NYU SCCPC

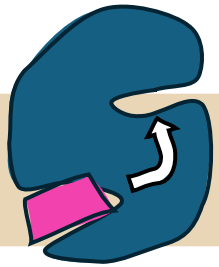
Tuning DiffDock for allosteric structure-based drug design to the kinome

Eric A. Chen

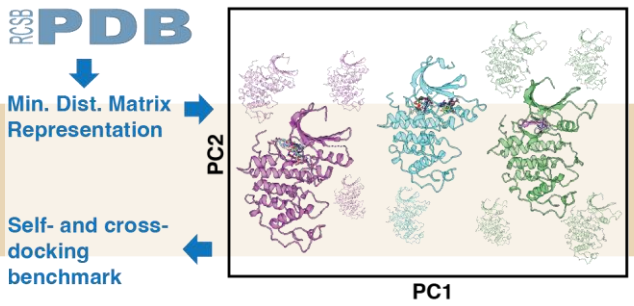
New York University – Yingkai Zhang

Gordon Research Seminar: 7/13/25

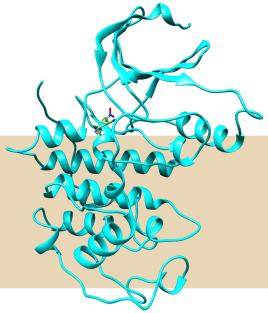
Application of Next-Gen Strategies to Address Challenging Drug Discovery Problems

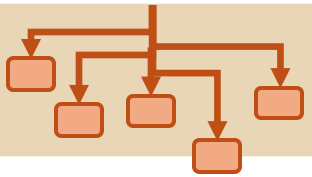


Yang, C., **Chen, E. A.** & Zhang, Y. *Molecules* **27**, 4568 (2022).
Xia, S., **Chen, E. A.** & Zhang, Y. *J. Chem. Theory Comput.* **19**, 7478–7495 (2023).



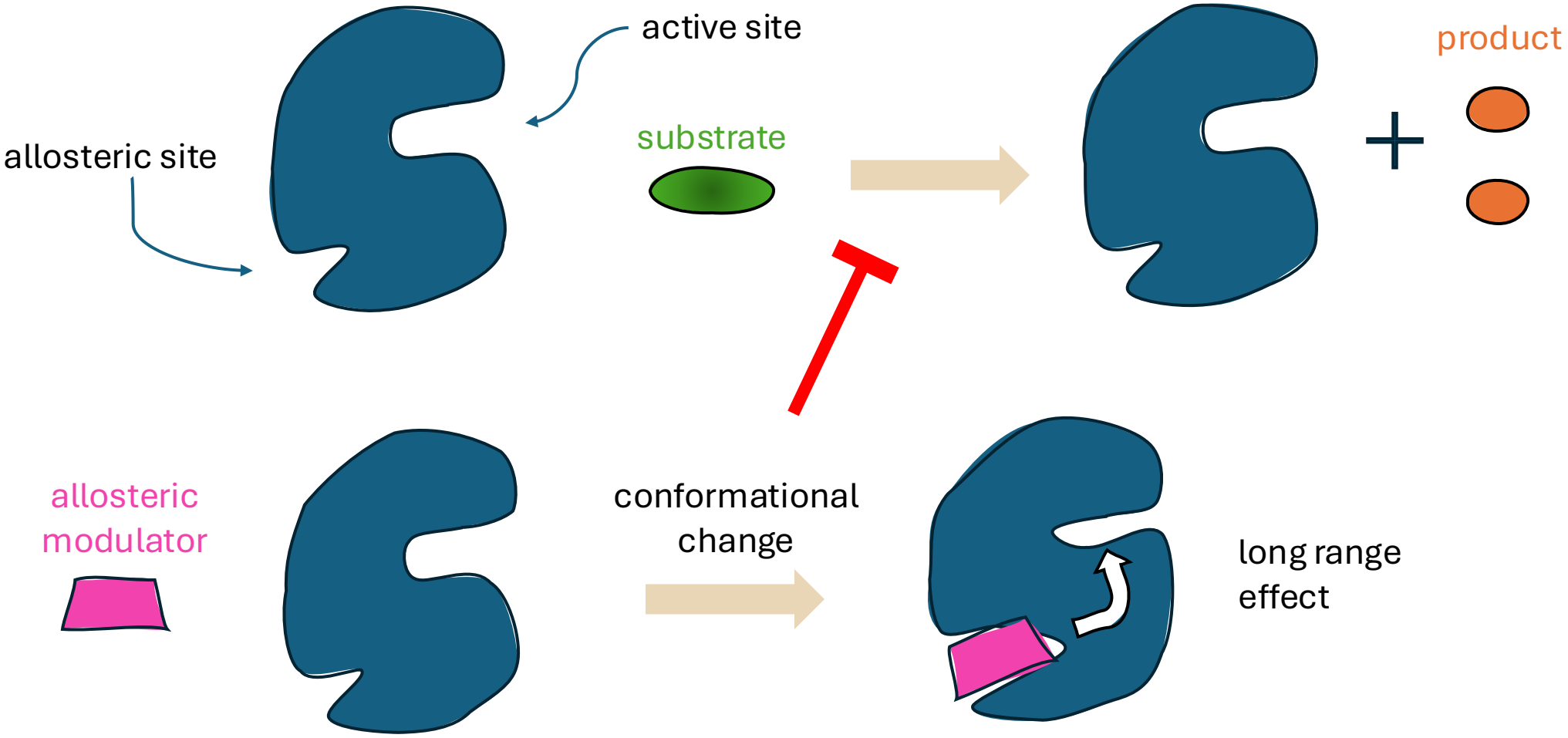
Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model.* **65**, 3737–3748 (2025).





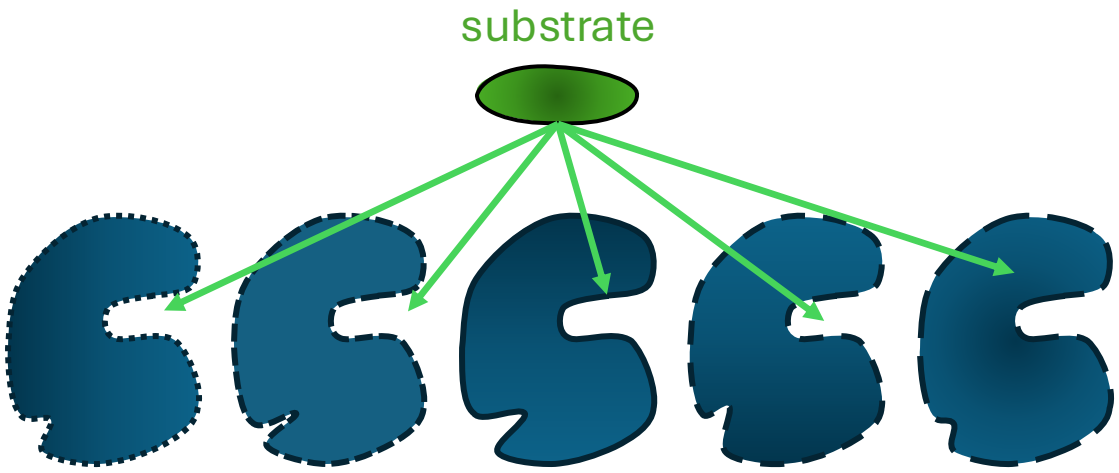
Chen, E.A., Green J., Zhang. Y. In prep

What is allosteric modulation?



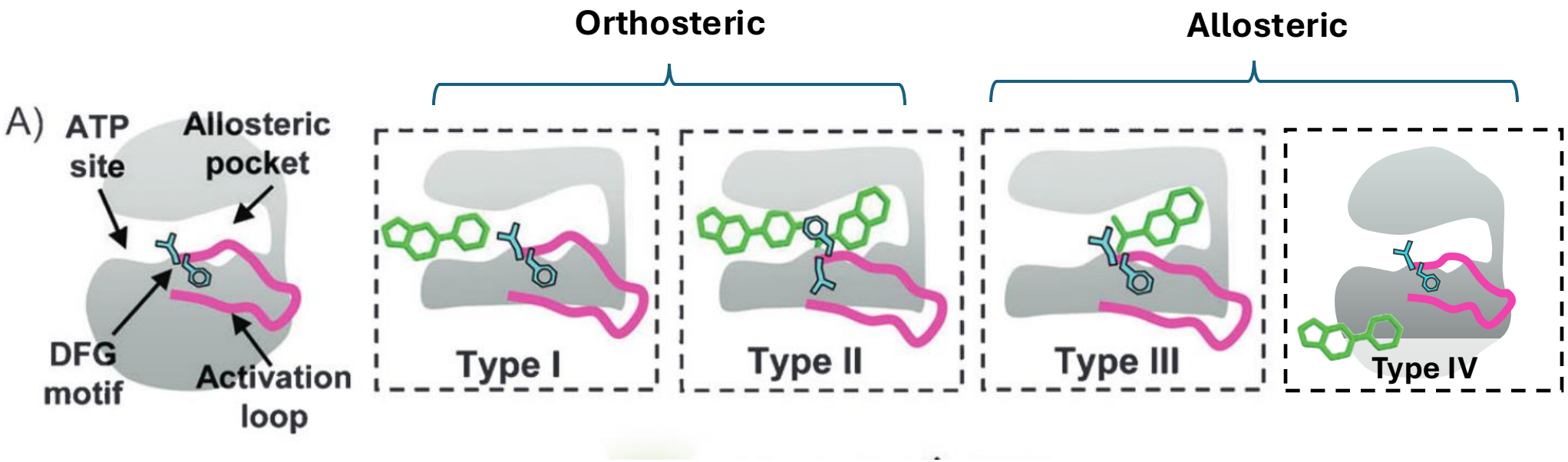
Allosteric modulators are promising therapeutics

- Increased specificity and reduced toxicity
  - Allosteric sites are more diverse
- Non/un-competitive inhibition
  - High concentrations of native substrate
- Saturable
- Can lead to cooperative regulation at multiple sites



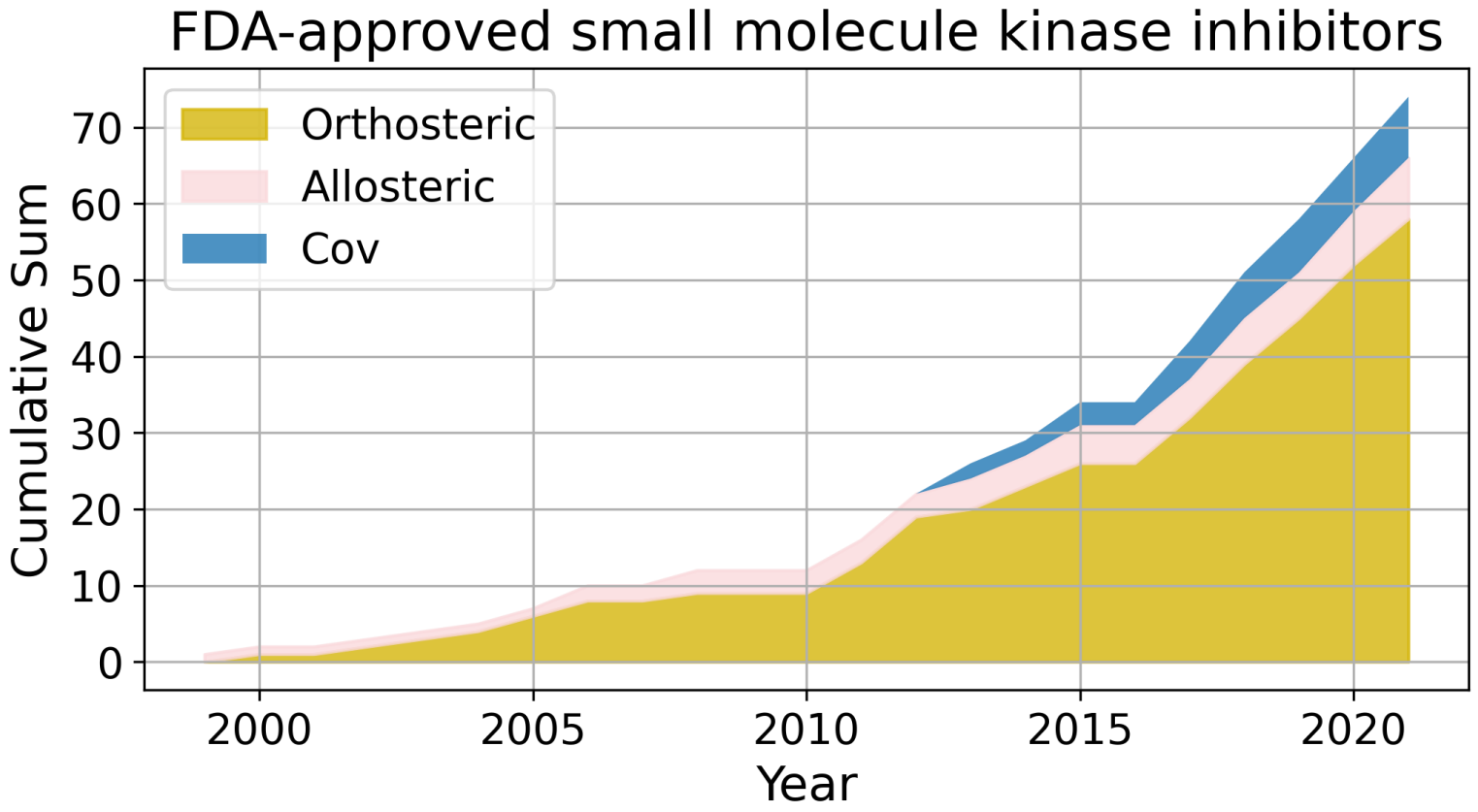
Taylor, S. S., Kornev, A. P. *Trends Biochem. Sci.* **2011**, *36* (2), 65–77.
Lu, S., He, X., Ni, D., Zhang, J. *J. Med. Chem.* **2019**, *62* (14), 6405–6421.

Kinase ligand binding types are well-defined



Martin, M. P. *et al*. *ChemBioChem* **13**, 2128–2136 (2012).

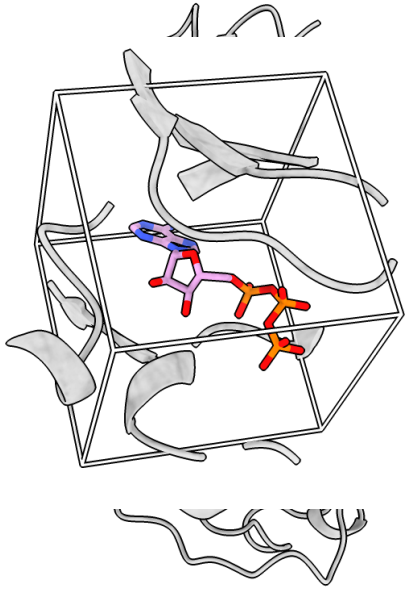
Orthosteric inhibitors make up the majority of FDA-approved kinase drugs



Pan, Y. & Mader, M. M., *J. Med. Chem.* **65**, 5288–5299 (2022).

Main challenges of allosteric SBDD

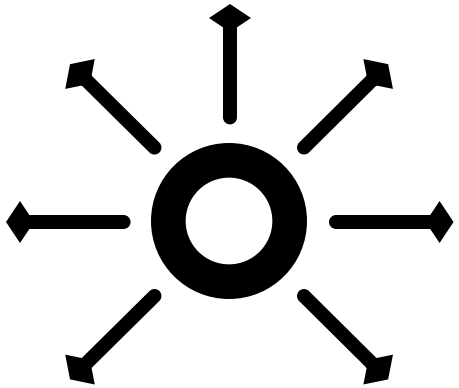
Limited to finding compounds at known sites



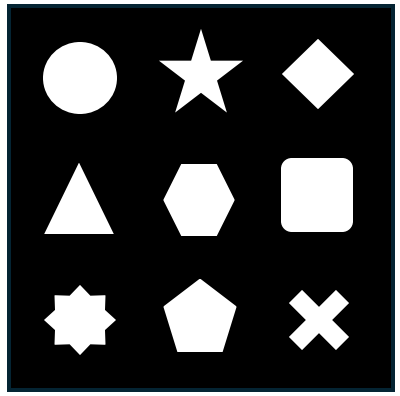
Selecting one structure may not be sufficient

Yang, C., **Chen, E. A.** & Zhang, Y. *Molecules* **27**, 4568 (2022).
Xia, S., **Chen, E. A.** & Zhang, Y. *J. Chem. Theory Comput.* **19**, 7478–7495 (2023).

Allosteric ligand design essentials

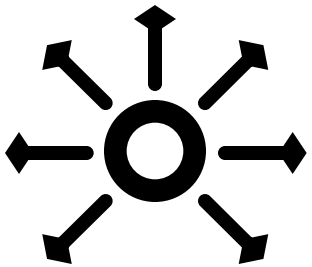


Expand beyond orthosteric site



Protein–ligand conformational heterogeneity

Can blind docking improve our chances at hit finding allosteric ligands?



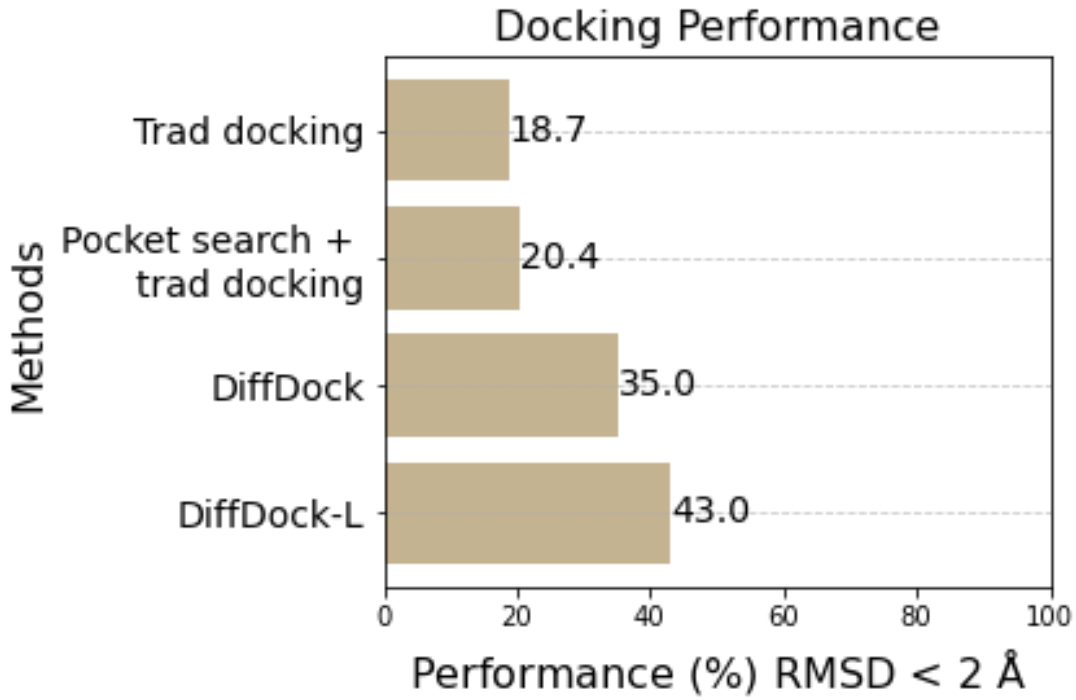
Docking

- Accurately predict binding pose of orthosteric and allosteric

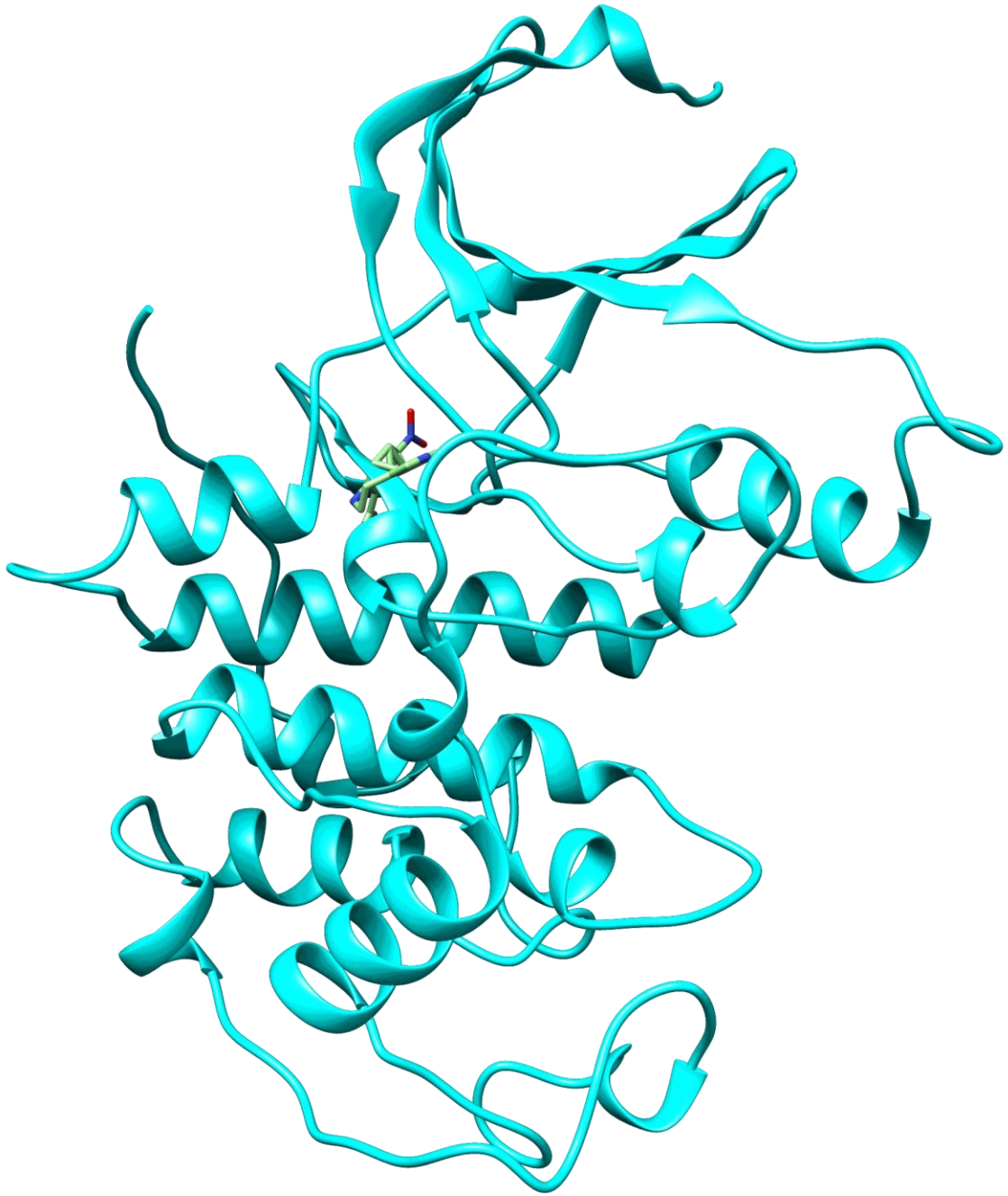
Screening

- Filter away orthosteric binders
  - Reduce the rate of false-positives

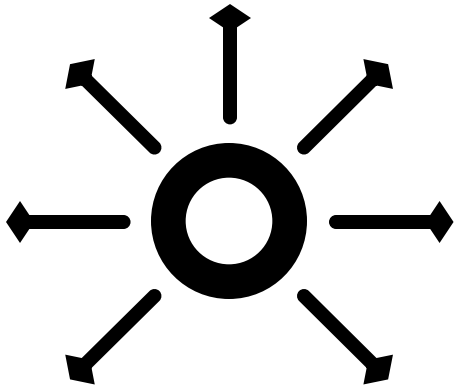
DiffDock reports state-of-the-art blind docking performance on PDBBind benchmark



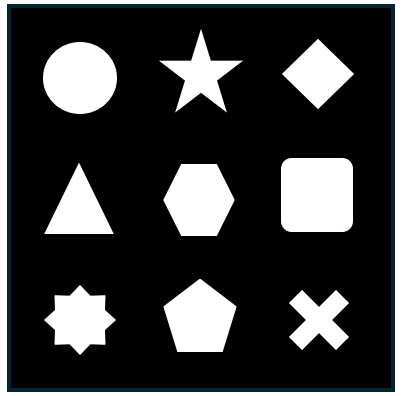
Corso, G., *et al.* arXiv:2210.01776 (2022)
Corso, G., *et al.* arXiv:2402.18396 (2024)



Allosteric ligand design essentials

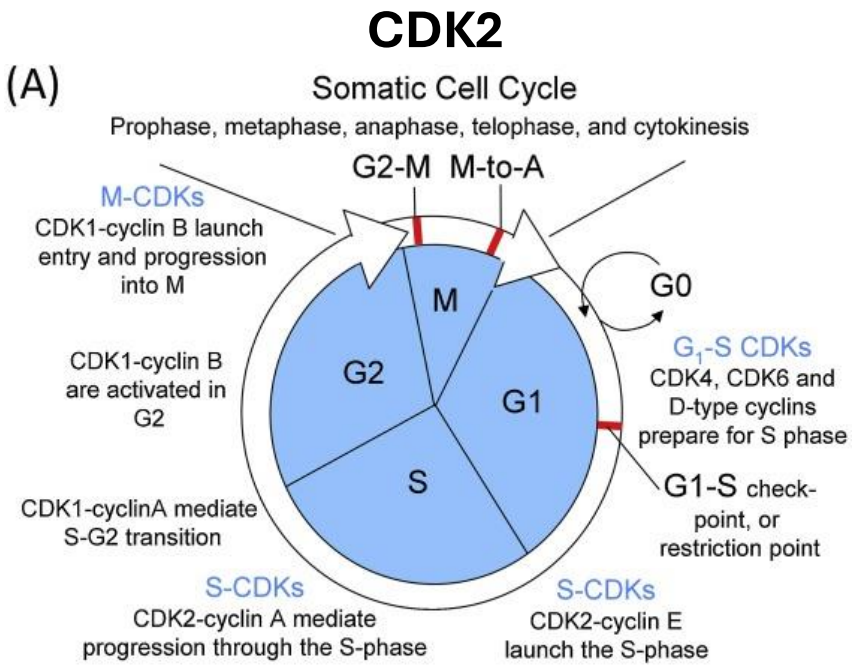


Expand beyond orthosteric site



Protein–ligand conformational heterogeneity

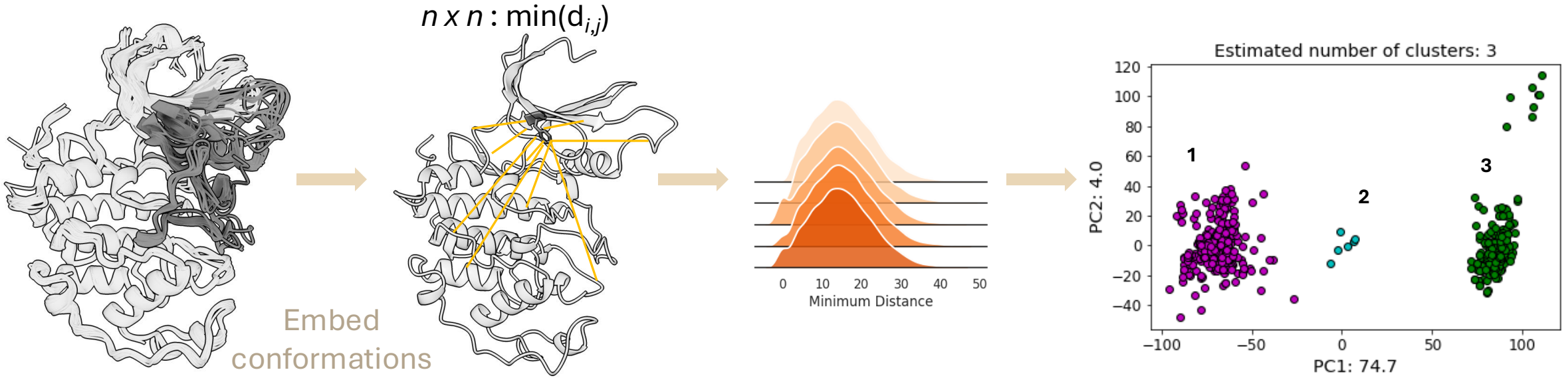
CDK2 is a rich oncogenic/contraceptive target



Can we develop a framework for analyzing CDK2 conformational heterogeneity?

Roskoski, R. *Pharmacol. Res.* **2019**, *139*, 471–488.
Faber, E. B., *et al. Biol. Reprod.* **2020**, *103* (2), 357–367.

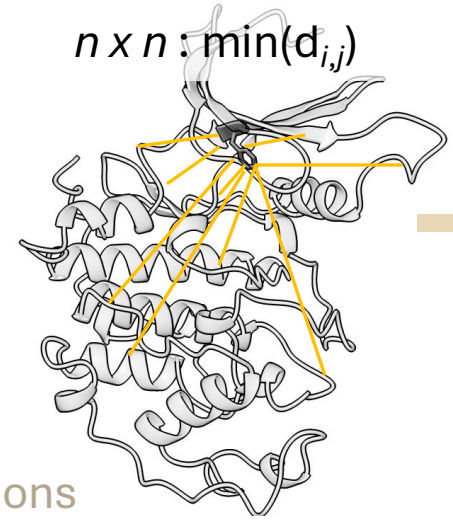
Minimum Distance Matrix Representation (MDMR) framework for conformational heterogeneity

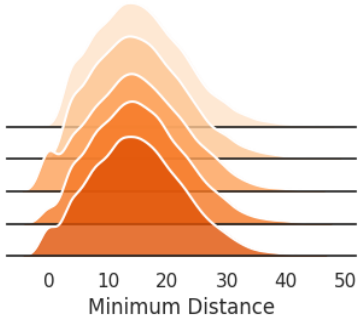


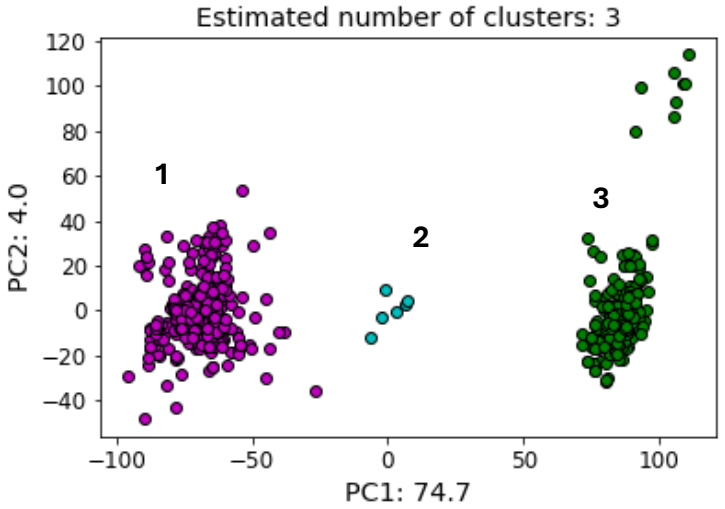
Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model*. **65**, 3737–3748 (2025).

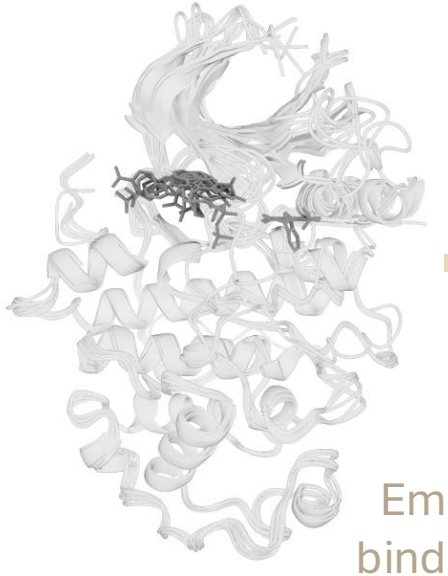
MDMR distinguishes ligand conformations

Embed conformations

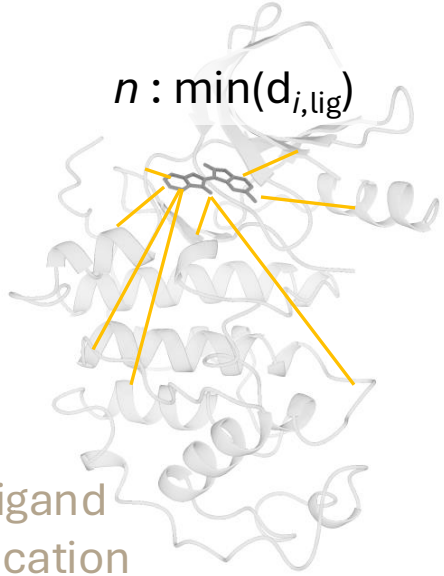


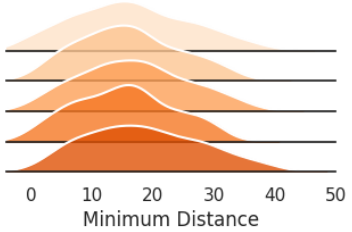


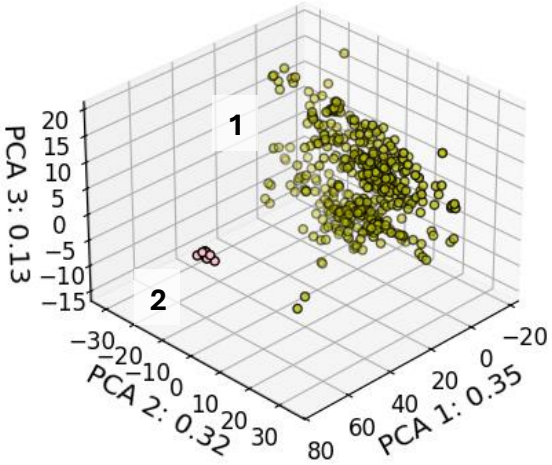




Embed ligand binding location

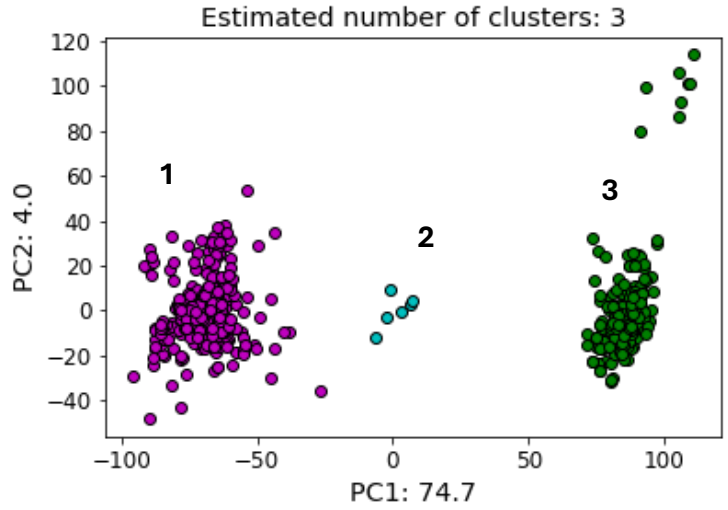


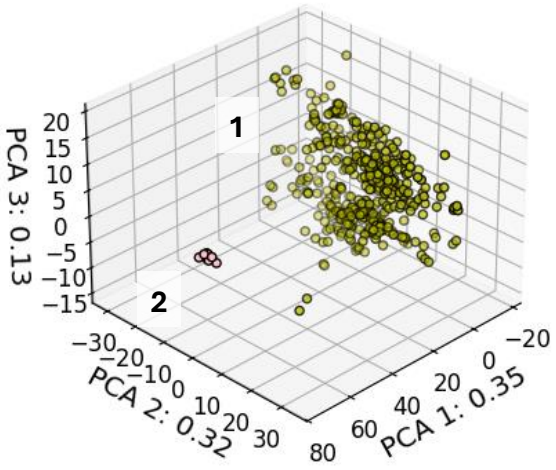


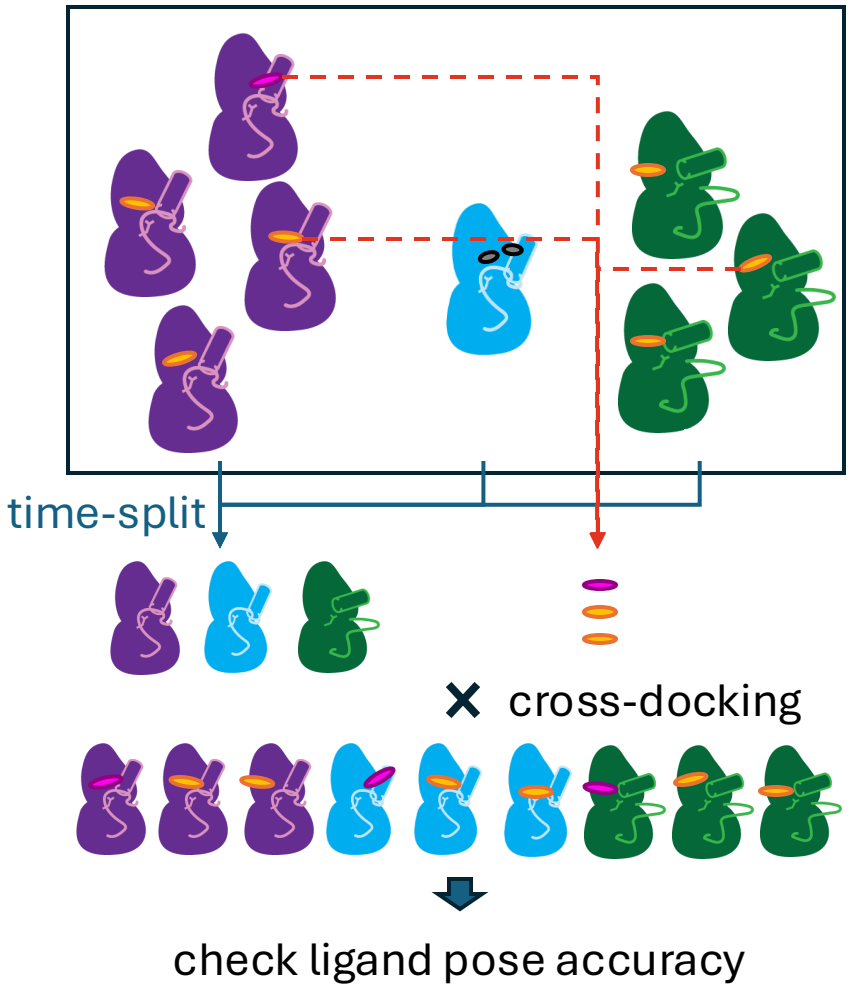


Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model*. **65**, 3737–3748 (2025).

MDMR builds cross-docking benchmark



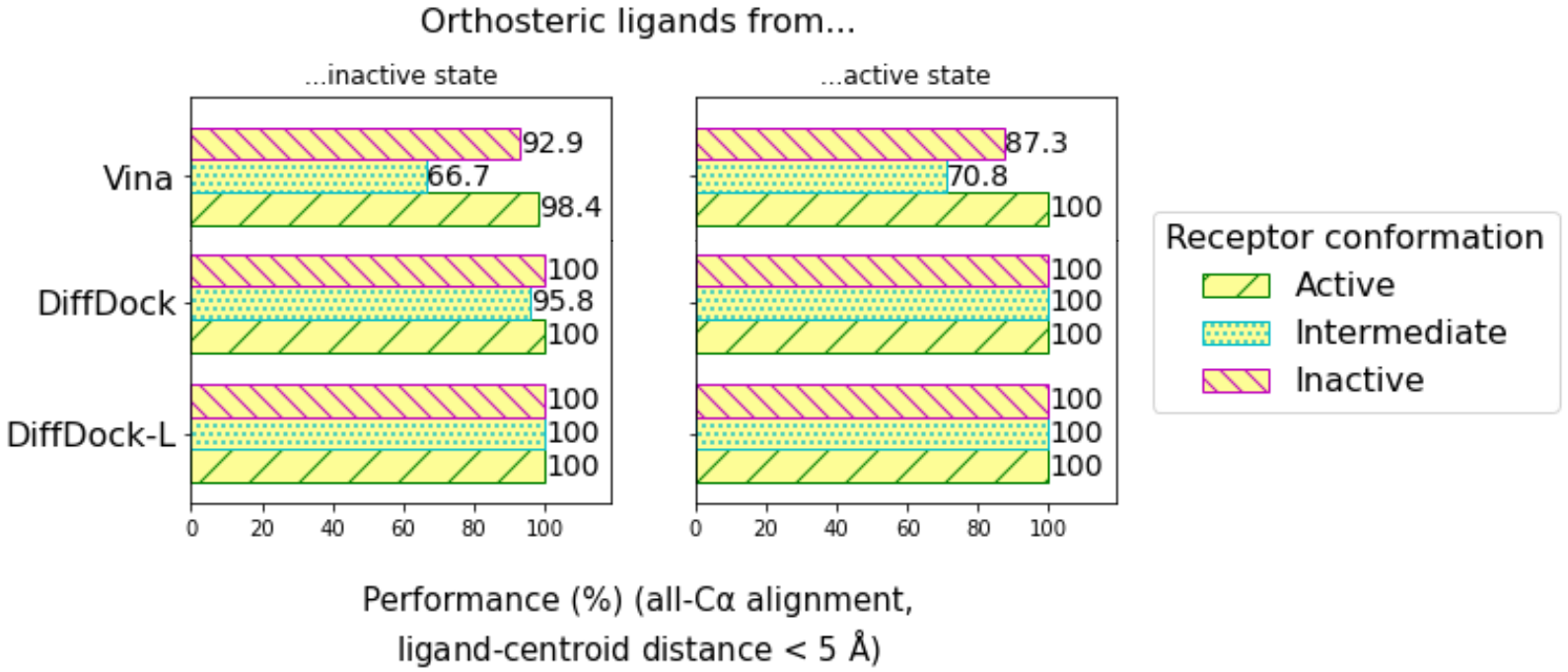


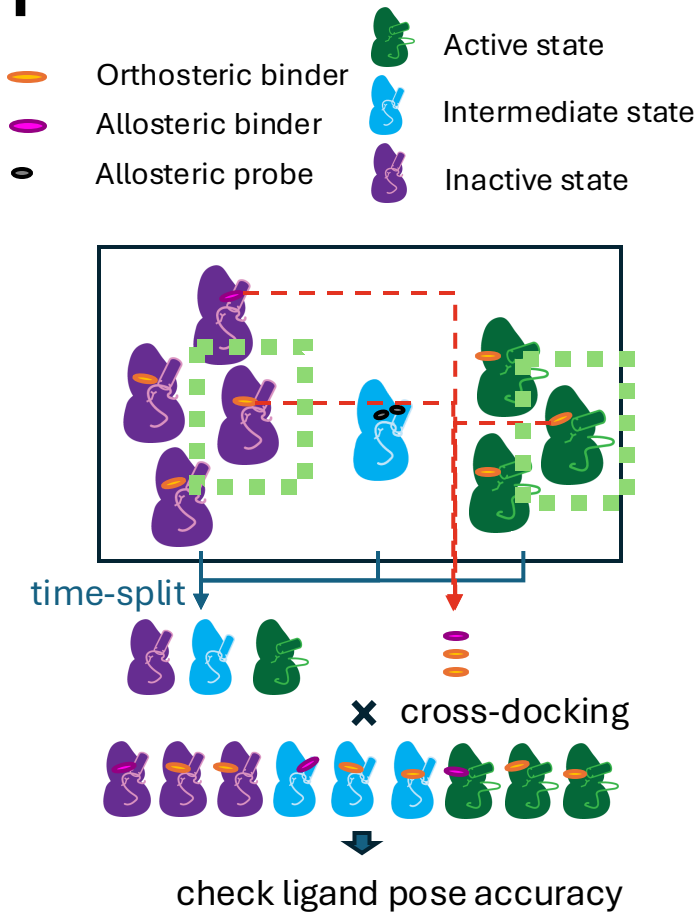


- Trad
- DL
- DL + trad

Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model.* **65**, 3737–3748 (2025).

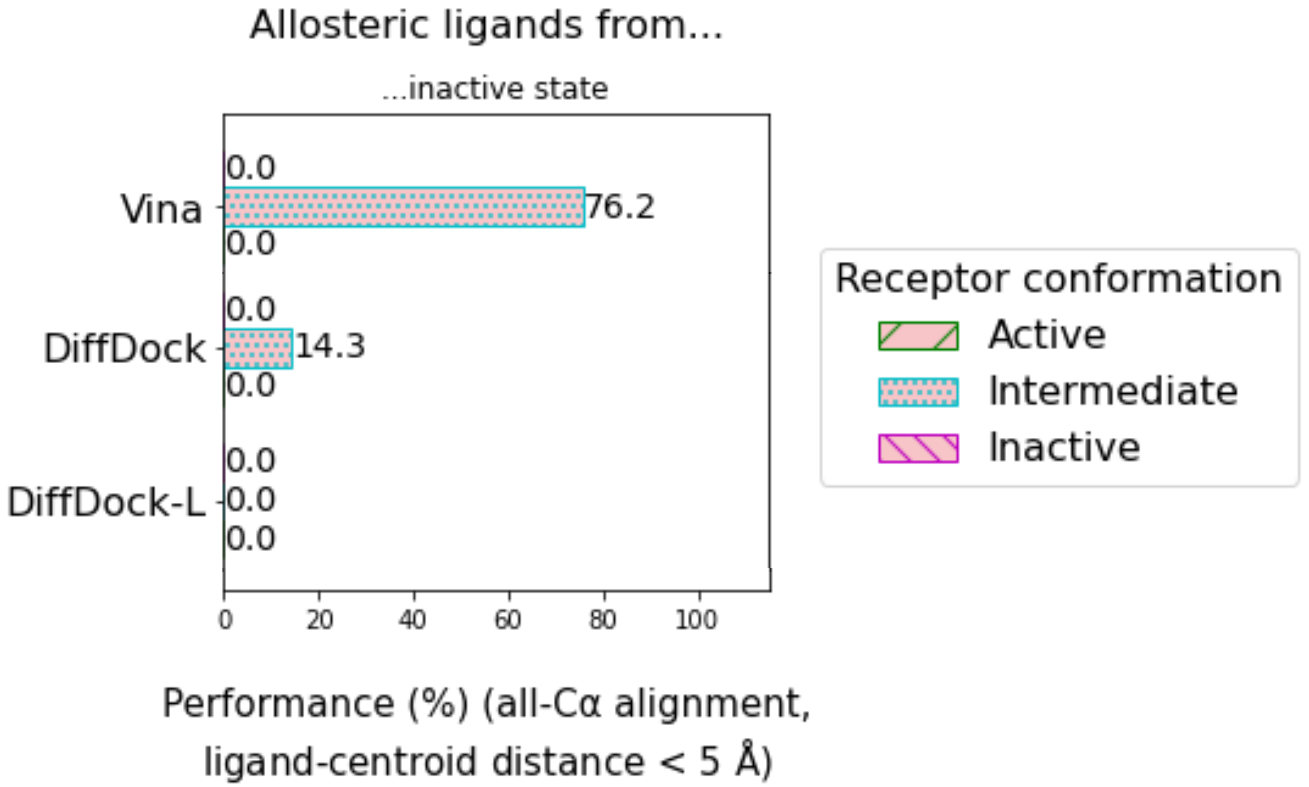
Orthosteric docking is independent of receptor conformation

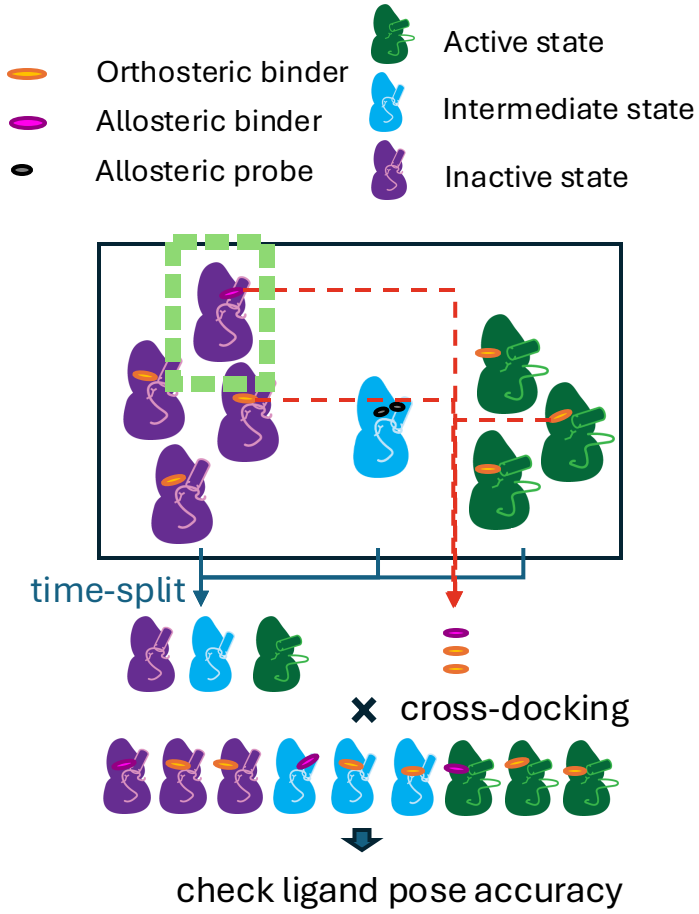




Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model*. **65**, 3737–3748 (2025).

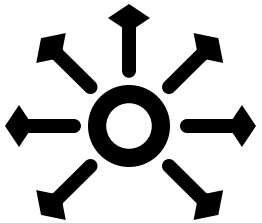
Allosteric docking depends on receptor conformation





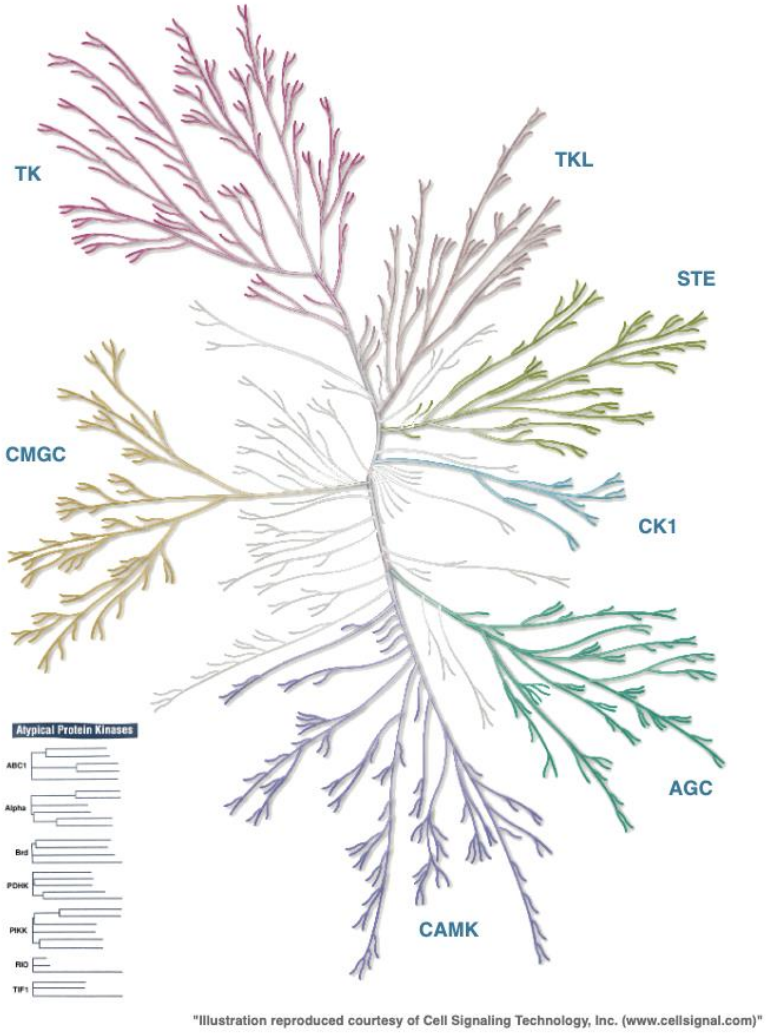
Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model*. **65**, 3737–3748 (2025).

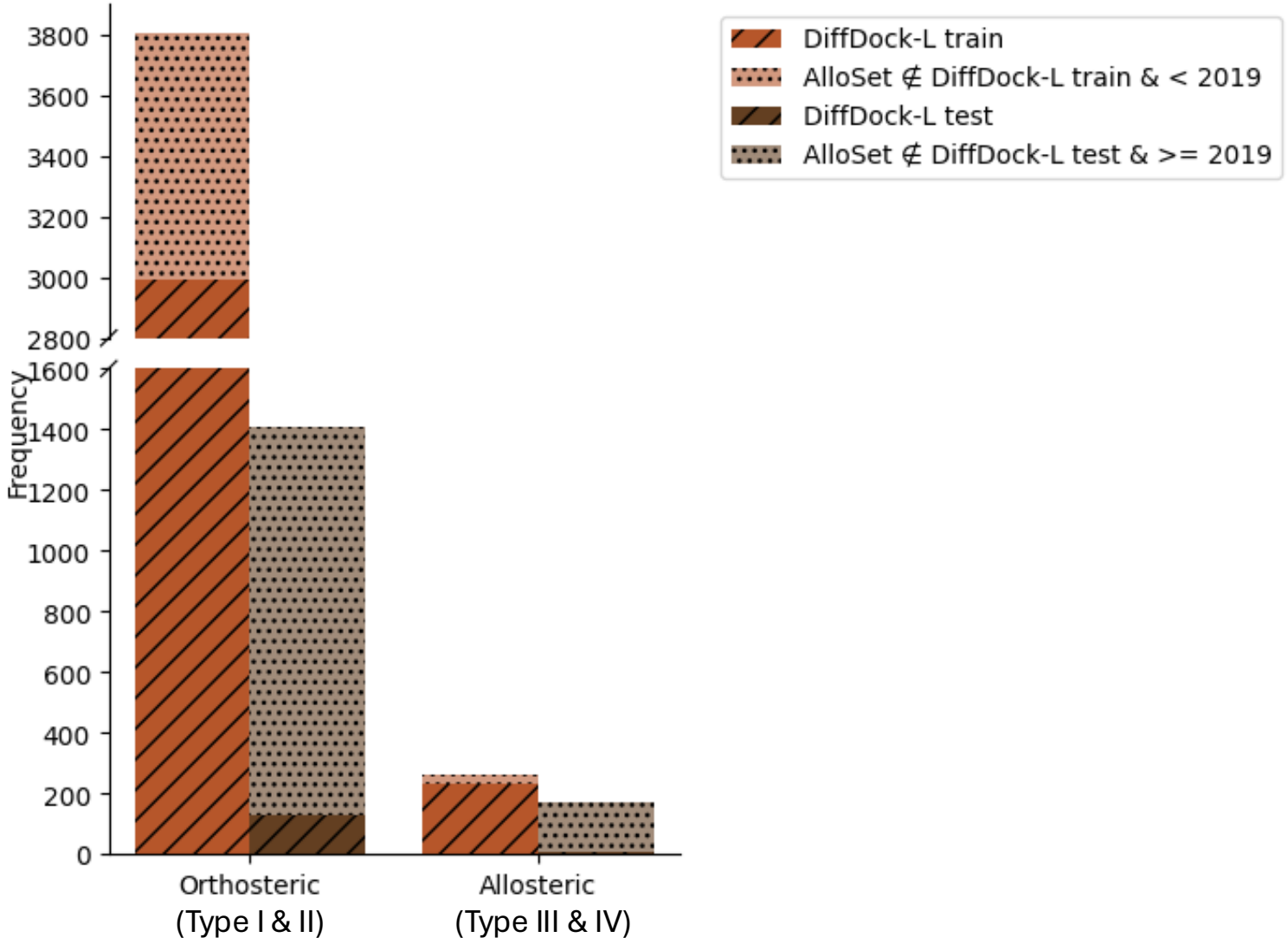
Improve state-of-the-art methods on allosteric kinase ligands



- Extend blind docking to allosteric sites
- Integrate within SBDD pipeline

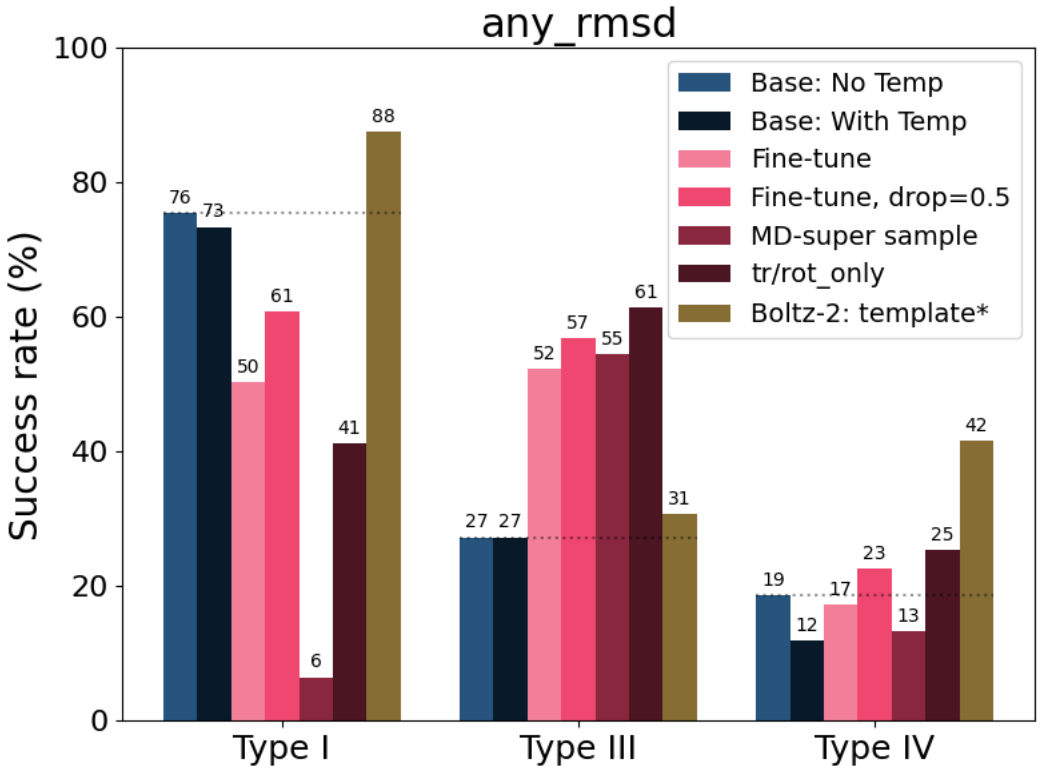
Curating a kinome-wide structural dataset

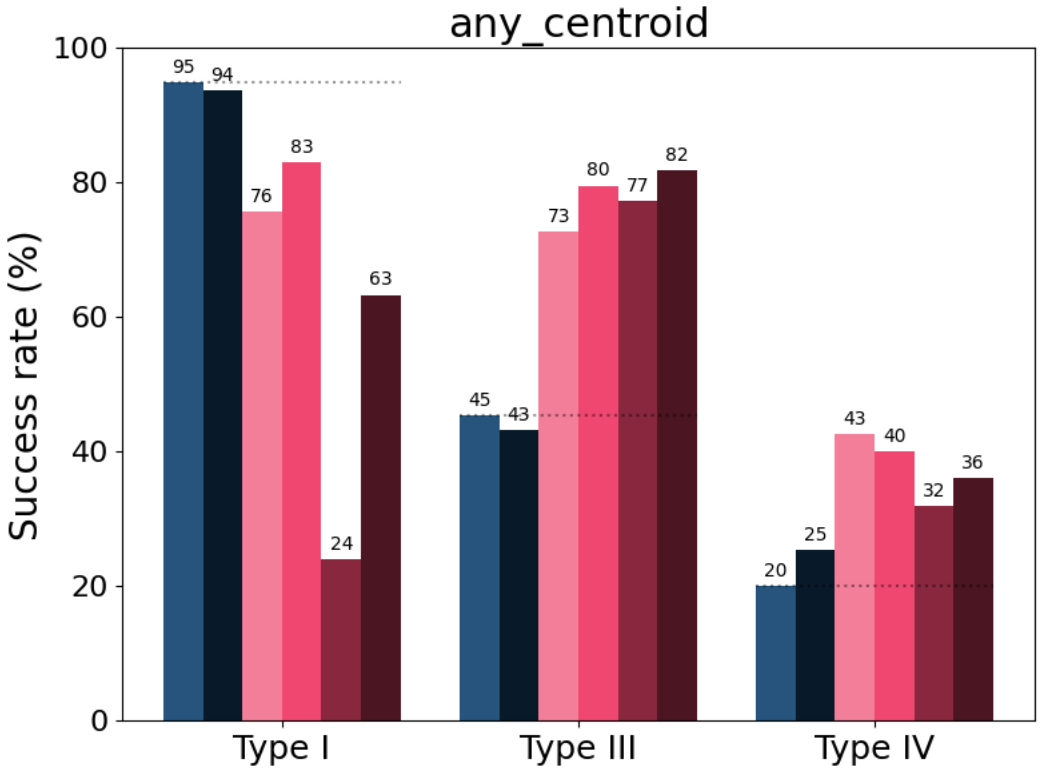




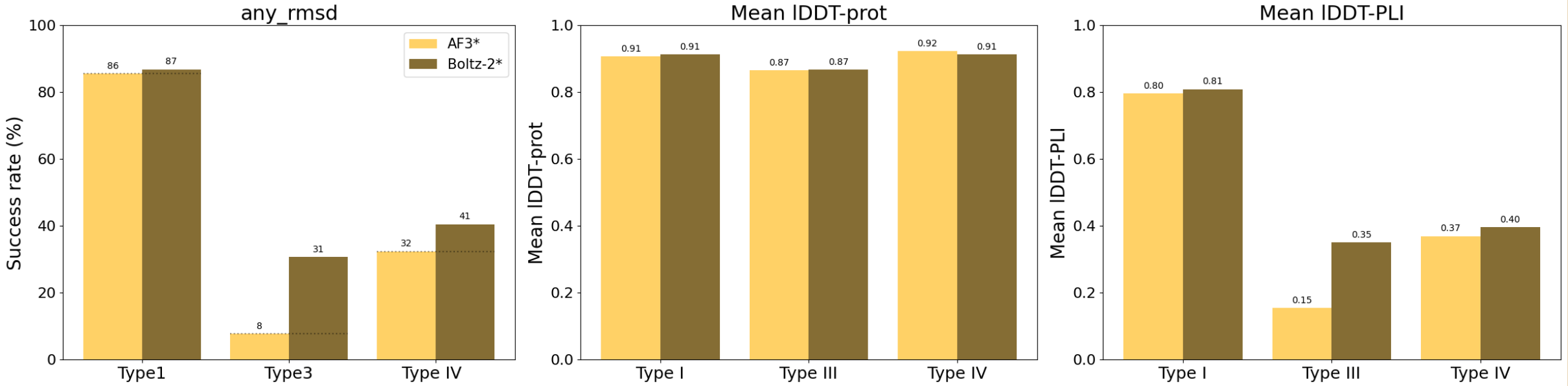
Kanev, G. K., *et al*. *Nucleic Acids Res*. **49** D562–D569 (2020).
Modi, V. & Dunbrack, R. L., Jr. *Nucleic Acids Res*. **50**, D654–D664 (2022).

Fine-tuning DiffDock-L on allosteric binders improves performance



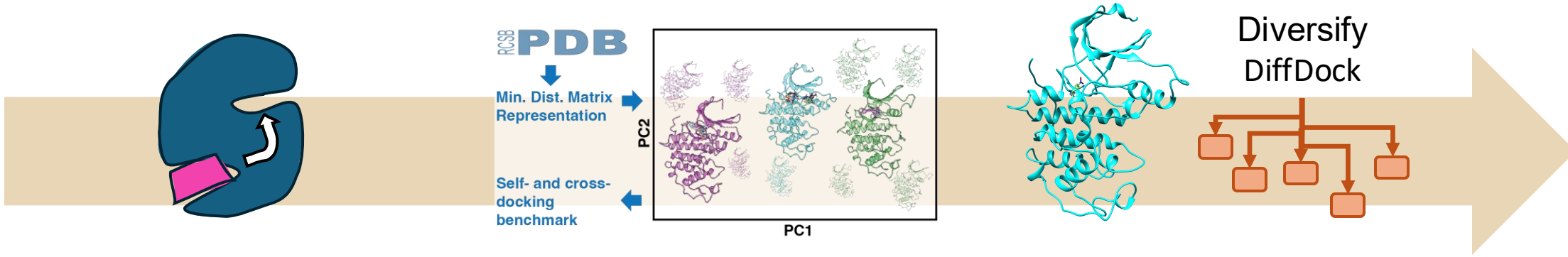


Co-folding methods also need improvement in this area



Future directions

- Receptor time-split, assess performance within the context of high-throughput screening with diverse library
- Improve docking pose accuracy and physical validity
- Development of scoring function that accurately ranks binders
- Which aspects of a model improve/detract from performance?



Yang, C., **Chen, E. A.** & Zhang, Y. *Molecules* **27**, 4568 (2022).
Xia, S., **Chen, E. A.** & Zhang, Y. *J. Chem. Theory Comput.* **19**, 7478–7495 (2023).

Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model.* **65**, 3737–3748 (2025).

Chen, E.A., Green J., Zhang. Y. In prep

NYU

SCCPC

Acknowledgements

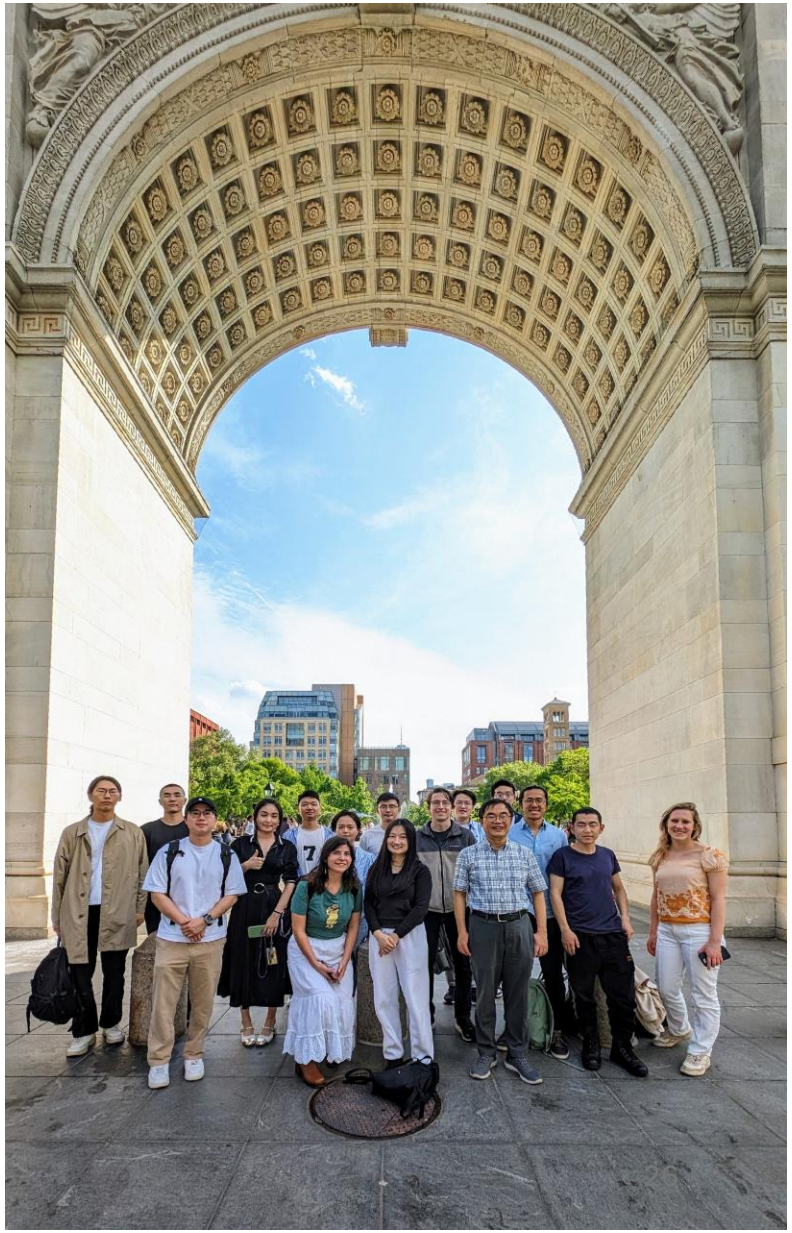
Prof. **Yingkai Zhang**

Zixuan Feng
Charlotte Infante
Thomas Kelly
Xuhang Dai
Yuting Wu
Shih-Chiang Lo
Chao Han
Weijun Zhou
Thomas Perone
Qi Ouyang
Yaowen Gu
Xudong Zhang
Fengyang Han
Seayoung Lee
Justin Green
Xiaolin Pan

Dr. Alida Besch
Dr. Song Xia
Dr. Chao Yang
Dr. Jieyu (Jocelyn) Lu
Dr. Dongdong Zhang
Dr. Xiaokang Guo

Dr. Federica Maschietto

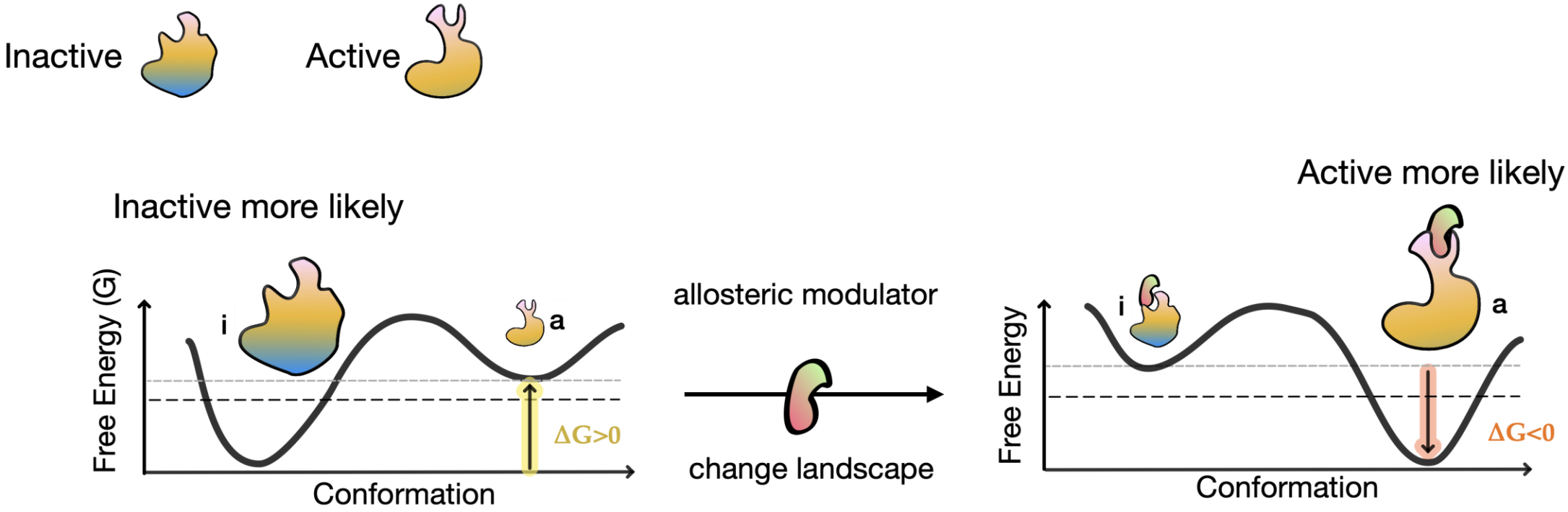
:eac709@nyu.edu
:123ric.bsky.social
:eric-chen-10
:echen1214



Thank you for listening!!

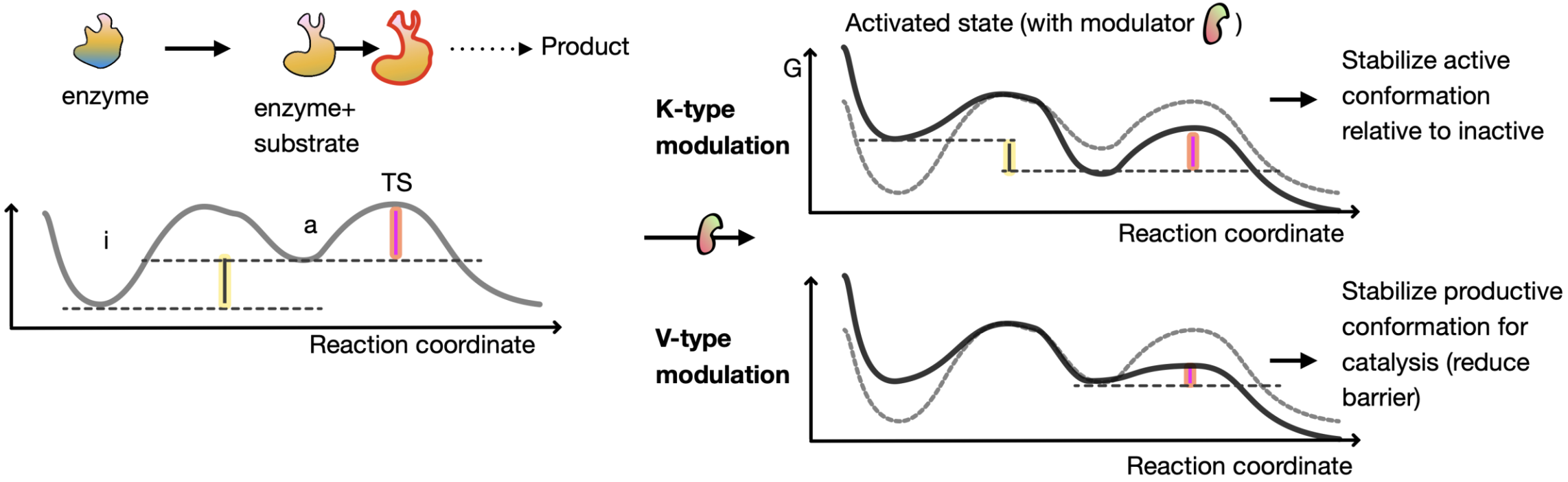
Appendix I.

Thermodynamic characterization of allostery



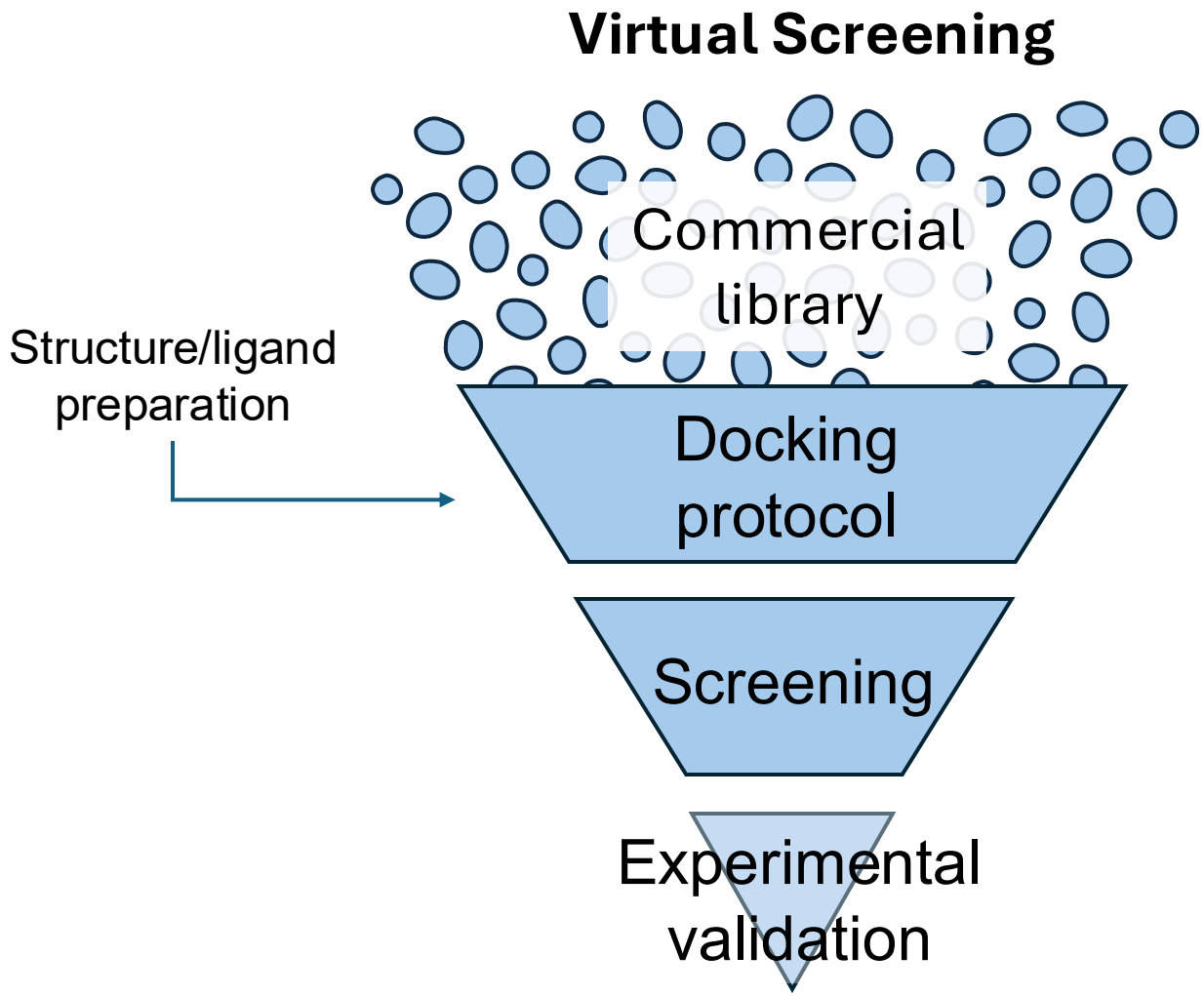
Measure ΔΔG

Allosteric response

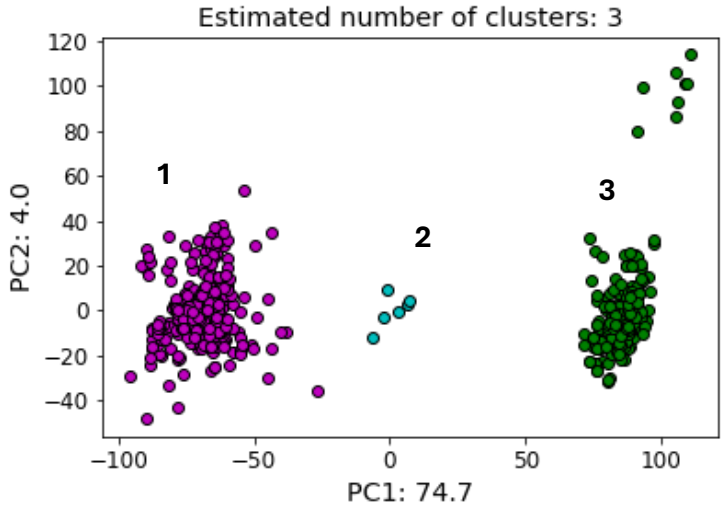


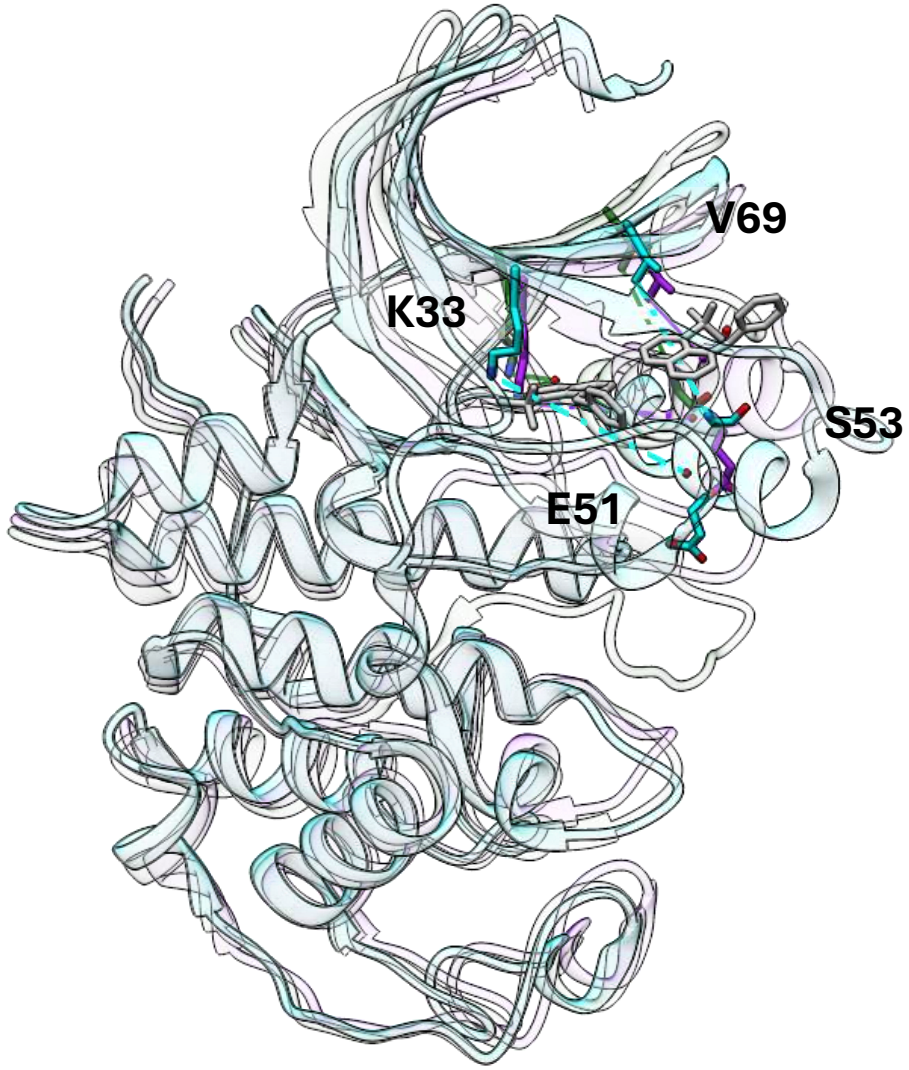
The few approved allosteric drugs act by K-type mode
K-type -> sensitive to external conditions
K-type to V-type -> extra control, more robust, less dependent on external conditions

Structure Based Drug Design: tool for drug discovery



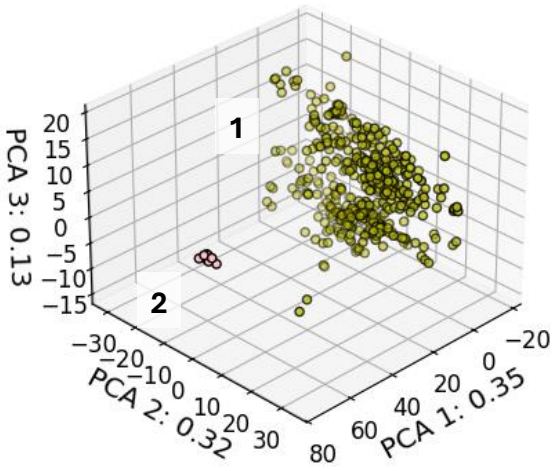
MDMR distinguishes receptor conformations

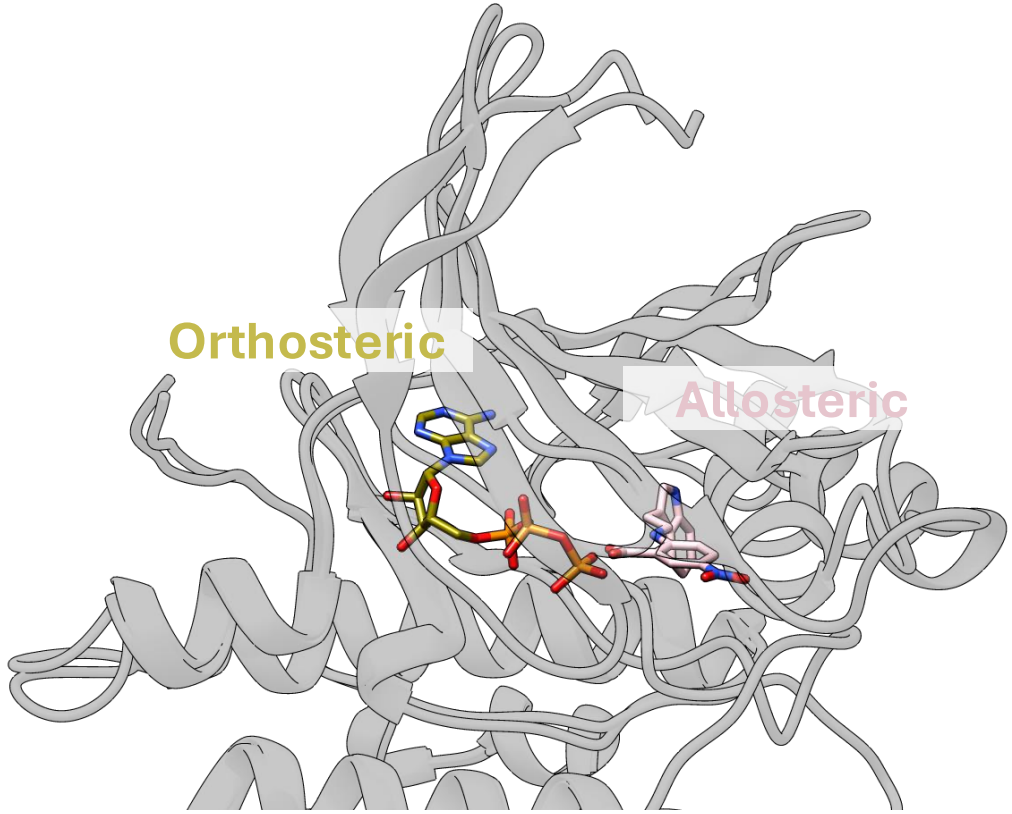




Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model*. **65**, 3737–3748 (2025).

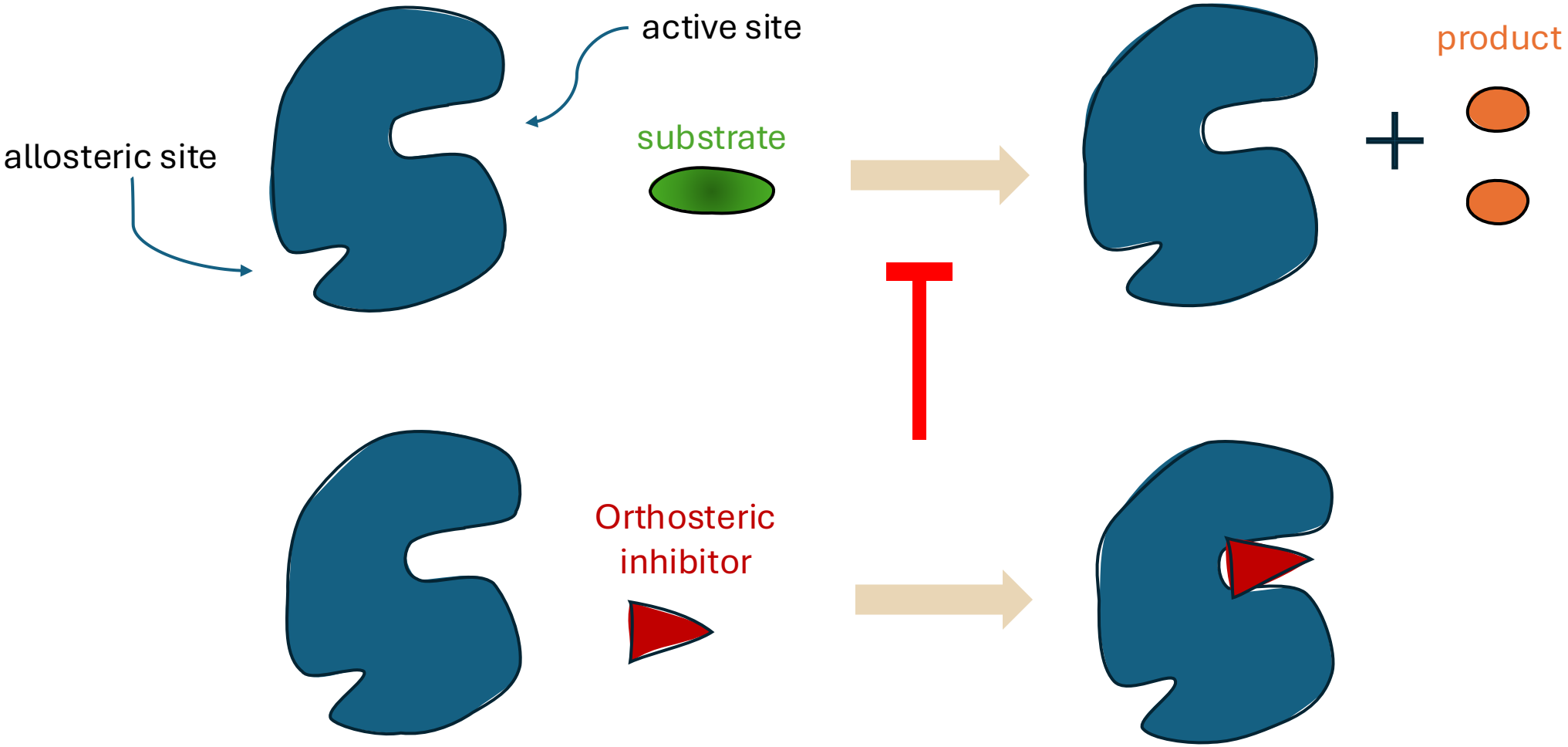
MDMR distinguishes ligand conformations





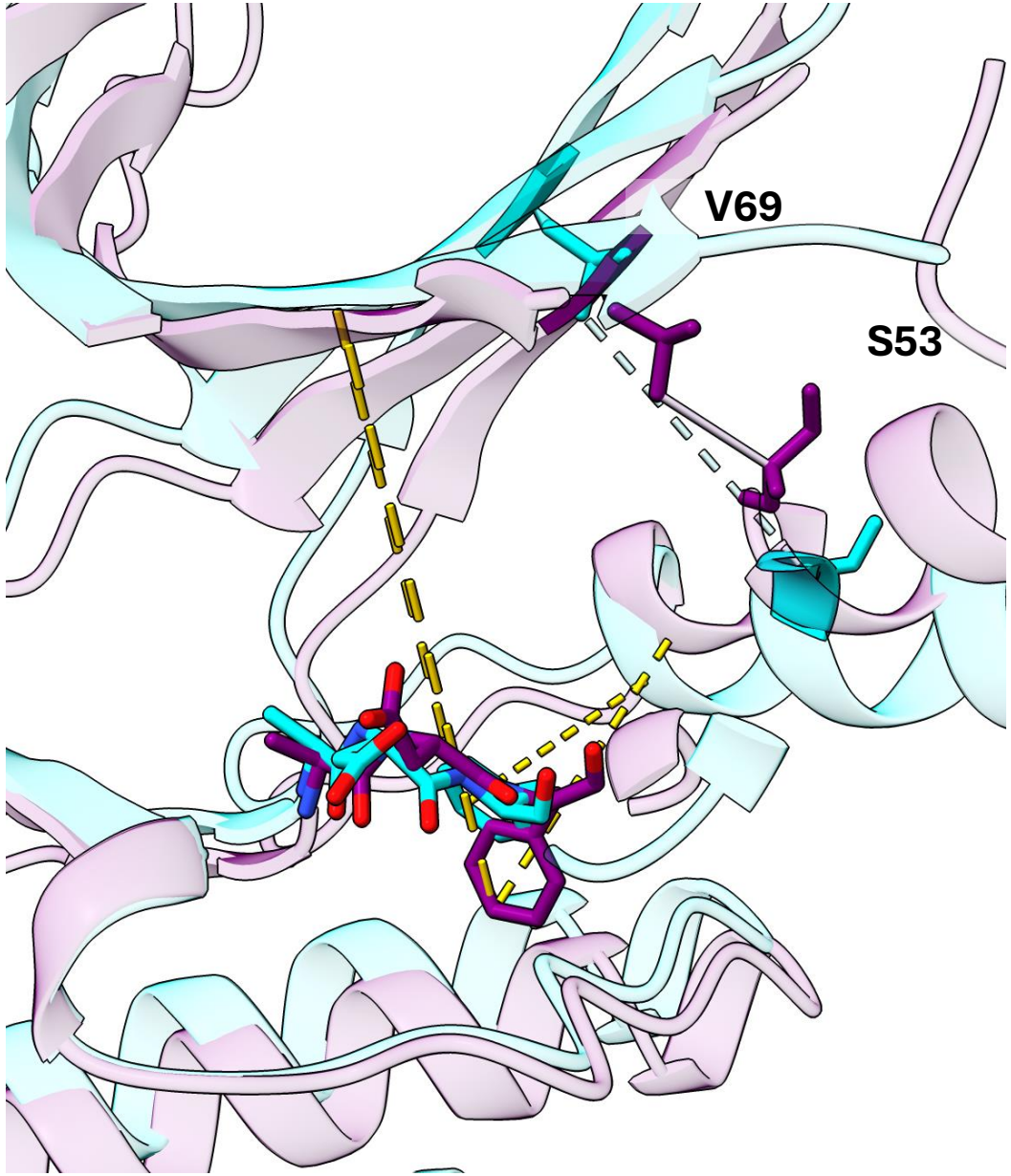
Chen, E. A. & Zhang, Y. *J. Chem. Inf. Model*. **65**, 3737–3748 (2025).

What is allosteric modulation?



Heuristic based methods do not distinguish cluster 2

PDB ID	Modi and Dunbrack	MDMR
2C5Y	DFGin–BLBminus	Cluster 1
3PXF	DFGin–BLBminus	Cluster 2



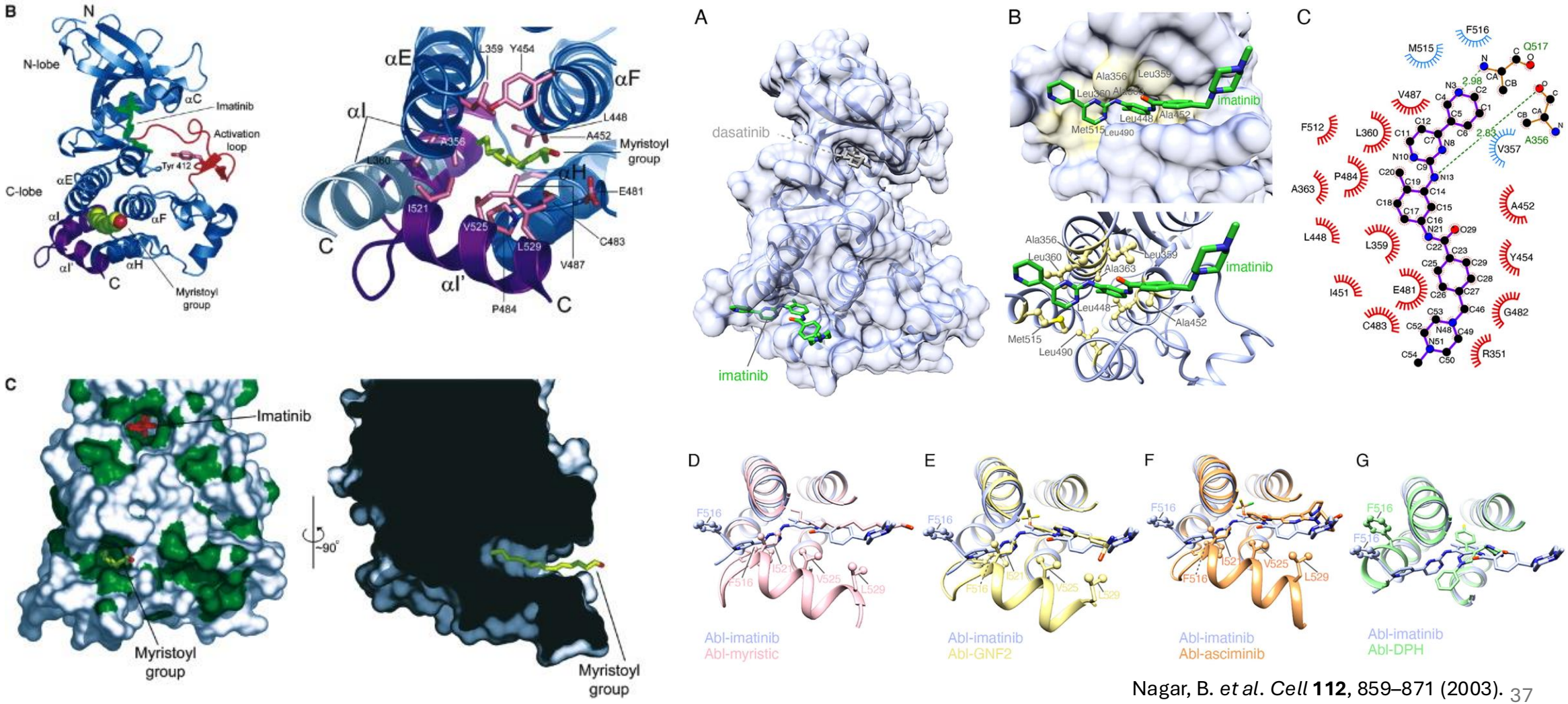
Modi, V. & Dunbrack, R. L. *Proc. Natl. Acad. Sci. U.S.A* **116**, 6818–6827 (2019).

FDA-approved small molecule protein and lipid kinase inhibitors

Year								
1999	Sirolimus							
2001	Imatinib							
2002	(none)							
2003	Gefinitib							
2004	Erlotinib							
2005	Sorafenib							
2006	Dasatinib	Sunitinib						
2007	Temsirolimus	Lapatinib	Nilotinib					
2008	(none)							
2009	Everolimus	Pazopanib						
2010	(none)							
2011	Crizotinib	Ruxolitinib	Vandetanib	Vemurafenib				
2012	Axitinib	Bosutinib	Cabozantinib	Ponatinib	Regorafenib	Tofacitinib		
2013	Afatinib	Ibrutinib	Trametinib	Dabrafenib				
2014	Ceritinib	Idelalisib	Nintedanib					
2015	Osimertinib	Cobimetinib	Alectinib	Lenvatinib	Palbociclib			
2016	(none)							
2017	Acalabrutinib	Neratinib	Abemaciclib	Brigatinib	Copanlisib	Midostaurin	Netarsudil	Ribociclib
2018	Dacomitinib	Binimetinib	Baricitinib	Duvelisib	Encorafenib	Fostamatinib	Gilteritinib	Larotrectinib
2019	Zanubrutinib	Alpelisib	Entrectinib	Erdafitinib	Fedratinib	Pexidartinib	Upadacitinib	Lorlatinib
2020	Selumetinib	Avapritinib	Capmatinib	Pemigatinib	Pralsetinib	Ripretinib	Selpercatinib	Tucatinib
2021	Mobocertinib	Asciminib	Belumosudil	Ingrafitinib	Tepotinib	Umbralisib	Tivozanib	Trilaciclib

Pan, Y. & Mader, M. M *J. Med. Chem.* **65**, 5288–5299 (2022).

Gleevec (Imatinib) is a type II inhibitor, with unexpected allosteric activating function of Abl



Nagar, B. *et al*. *Cell* **112**, 859–871 (2003).

Xie, T., Saleh, T., Rossi, P., Miller, D. & Kalodimos, C. G. *Journal of Molecular Biology* **434**, 167349 (2022).

Asciminib Type IV inhibitor

- Type I/II binding modes may not be so selective and are susceptible to resistance utates (eg T315I gatekeep mutation). Myristoyl site "conceptually … more selective than ATP-competitive TKIs"[1]
- 2003 – ABl1 has a regulatory myristoyl site[2]
- 2006 - Lead GNF-2 was discovered during a phenotype cytotoxicity based screen but later biochemical studies showed that it was ATP noncompetitive[3] – Did not lead to clinical candidates
- NMR fragment screen identified new starting points with NMR and X-ray to culminate in Asciminib[1]
  - Novartis co-administer asciminib and orthosteric inhibitor – might enhance target coverage and prevent resistance

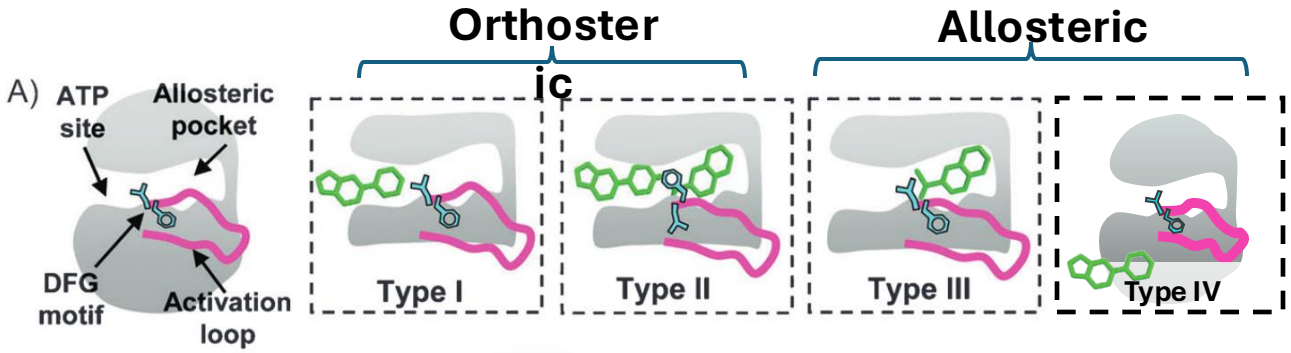
[1]Schoepfer, J. *et al*. *J. Med. Chem.* **61**, 8120–8135 (2018).
[2]Nagar, B. *et al*. *Cell* **112**, 859–871 (2003).
[3]Adrián, F. J. *et al*. *Nat Chem Biol* **2**, 95–102 (2006).

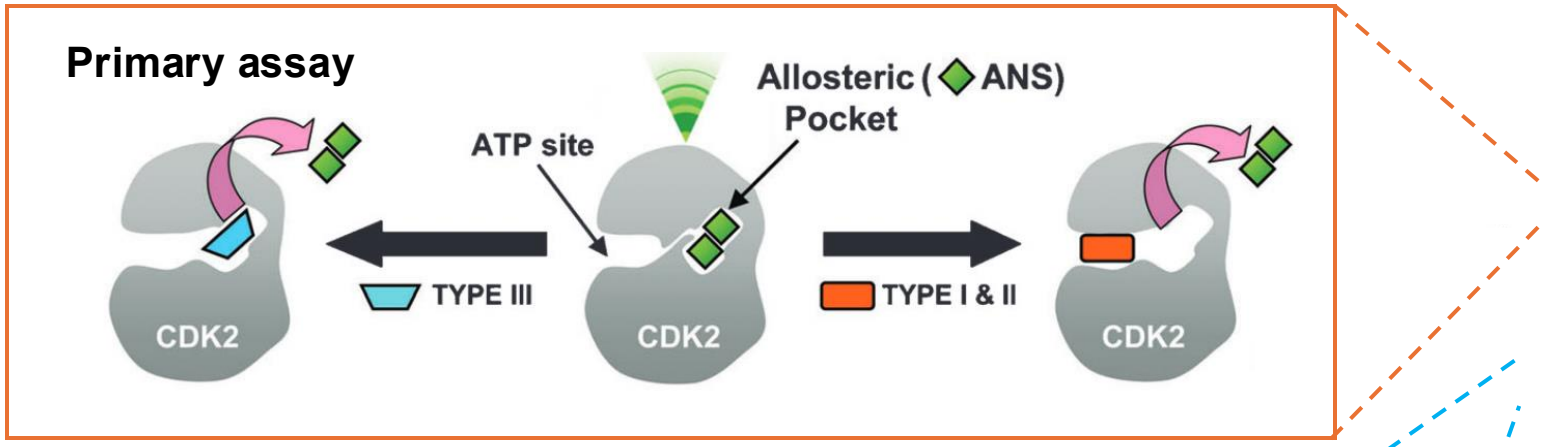
MEK1/2 Type III inhibitor

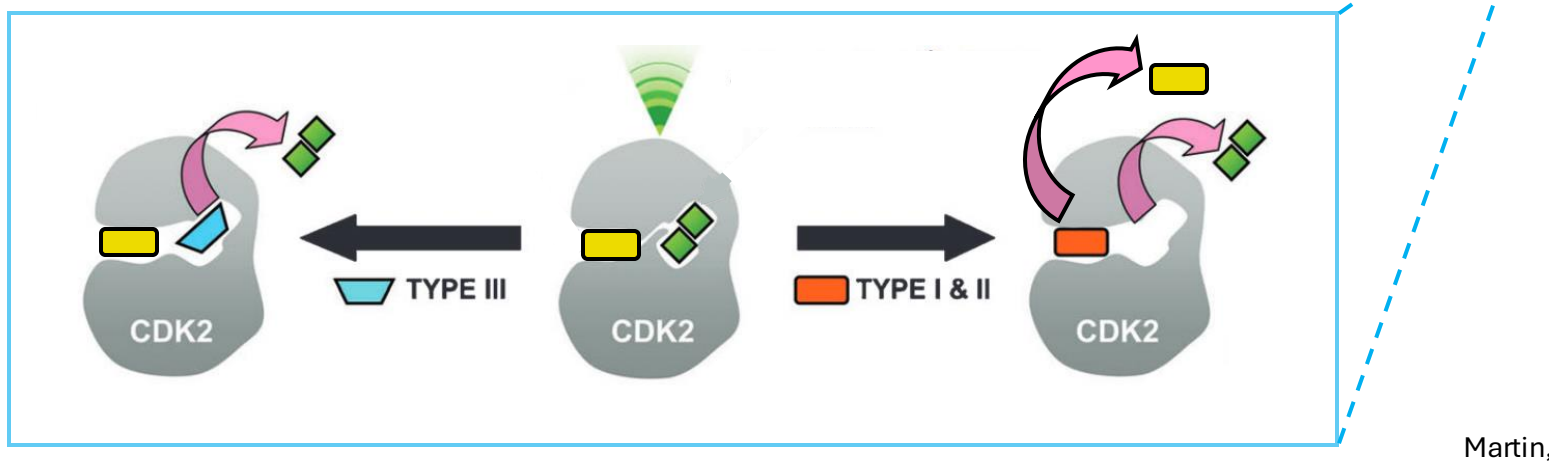
- 1995 – kinase cascade assay identify compounds that were later determined to be noncompetitive with ATP…optimized to get (CI-1040/PD-184352)[1]
- 2007 – separately, Japan Tobacco discovered JTP-74057 through phenotypic screen to identify compounds that induce downstream cellular effects [2] → later rationally developed by GSK (GSK1120212, Trametinib)
- 2008 – CI-1040 compound terminated due to limited efficacy, but lead to optimization of other compounds (PD-0325901/Mirdametinib clinical trials)[3]

[1]Dudley, D. T., Pang, L., Decker, S. J., Bridges, A. J. & Saltiel, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7686–7689 (1995).
[2]Yamaguchi, T. *et al*. *Cancer Science* **98**, 1809–1816 (2007).
[3]Barrett, S. D. *et al*. *Bioorganic & Medicinal Chemistry Letters* **18**, 6501–6504 (2008).

Can we improve hit rates of allosteric inhibitors?







Martin, M. P. *et al*. *ChemBioChem* **13**, 2128–2136 (2012).
Faber, E. B. *et al*. *J. Med. Chem.* **66**, 1928–1940 (2023).

vs. Möbitz et al classification on the partial CDK2 dataset

DFG	Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4
DFGin	Active	203			
	Active αC helix out				
	FG down	1			
	FG down αC helix out		60		118
	G down				
	G down αC helix out			5	53
	Aup Met				
DFGout	DFG Flipped				
	DFGout Type2				
	AuP BRAF				
	AuP FMS				
	AuP IGF1R				
	Disordered				8
	Other				3

Möbitz, H. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1854**, 1555–1566 (2015).

vs. Ung et al classification on the partial CDK2 dataset

Spatial groups	Cluster 1	Cluster 2	Cluster 3	Cluster 4
CIDI	35			
CODI		39	5	123
CIDO				
CODO				
ωCD				

← Active

Inactive

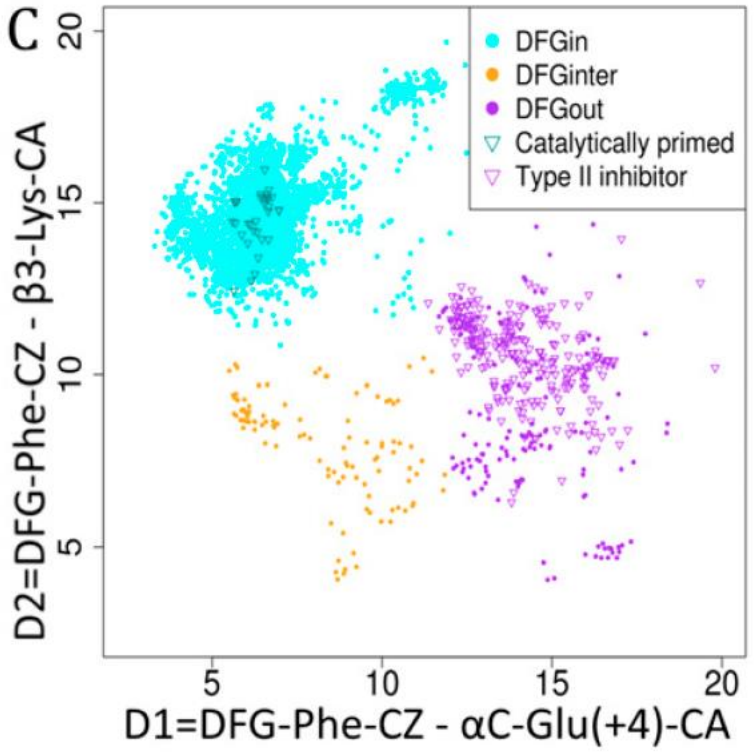
← Distorted/intermediate

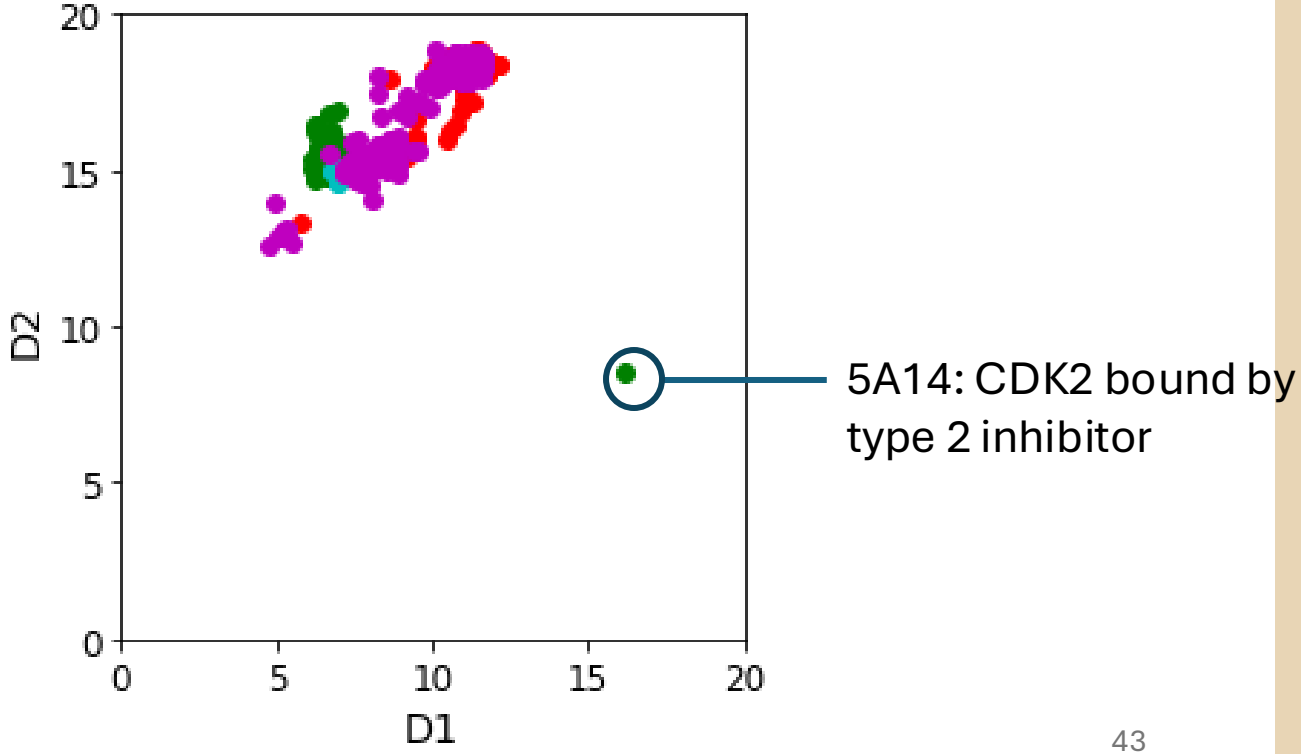
CIDI: C-helix-in - DFGin
CIDO: C-helix-in - DFGout
CODI: C-helix-out – DFGin
CODO: C-helix-out - DFGout
ωCD: DFGintermediate

Ung, P. M.-U., Rahman, R. & Schlessinger, A. *Cell Chemical Biology* **25**, 916-924.e2 (2018).

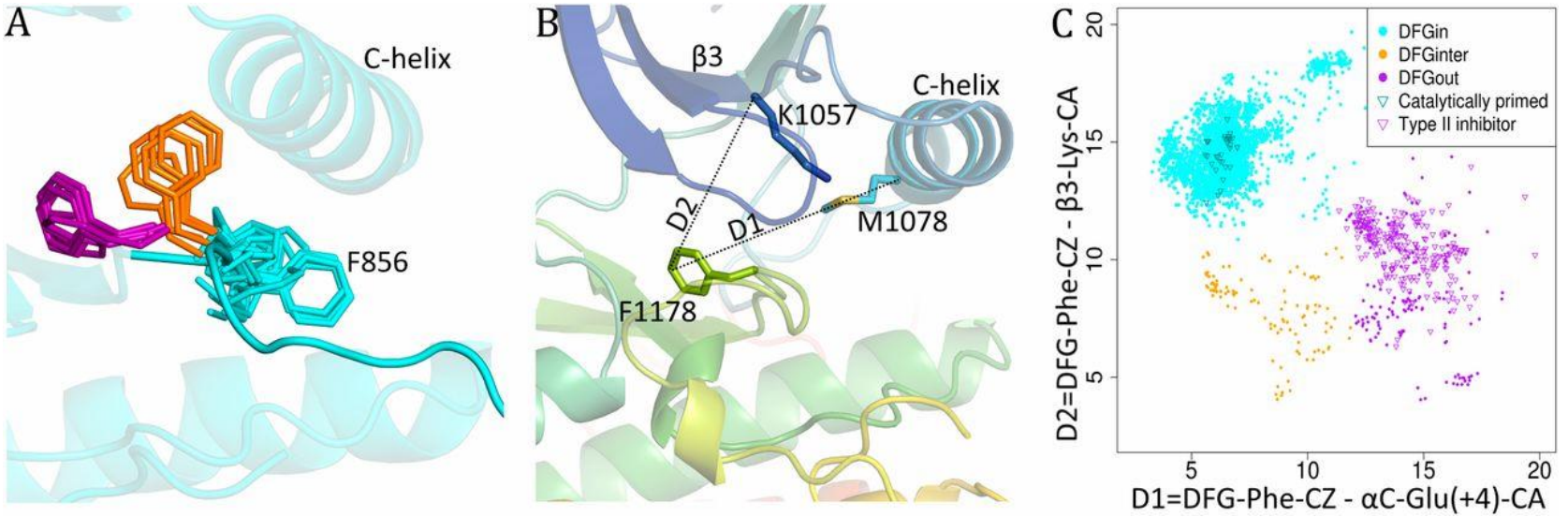
vs. Modi and Dunbrack classification on CDK2 dataset

- Dist(αC-Glu(+4)-Cα, DFG-Phe-Cζ) & Dist(β3-Lys-Cα, DFG-Phe-Cζ)
- CDK2:CDK2:Dist(L55 Cα–F146 Cζ) & Dist(K33 Cα–F146 Cζ)

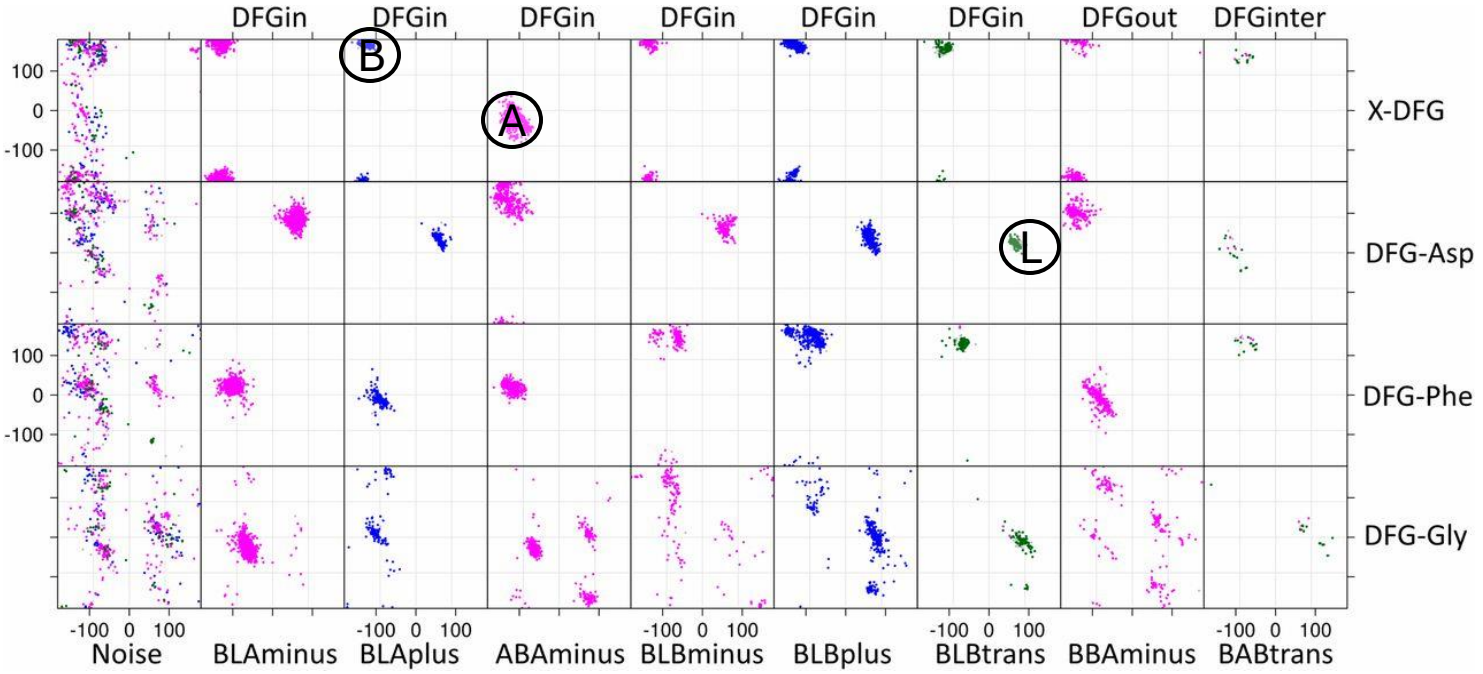




Modi, V. & Dunbrack, R. L. *PNAS* **116**, 6818–6827 (2019).



Clusters are named by the Ramachandran regions of the X-DFG, DFG-Asp, and DFG-Phe residues and the χ1 rotamer of DFG-Phe: g− (χ1 ~ −60°, minus, magenta); g+ (χ1 ~ +60°, plus, blue); trans rotamer (χ1 ~ 180°, trans, green).



Modi, V. & Dunbrack, R. L. *Proc. Natl. Acad. Sci. U.S.A* **116**, 6818–6827 (2019).

Heuristic based methods do not distinguish cluster 2

Modi and Dunbrack			MDMR		
	Spatial groups	Clusters	Cluster 1	Cluster 2	Cluster 3
	DFGin	BLAminus	9		127
		BLAplus			
		ABAminus			2
		BLBplus	19		
		BLBminus	79	6	
		BLBtrans	123		
		Noise	14		2
	DFGout	BBAminus			
		Noise			1
	DFGinter	BABtrans			
		Noise			

Most frequent active conformation

Most frequent DFG-in inactive conformation

Modi, V. & Dunbrack, R. L. *Proc. Natl. Acad. Sci. U.S.A* **116**, 6818–6827 (2019).

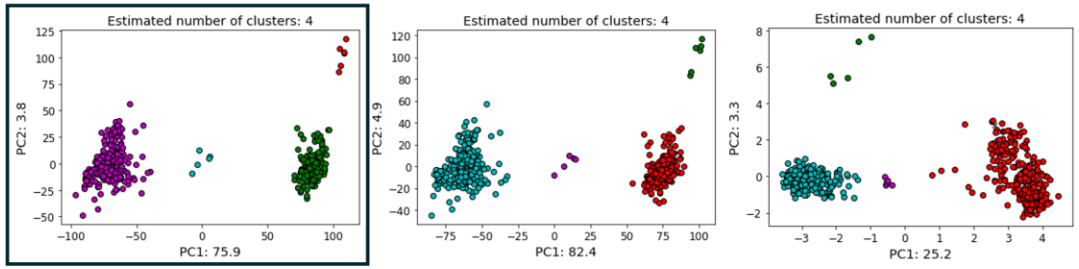
Shortest distance

Cα–Cα distance

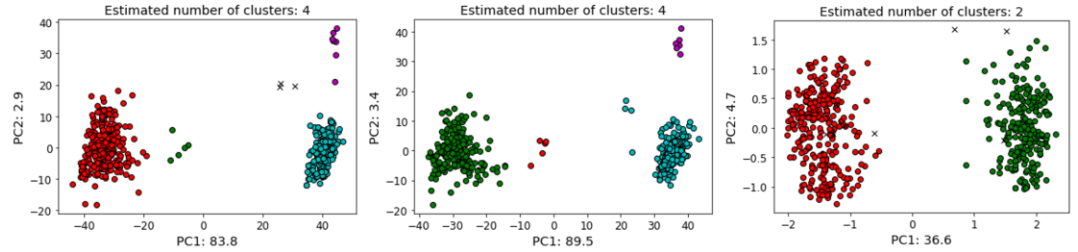
Contact map

$$\text{map} = \begin{cases} 1\,;\, d_{ij}^{shortest} < 5 \\ 0\,;\, d_{ij}^{shortest} > 5 \end{cases}$$

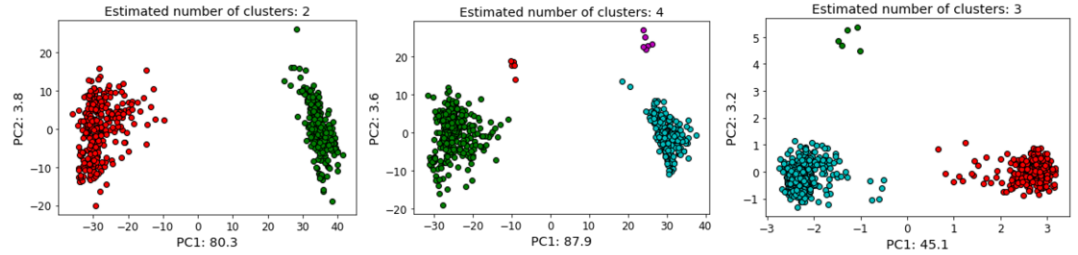
All residues



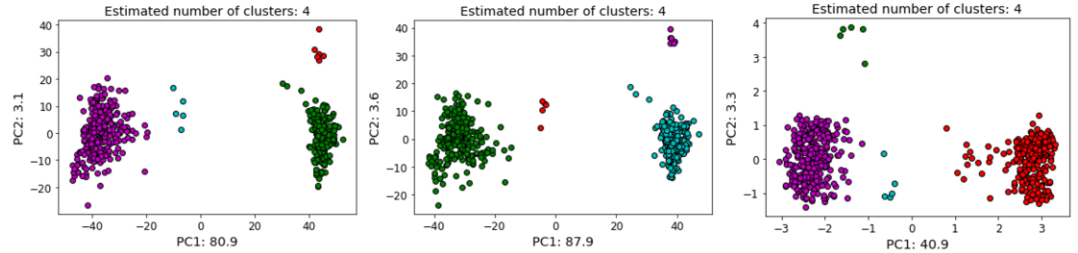
Identically-, conservatively- and semi-conserved residues

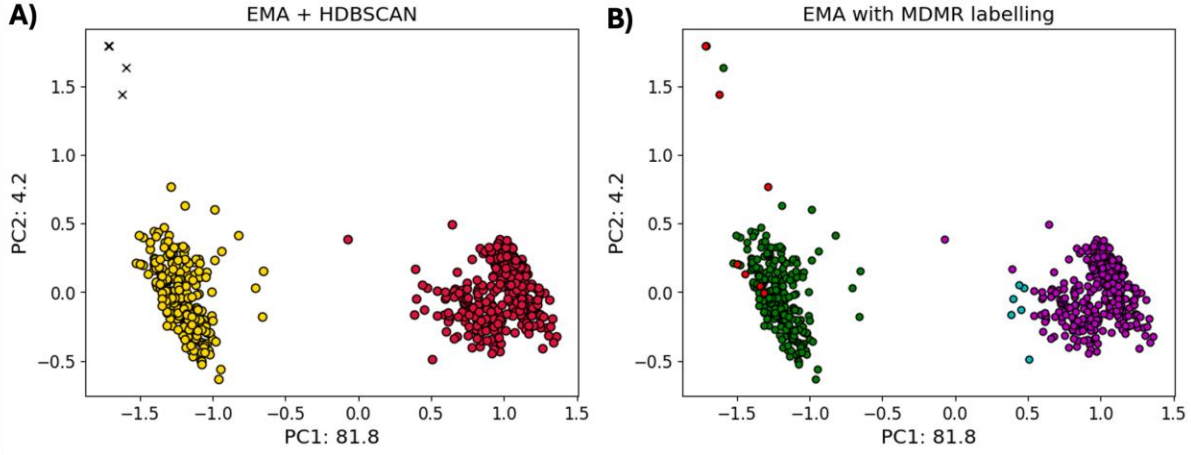


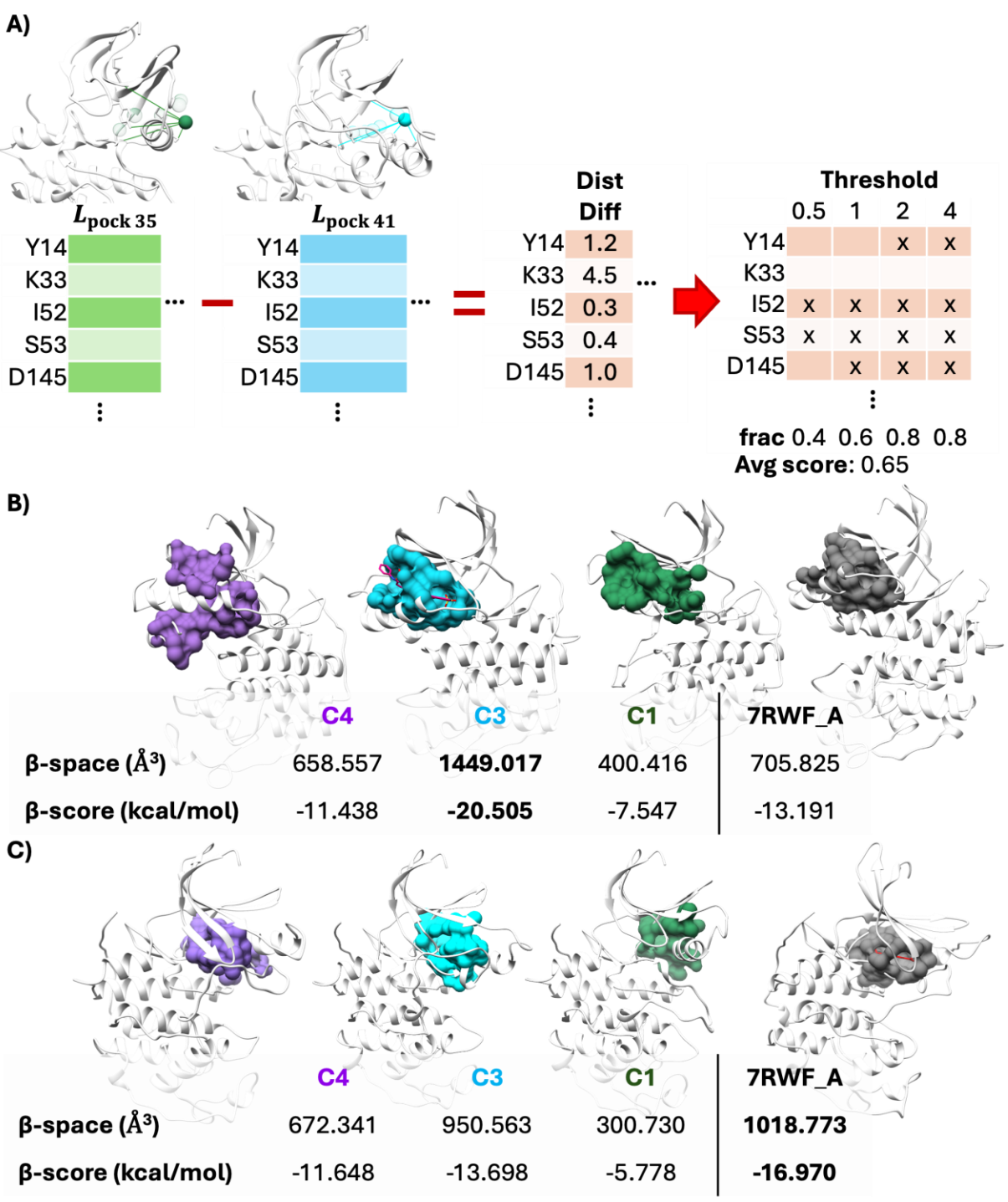
KLIFS only

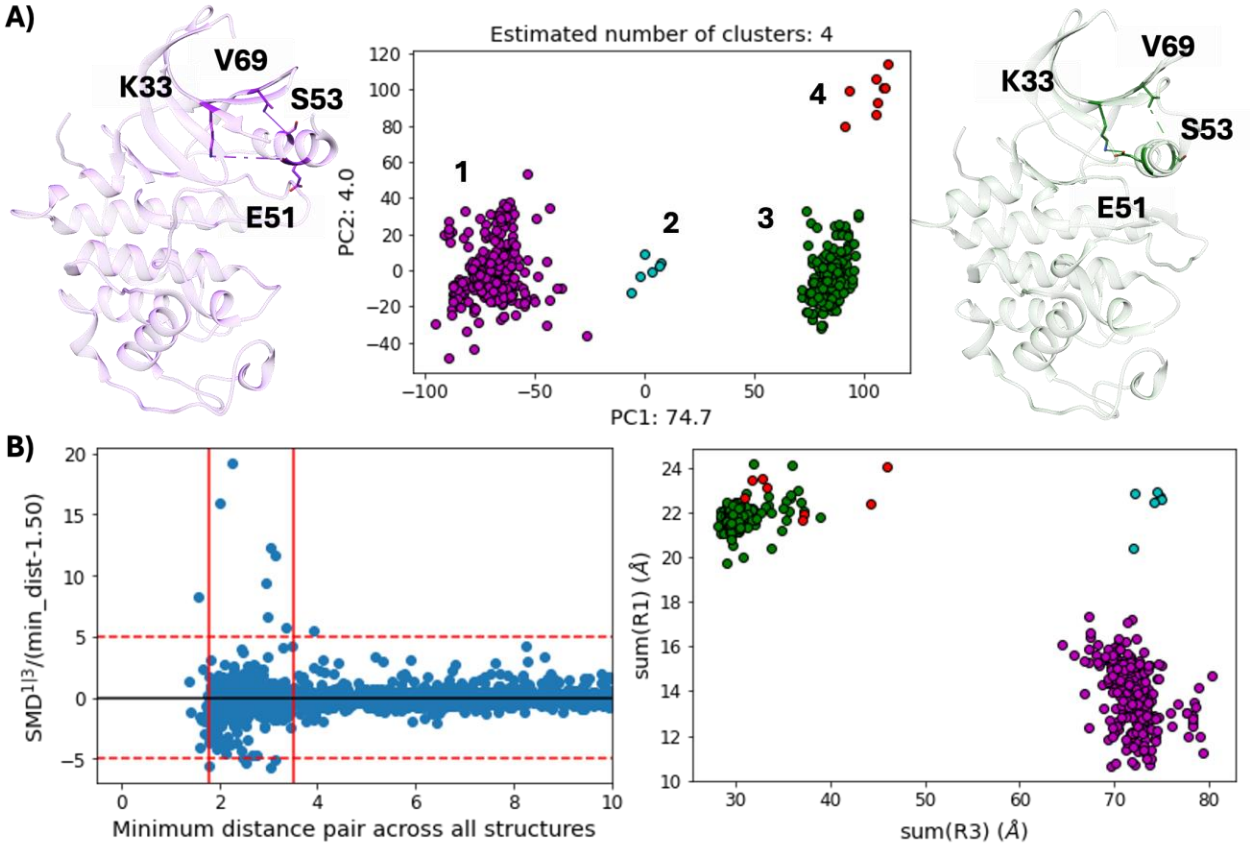


KLIFS + identically conserved






A)
$L_{pock\ 35}$
Y14
K33
I52
S53
D145
$L_{pock\ 41}$
Y14
K33
I52
S53
D145
Dist Diff
Y14 1.2
K33 4.5
I52 0.3
S53 0.4
D145 1.0
Threshold
0.5 1 2 4
Y14 x x
K33
I52 x x x x
S53 x x x x
D145 x x x
frac 0.4 0.6 0.8 0.8
Avg score: 0.65
B)
C4
C3
C1
7RWF_A
β-space (Å³) 658.557 1449.017 400.416 705.825
β-score (kcal/mol) -11.438 -20.505 -7.547 -13.191
C)
C4
C3
C1
7RWF_A
β-space (Å³) 672.341 950.563 300.730 1018.773
β-score (kcal/mol) -11.648 -13.698 -5.778 -16.970



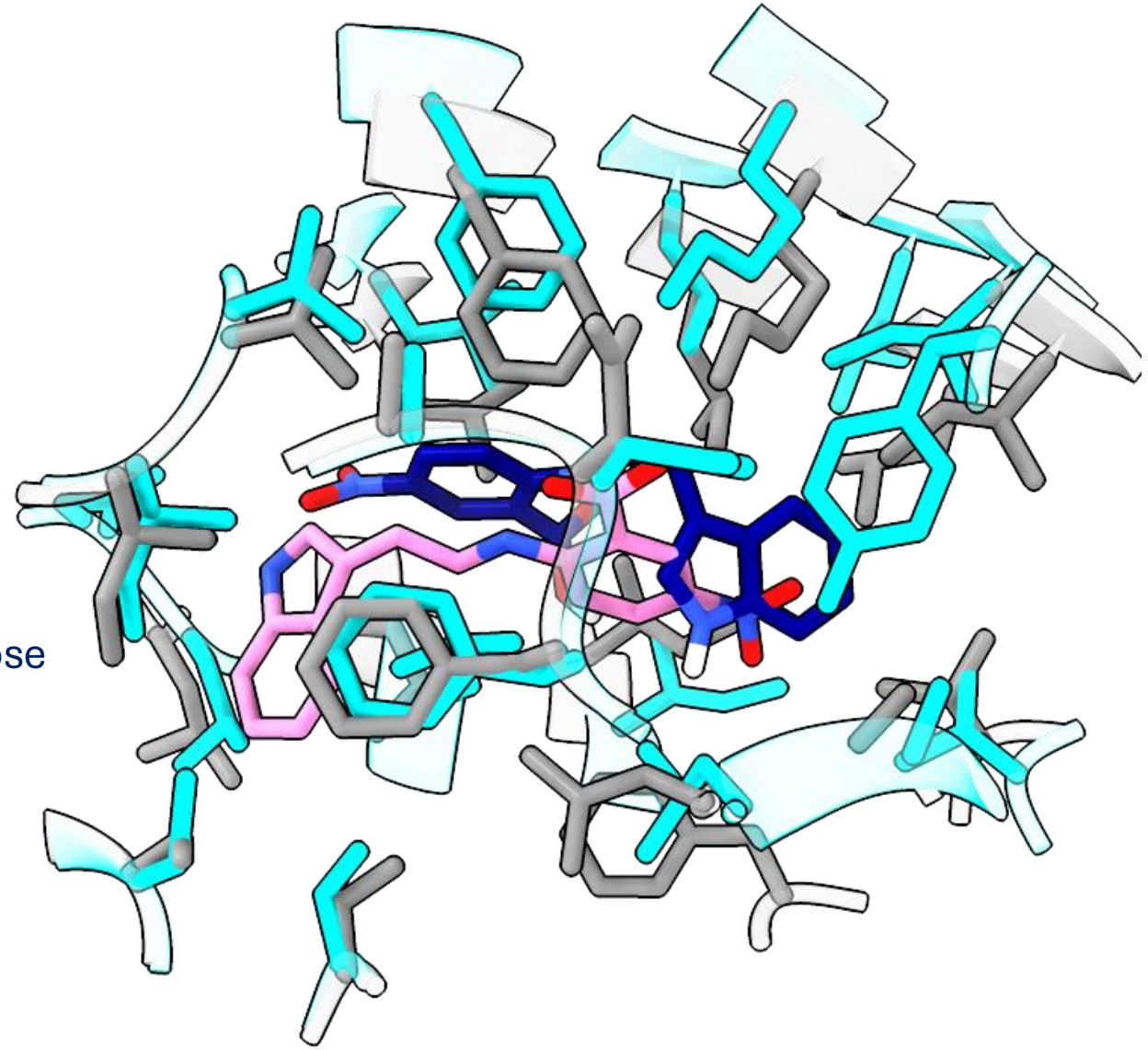
R1: Interactions stabilizing inactive state			
Res 1	Res 2	Min dist (Å)	nSMD
SER 53	VAL 69	3.034	-5.678
LYS 56	ASP 65	2.763	-5.269
LEU 55	VAL 123	2.546	-5.012

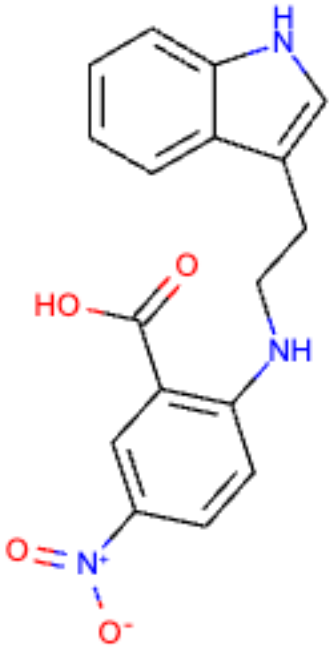
R3: Interactions stabilizing active states			
Res 1	Res 2	Min dist (Å)	nSMD
LYS 33	GLU 51	2.264	19.185
GLU 57	ARG 122	2.204	15.951
GLU 51	LEU 78	3.043	12.247
GLU 51	PHE 80	3.147	11.666
LYS 34	GLU 51	2.938	9.462
LEU 55	PHE 80	2.993	6.614
LEU 55	VAL 64	3.364	5.695

$$\mathrm{SMD}_{ij}^{x|y} = \frac{\overline{d_{ij}^{x}} - \overline{d_{ij}^{y}}}{\sqrt{\sigma_{ij}^{x} * \sigma_{ij}^{y}}}$$

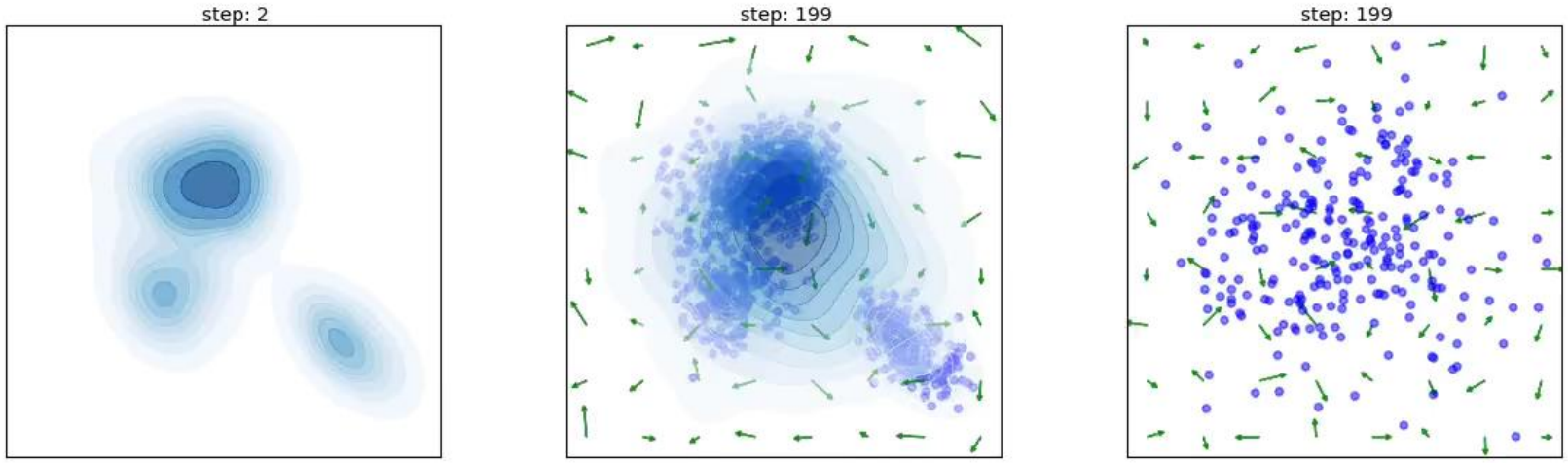
$$\text{Normalized SMD}_{ij}^{x|y} = \frac{\mathrm{SMD}_{ij}^{x|y}}{\min(d_{ij}) - 1.5}$$

7RWF_A
7RWF_A_7TW
3PXF_A
DD+LRD docked pose





DiffDock: diffusion models for blind docking: learn protein–ligand interaction space and generate pose



Forward Process (1) - add noise

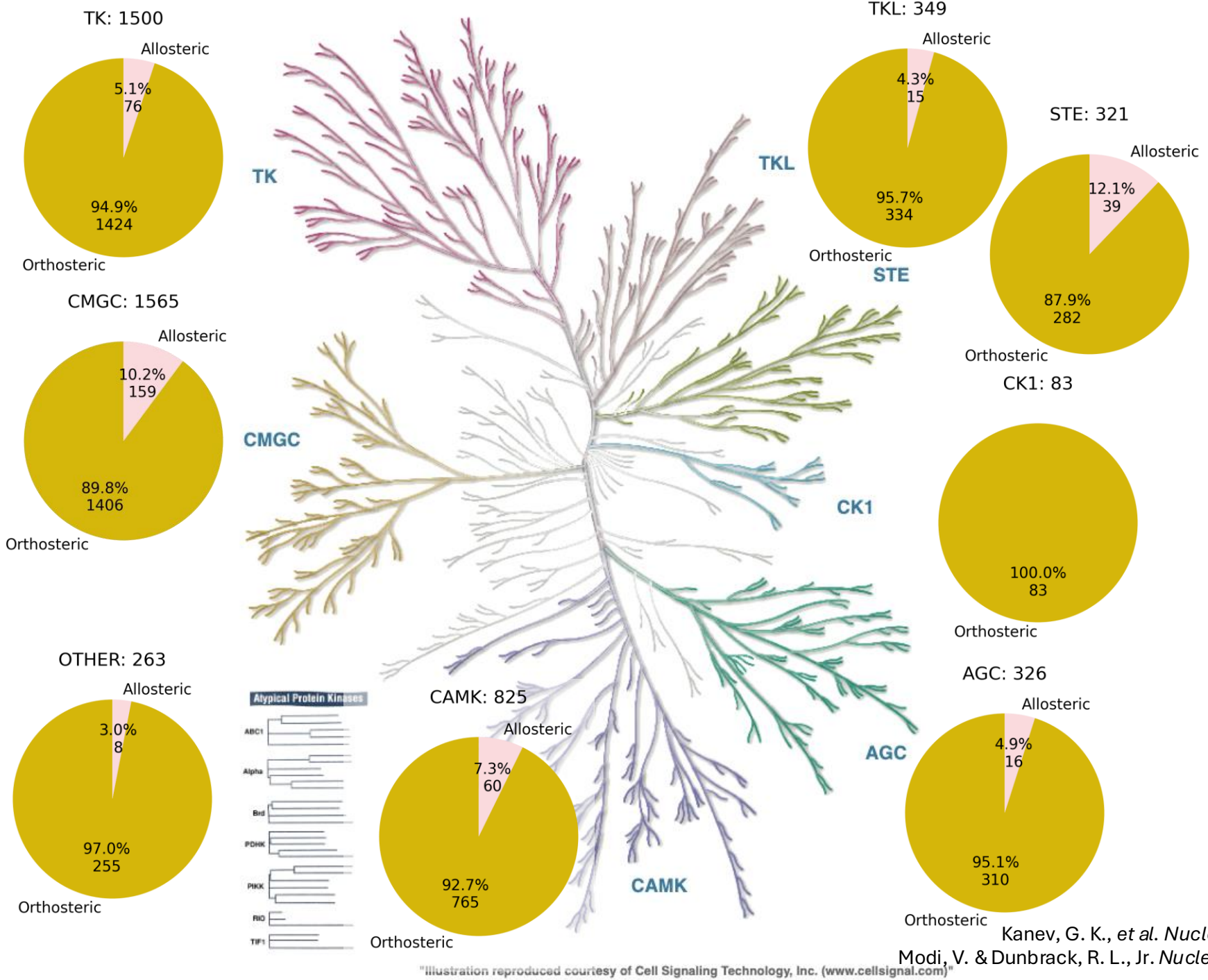
x(0) ground truth → $\mathbf{dx} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$ → x(T) noisy state

Reverse process (2) - learn score $\nabla_x \log(p_t(x))$

Inference (2) - sample learned distribution by denoising from random

x(0) ground truth ← $\mathbf{dx} = [\mathbf{f}(\mathbf{x}, t) - g^2(t)\nabla_x \log(p_t(x))]\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$ ← x(T) noisy state

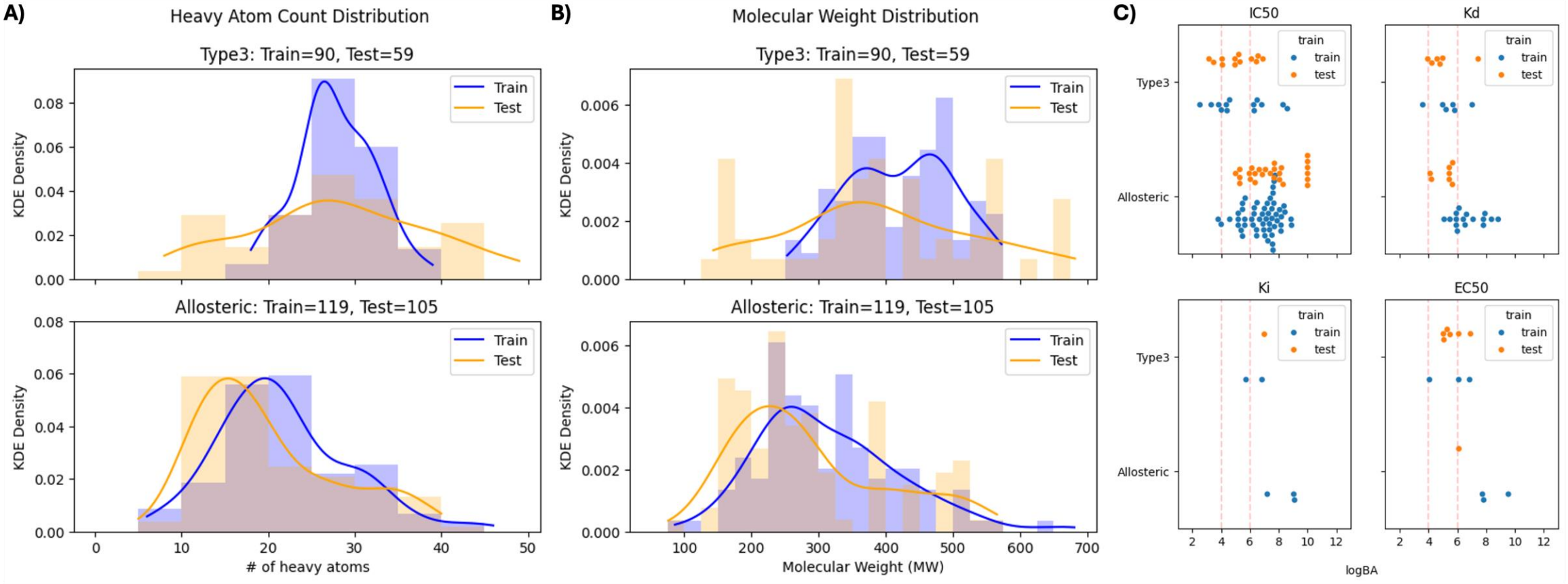
Corso, G., *et al.* arXiv:2210.01776 (2022)



"Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com)"

Curating a kinome-wide structural dataset

Kanev, G. K., *et al*. *Nucleic Acids Res*. **49** D562–D569 (2020).
Modi, V. & Dunbrack, R. L., Jr. *Nucleic Acids Res*. **50**, D654–D664 (2022).


A)
Heavy Atom Count Distribution
Type3: Train=90, Test=59
Allosteric: Train=119, Test=105
KDE Density
# of heavy atoms
Train
Test
B)
Molecular Weight Distribution
Type3: Train=90, Test=59
Allosteric: Train=119, Test=105
KDE Density
Molecular Weight (MW)
Train
Test
C)
IC50
Kd
Ki
EC50
Type3
Allosteric
train
test
logBA