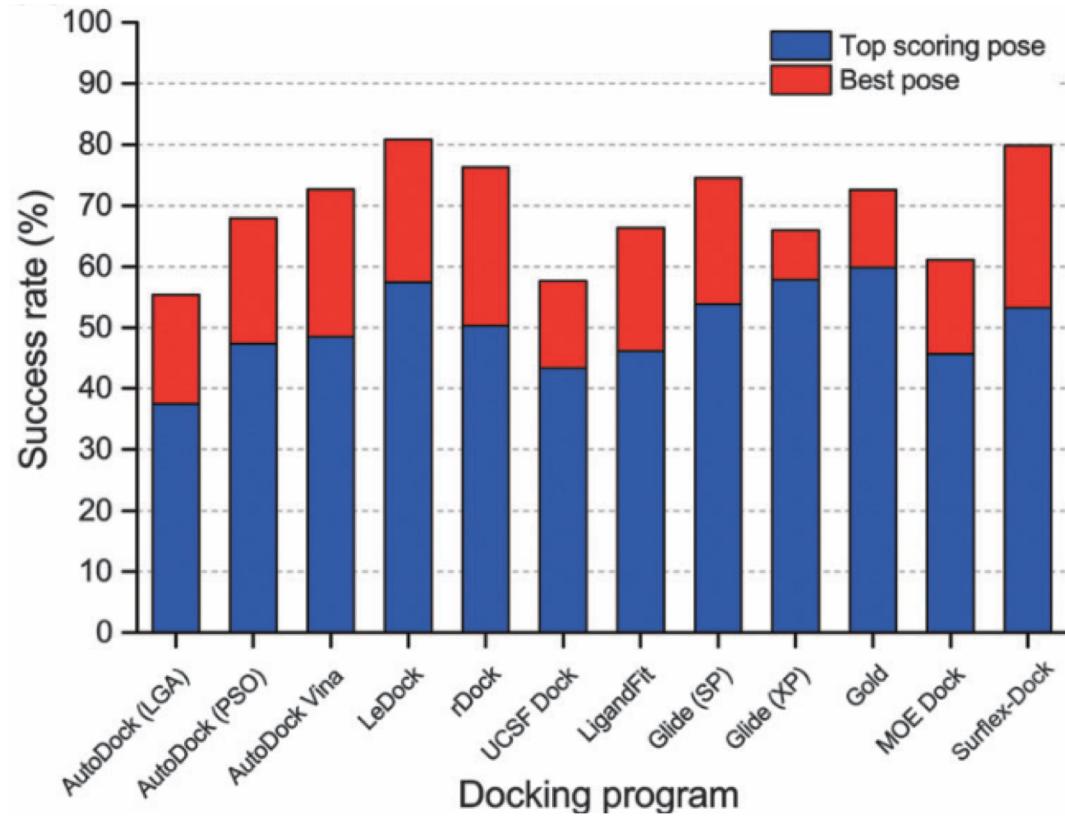
Free web services



SwissDock



Performance Comparison



Suggestion: consensus scoring

Docking program	Correlation coefficient	Top scored pose	Best pose
AutoDock (LGA)	r_p^a	0.433 ± 0.009^c	0.404 ± 0.009
	r_s^b	0.477 ± 0.008	0.450 ± 0.009
AutoDock (PSO)	r_p	0.492 ± 0.008	0.466 ± 0.008
	r_s	0.534 ± 0.007	0.513 ± 0.008
AutoDock Vina	r_p	0.564 ± 0.008	0.569 ± 0.008
	r_s	0.580 ± 0.008	0.584 ± 0.008
LeDock	r_p	0.442 ± 0.009	0.463 ± 0.009
	r_s	0.462 ± 0.010	0.486 ± 0.009
rDock	r_p	-0.015 ± 0.011	-0.021 ± 0.011
	r_s	-0.017 ± 0.011	-0.005 ± 0.011
UCSF DOCK	r_p	0.291 ± 0.010	0.276 ± 0.011
	r_s	0.331 ± 0.011	0.323 ± 0.011
LigandFit	r_p	-0.132 ± 0.011	-0.105 ± 0.011
	r_s	-0.221 ± 0.012	-0.192 ± 0.012
Glide (SP)	r_p	0.444 ± 0.008	0.402 ± 0.009
	r_s	0.473 ± 0.009	0.419 ± 0.010
Glide (XP)	r_p	0.367 ± 0.010	0.356 ± 0.010
	r_s	0.389 ± 0.010	0.374 ± 0.010
GOLD	r_p	-0.500 ± 0.008	-0.494 ± 0.008
	r_s	-0.515 ± 0.008	-0.513 ± 0.008
MOE Dock	r_p	0.564 ± 0.008	0.411 ± 0.009
	r_s	0.589 ± 0.009	0.457 ± 0.009
Surflex-Dock	r_p	-0.340 ± 0.009	-0.350 ± 0.009
	r_s	-0.370 ± 0.009	-0.382 ± 0.009

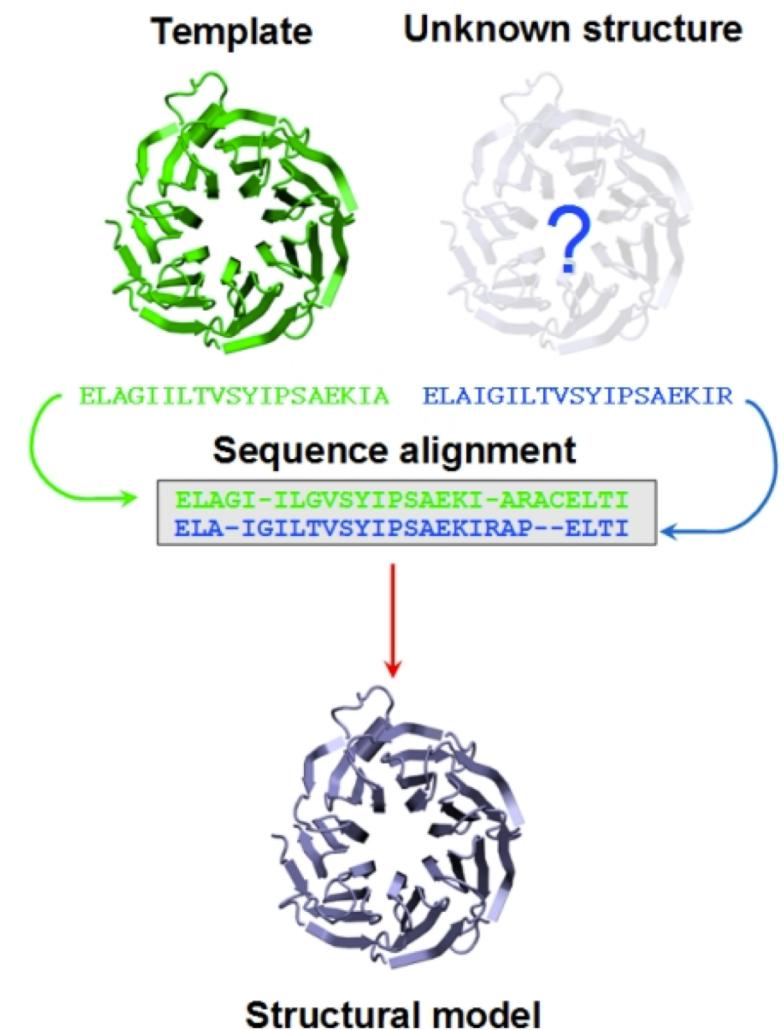
Homology Modeling

Protein 3D structure not available

Homology modeling. Sequence homology > 30% PDB samples.

Programs: MODELLER, I-TASSER, SWISS-MODEL...

The screenshot shows the SWISS-MODEL web interface. At the top, there is a logo for SIB BIOZENTRUM University of Basel, The Center for Molecular Life Sciences. Below the logo, the text "SWISS-MODEL" is displayed. A navigation bar with tabs "Modelling" (which is selected), "Repository", "Tools", and "Doc" follows. The main area is titled "Start a New Modelling Project". It contains fields for "Target Sequence(s)" (with a note about the format: FASTA, Clustal, plain string, or valid UniProtKB AC) and "Project Title" (set to "Untitled Project"). There is also an "Email" field (set to "Optional"). Below these fields are two buttons: "Search For Templates" and "Build Model". To the right of the "Target Sequence(s)" field is a text input placeholder: "Paste your target sequence(s) or UniProtKB AC here". Below this input are two buttons: "+ Upload Target Sequence File..." and "Validate".

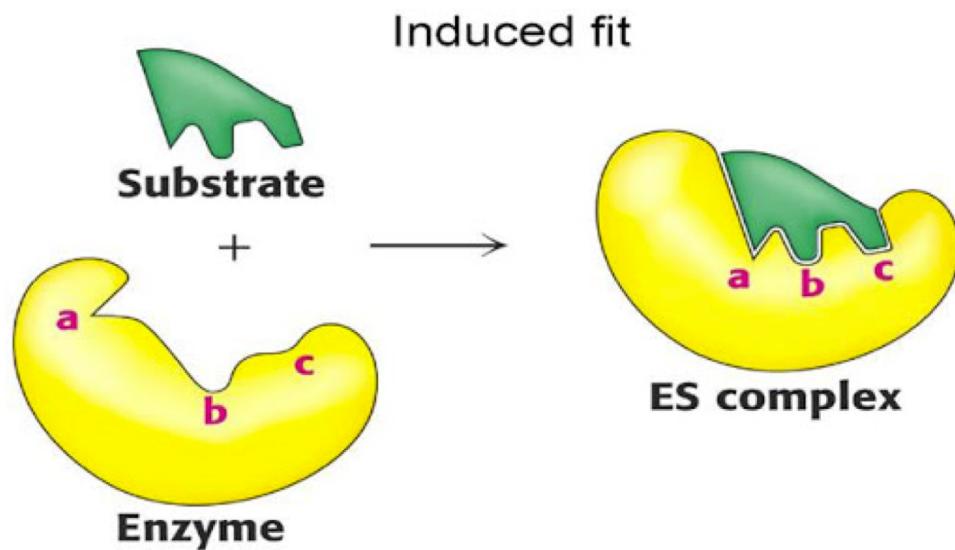


Induced-Fit Docking

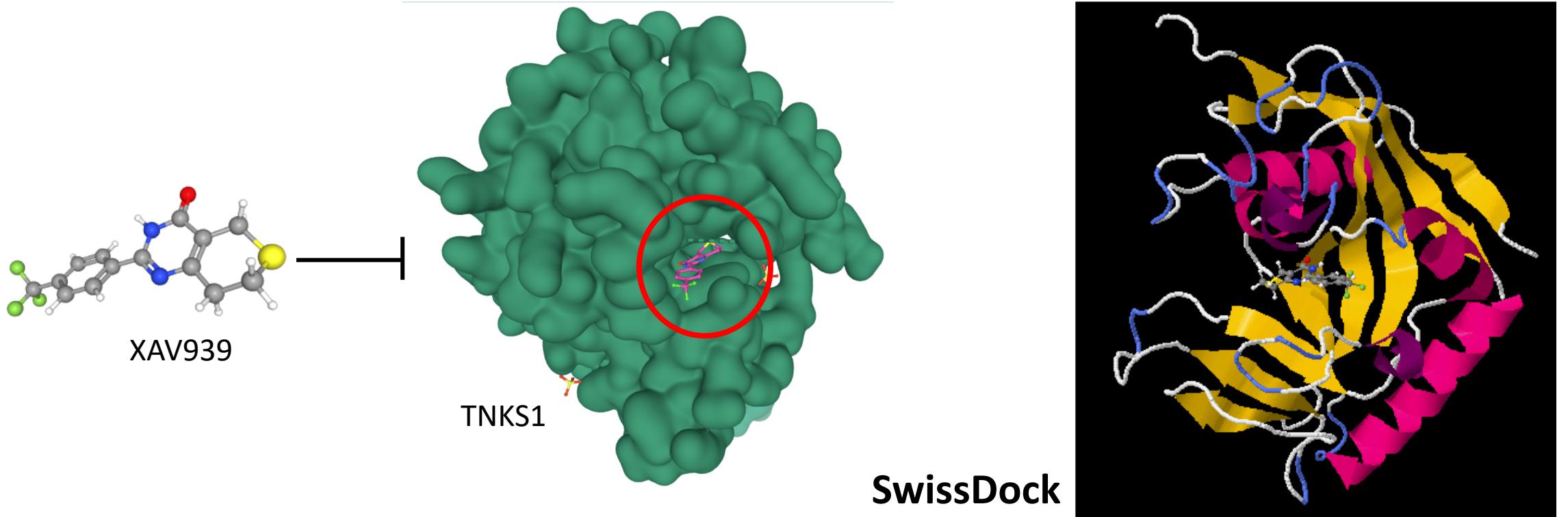
Can't obtain the right binding pose: try induced-fit docking (or flexible docking).

Generally, docking programs treat the receptor as rigid and ignore the induced-fit effect of receptor-ligand interactions. In fact, the orientation of receptor pocket amino acids change to allow better fit with the ligand.
~30 fold time-consuming of rigid docking.

Programs: DOCK, AutoDock Vina, Schrödinger's induced-fit docking...



Case: Tankyrase1 XAV-939 Docking



SwissDock

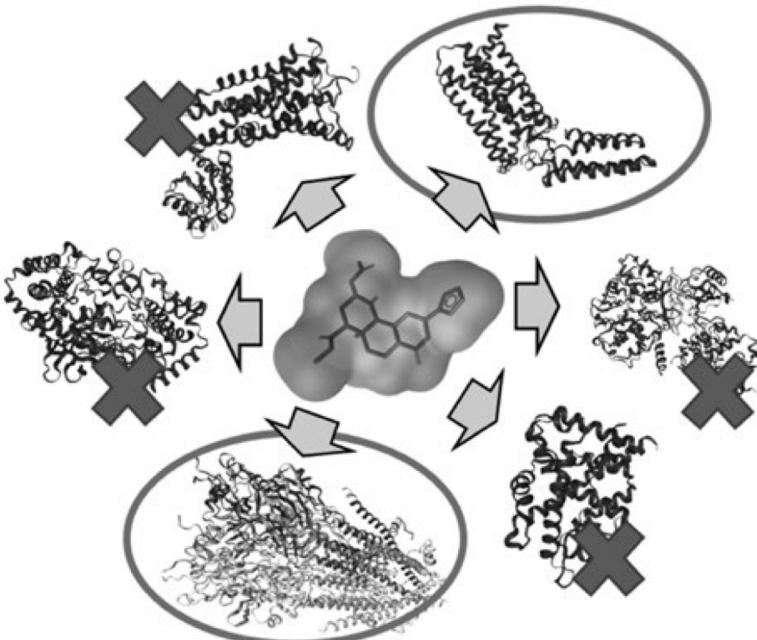


Swiss Institute of
Bioinformatics

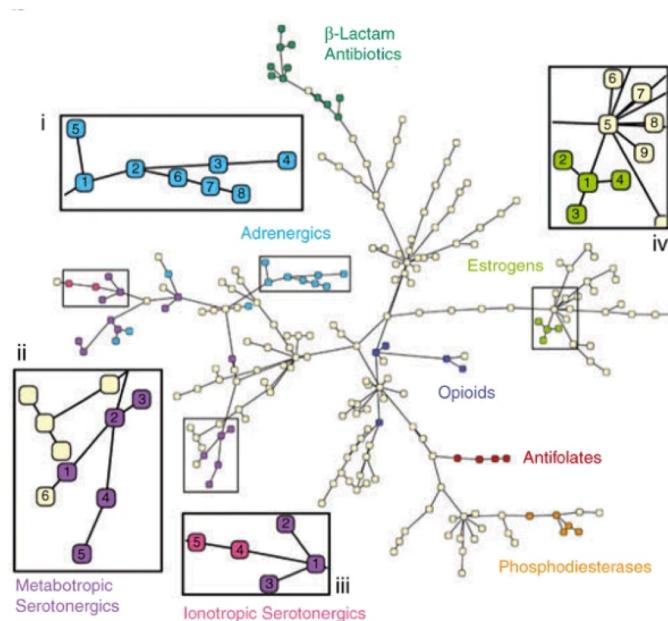
Show	Cluster	Element	FullFitness (kcal/mol)	Estimated ΔG (kcal/mol)
<input checked="" type="radio"/>	0	0	-1189.94	-9.09

Target Fishing

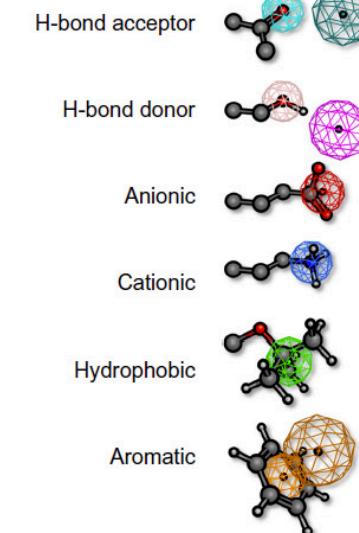
Given an active small molecule derived from phenotypic screening, find its binding target protein(s).



Reverse docking
Dock the query compound to each target with known 3D structure



Similarity ensemble approach
Compare similarities between query compound and known ligands of each target



Pharmacophore mapping
Pharmacophores extracted from either protein structures or ligands alignment

Programs for Target Fishing

Reverse
Docking



809 targets

Consensus Reverse Docking System
<http://pbil.kaist.ac.kr/CRDS/>
5254 targets

Ligand
Similarity

Similarity Ensemble Approach (SEA)
<http://sea.bkslab.org/>
~5000 targets; run < 1 minute

Pharmacophore
Mapping

PharmMapper
<http://www.lilab-ecust.cn/pharMapper/>
23236 proteins; run < 1 day

Case: Aspirin Target Fishing with SEA

Similarity ensemble approach (SEA)

The Similarity ensemble approach relates proteins based on the set-wise chemical similarity among their ligands. It can be used to rapidly search large compound databases and to build cross-target similarity maps.

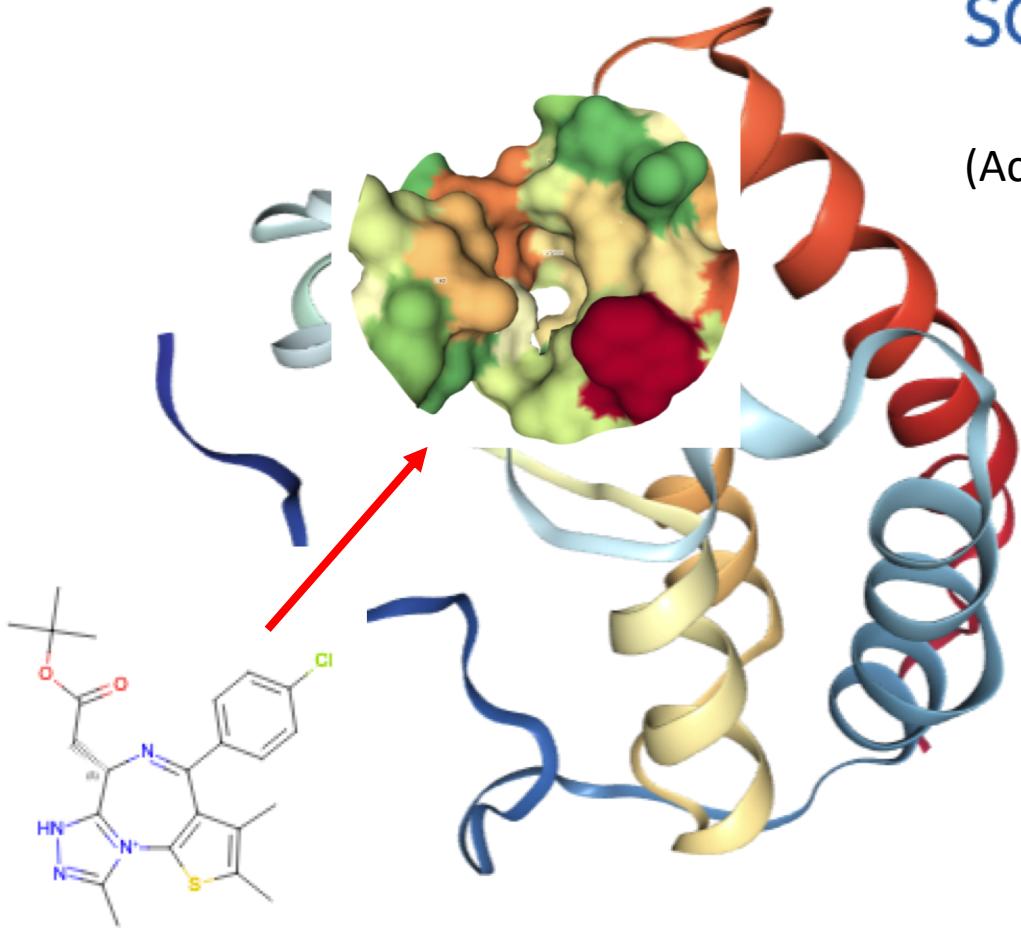
CC(=O)OC1=CC=CC=C1C(=O)O|

Try SEA

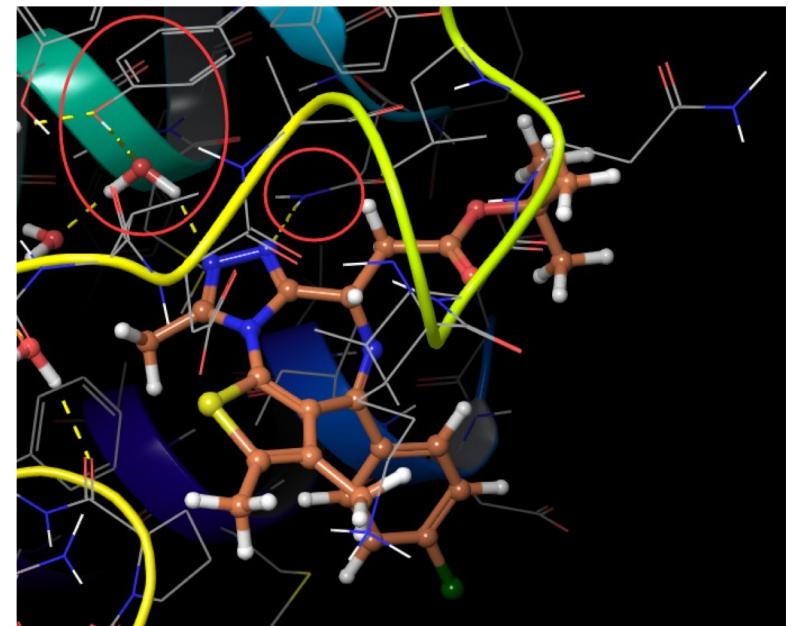
Query	Target Key	Target Name	Description	P-Value	MaxTC
 Aspirin	PGH1_BOVIN	PTGS1	Prostaglandin G/H synthase 1	1.652e-13	1.00
	PGH1_SHEEP	PTGS1	Prostaglandin G/H synthase 1	1.528e-08	1.00
	PGH2_SHEEP	PTGS2	Prostaglandin G/H synthase 2	0.0005039	1.00
	ITA2B_HUMAN	ITGA2B	Integrin alpha-IIb	0.1645	1.00
	ITB3_HUMAN	ITGB3	Integrin beta-3	0.4027	1.00
	AK1C2_HUMAN	AKR1C2	Aldo-keto reductase family 1 member C2	1.003e-79	0.47
	Q6TUJ4_PSEAI	blalMP-1	Metallo beta-lactamase	2.875e-57	0.41
	FABH_ENTFA	fabH	3-oxoacyl-[acyl-carrier-protein] synthase 3	1.55e-47	0.31

The End
Q & A

Case: BRD4-JQ1 Docking



SCHRÖDINGER®
Glide
(Accessible via HPCC@MSU)



M Project Table --- Scratch Project

Row	In	Stars	Title	DNUMBER	docking score	H_ACCEPTORS	I
17	—	★★★★★	1 - SP_brd4_testLigands_pvl [36]				
18	—	★★★★★	3mxif_prep				
			JQ1		-7.662		

A very detailed step-by-step tutorial:

https://drive.google.com/file/d/1358lwihwRyOc_3OOXg75_pgdWP3lhNKH/view?usp=sharing