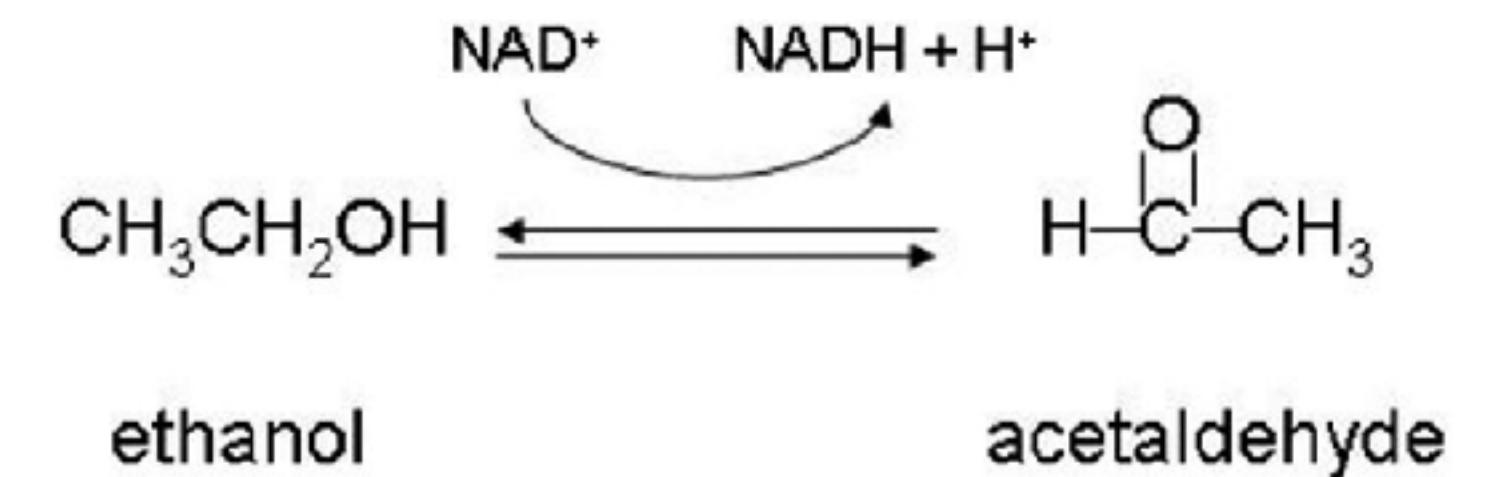
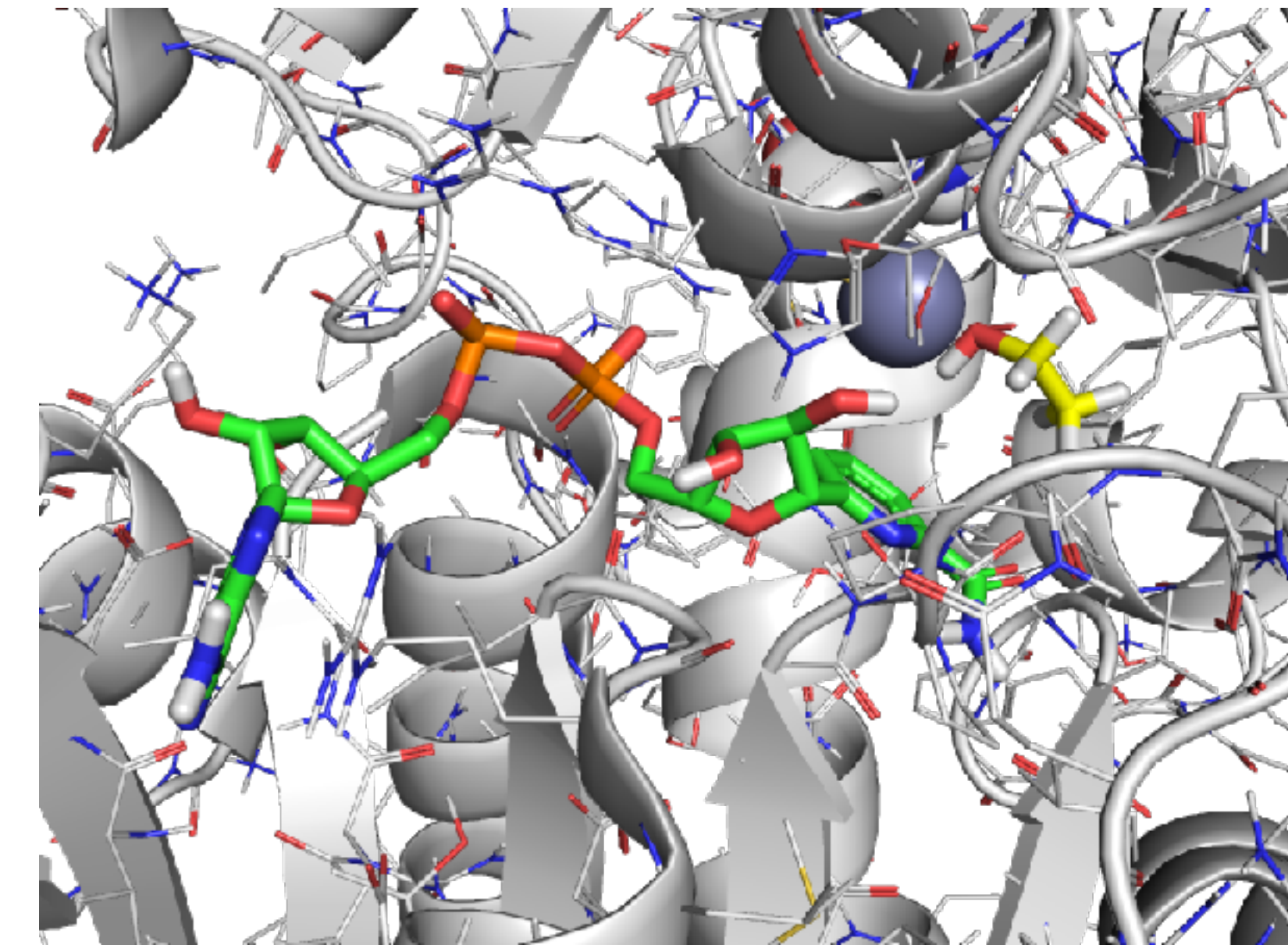
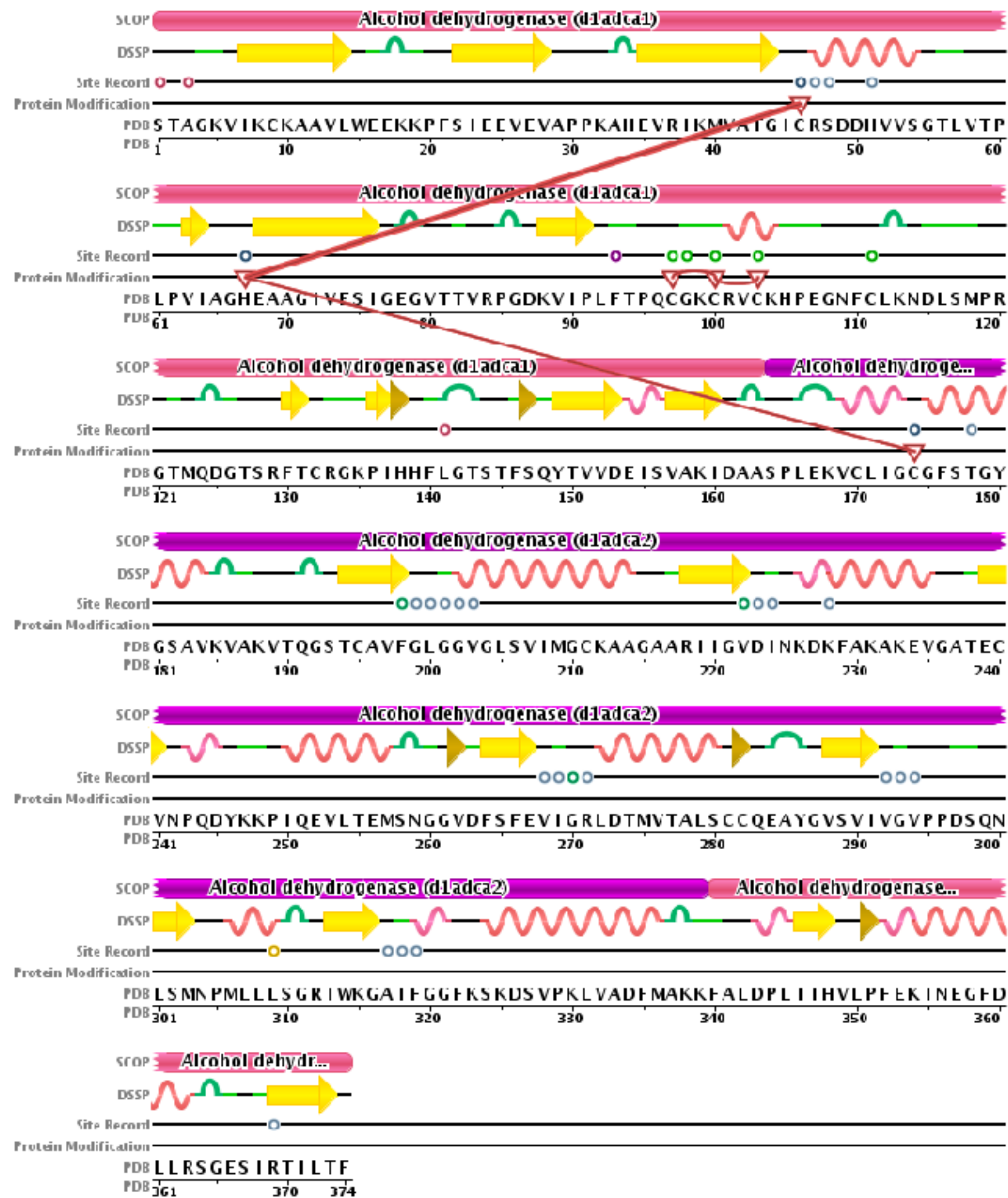




Computational Structural Biology and Drug Discovery

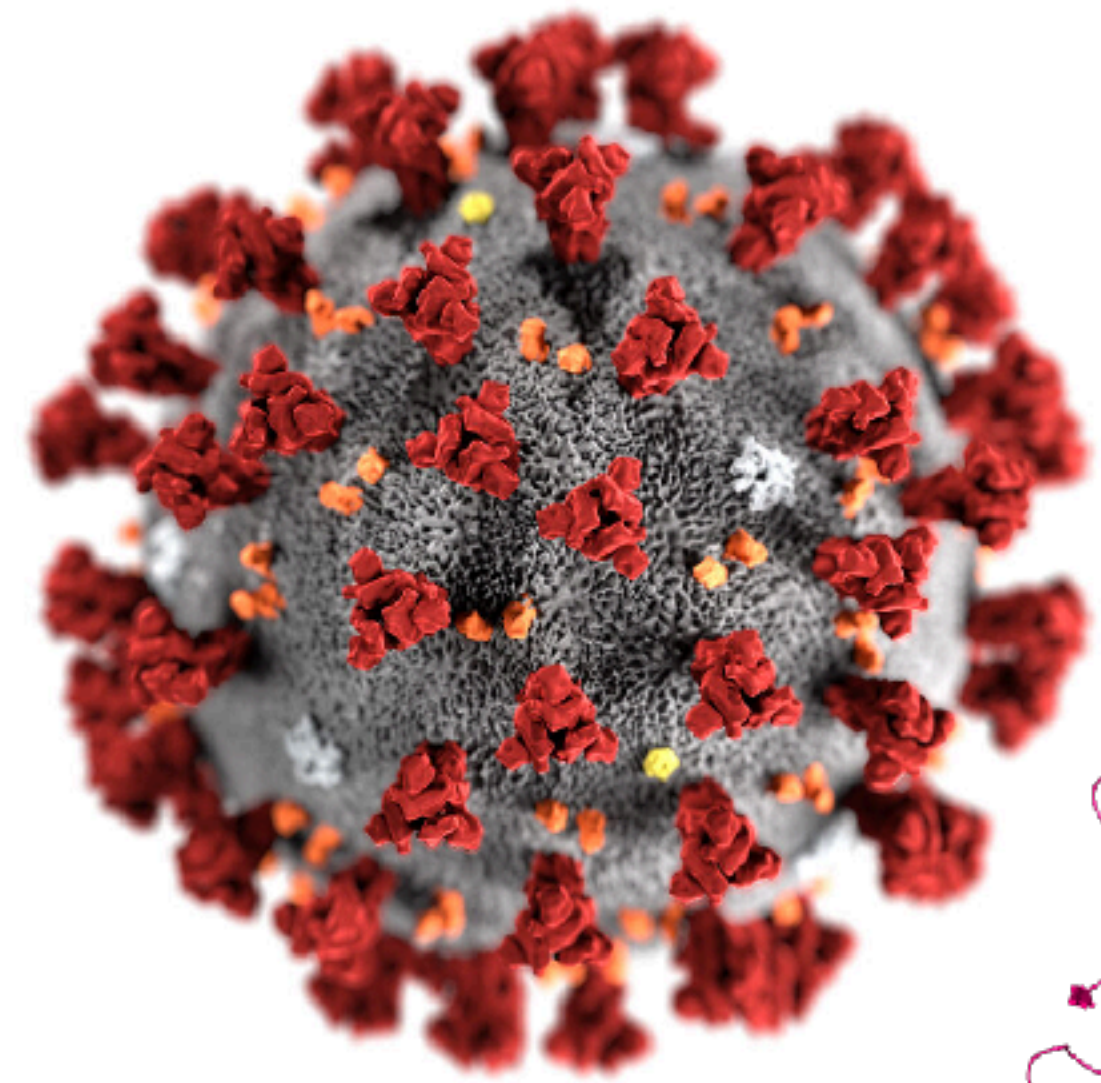
David Ryan Koes
6/2/2020

Sequence → Structure → Function

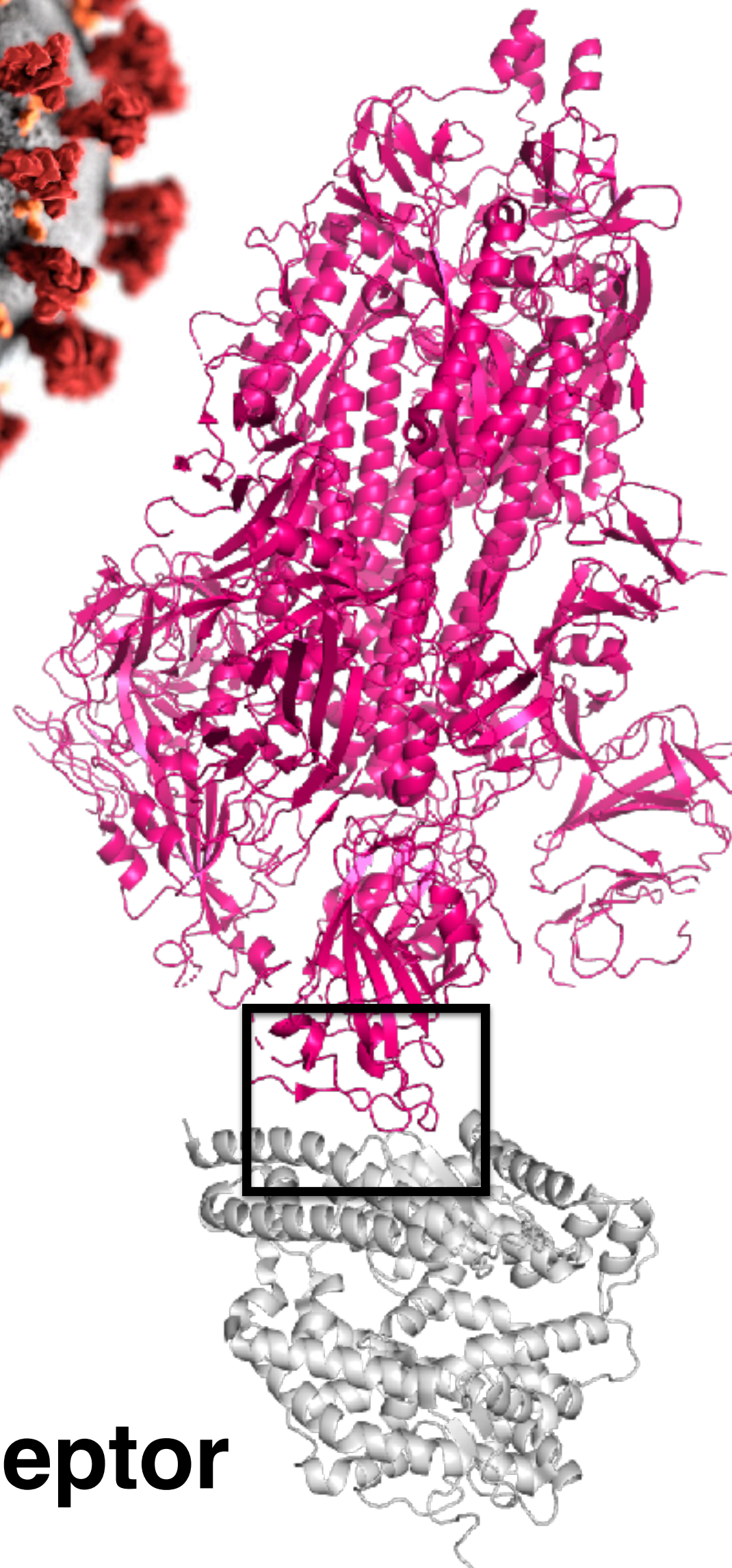


Alcohol Dehydrogenase

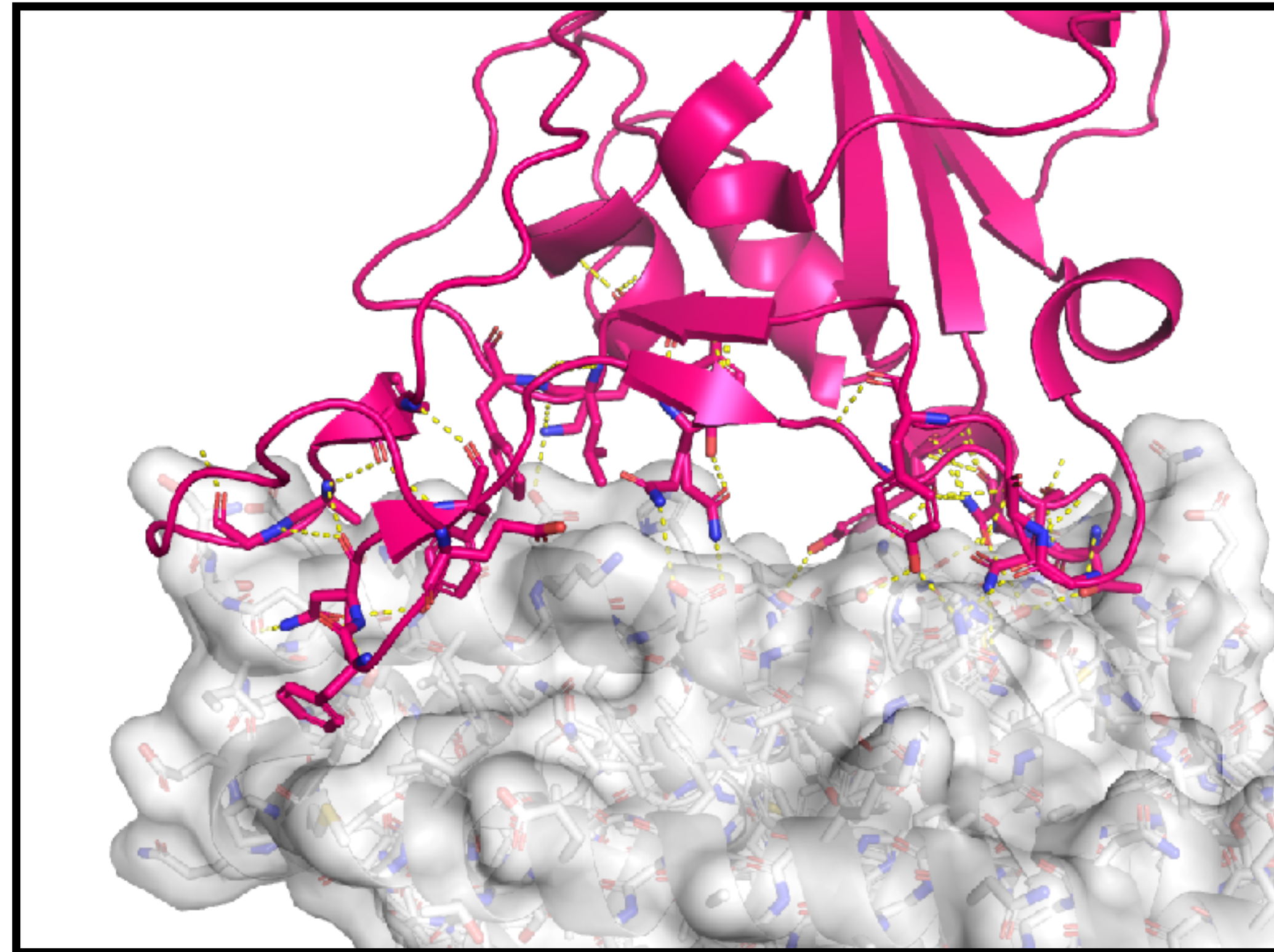
COVID-19



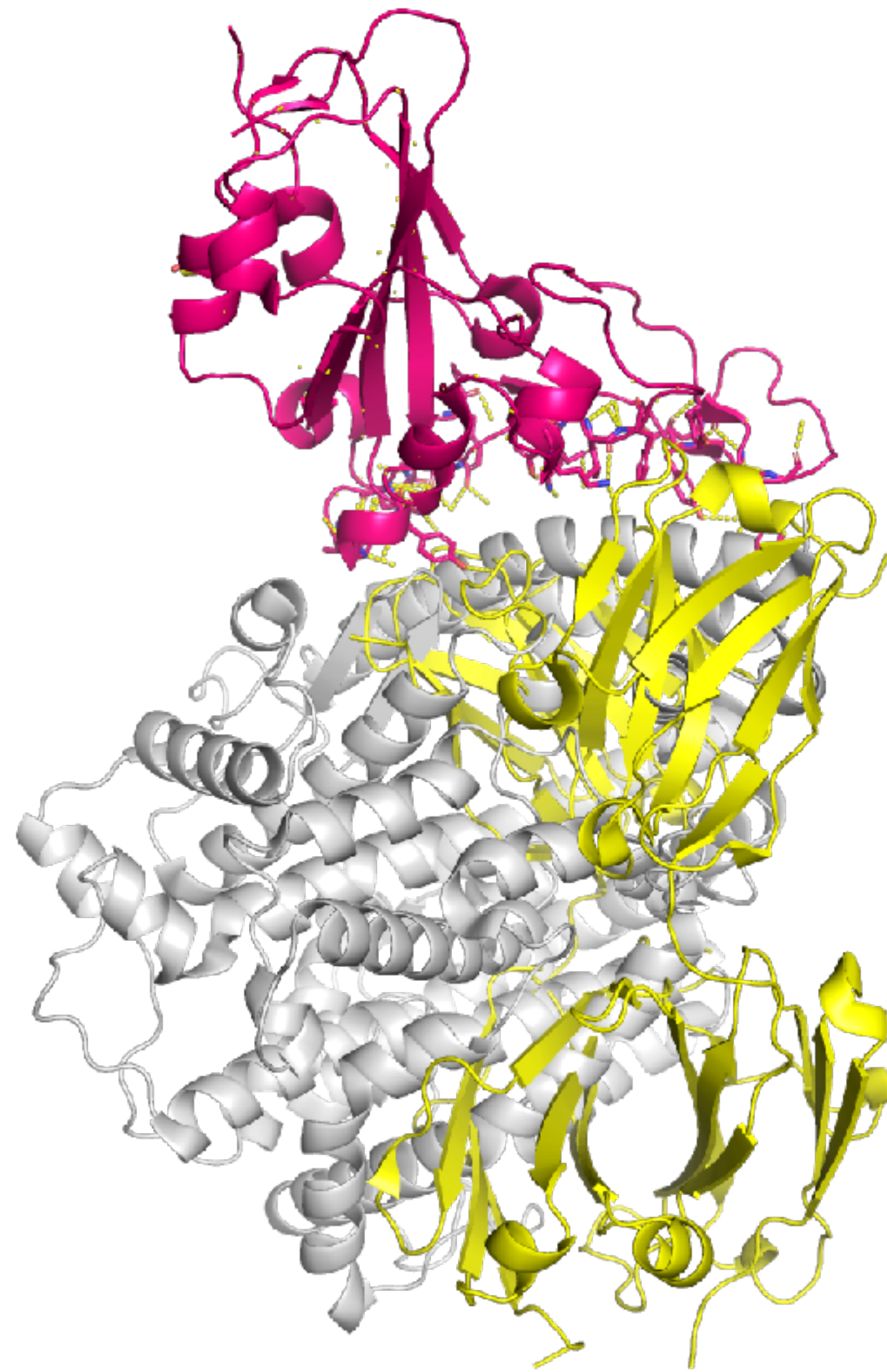
Spike



ACE2 Receptor



COVID-19

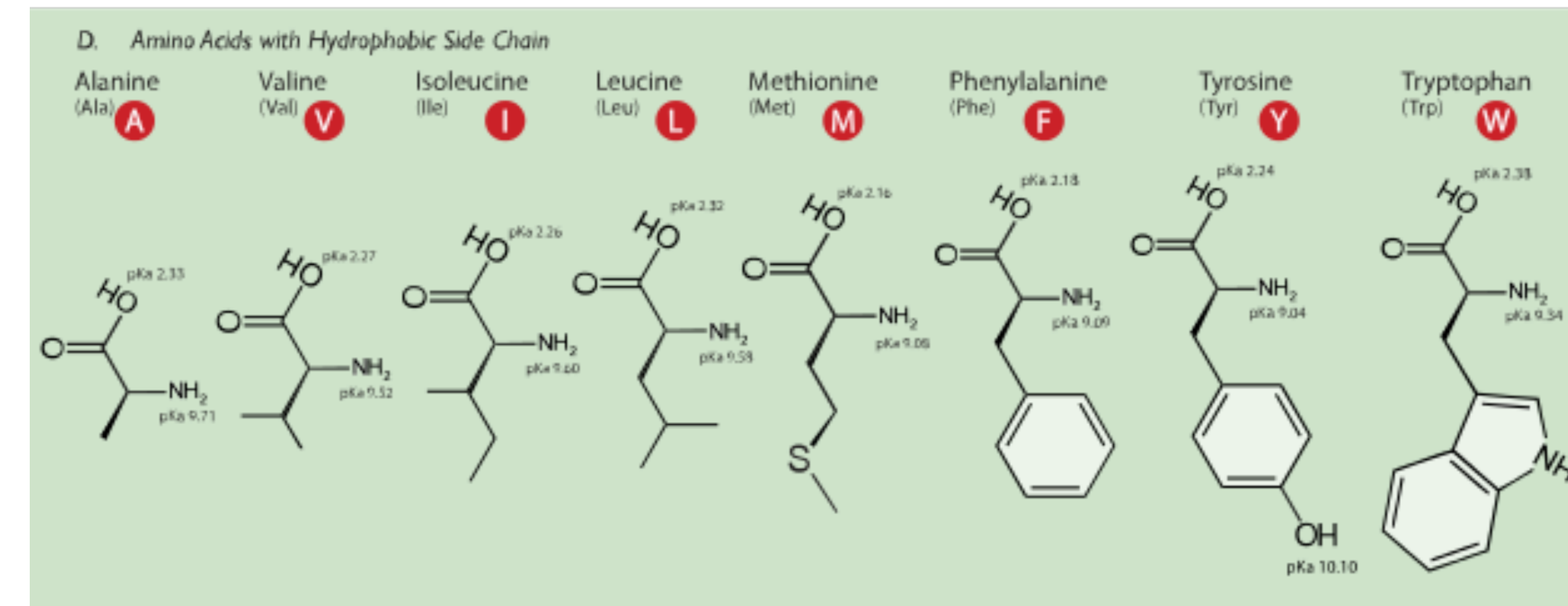
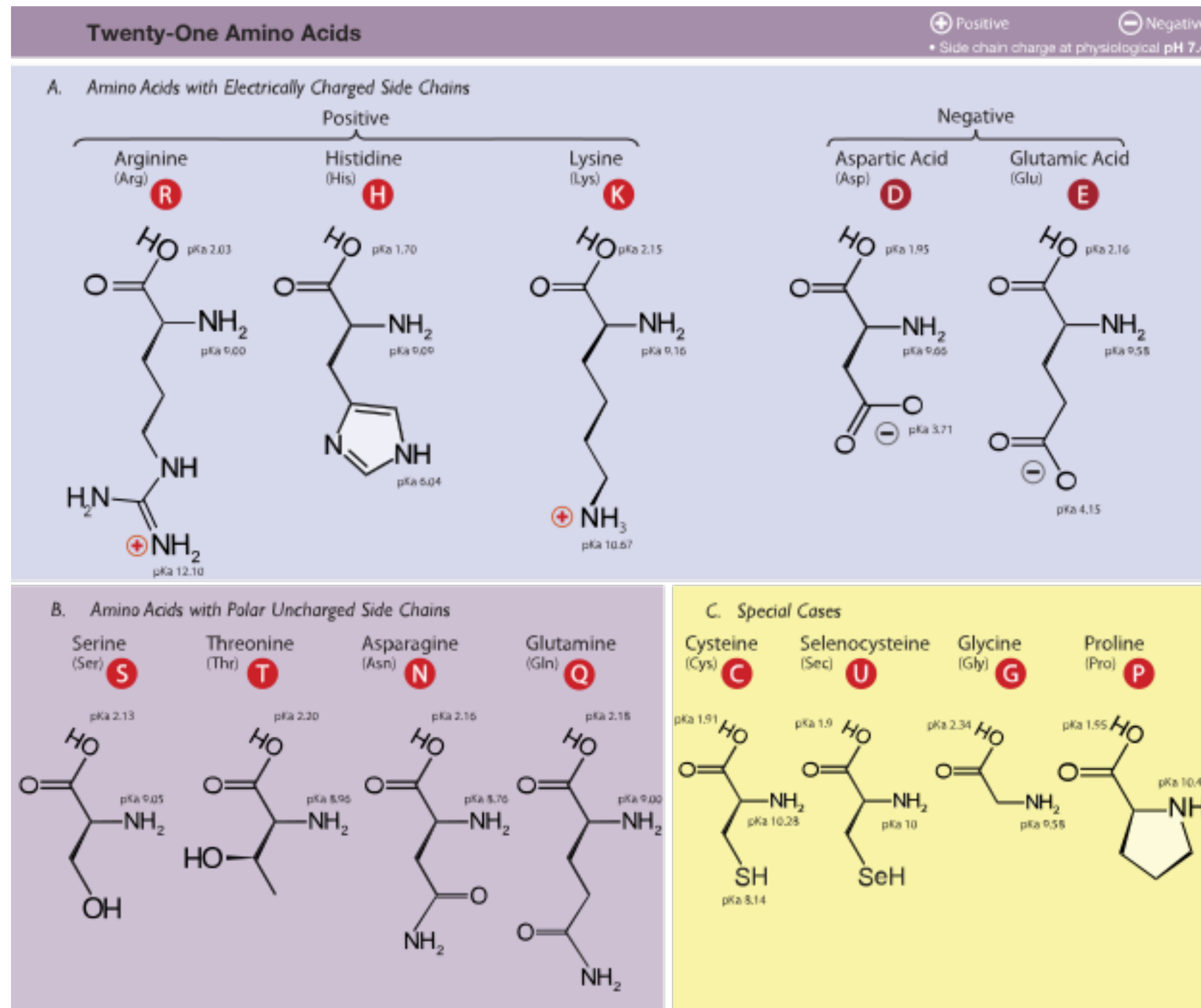


Antibody

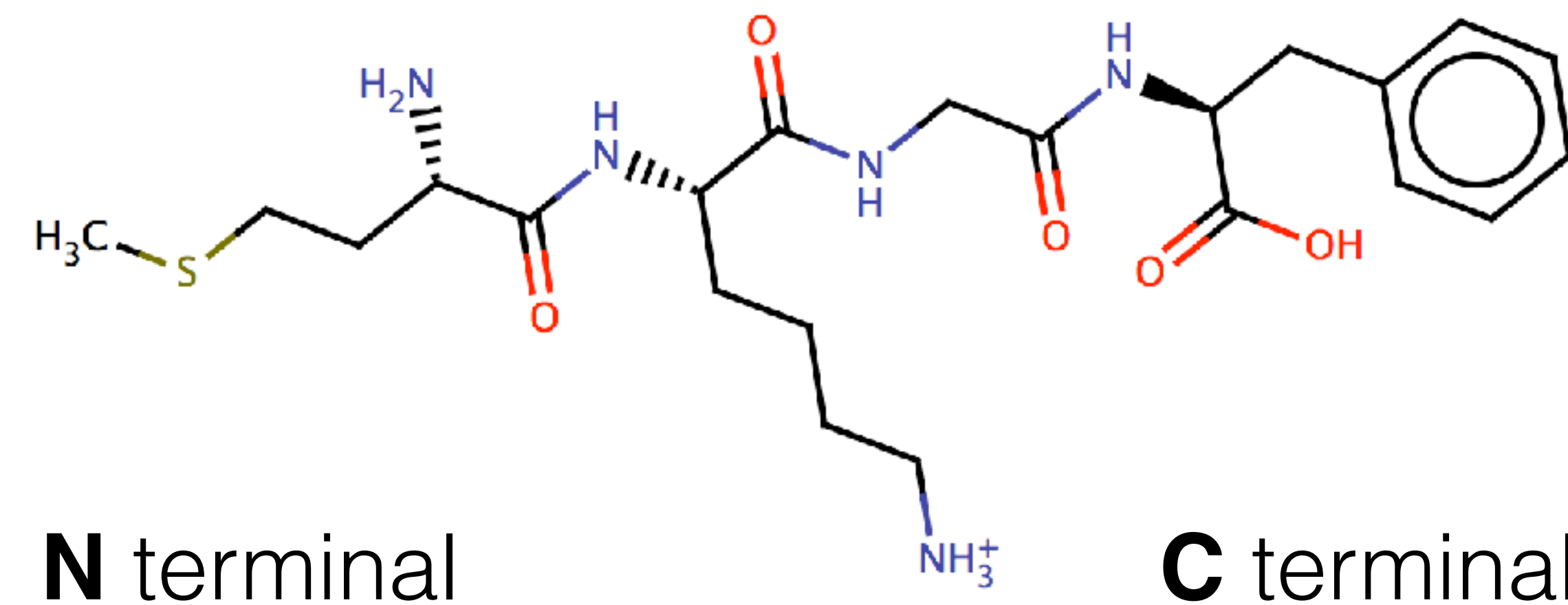


If the trial ultimately shows the treatment is effective against Covid-19, it could be available by autumn, according to the Indianapolis-based company.

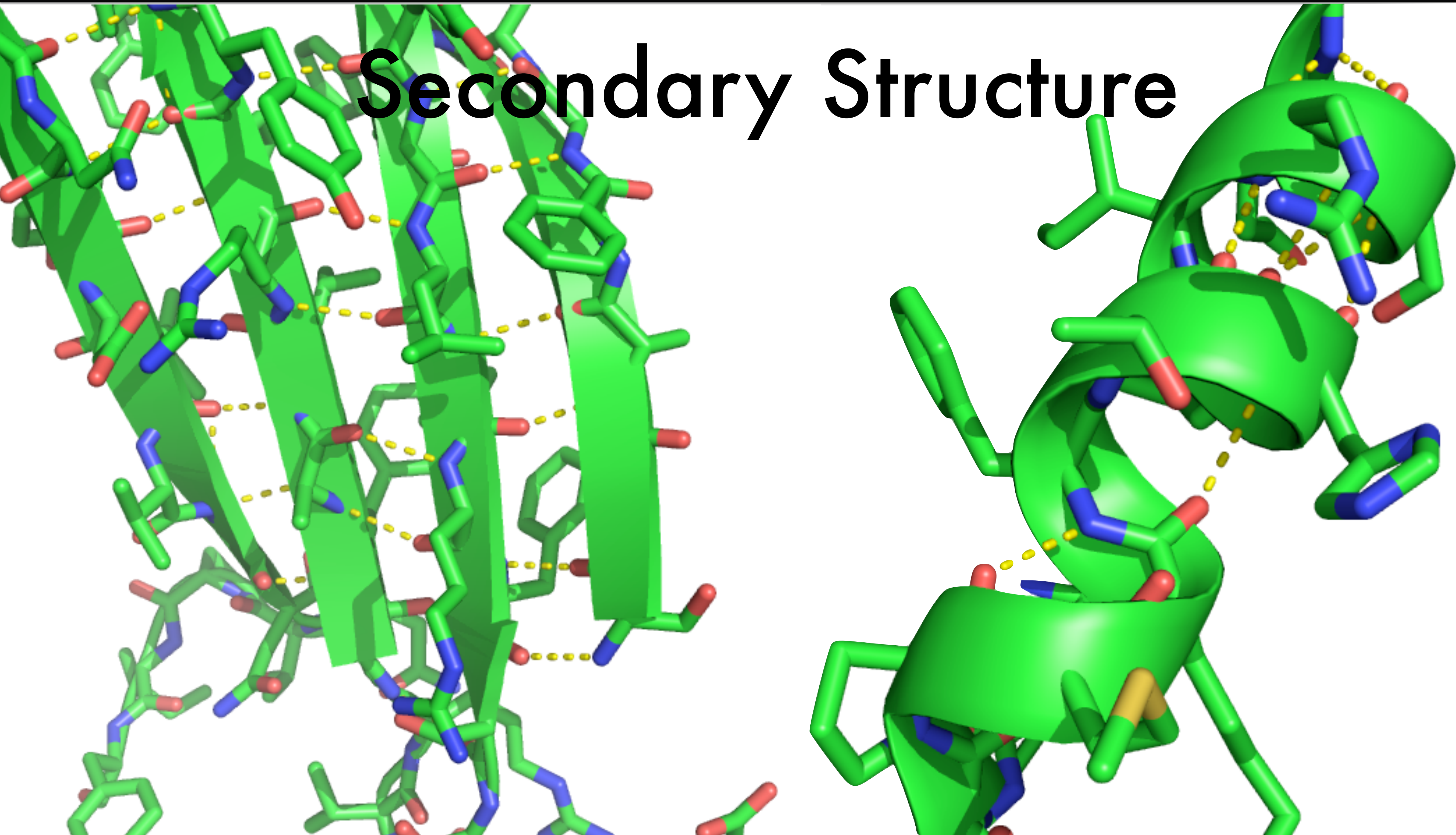
Primary Structure



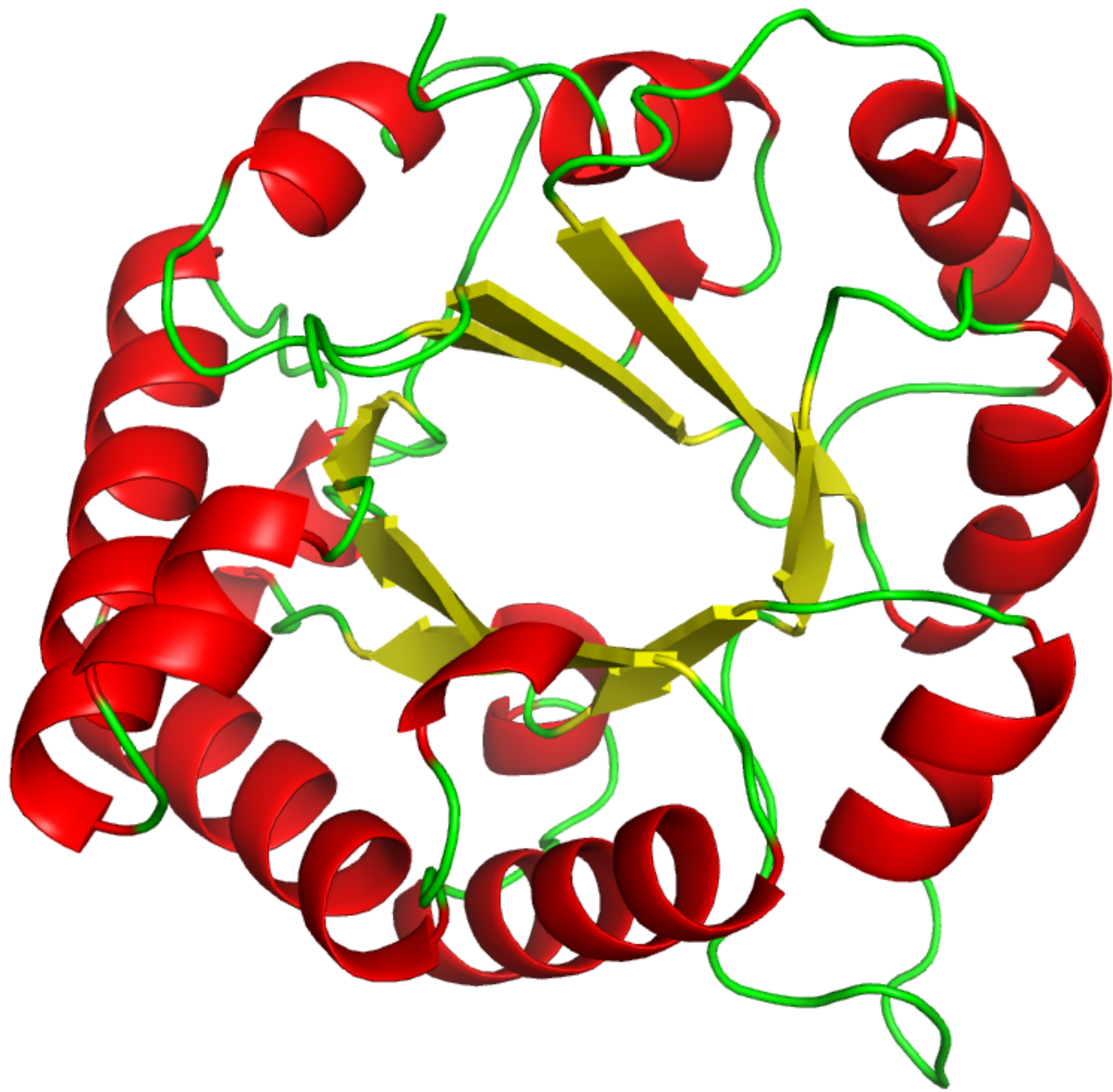
https://en.wikipedia.org/wiki/Amino_acid



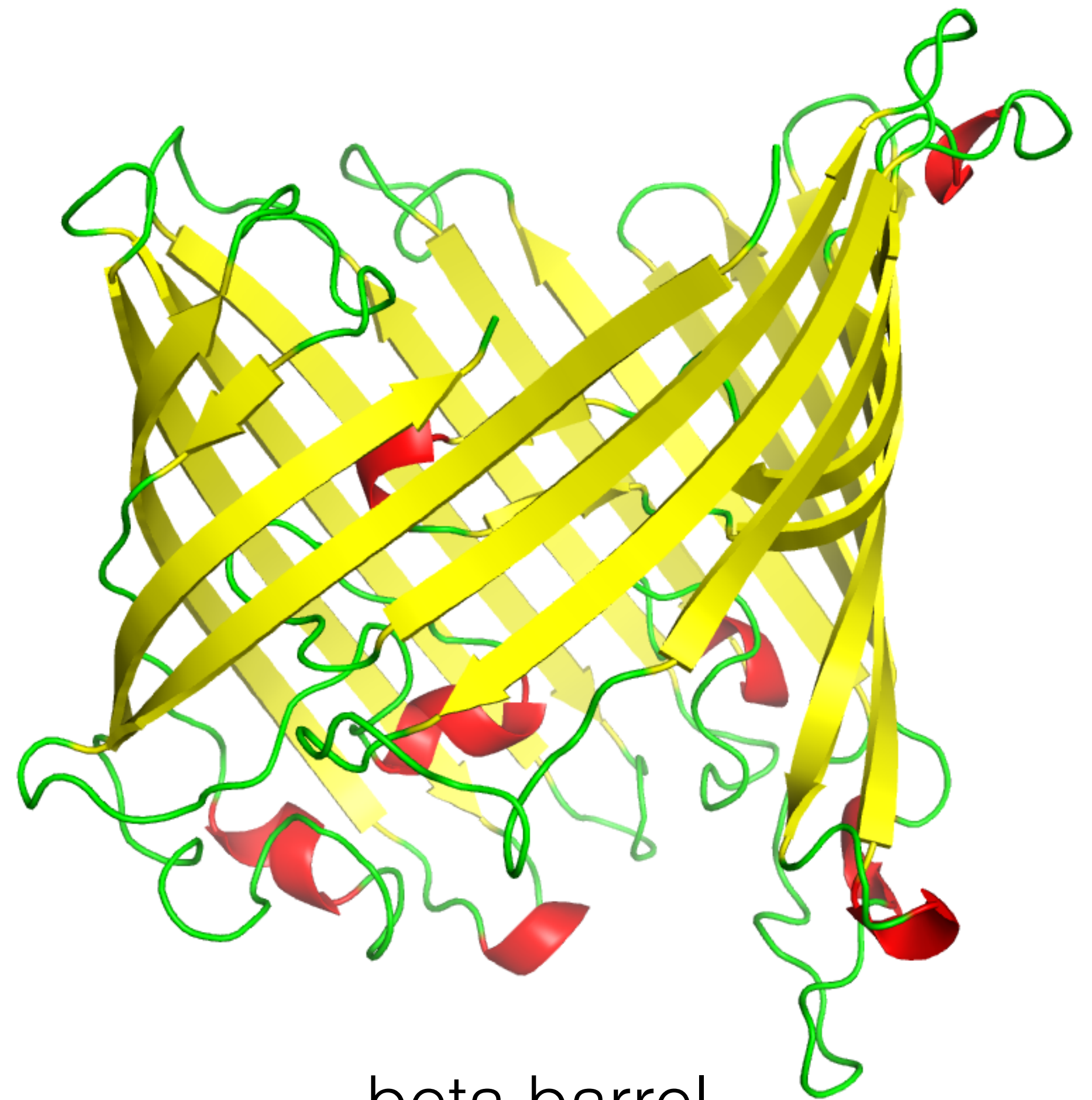
Secondary Structure



Tertiary Structure

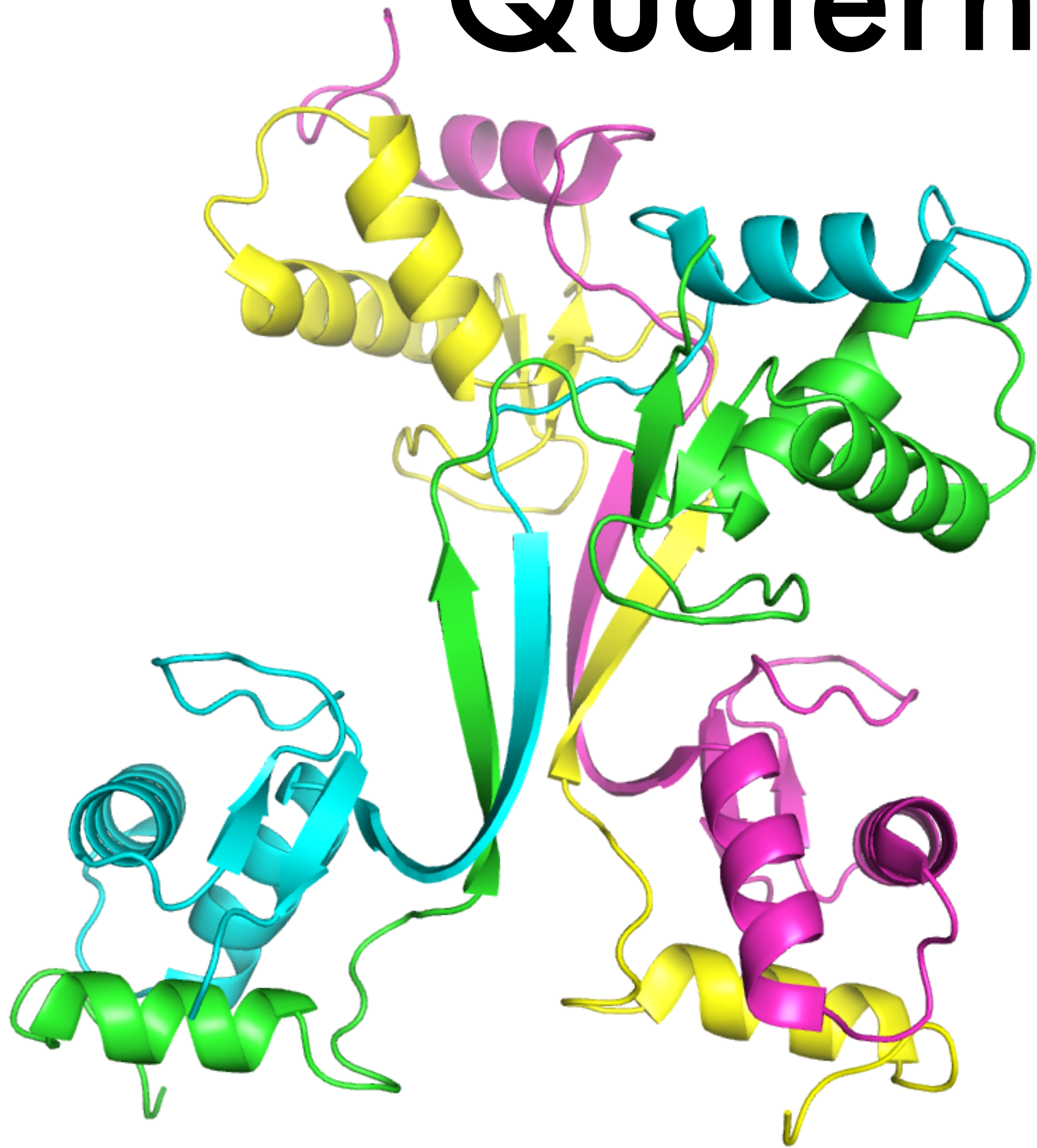


TIM barrel fold

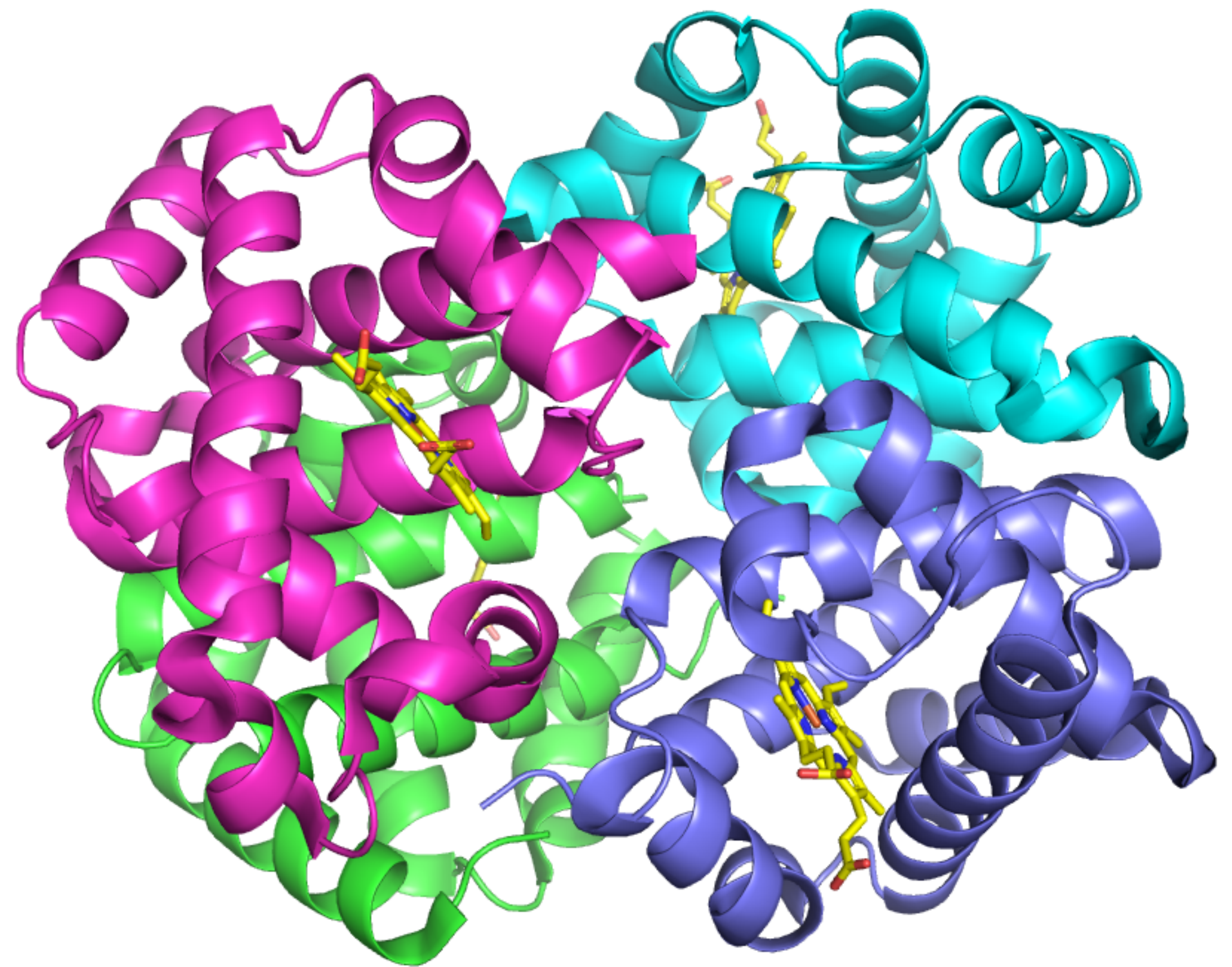


beta barrel

Quaternary Structure

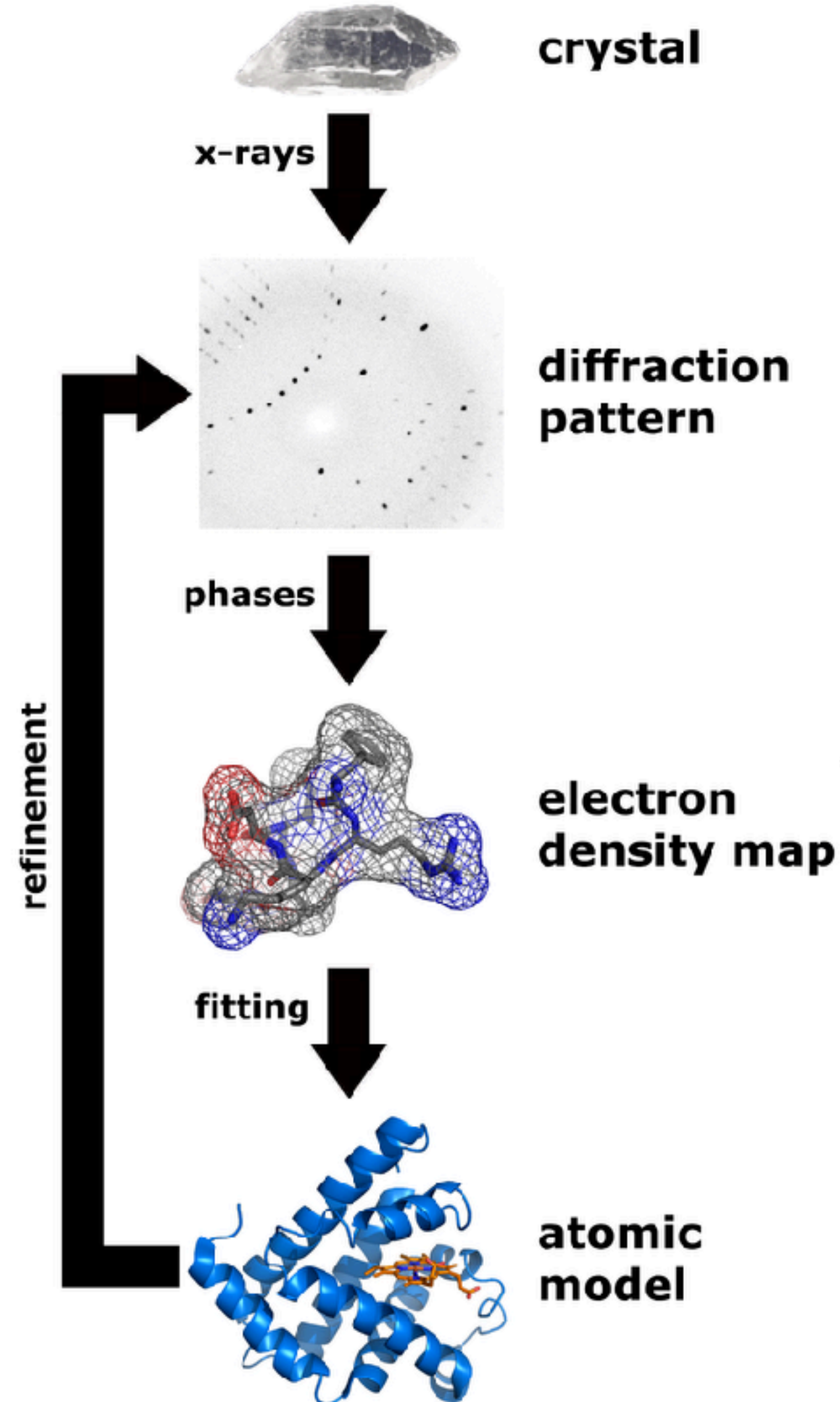


domain swapping



hemoglobin

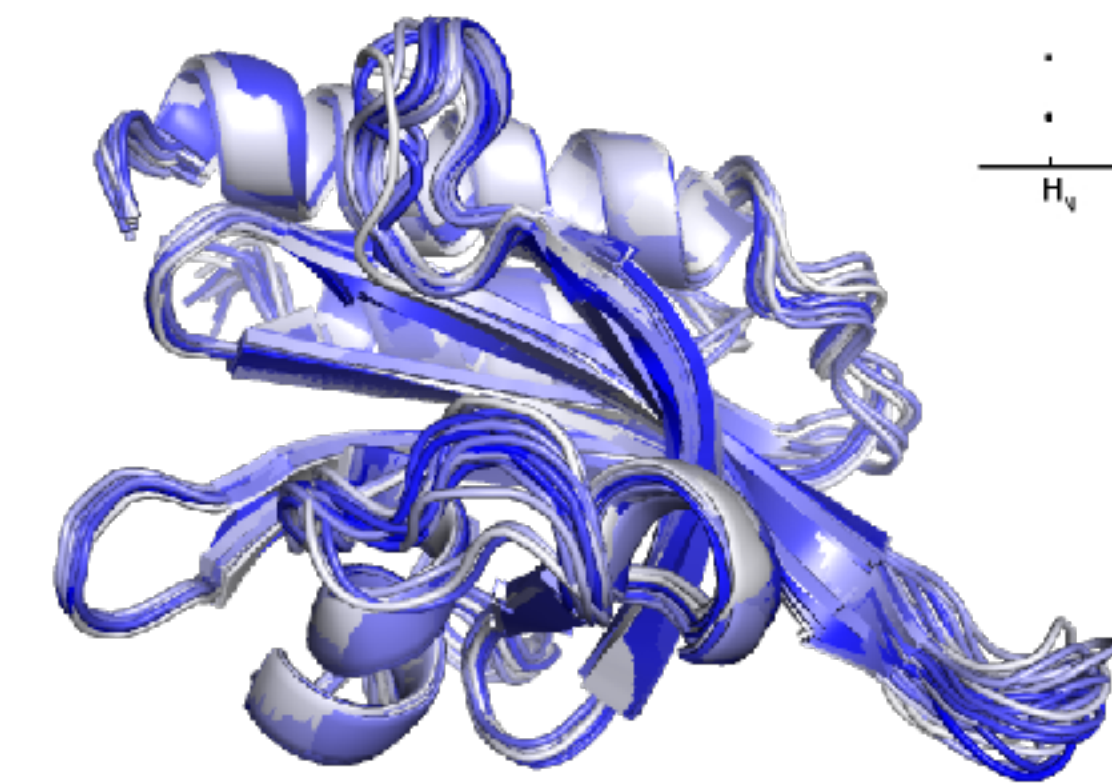
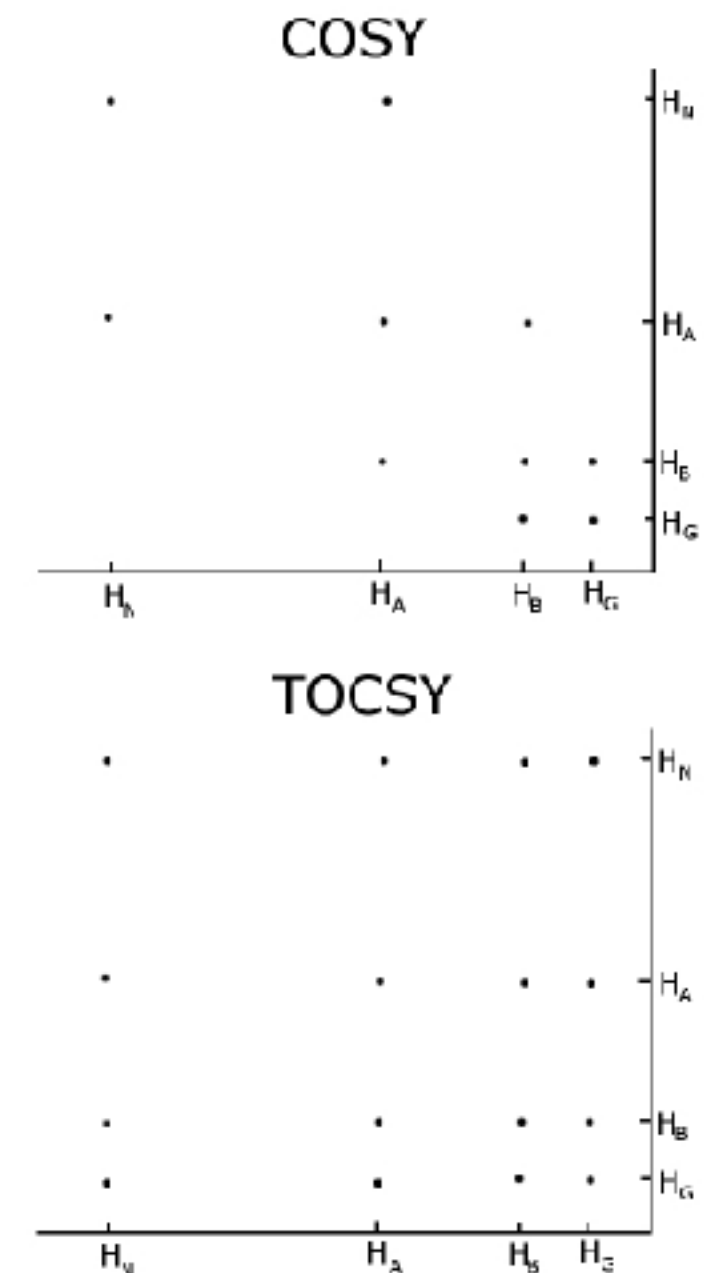
Structure Determination



$$\mathcal{F}\{g(t)\} = G(f) = \int_{-\infty}^{\infty} g(t)e^{-2\pi ift} dt$$

$$\mathcal{F}^{-1}\{G(f)\} = \int_{-\infty}^{\infty} G(f)e^{2\pi ift} df = g(t)$$

x-ray crystallography

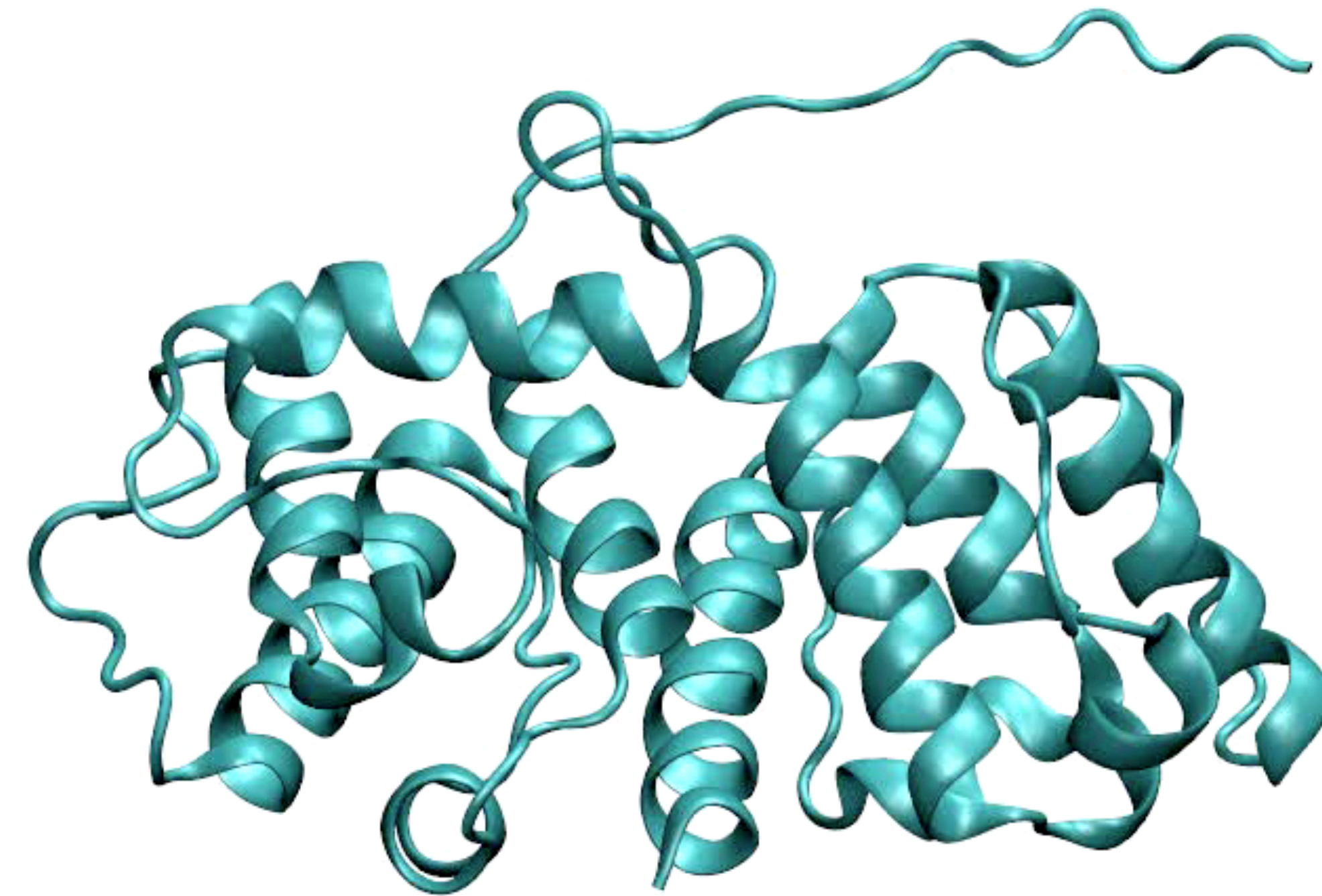
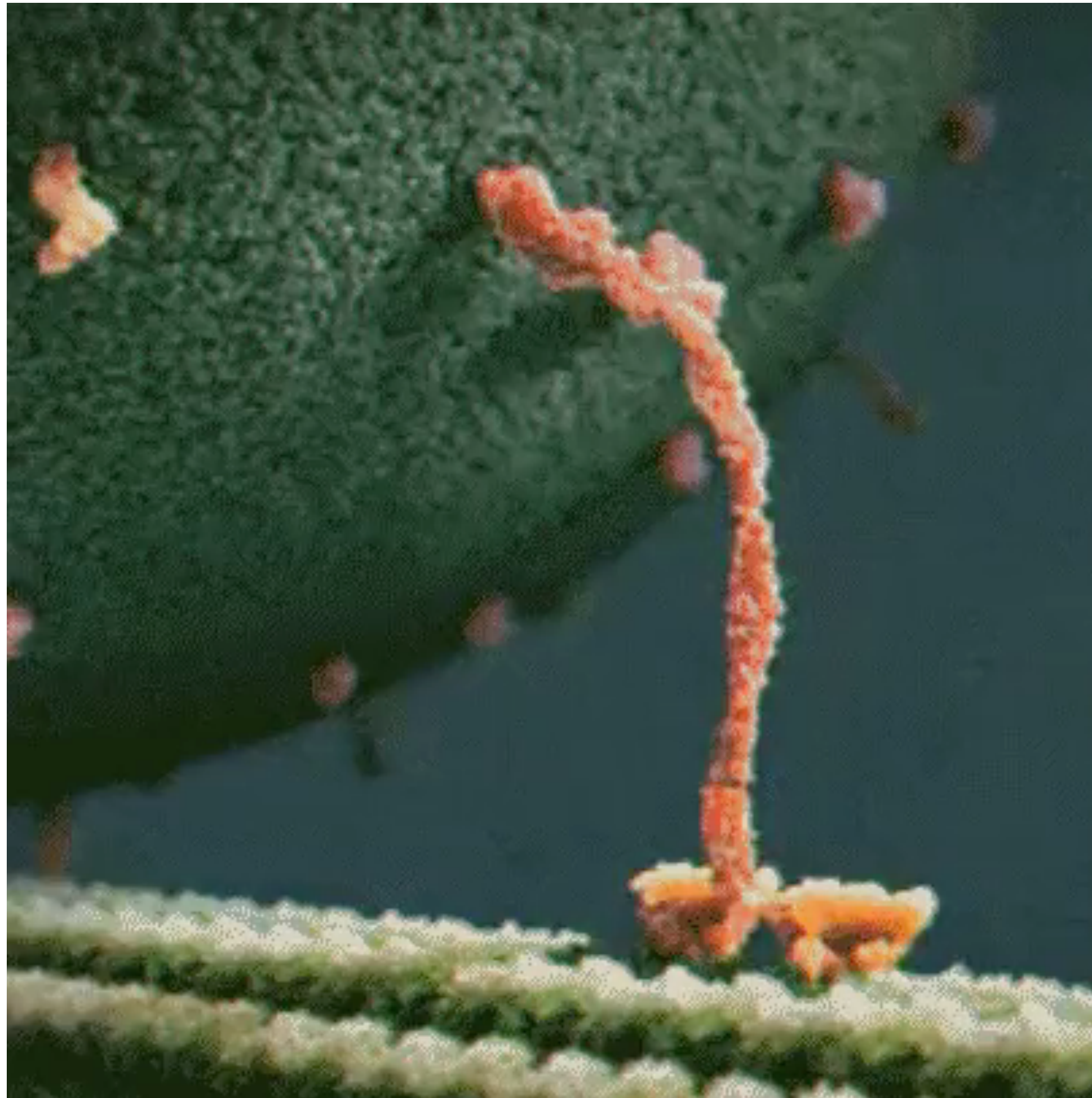


nuclear magnetic resonance

Sequence → Structure → Function

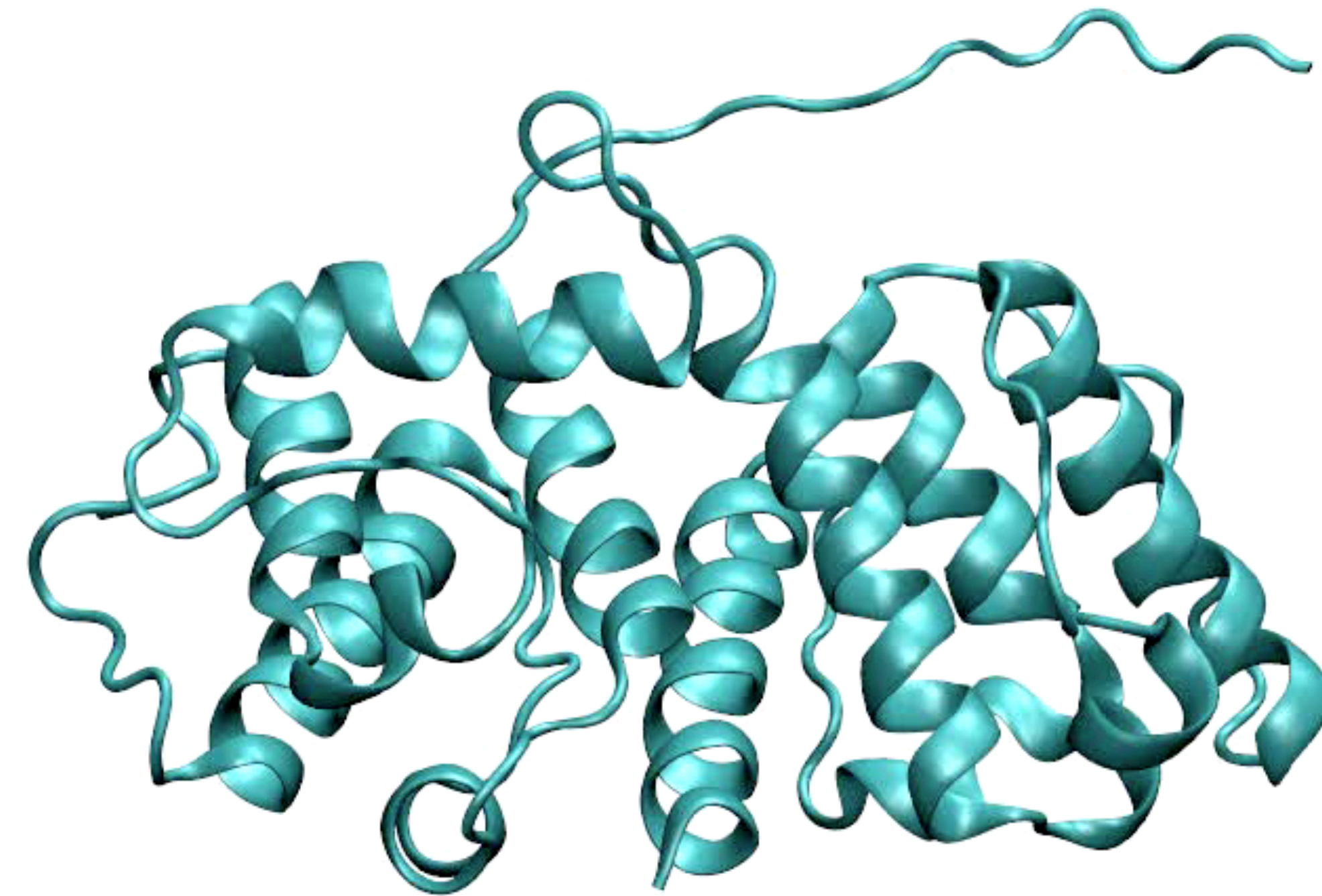
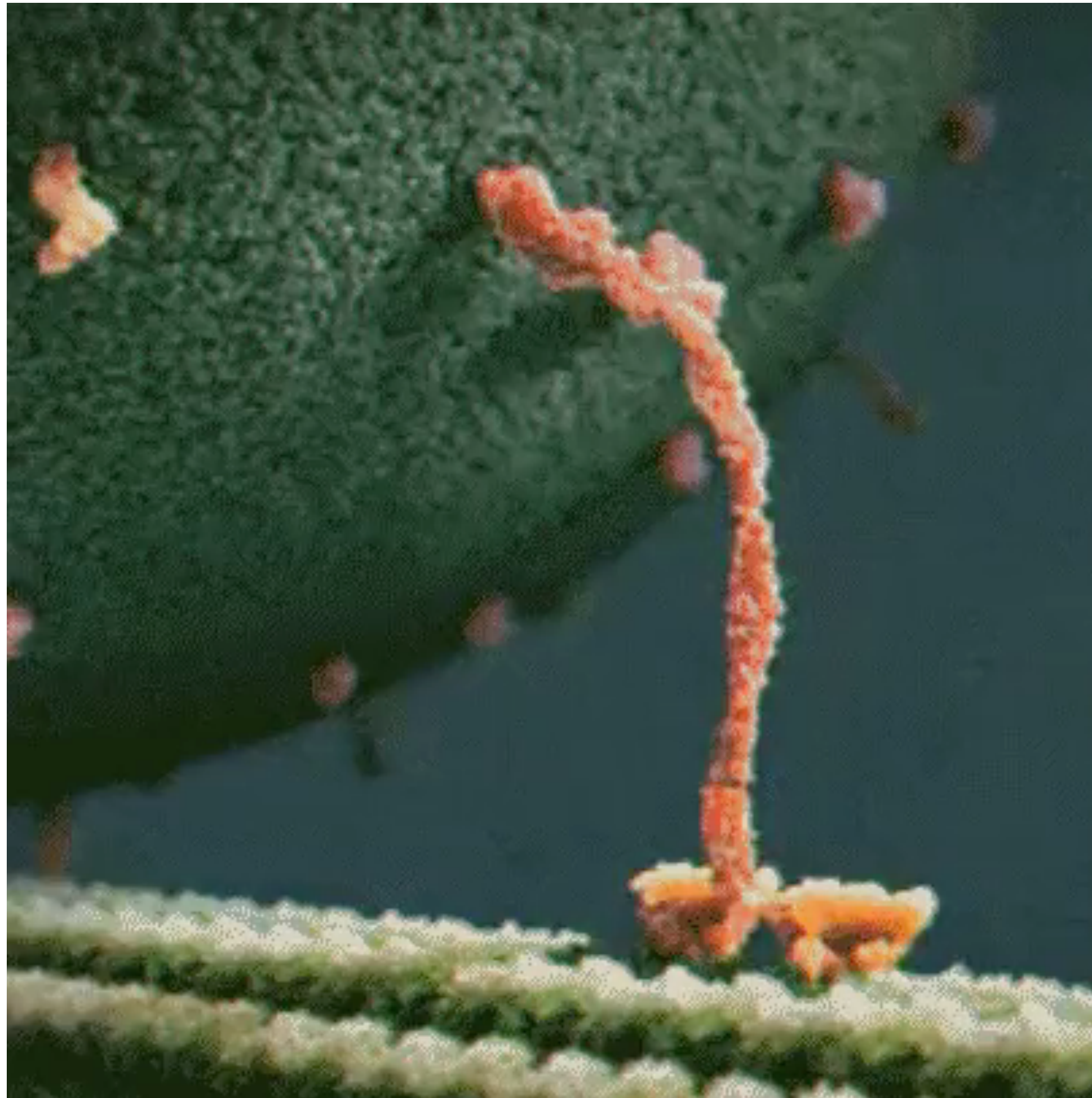
↑

Motion

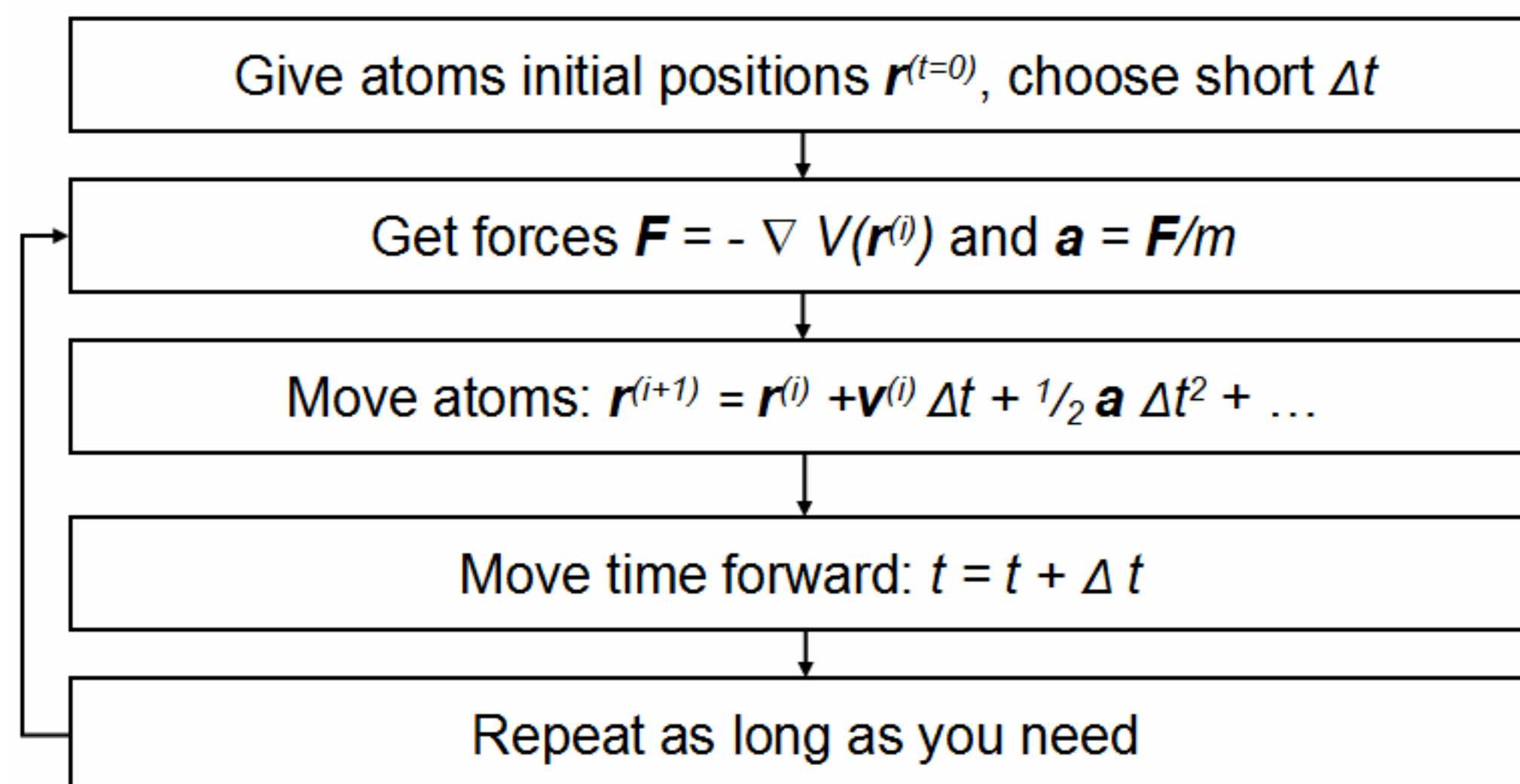


Sequence → Structure → Function

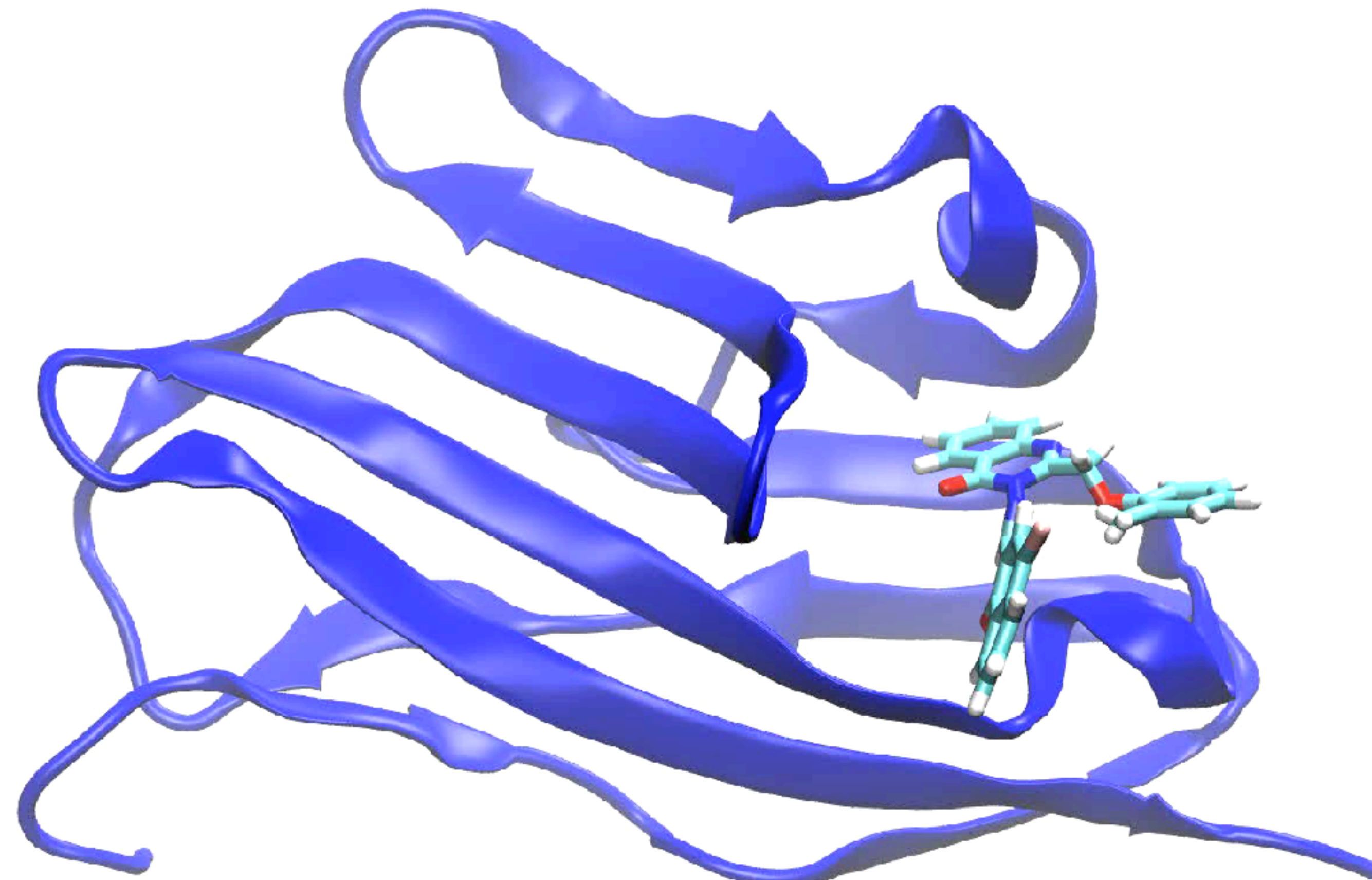
↑
Motion



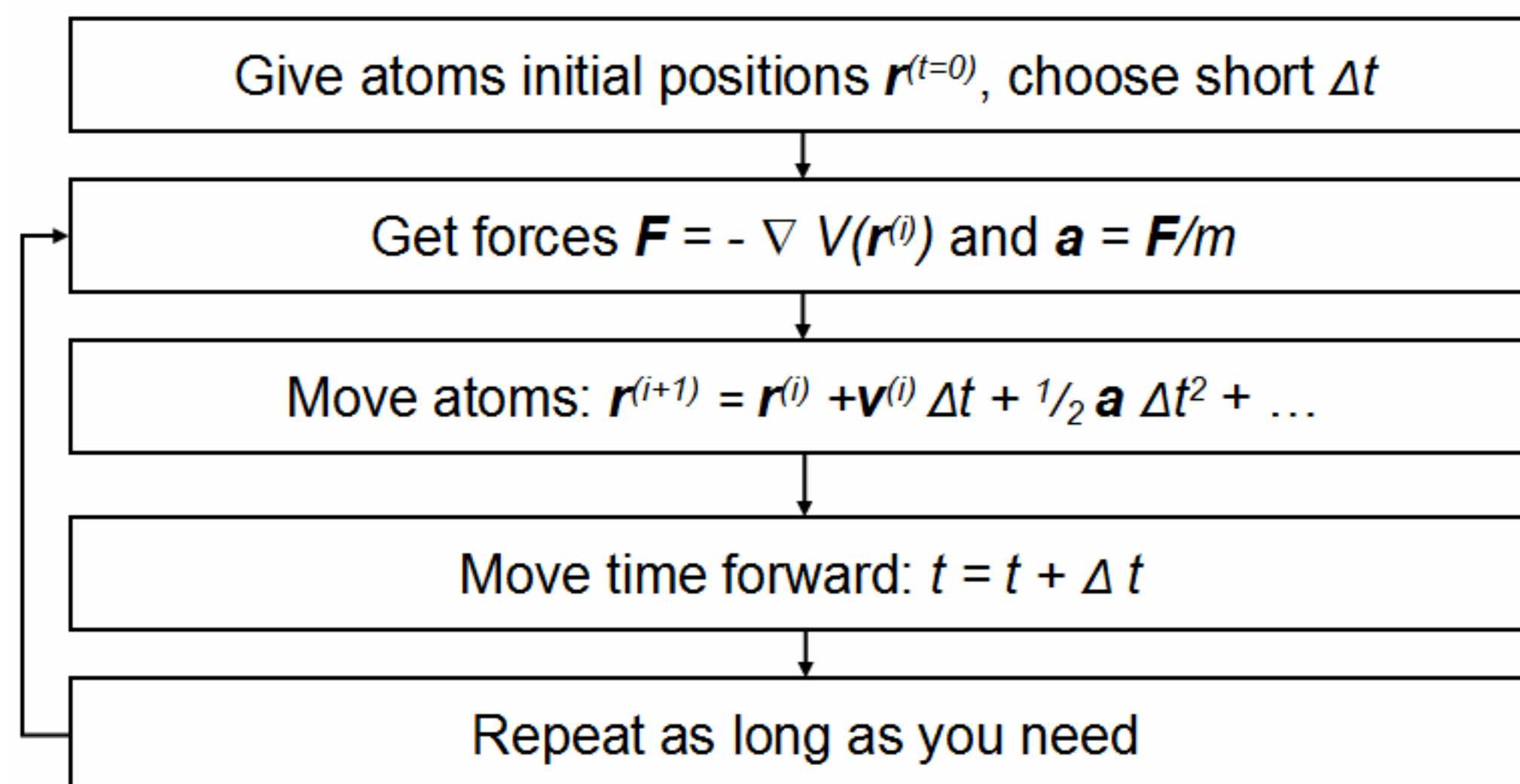
Molecular Dynamics



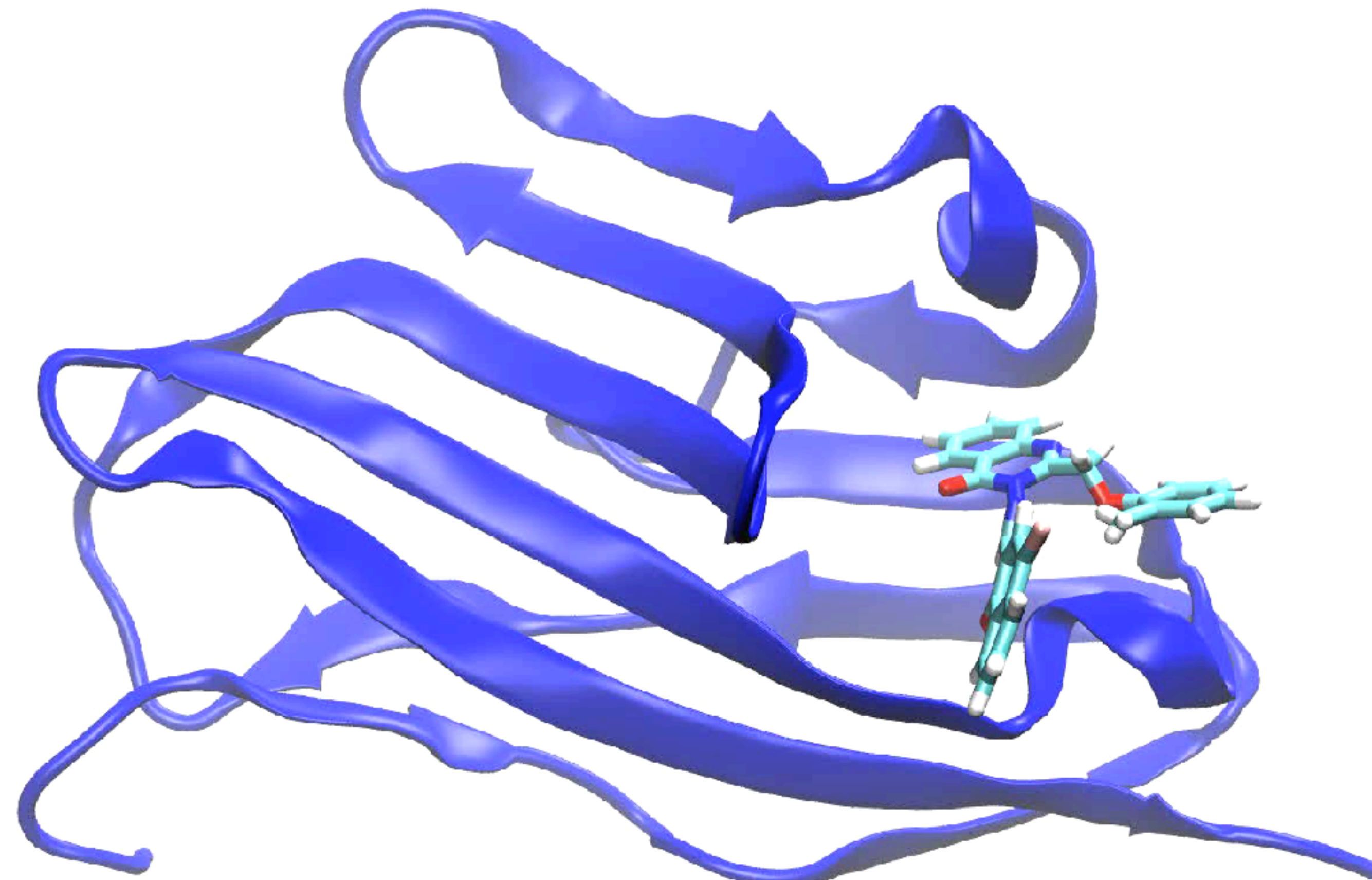
$$\begin{aligned}
 V(\mathbf{r}^N) = & \sum_{\text{bonds}} k_b (l - l_0)^2 + \sum_{\text{angles}} k_a (\theta - \theta_0)^2 \\
 & + \sum_{\text{torsions}} \sum_n \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\
 & + \sum_{j=1}^{N-1} \sum_{i=j+1}^N f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}
 \end{aligned}$$



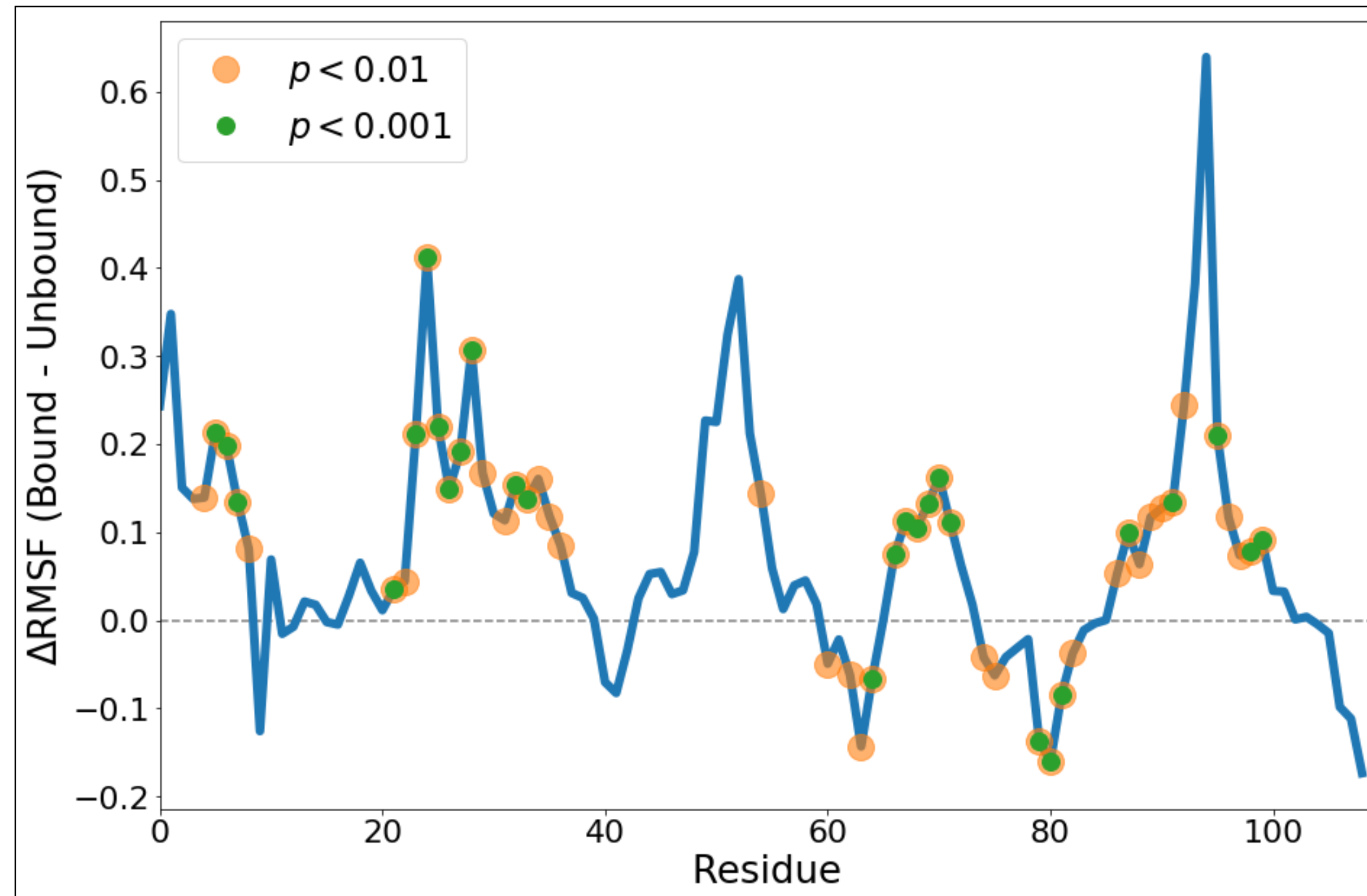
Molecular Dynamics



$$\begin{aligned}
 V(\mathbf{r}^N) = & \sum_{\text{bonds}} k_b (l - l_0)^2 + \sum_{\text{angles}} k_a (\theta - \theta_0)^2 \\
 & + \sum_{\text{torsions}} \sum_n \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\
 & + \sum_{j=1}^{N-1} \sum_{i=j+1}^N f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}
 \end{aligned}$$



Dynamics Analysis



more stable

less stable

Protein Folding



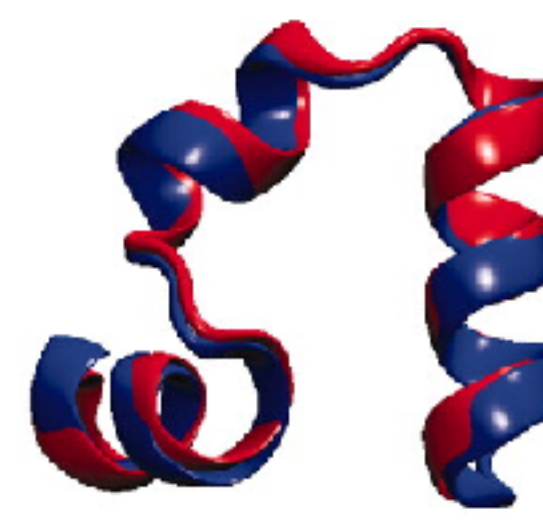
Chignolin 106 μ s
cln025 1.0 Å 0.6 μ s



Trp-cage 208 μ s
2JOF 1.4 Å 14 μ s



BBA 325 μ s
1FME 1.6 Å 18 μ s



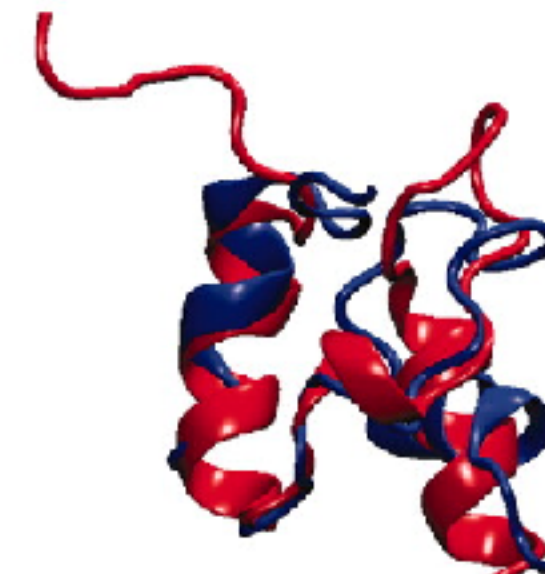
Villin 125 μ s
2F4K 1.3 Å 2.8 μ s



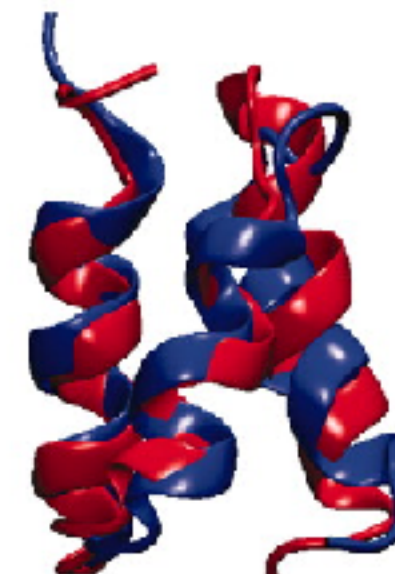
WW domain 1137 μ s
2F21 1.2 Å 21 μ s



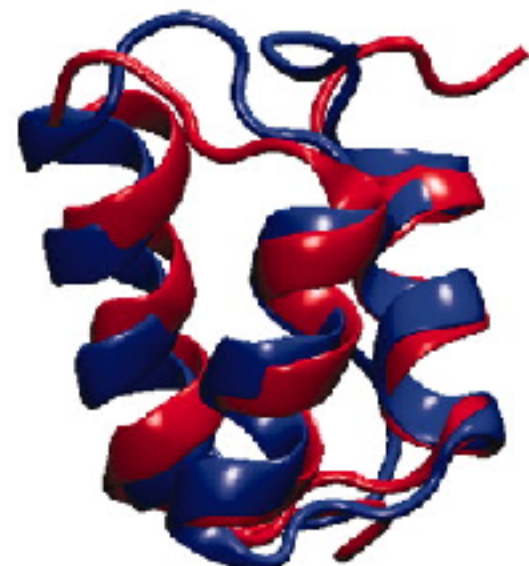
NTL9 2936 μ s
2HBA 0.5 Å 29 μ s



BBL 429 μ s
2WXC 4.8 Å 29 μ s



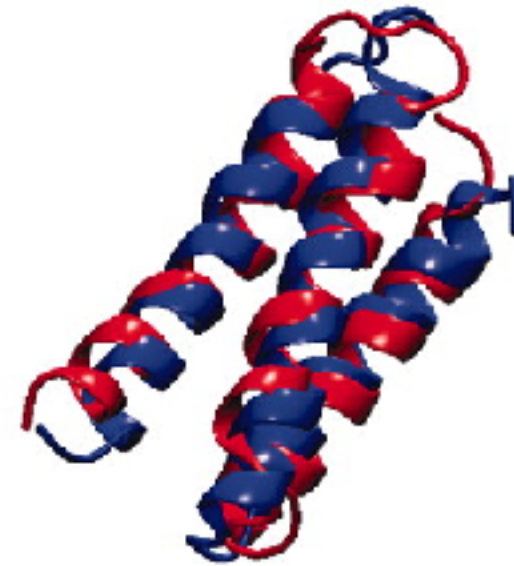
Protein B 104 μ s
1PRB 3.3 Å 3.9 μ s



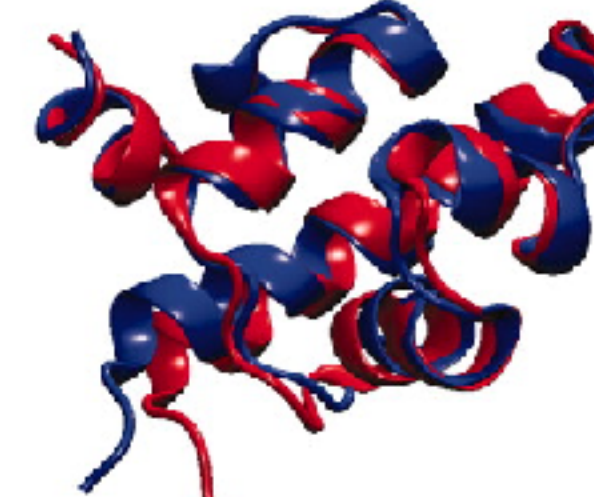
Homeodomain 327 μ s
2P6J 3.6 Å 3.1 μ s



Protein G 1154 μ s
1MIO 1.2 Å 65 μ s



α 3D 707 μ s
2A3D 3.1 Å 27 μ s



λ -repressor 643 μ s
1LMB 1.8 Å 49 μ s



[https://en.wikipedia.org/wiki/Anton_\(computer\)](https://en.wikipedia.org/wiki/Anton_(computer))

How Fast-Folding Proteins Fold

Kresten Lindorff-Larsen^{1,*†}, Stefano Piana^{1,*†}, Ron O. Dror¹, David E. Shaw^{1,2†}

¹D. E. Shaw Research, New York, NY 10036, USA.

²Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA.

[†]To whom correspondence should be addressed. E-mail: david.shaw@DEShawResearch.com (D.E.S.); kresten.lindorff-larsen@DEShawResearch.com (K.L.-L.); stefano.piana-agostinetti@DEShawResearch.com (S.P.)

* These authors contributed equally to the manuscript.

+ See all authors and affiliations

Science 28 Oct 2011:
Vol. 334, Issue 6055, pp. 517-520
DOI: 10.1126/science.1208351

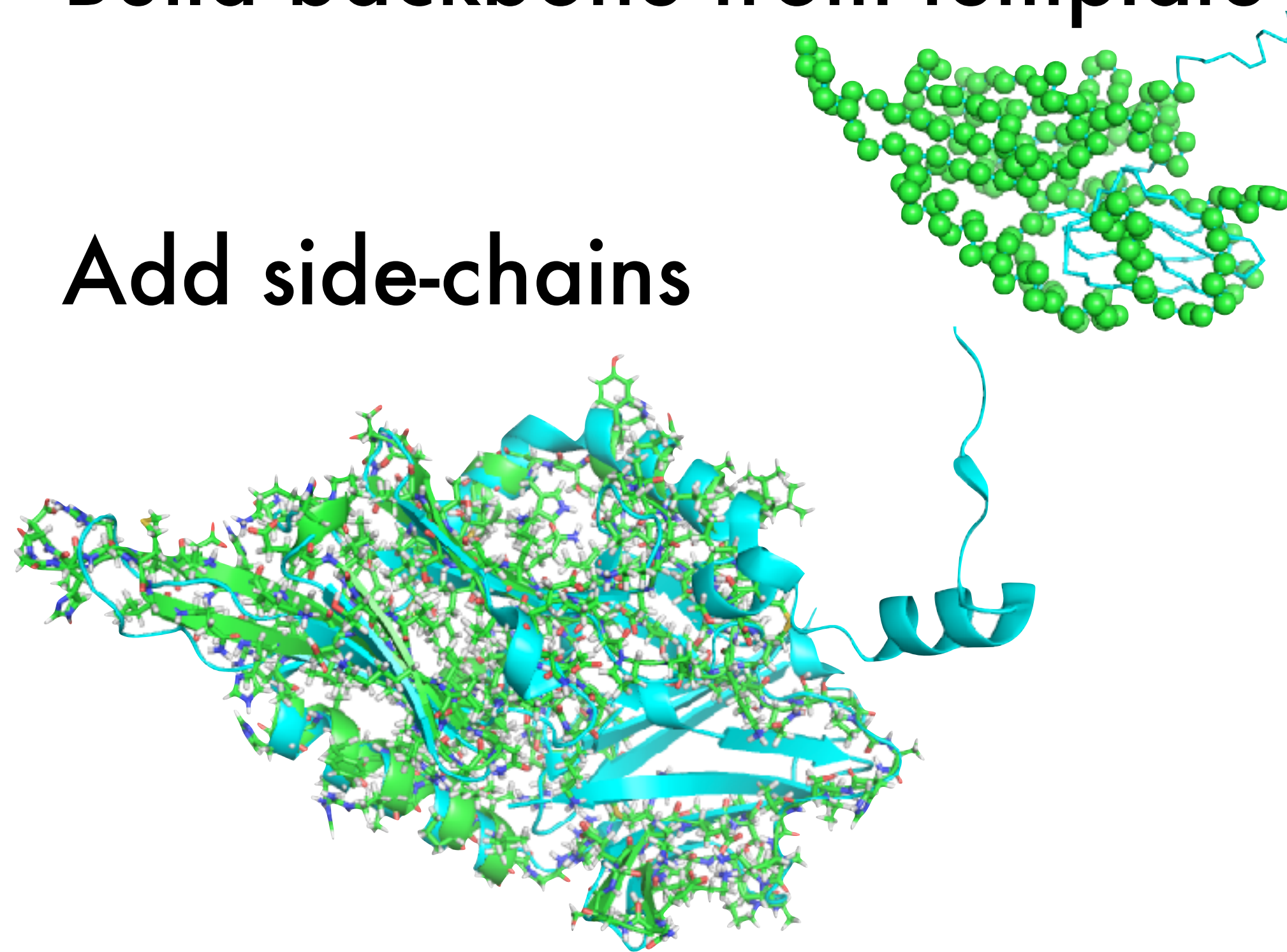
Homology Modeling

Find sequence alignment to template

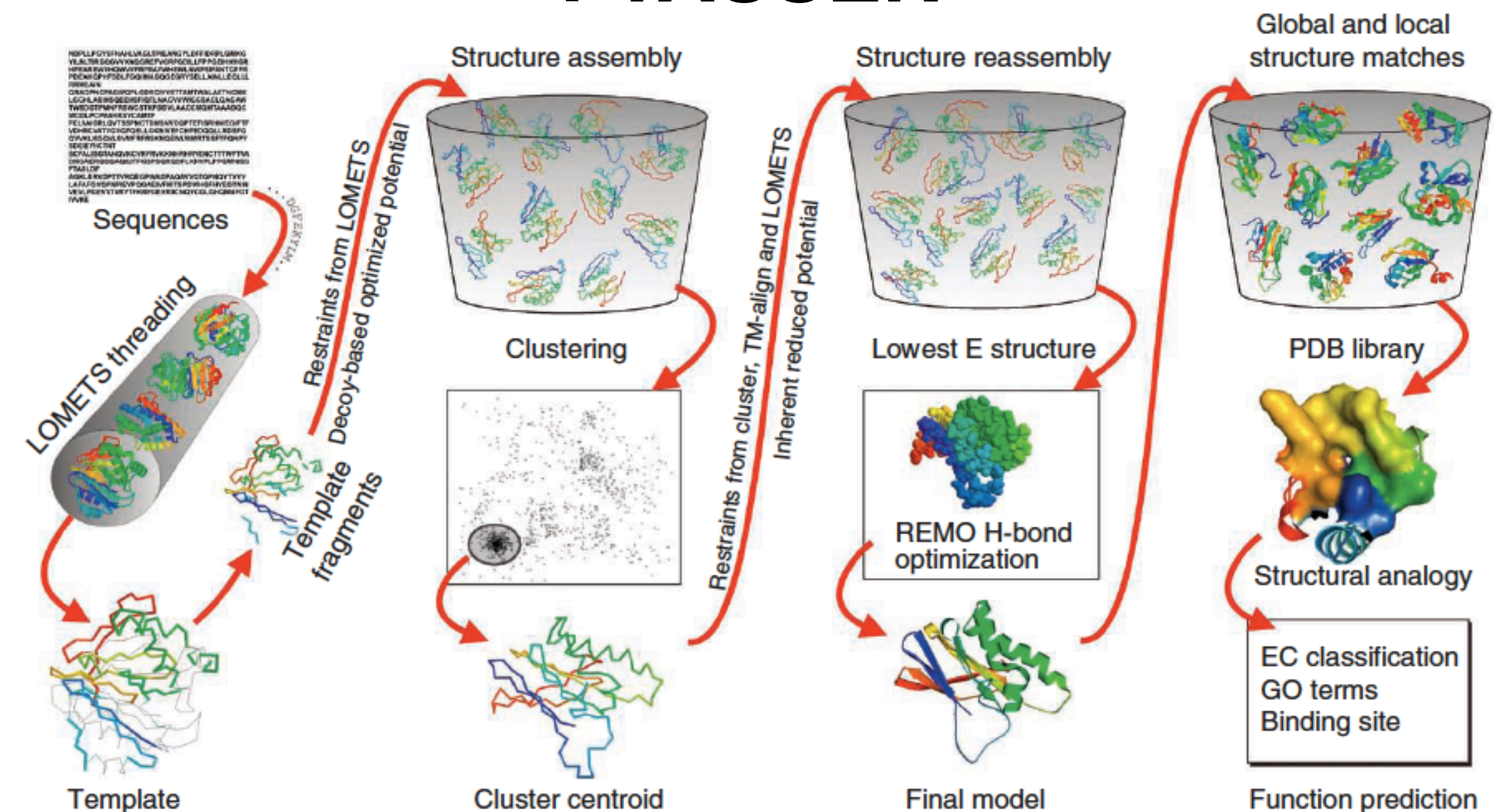
template `HPRRLLLPNTLGLGRRRYSPIYERSHGHHNQMSRRASGGPNALLPAVGKDGQVCMQVDSQFKPNELTVKVVDNTV`
`NAFESVMKEMSAIQPREFHPELEYTQP----GELDFLKDAYEVGKDGRLHFQVYFNVKNFKAEIITIKADKNKL`

Build backbone from template (loops??)

Add side-chains



I-TASSER



<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>

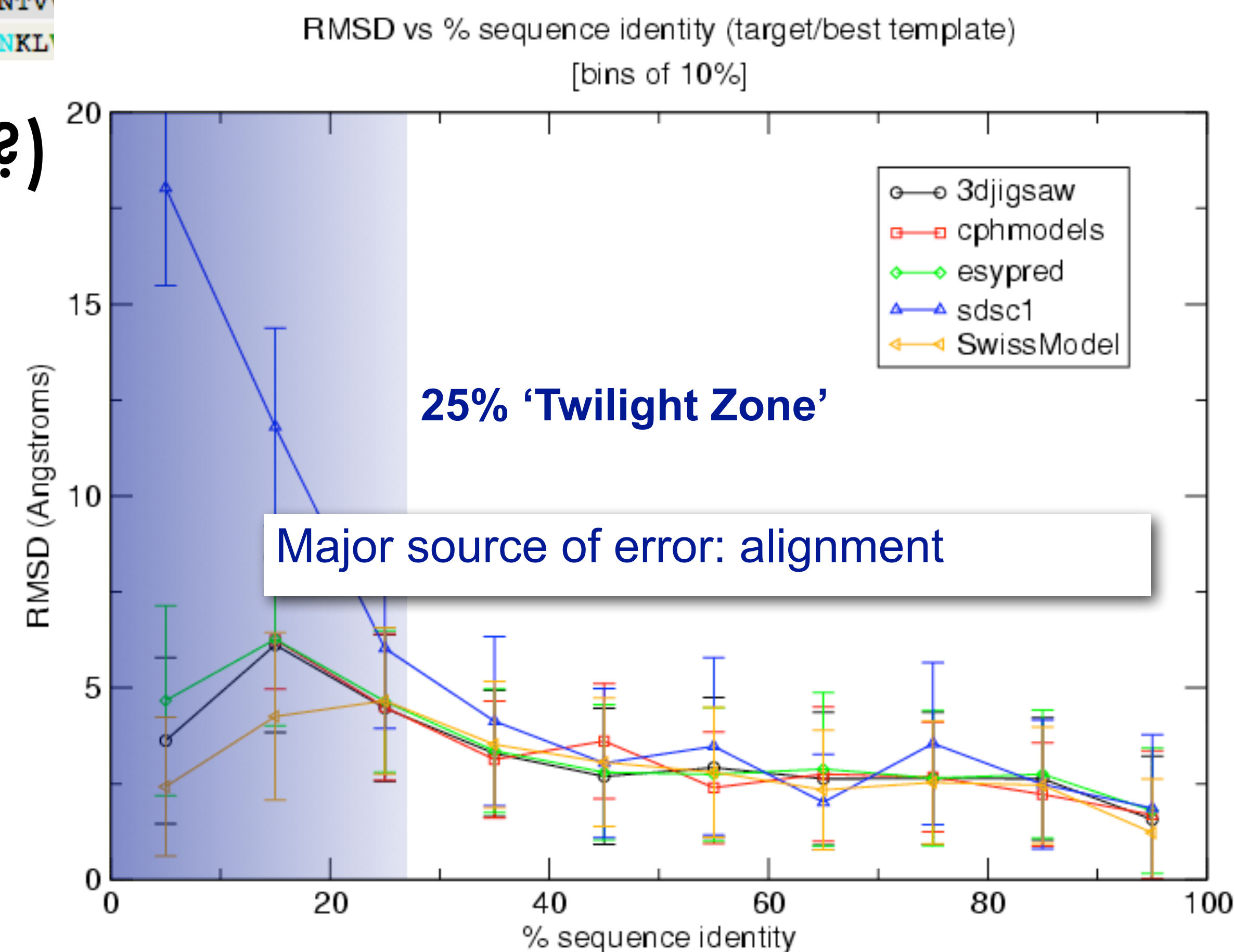
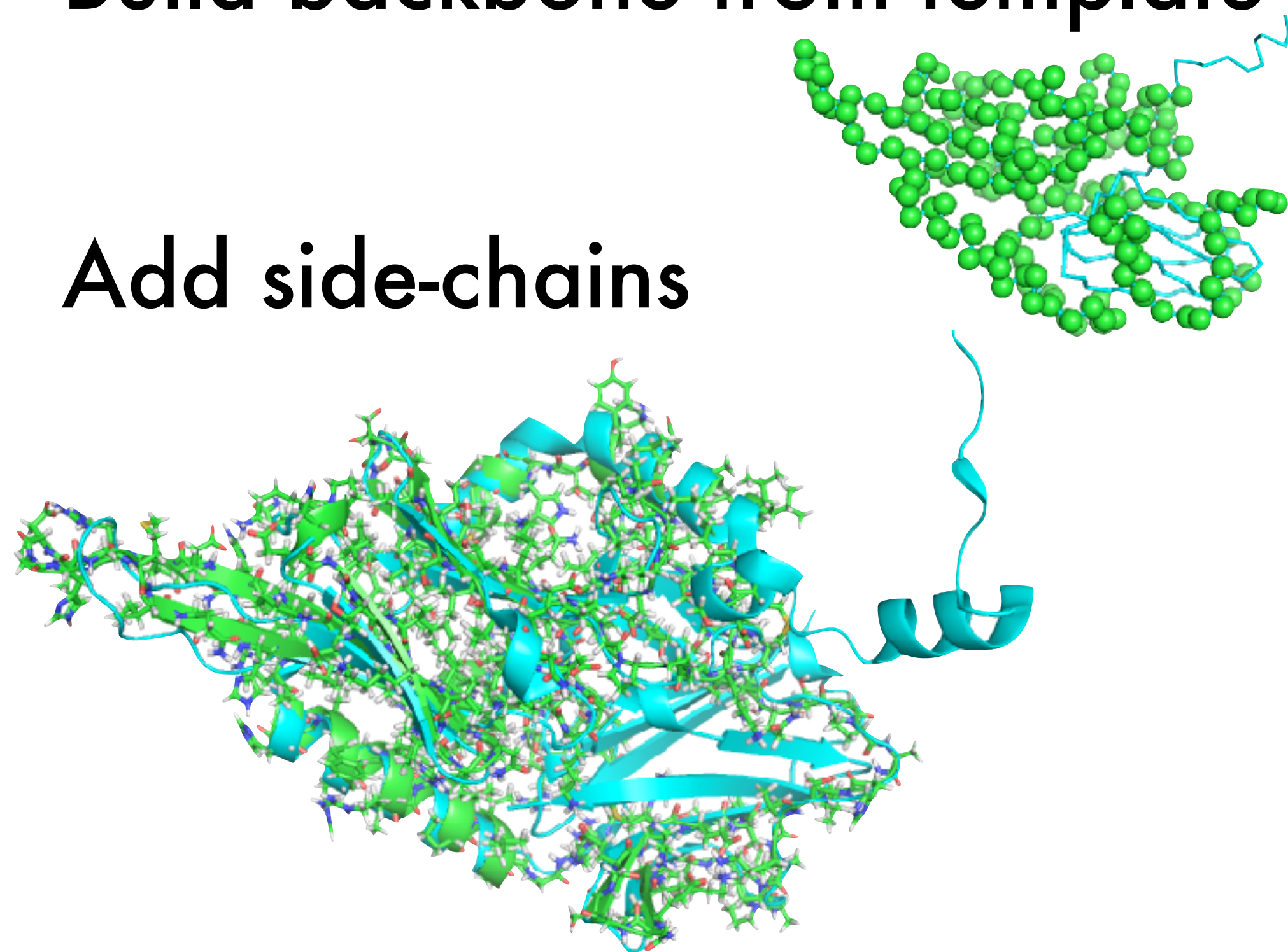
Homology Modeling

Find sequence alignment to template

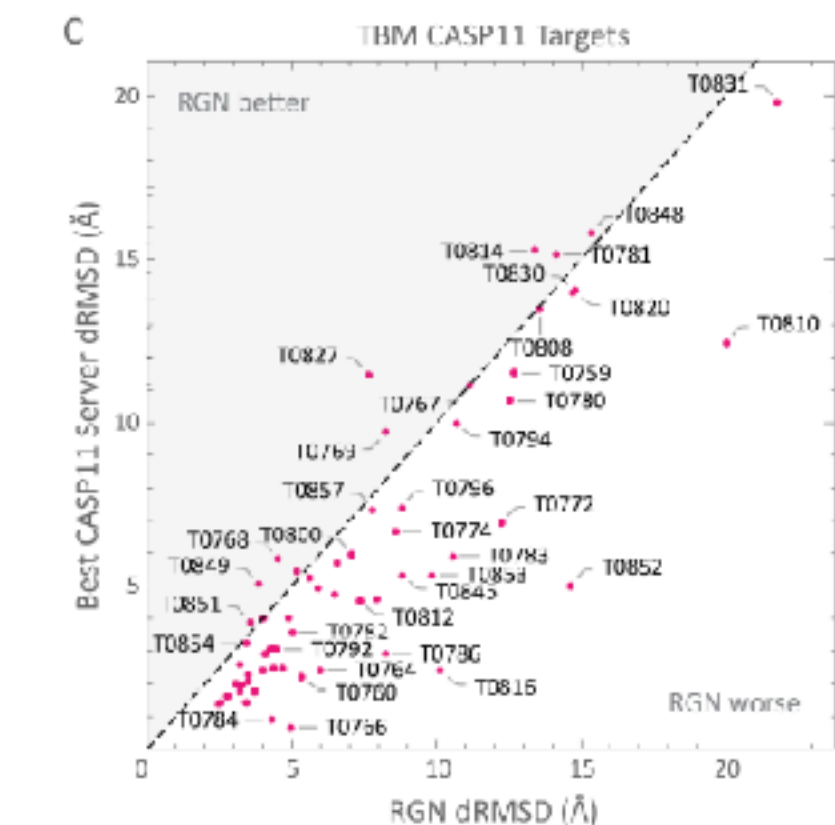
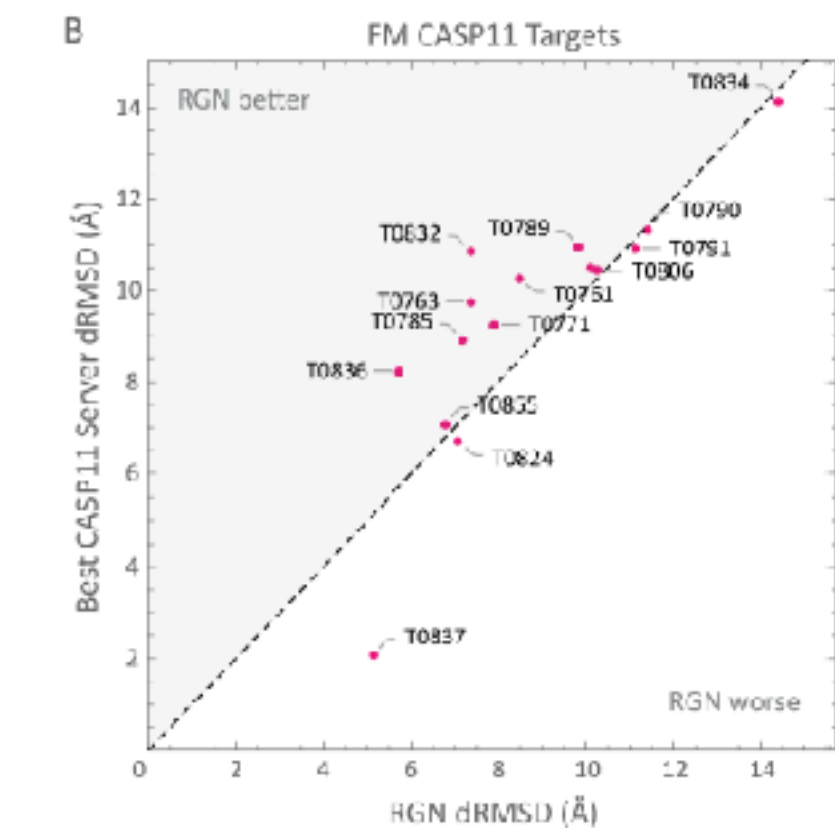
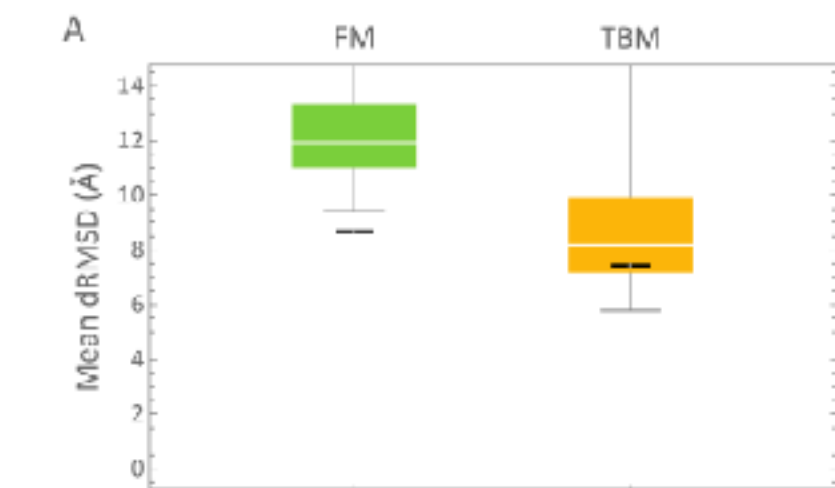
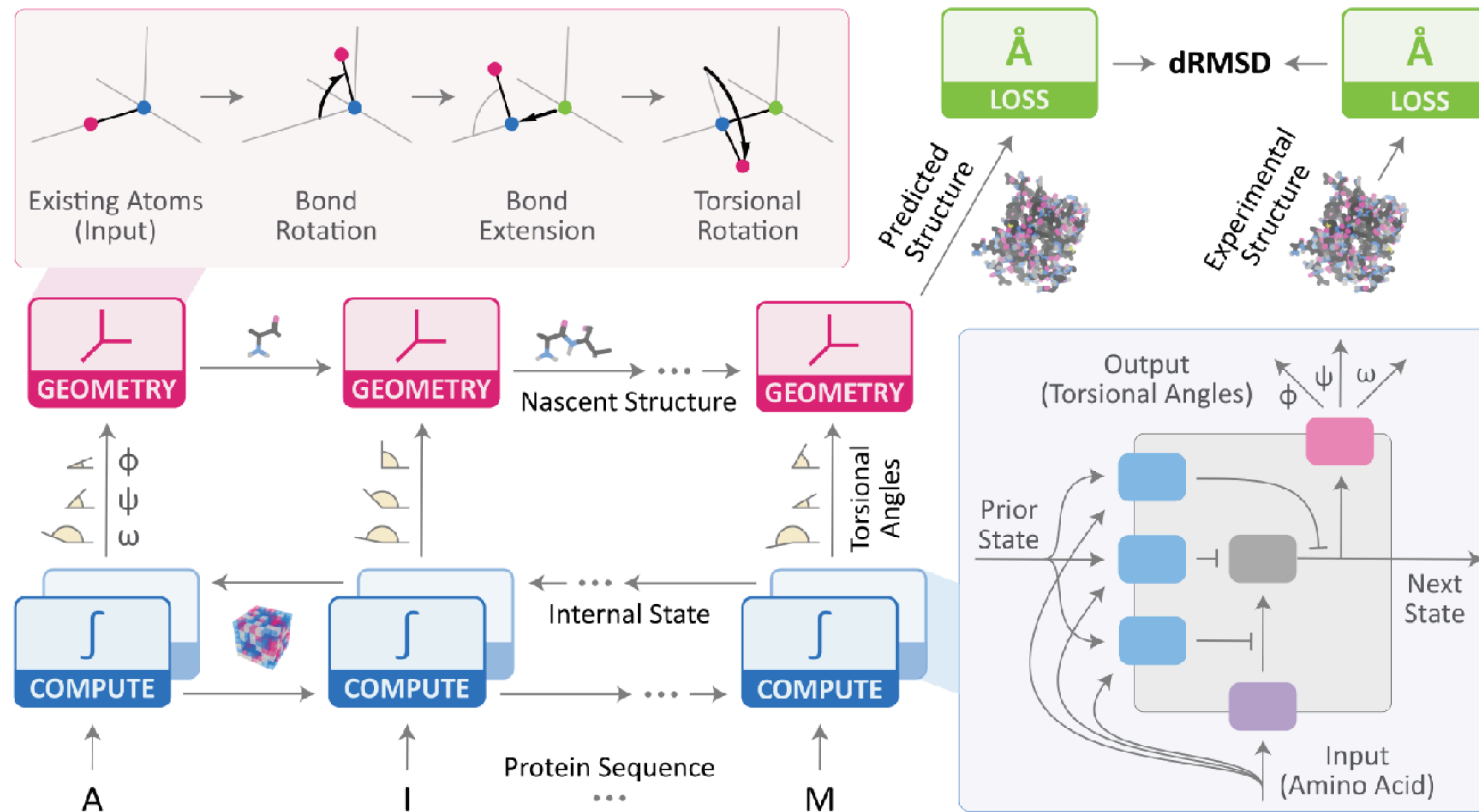
template HPRRLLLPNTLGLGRRRYSPIYERSHGHHNQMSRRASGGPNALLPAVGKDGQVCMDVSQFKPNELTVKVVDNTV
NAFESVMKEMSAIQPREFHPELEYTQP----GELDFLKDAYEVGKDGRLHFKVYFNVKNFKAEITIKADKNKL

Build backbone from template (loops??)

Add side-chains

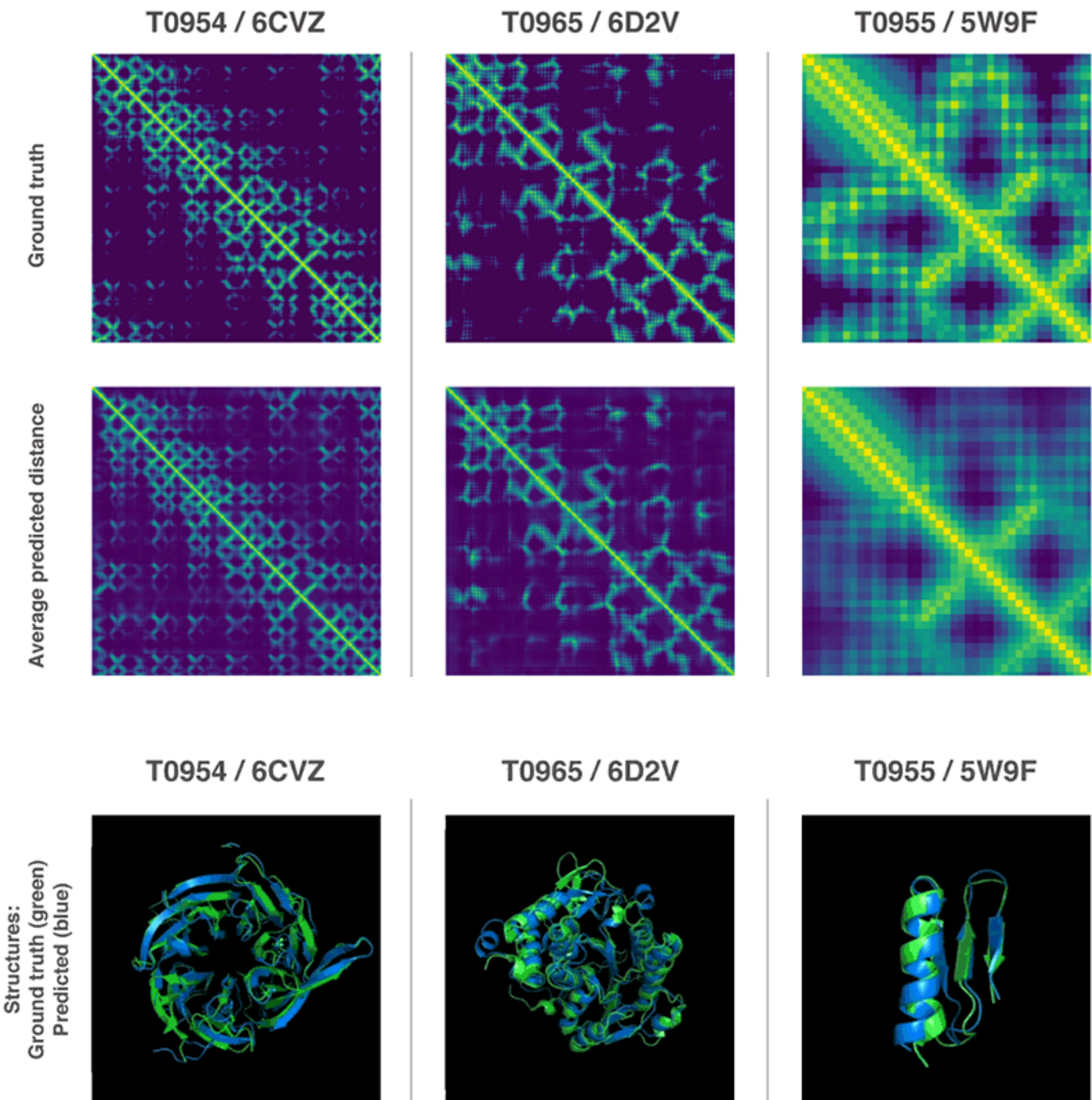
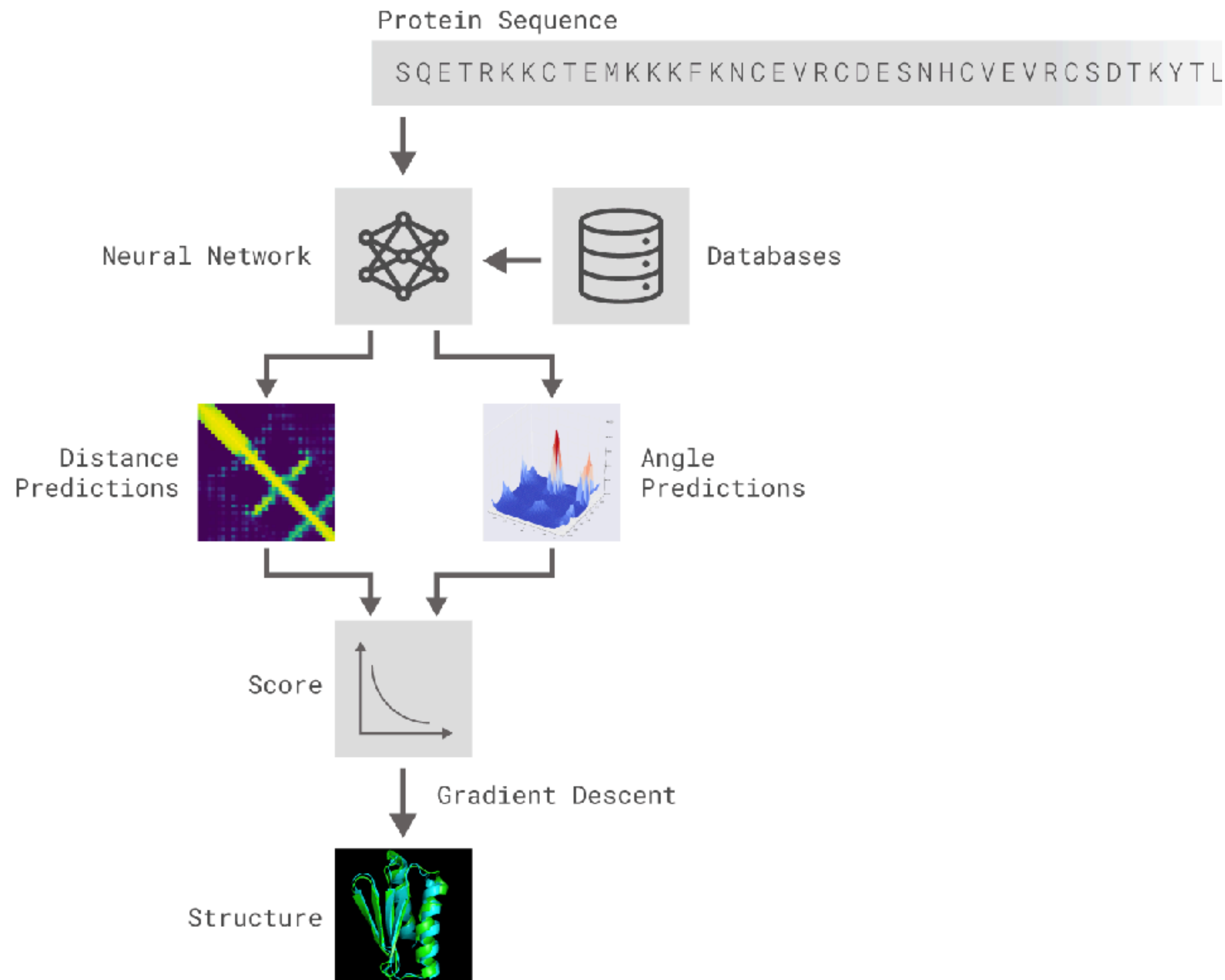


Learning to Predict Structure

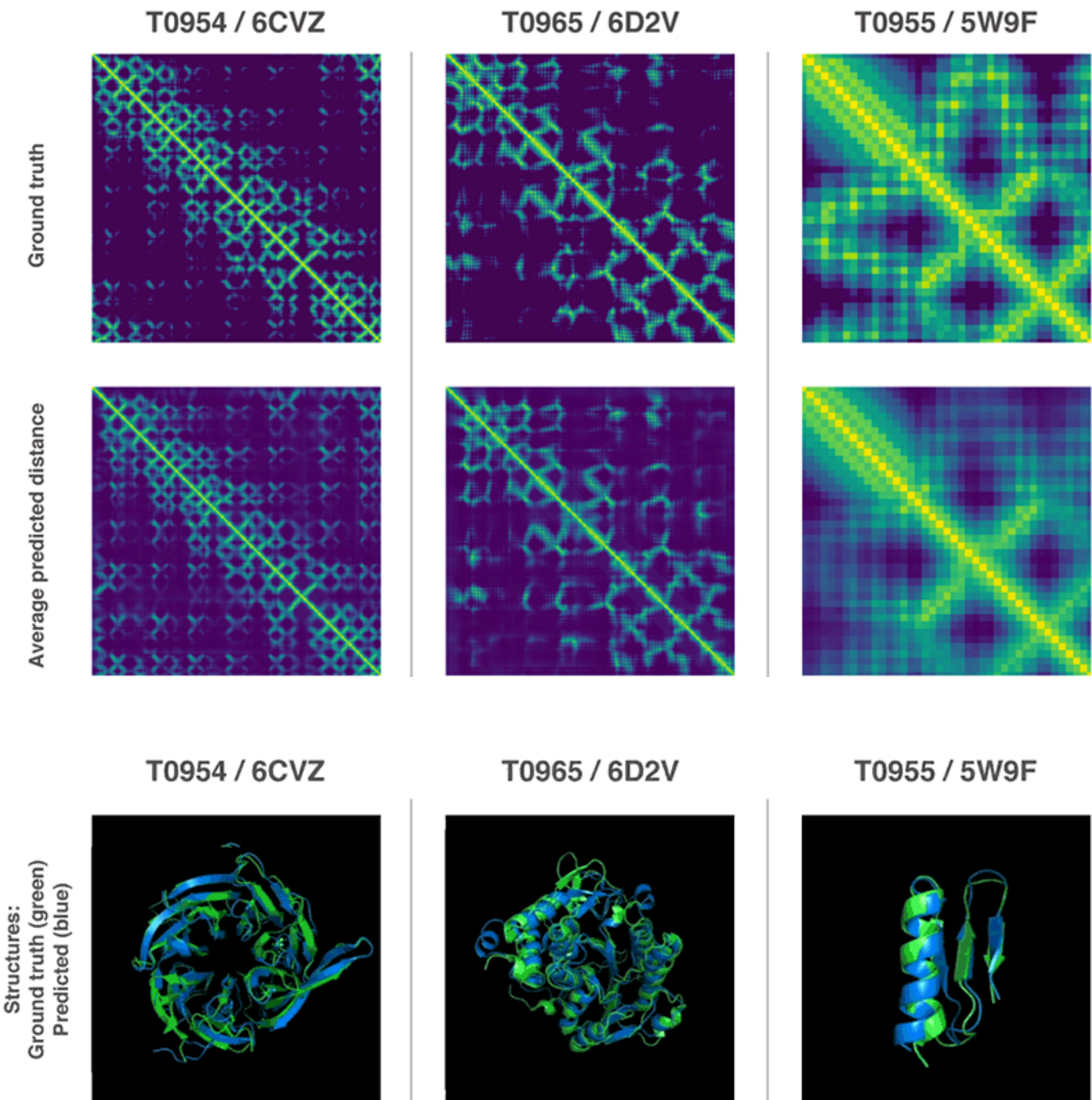
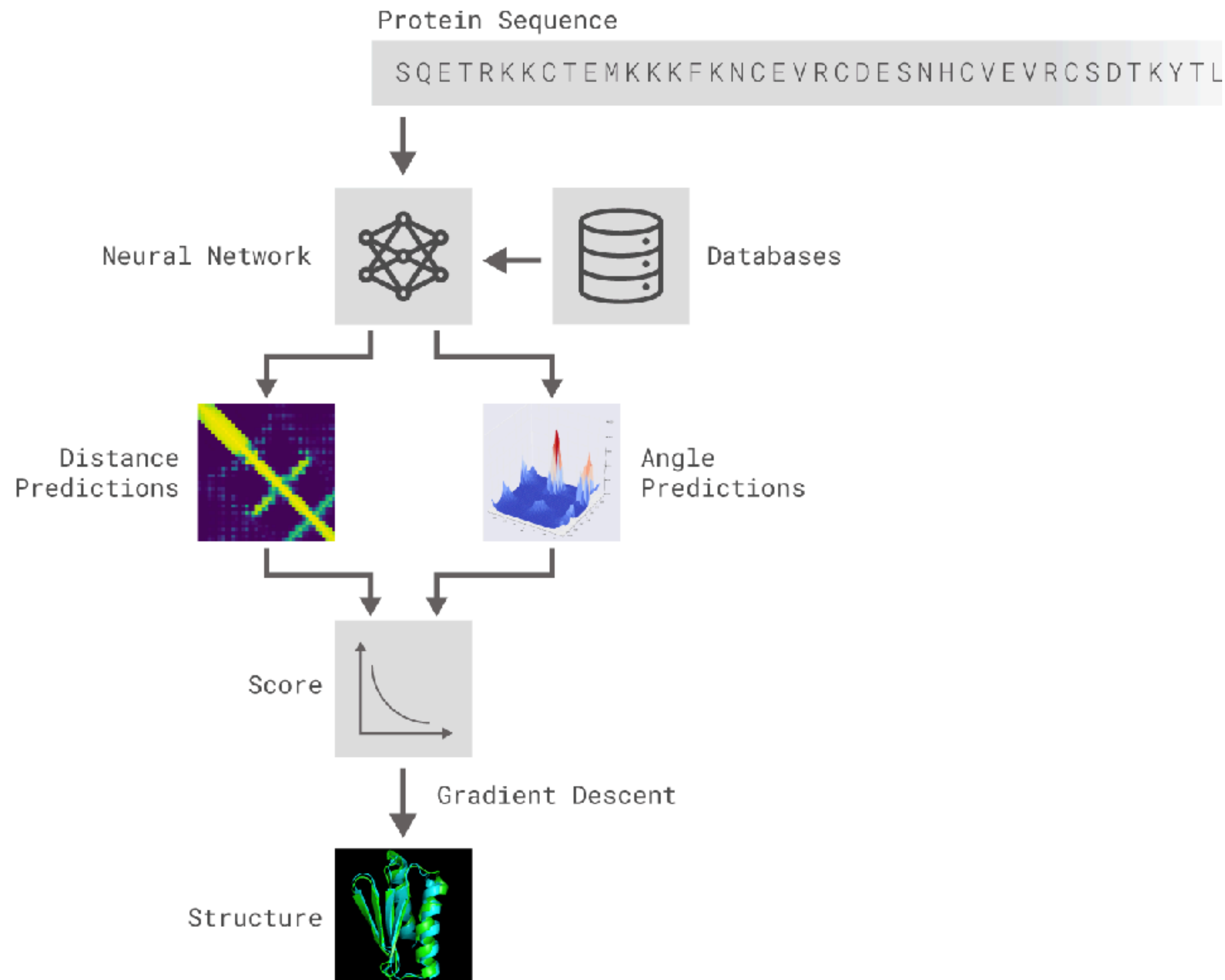


<https://www.biorxiv.org/content/early/2018/02/14/265231>

AlphaFold



AlphaFold

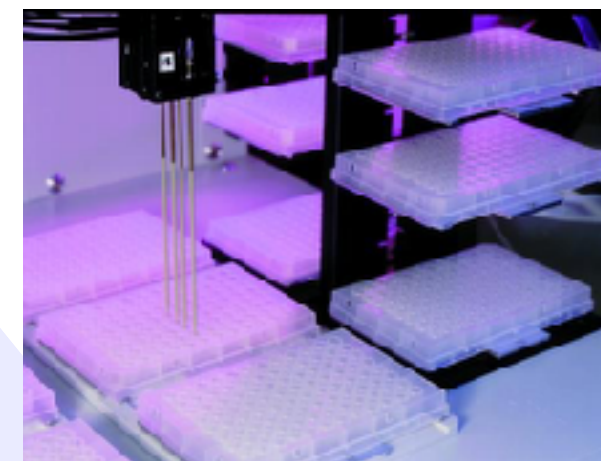


Computational Drug Discovery

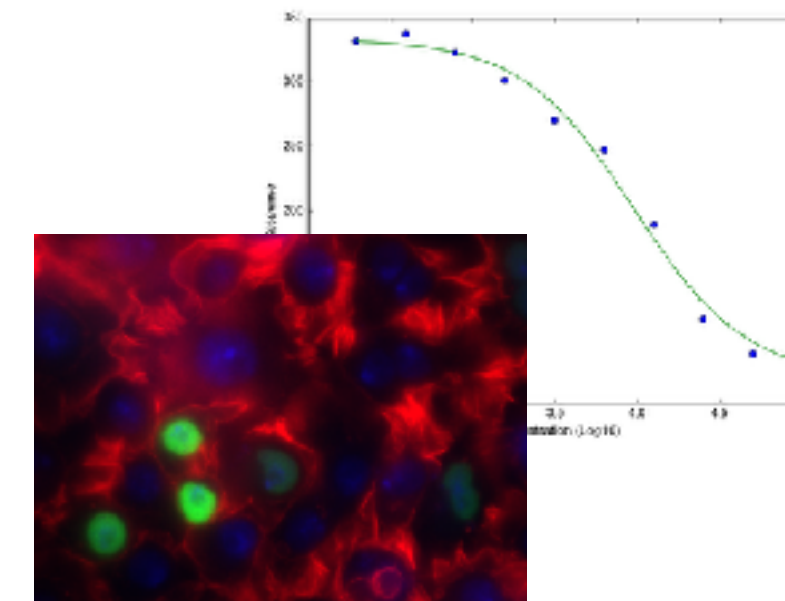
Drug Discovery

Omics

Target
Identification



Screening



Lead
Identification



Lead
Optimization

Compounds

Hits

Leads

**Clinical
Candidates**

Cost

Computational Drug Discovery

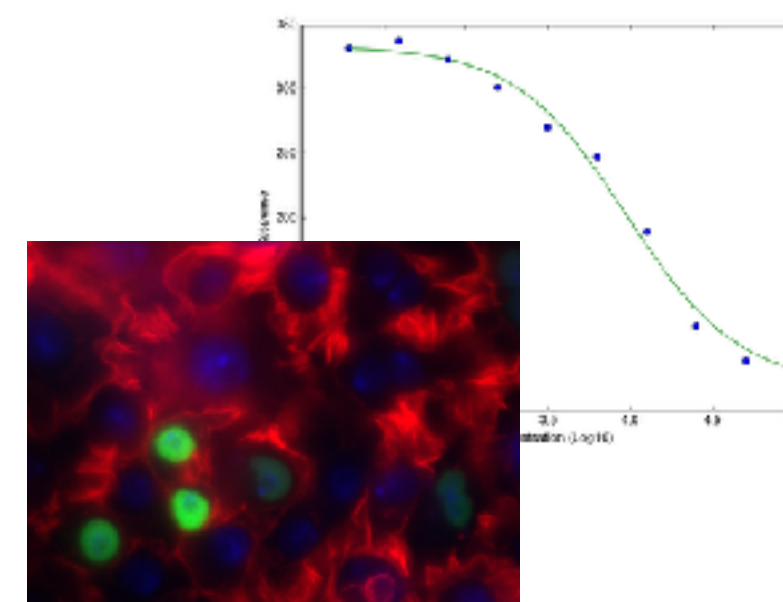
Omics

Target
Identification

Virtual

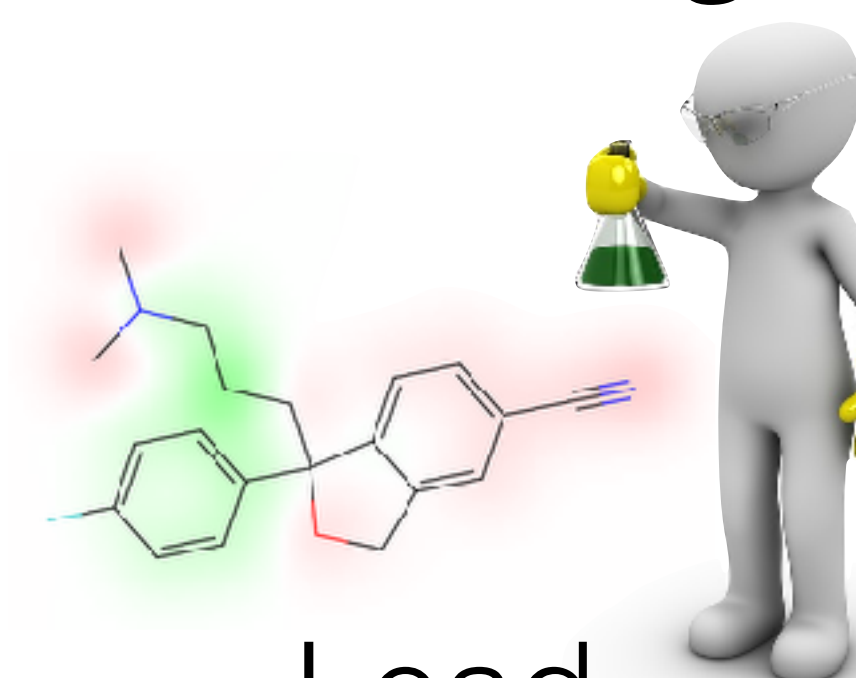


Screening



Lead
Identification

Modeling



Lead
Optimization

Compounds

Hits

Leads

Clinical
Candidates

Cost

Ligand Based Drug Design

Fingerprint Methods

- map molecules to a descriptor space:



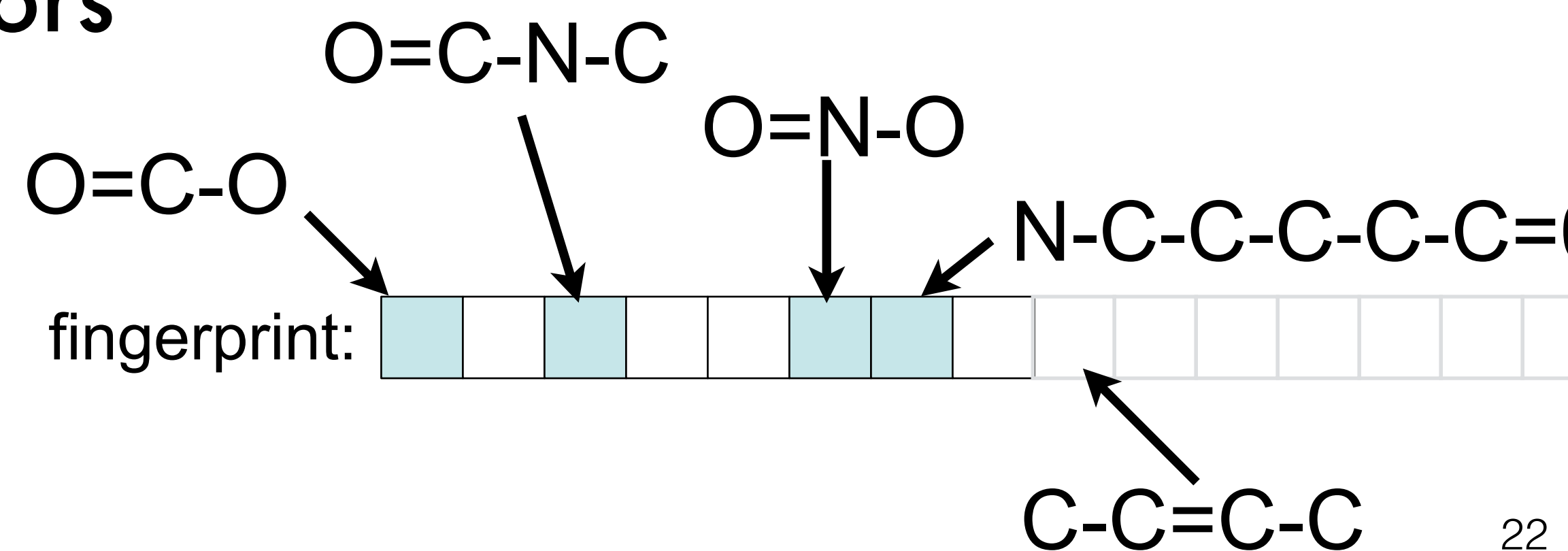
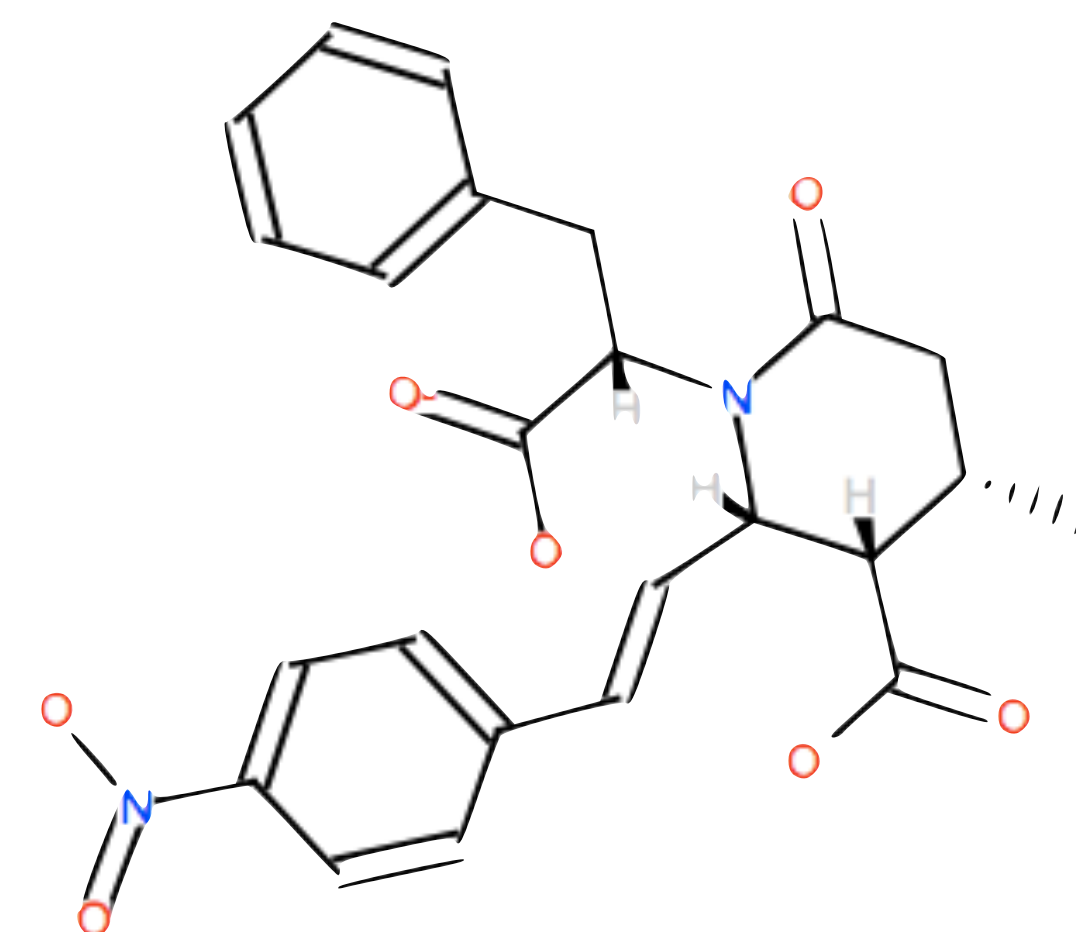
1D: molecule weight, #h-bonds, etc.

2D: paths, bond distances between atom-pairs

Example: Daylight/FP2

- all paths up to 7 bonds long
 - each path corresponds to a bit
- similarity is "distance" between descriptors
 - for bit vectors, Tanimoto distance used

$$T(A, B) = \frac{|A \cap B|}{|A \cup B|}$$



Ligand Based Drug Design

Fingerprint Methods

- map molecules to a descriptor space:



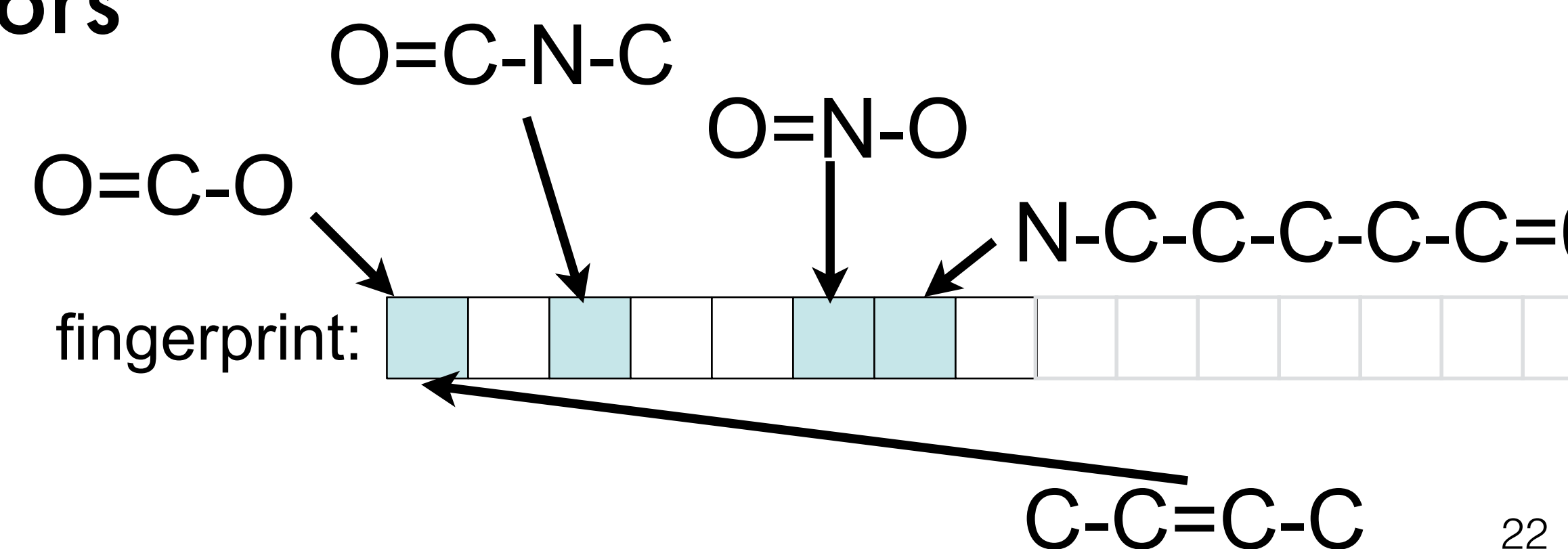
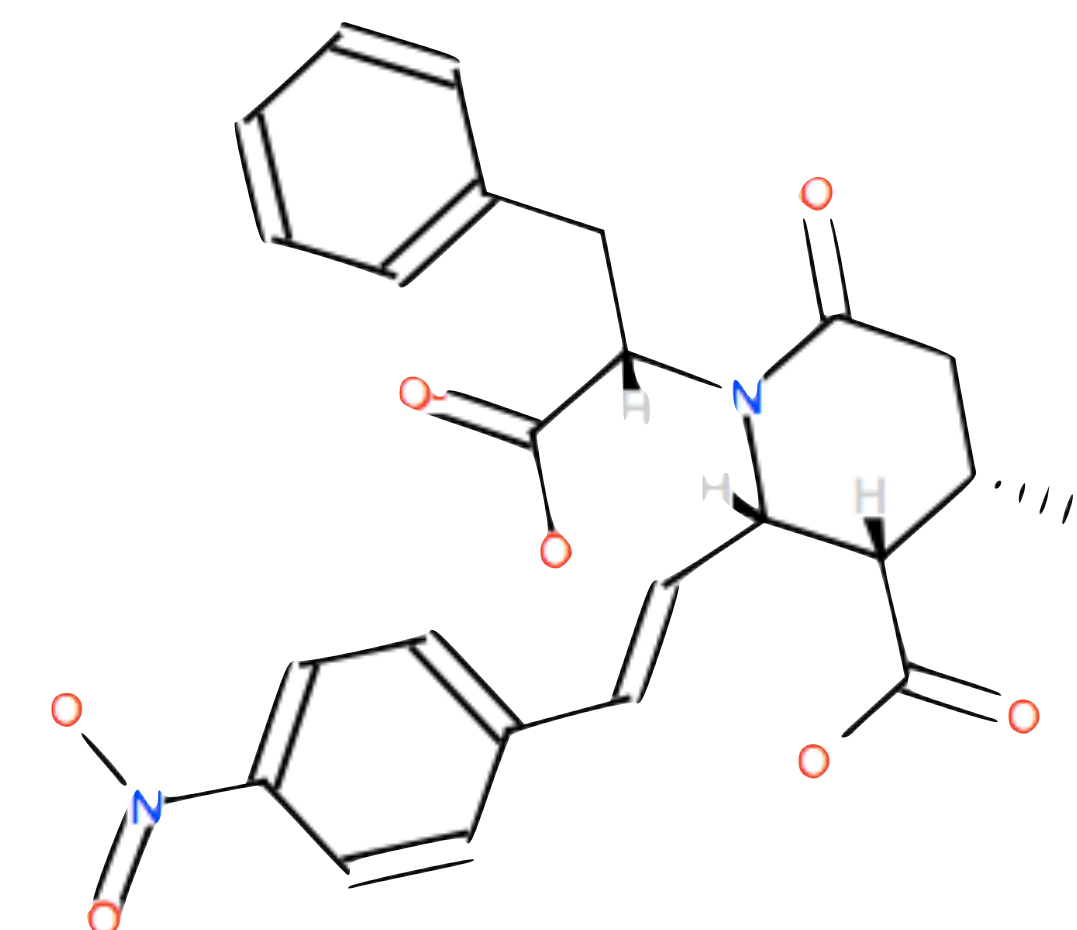
1D: molecule weight, #h-bonds, etc.

2D: paths, bond distances between atom-pairs

Example: Daylight/FP2

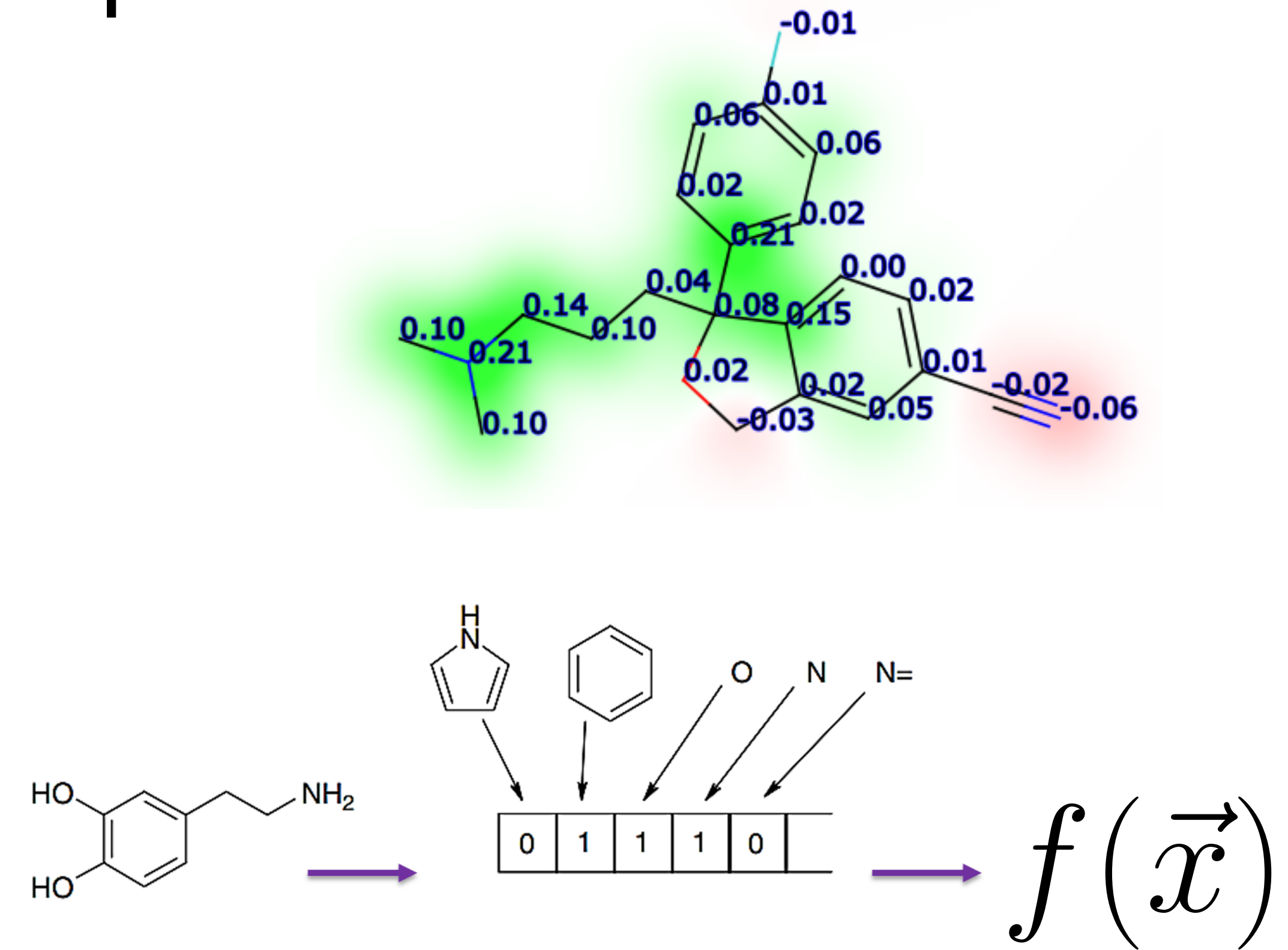
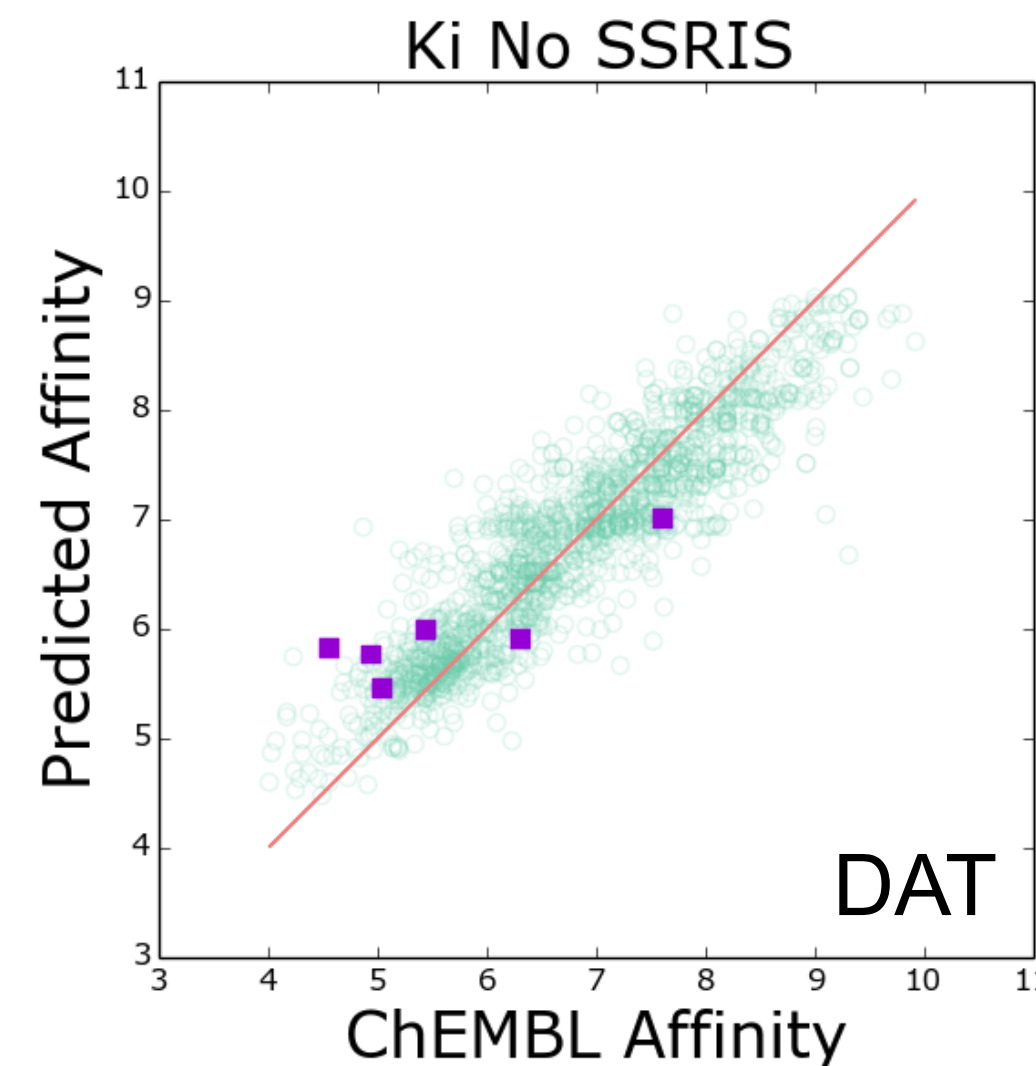
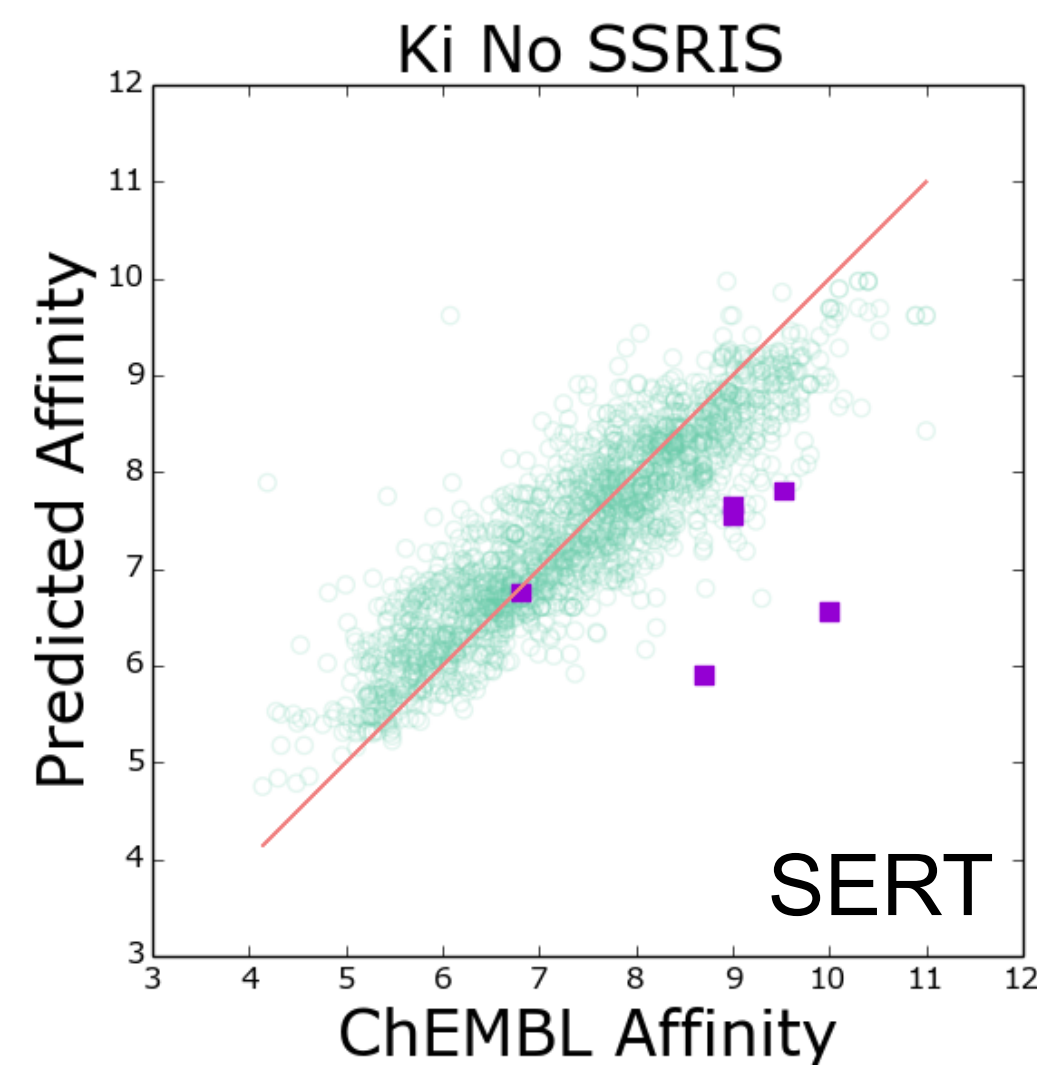
- all paths up to 7 bonds long
 - each path corresponds to a bit
- similarity is "distance" between descriptors
 - for bit vectors, Tanimoto distance used

$$T(A, B) = \frac{|A \cap B|}{|A \cup B|}$$



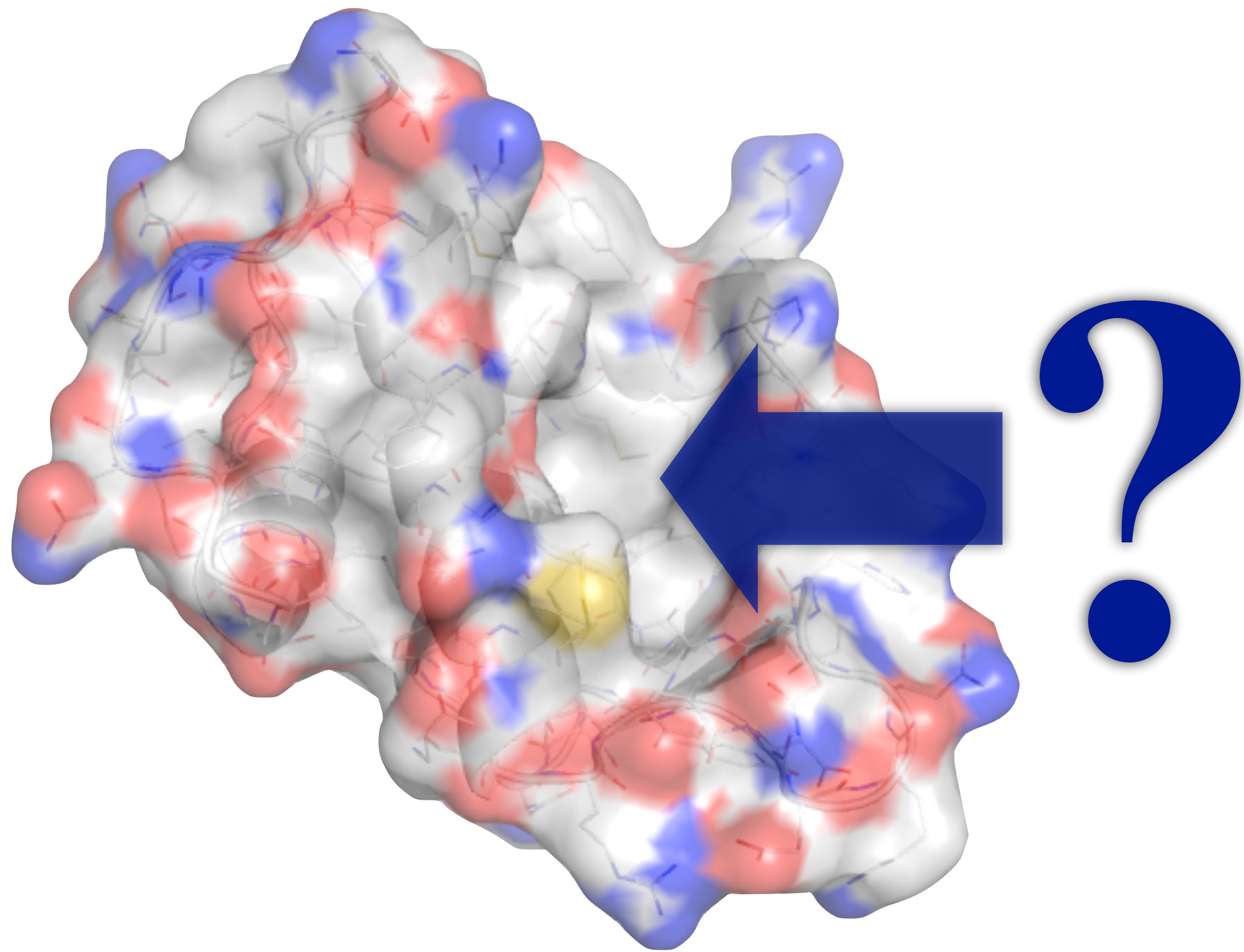
Ligand Based: QSAR

Quantitative Structure/Activity Relationships



$$f(\vec{x}) = w_1\vec{x}_1 + w_2\vec{x}_2 + w_3\vec{x}_3 + \dots + b$$

Structure Based Drug Design

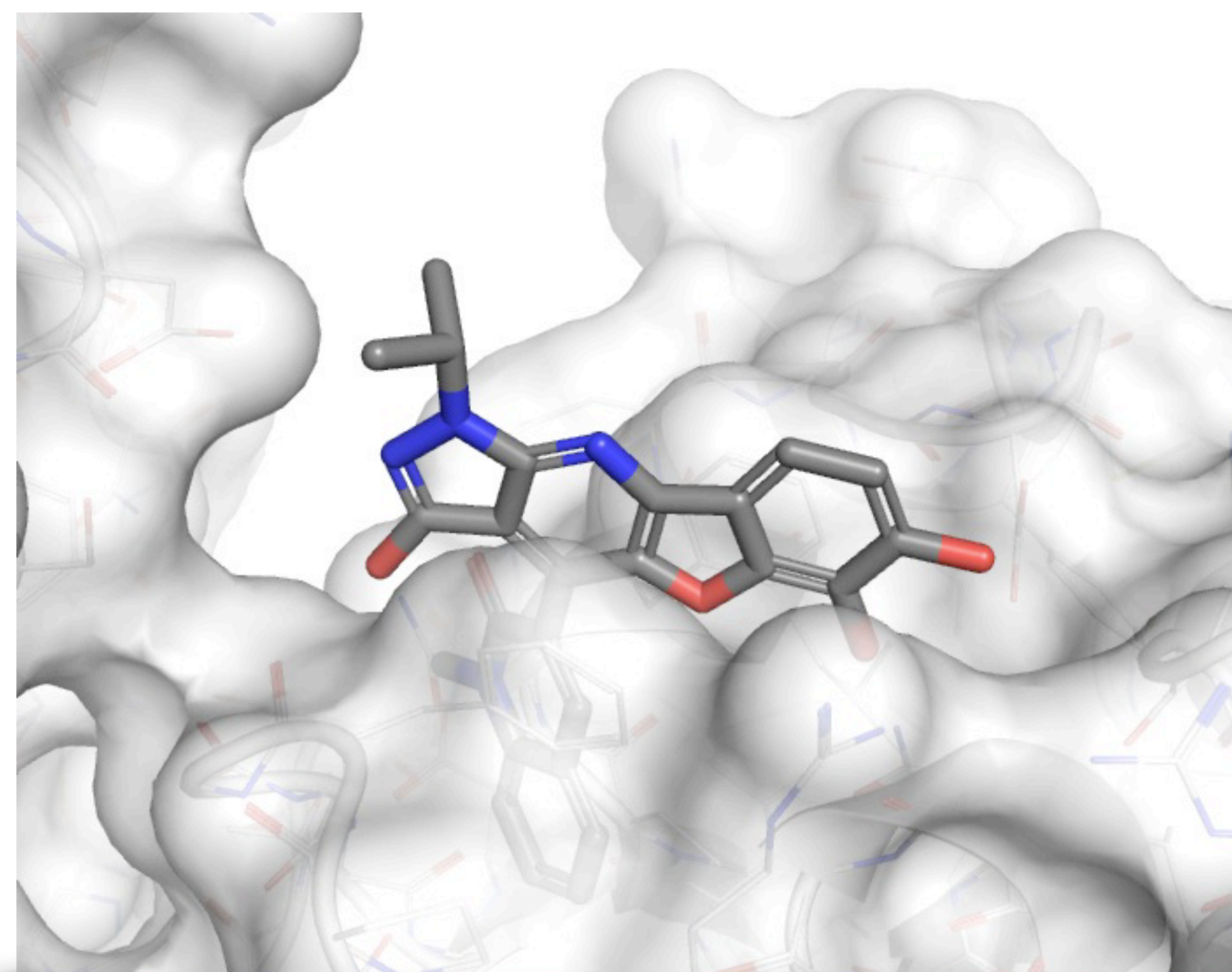


Unlike ligand based approaches,
generalizes to new targets

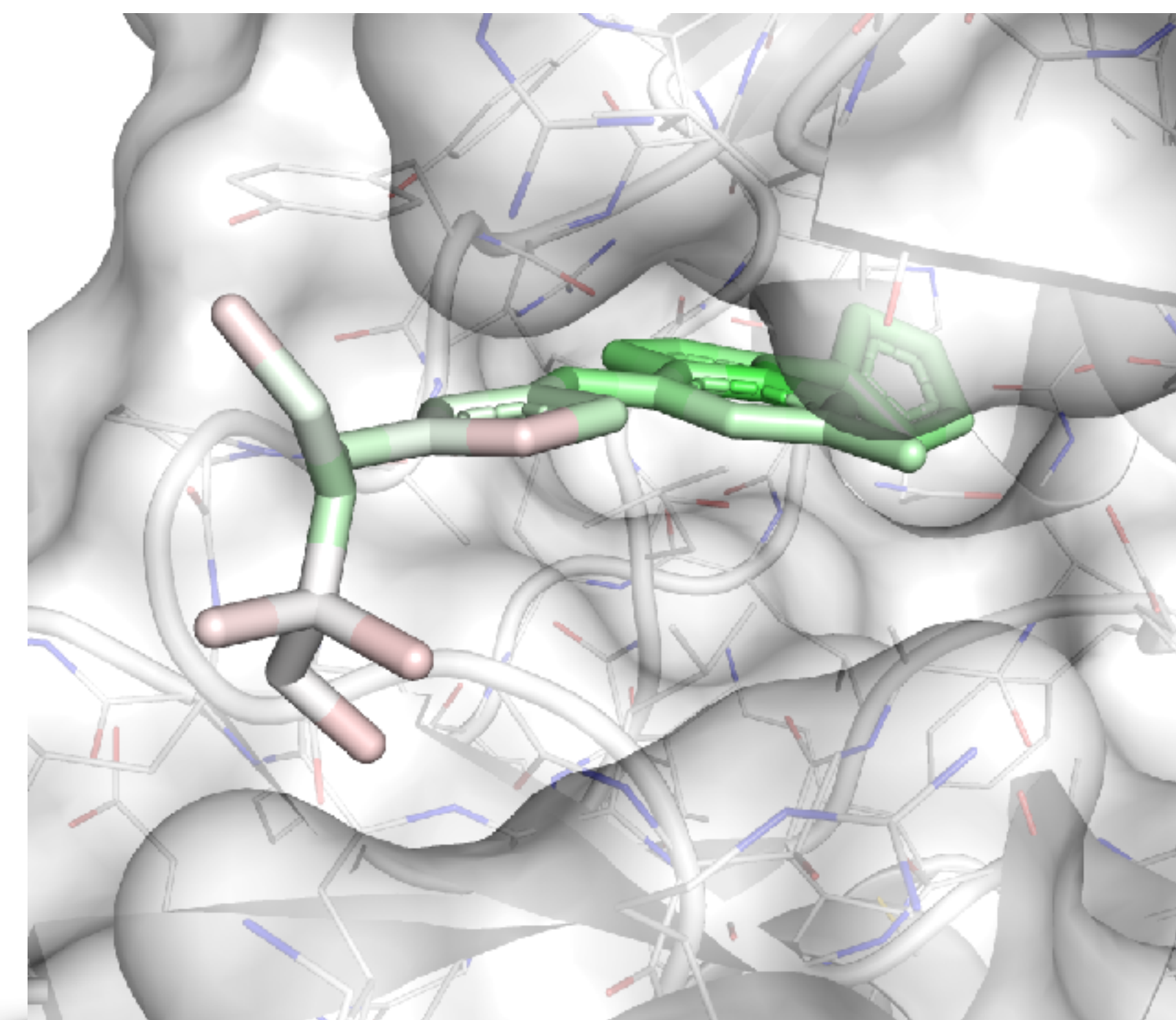
Requires **molecular target** with
known structure and **binding site**

Structure Based Drug Design

Virtual Screening



Lead Optimization



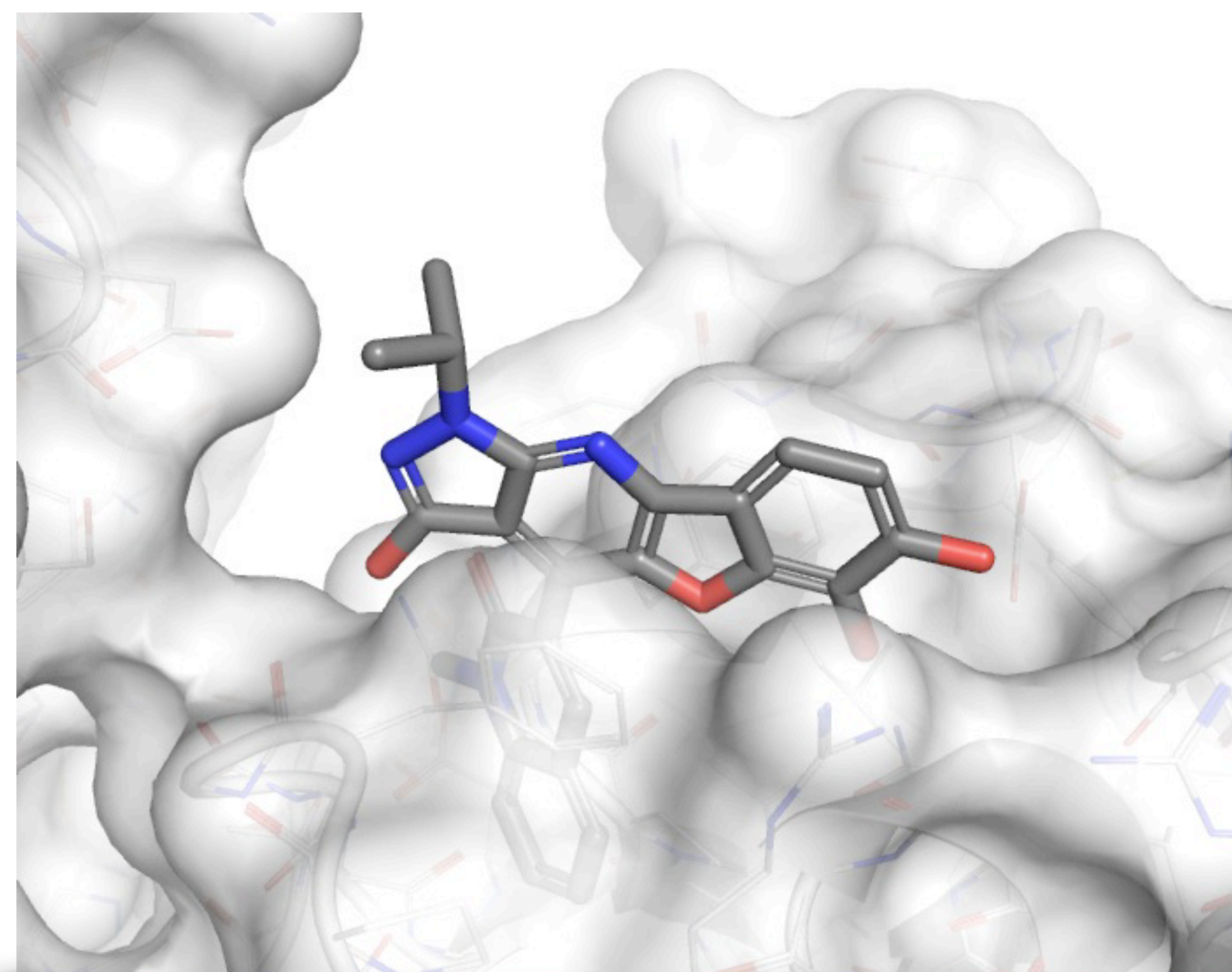
Pose Prediction

Binding Discrimination

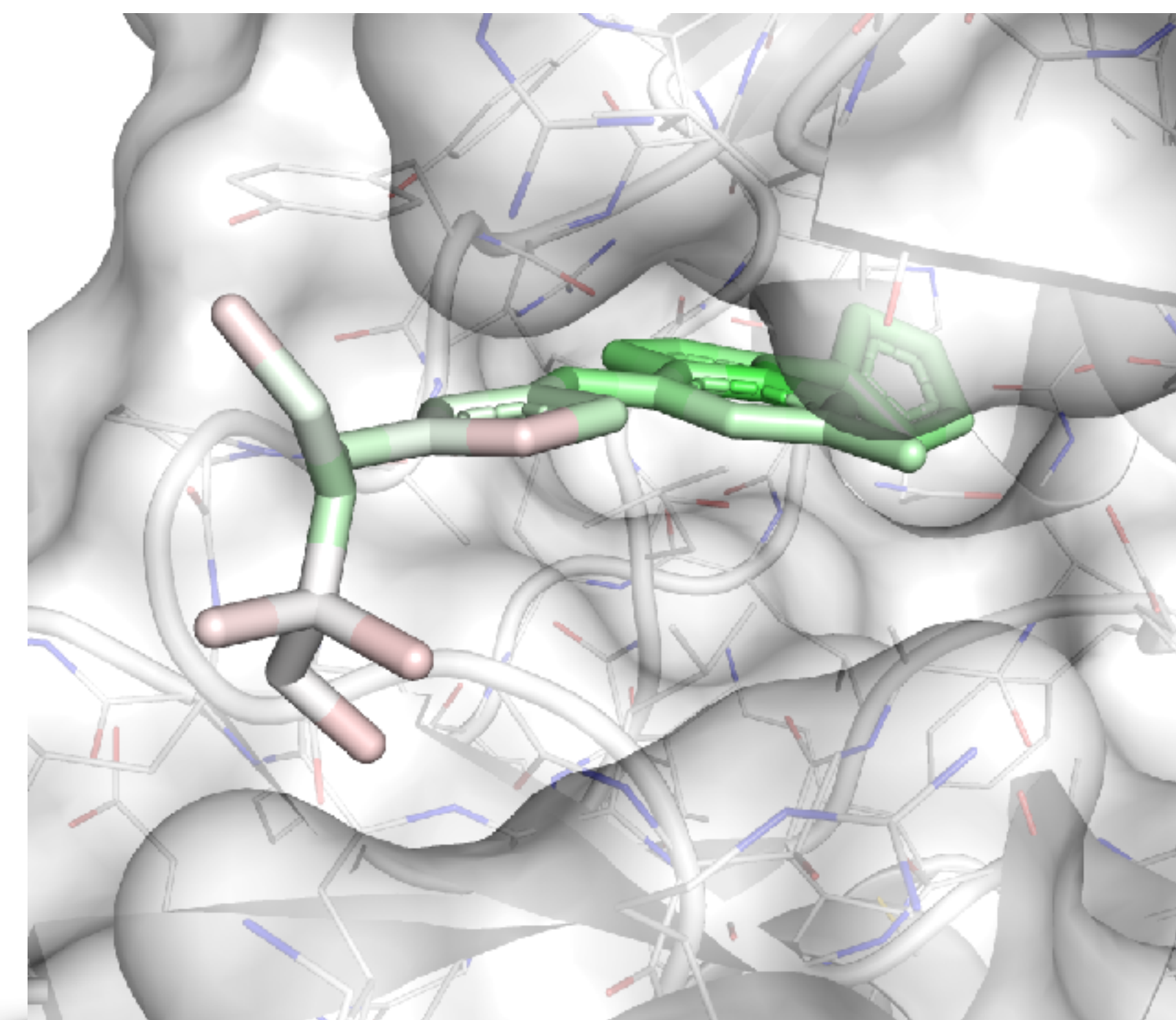
Affinity Prediction

Structure Based Drug Design

Virtual Screening



Lead Optimization

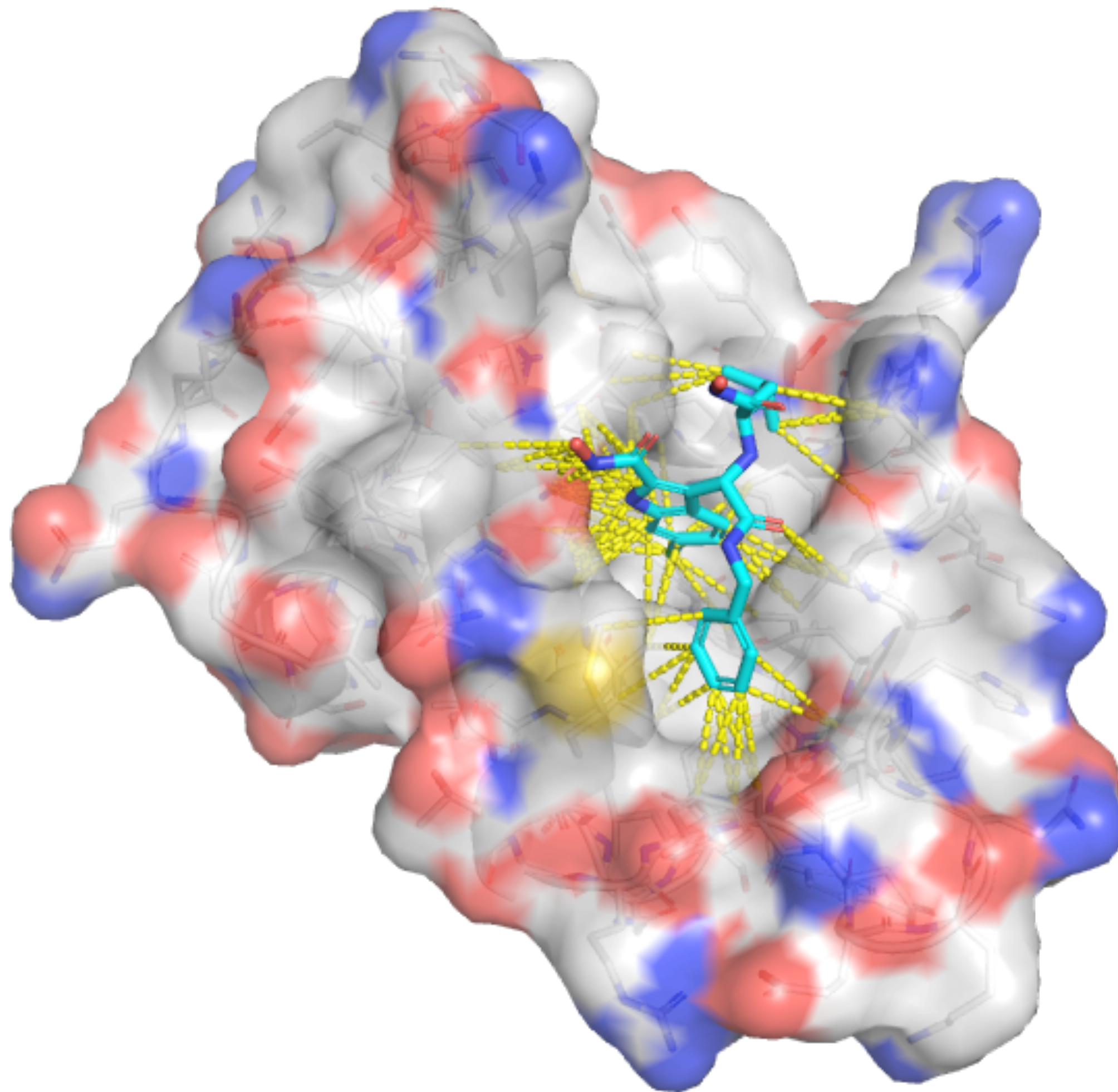


Pose Prediction

Binding Discrimination

Affinity Prediction

Protein-Ligand Scoring

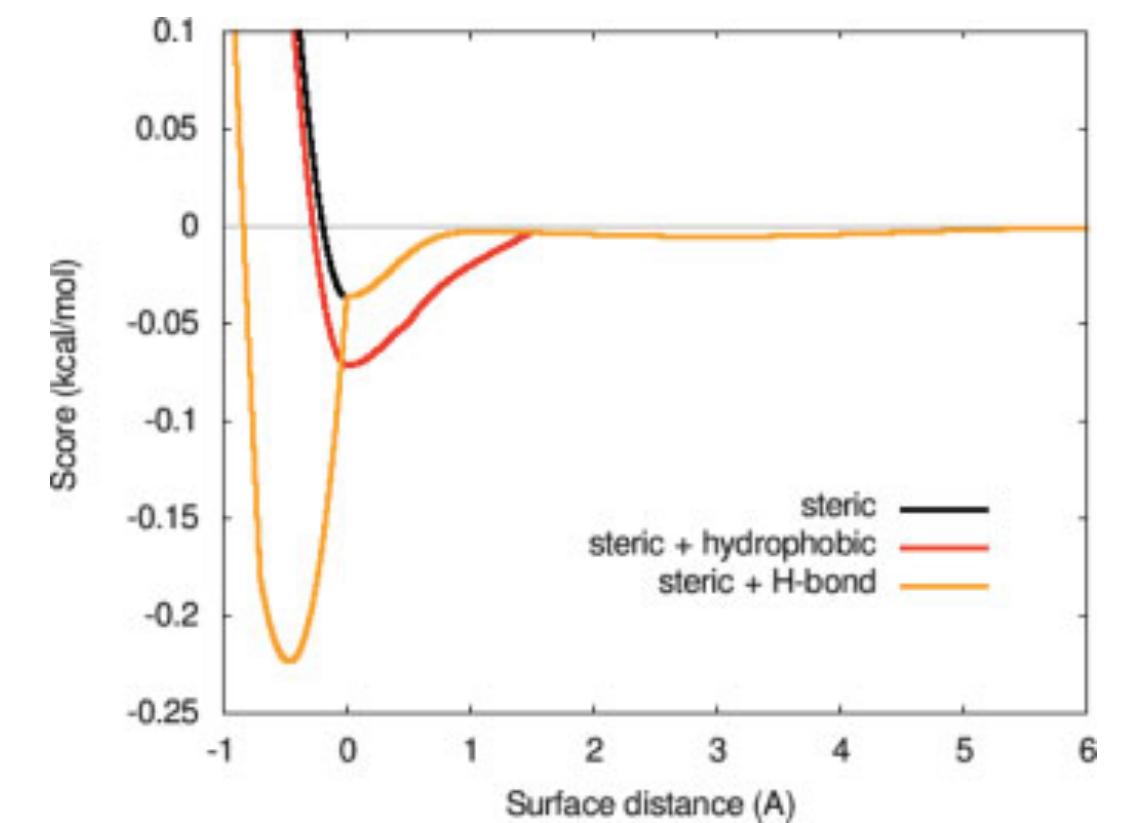
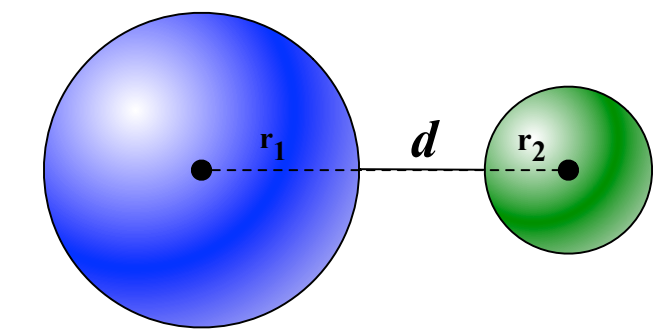


AutoDock Vina

$$\begin{aligned}\text{gauss}_1(d) &= w_{\text{guass}_1} e^{-(d/0.5)^2} \\ \text{gauss}_2(d) &= w_{\text{guass}_2} e^{-((d-3)/2)^2} \\ \text{repulsion}(d) &= \begin{cases} w_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}\end{aligned}$$

$$\text{hydrophobic}(d) = \begin{cases} w_{\text{hydrophobic}} & d < 0.5 \\ 0 & d > 1.5 \\ w_{\text{hydrophobic}}(1.5 - d) & \text{otherwise} \end{cases}$$

$$\text{hbond}(d) = \begin{cases} w_{\text{hbond}} & d < -0.7 \\ 0 & d > 0 \\ w_{\text{hbond}}(-\frac{10}{7}d) & \text{otherwise} \end{cases}$$

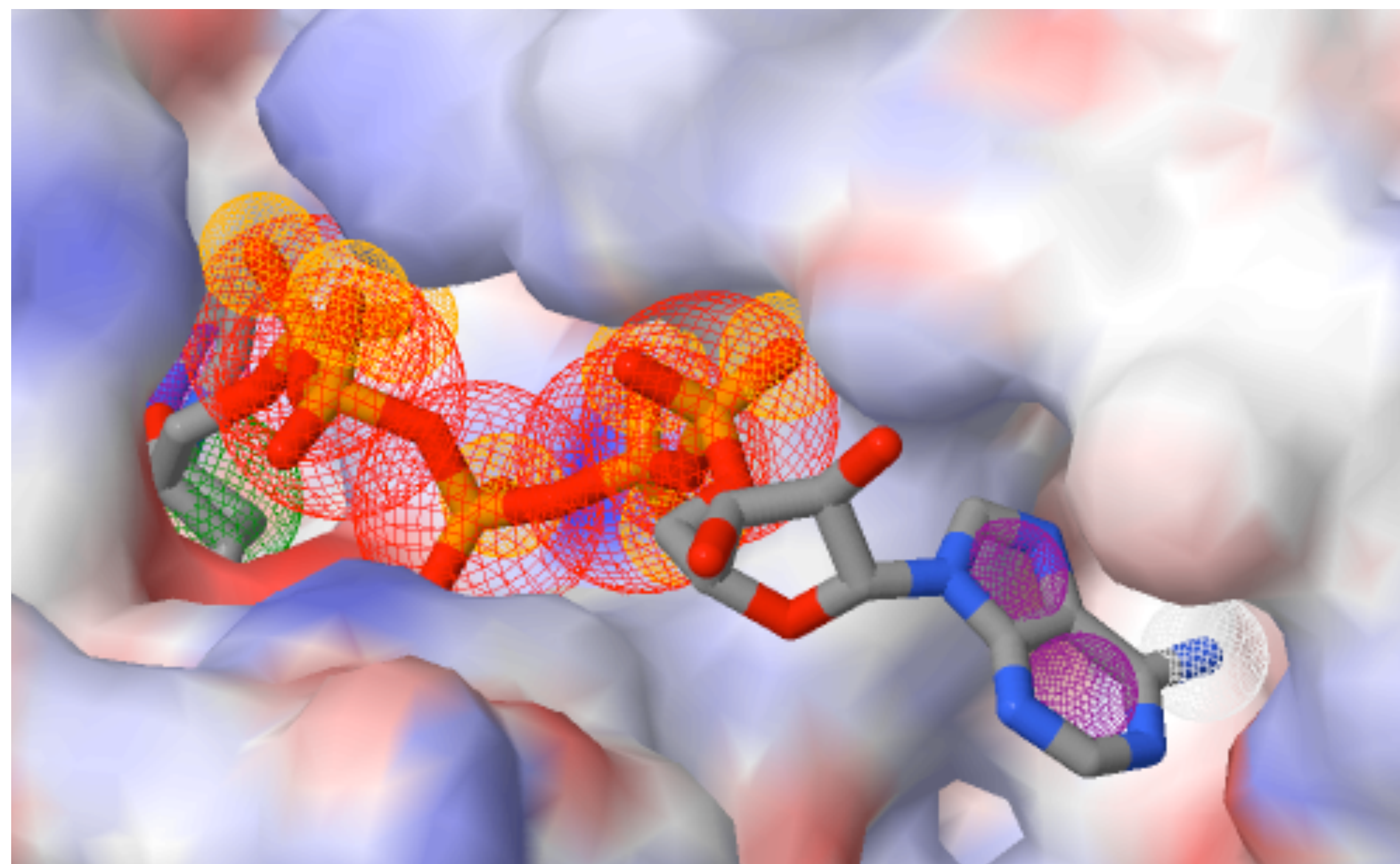


Pharmacophore

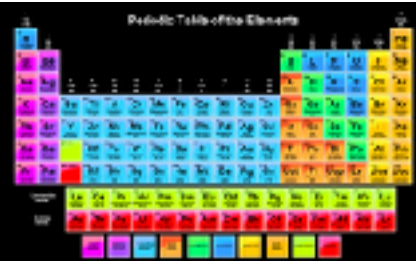
IUPAC: The ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger (or to block) its biological response.

Common Features:

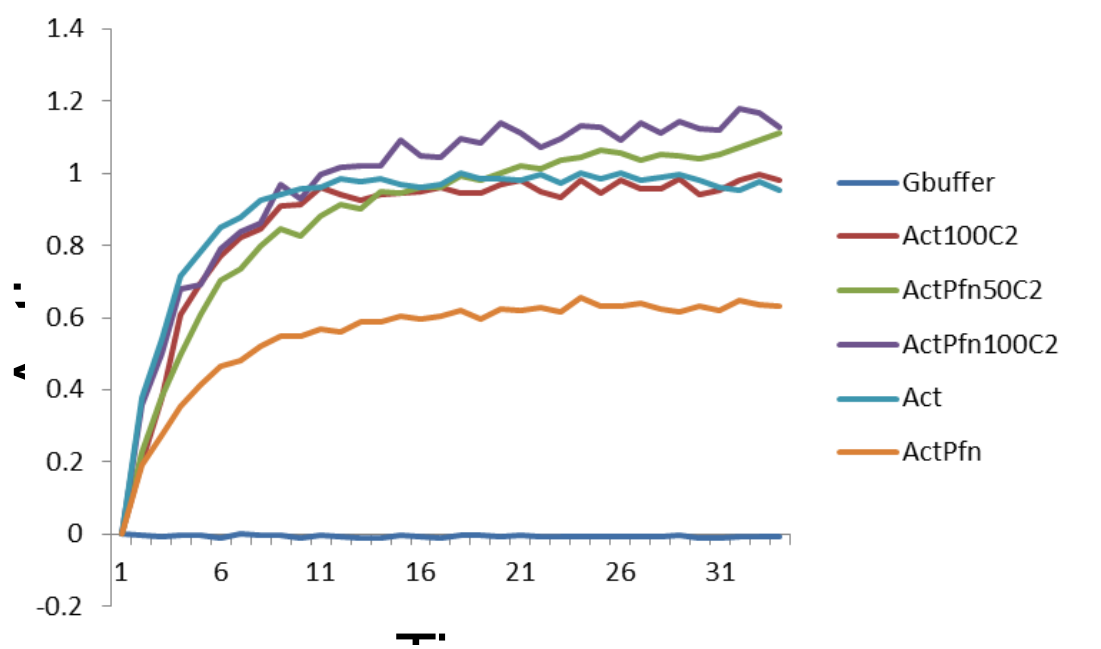
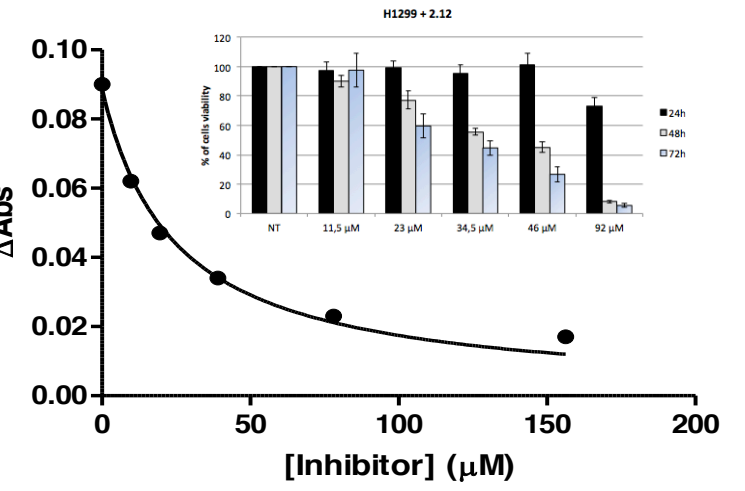
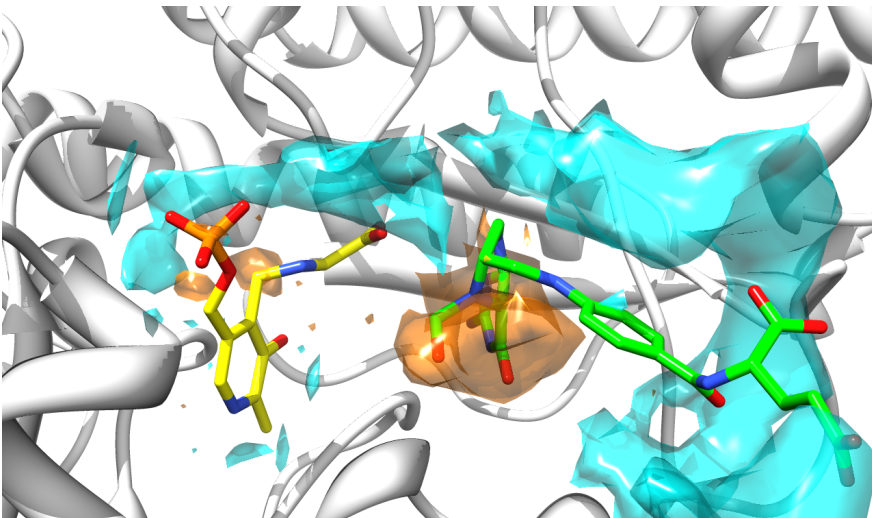
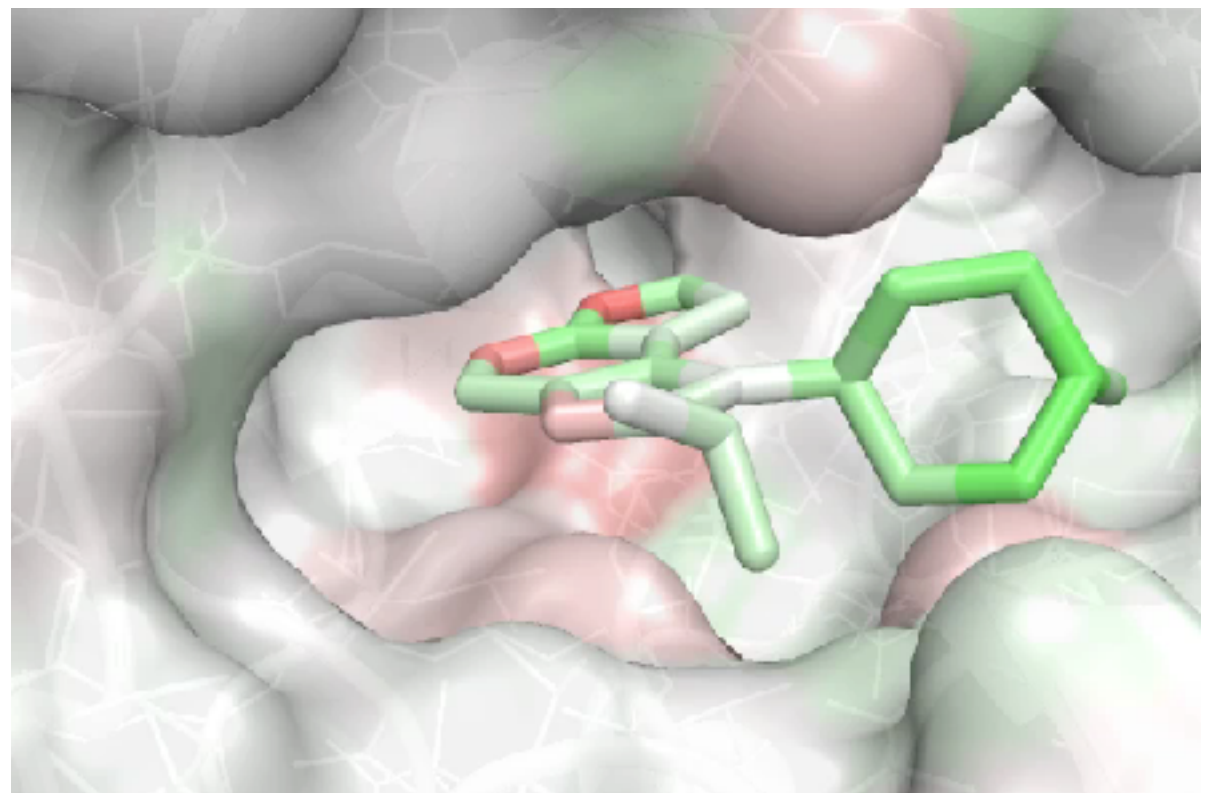
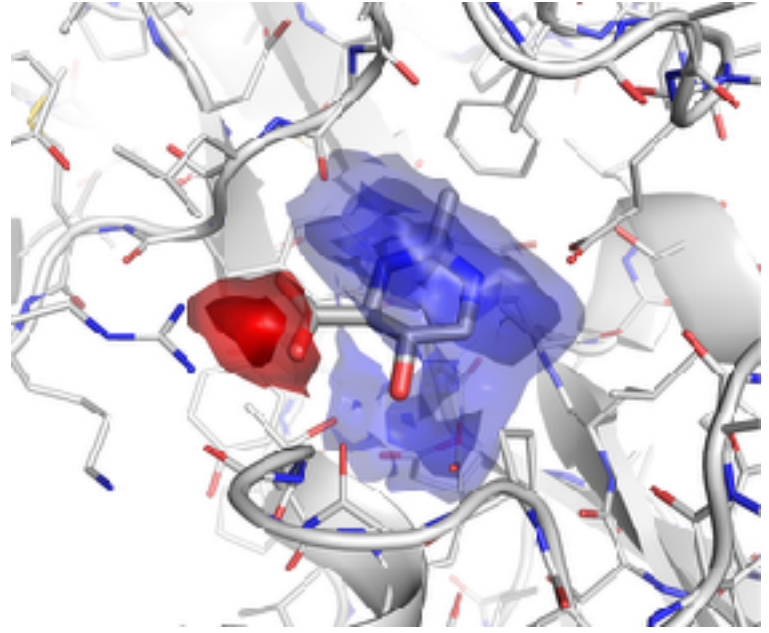
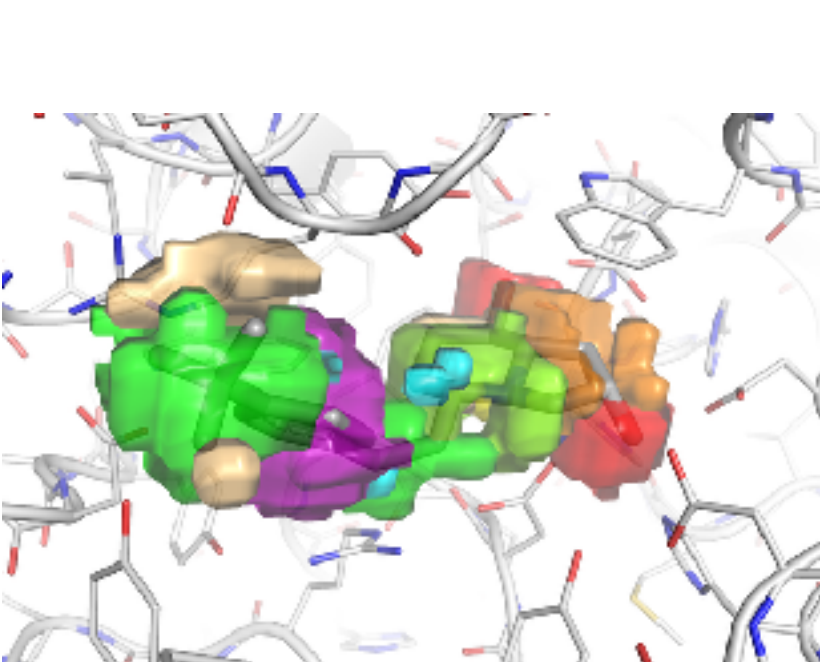
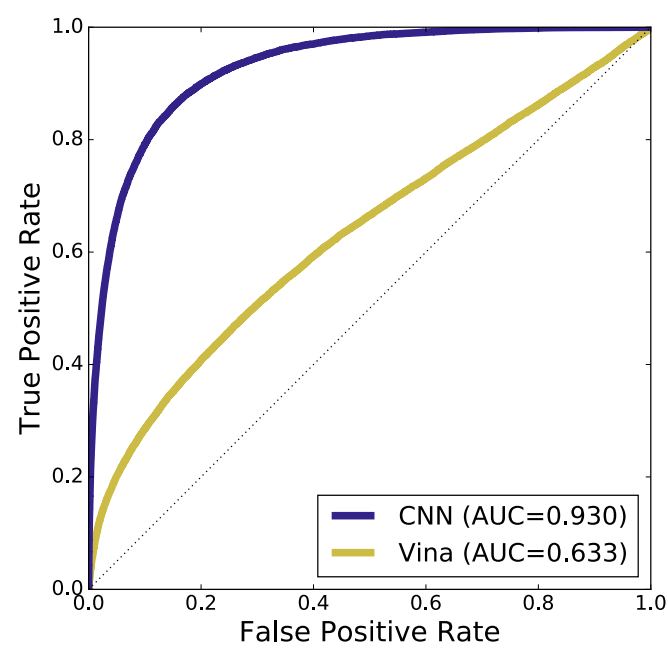
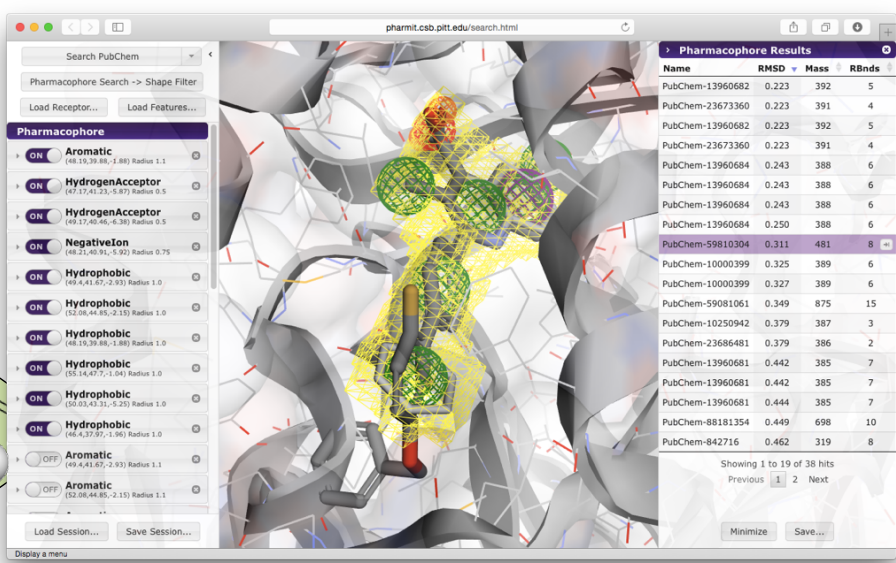
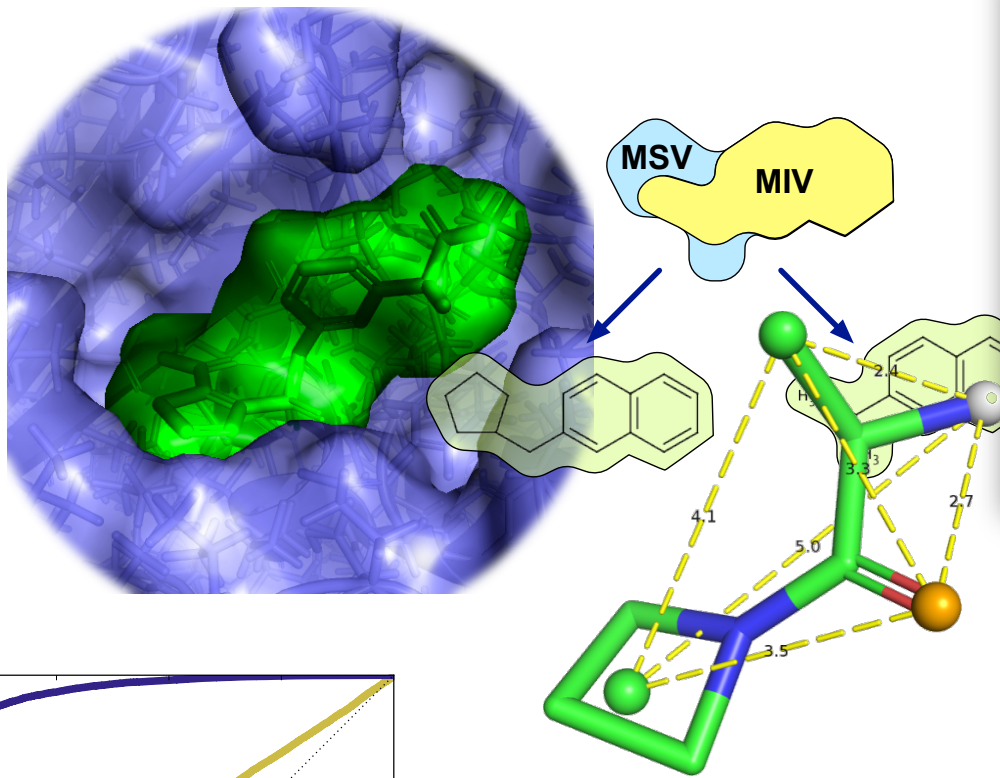
- aromatic ring
- hydrophobic area
- positive ionizable
- negative ionizable
- hydrogen bond donor
- hydrogen bond acceptor



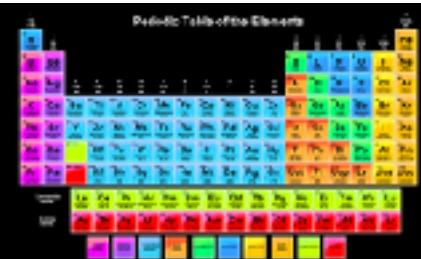
Purchasable



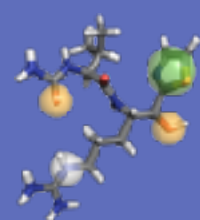
Accessible



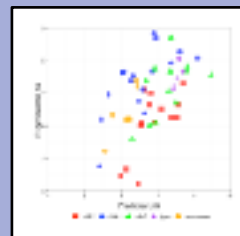
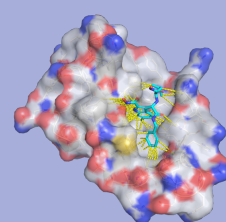
Accessible



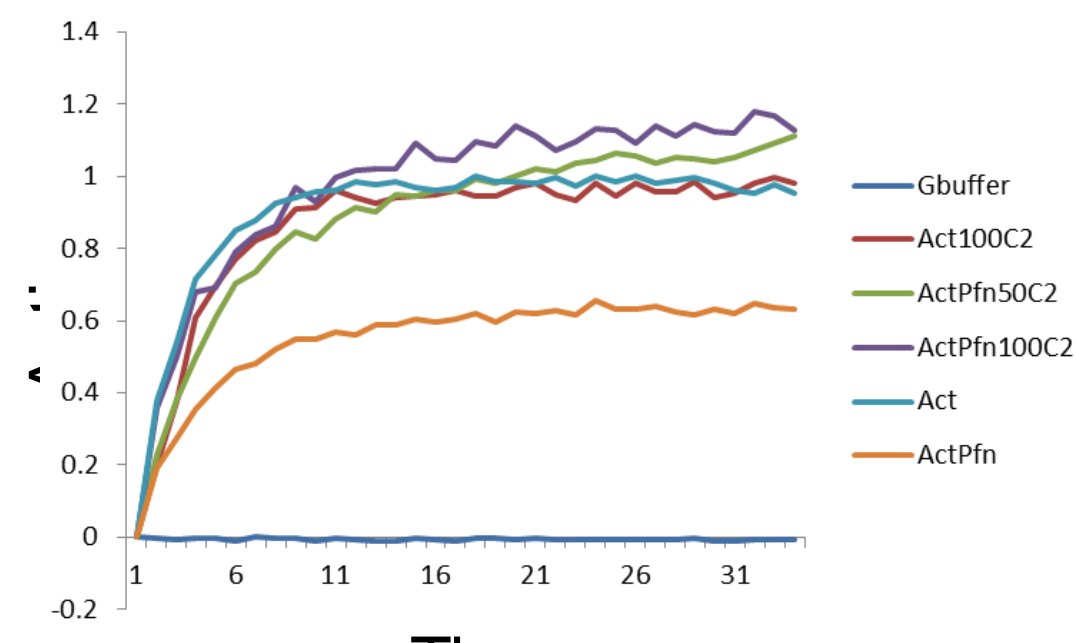
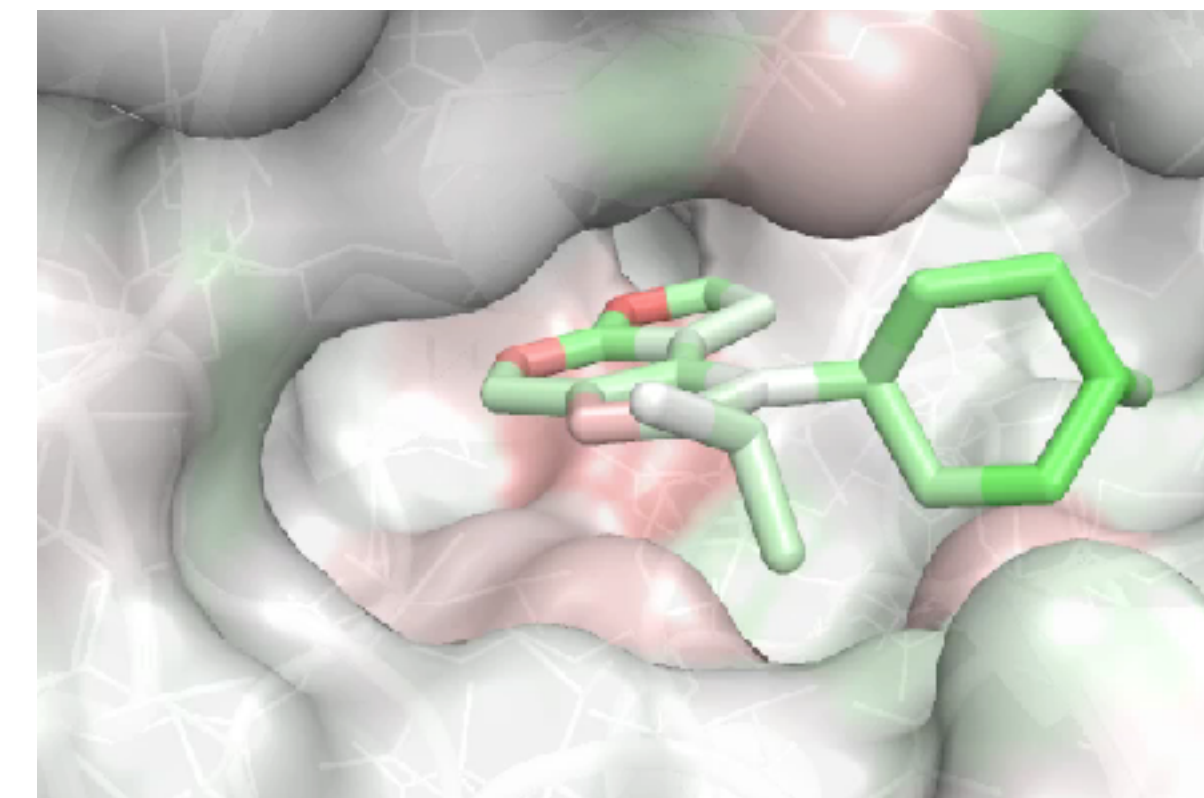
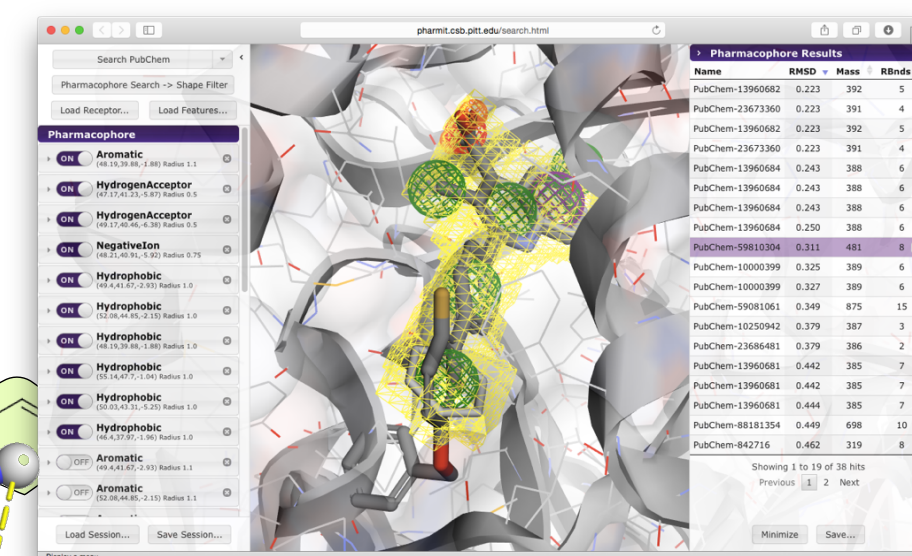
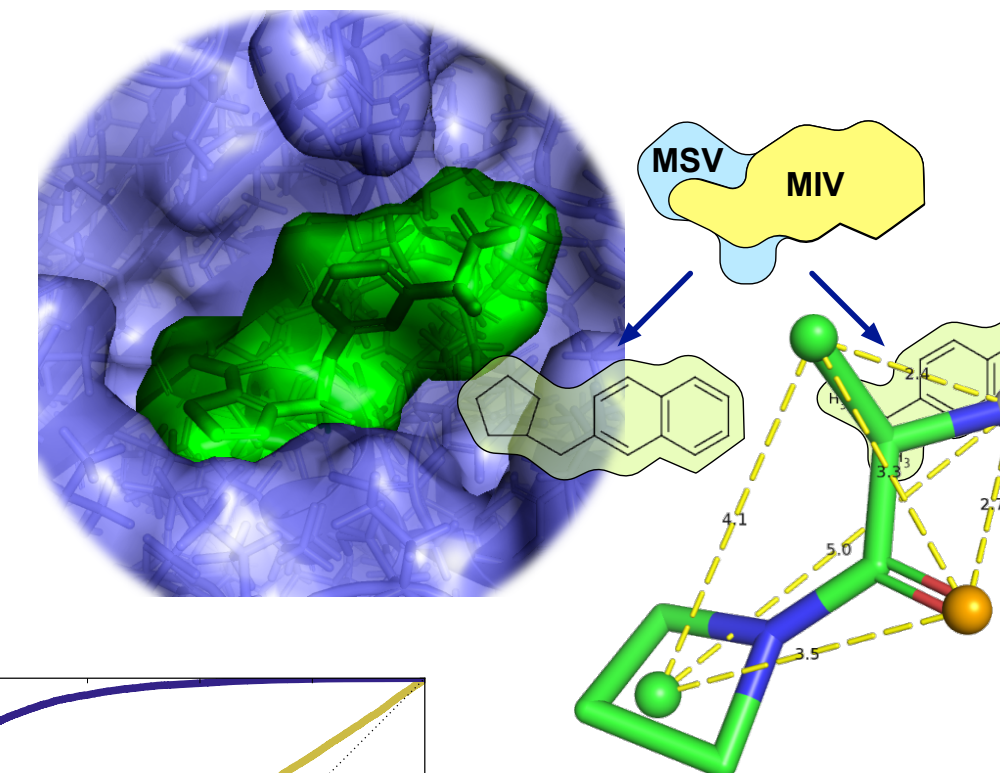
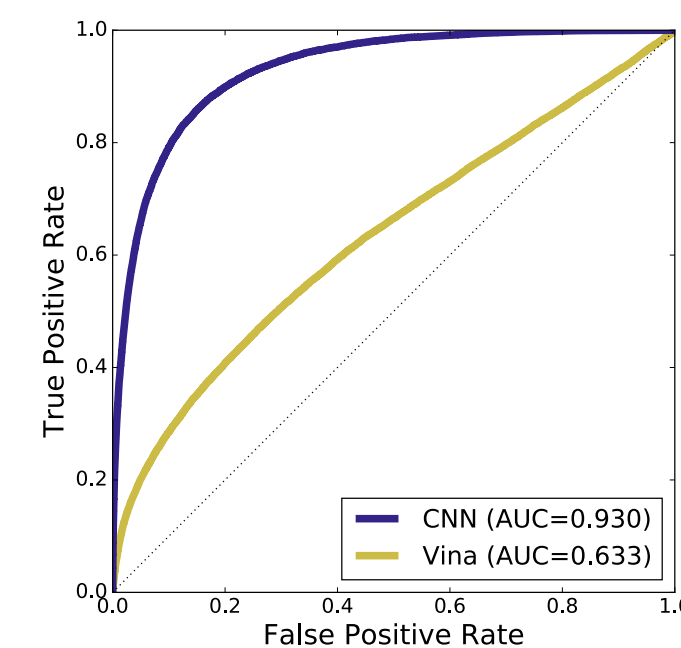
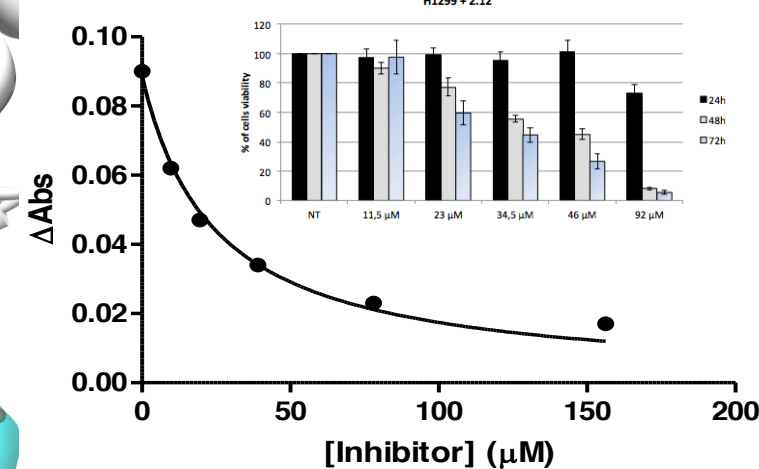
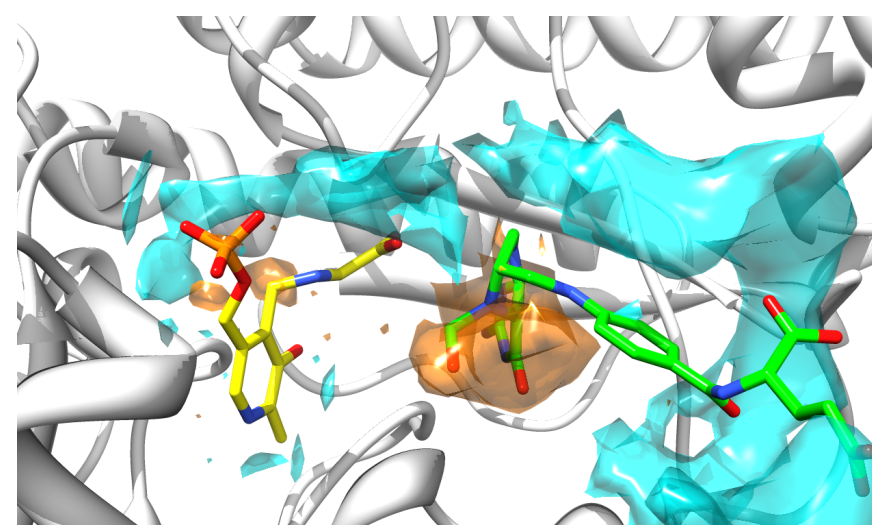
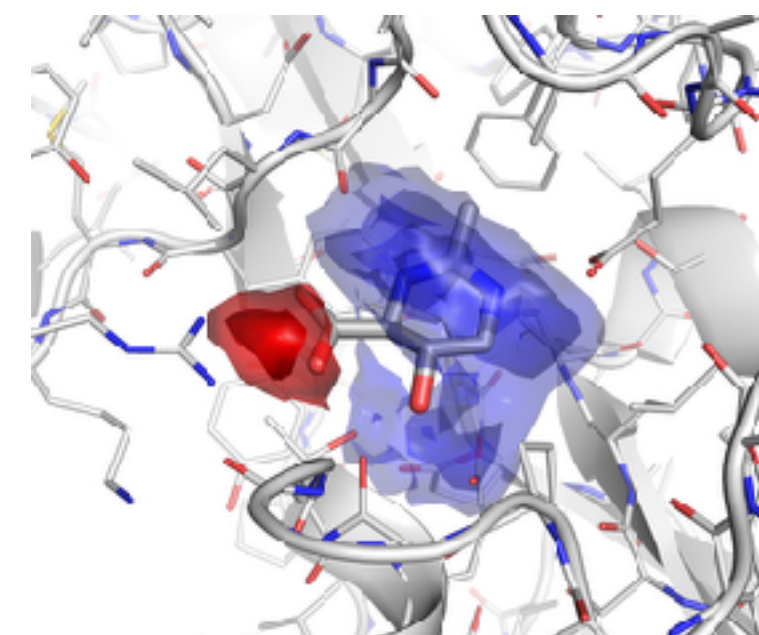
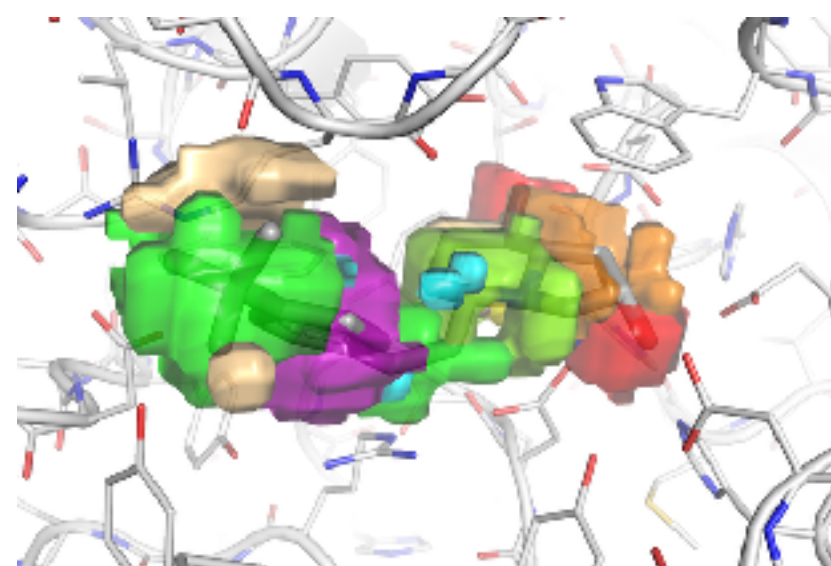
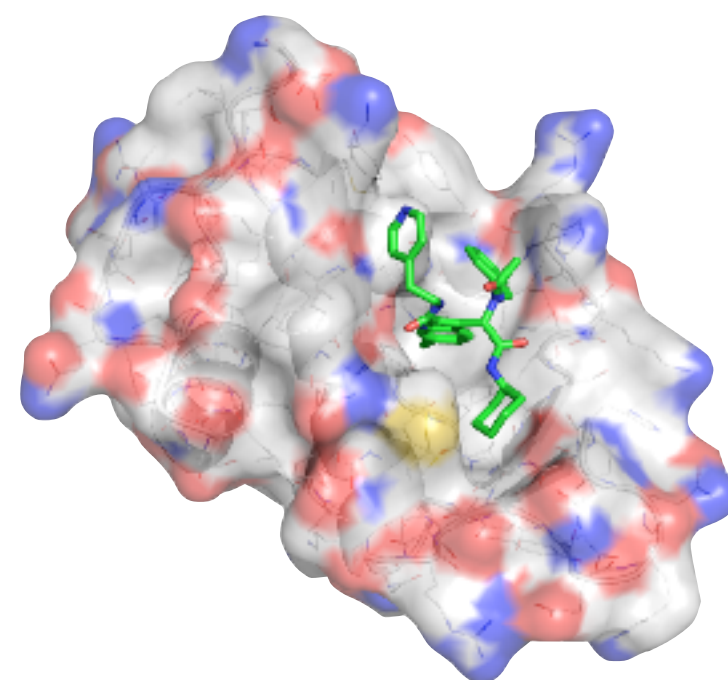
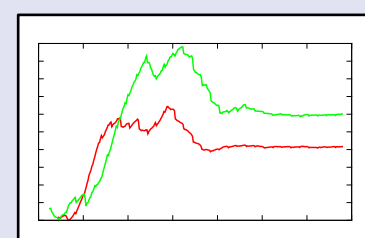
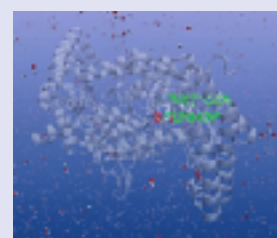
Matching



Scoring



Dynamics

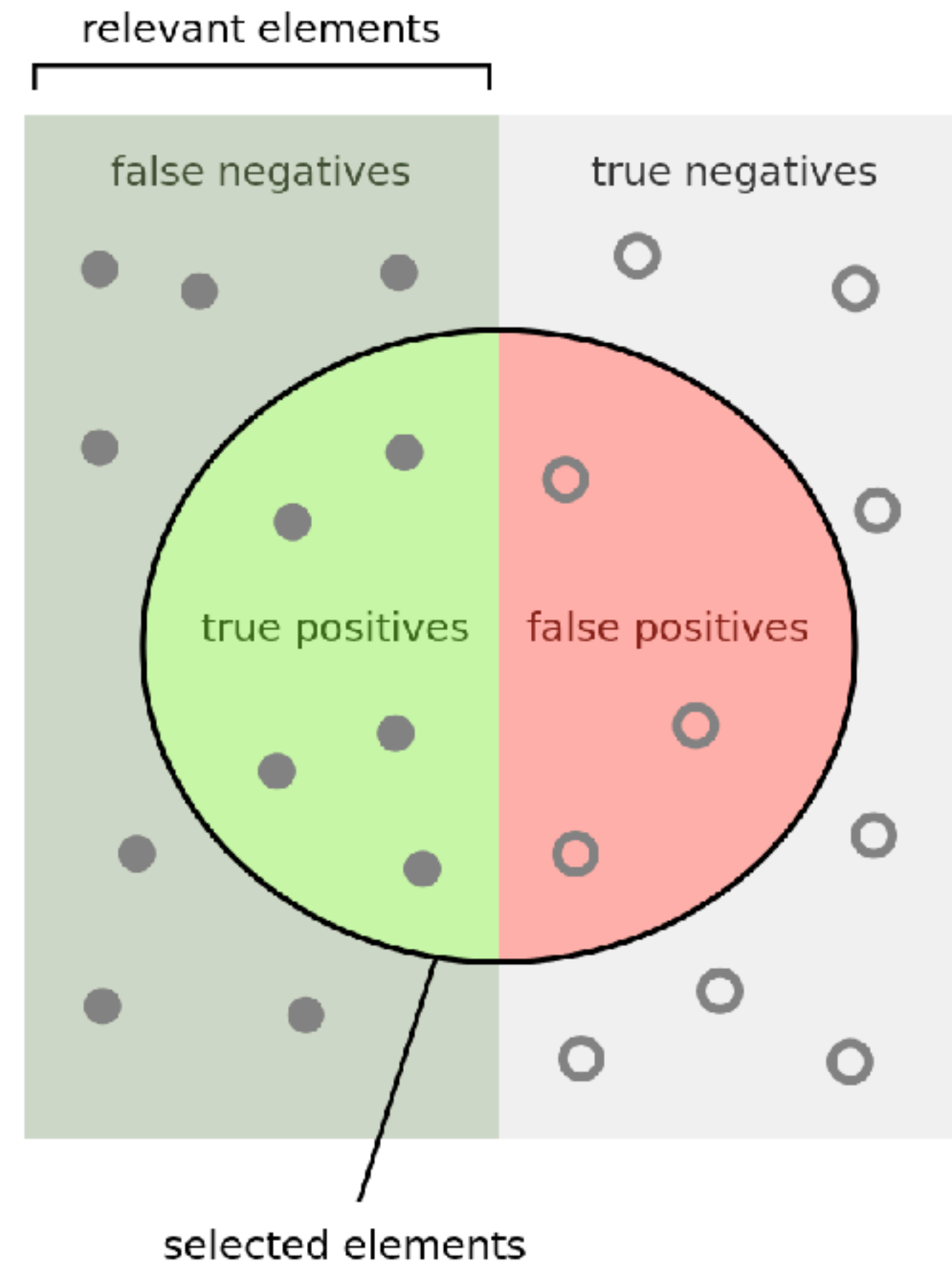


<http://pharmit.csb.pitt.edu>

4PPS

DUDe ER alpha benchmark

Metrics



How many selected items are relevant?

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\text{TPR} = \frac{\text{TP}}{P} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 1 - \text{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\text{TNR} = \frac{\text{TN}}{N} = \frac{\text{TN}}{\text{TN} + \text{FP}} = 1 - \text{FPR}$$

precision or positive predictive value (PPV)

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = 1 - \text{FDR}$$

accuracy (ACC)

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{P + N} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

F1 score

is the **harmonic mean** of **precision** and **sensitivity**

$$F_1 = 2 \cdot \frac{\text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$