# Computational Biology Workshop
# Day-3

**Vithurshan Varenthirajah, Harshit Balantrapu, Utkarsh Kapoor***

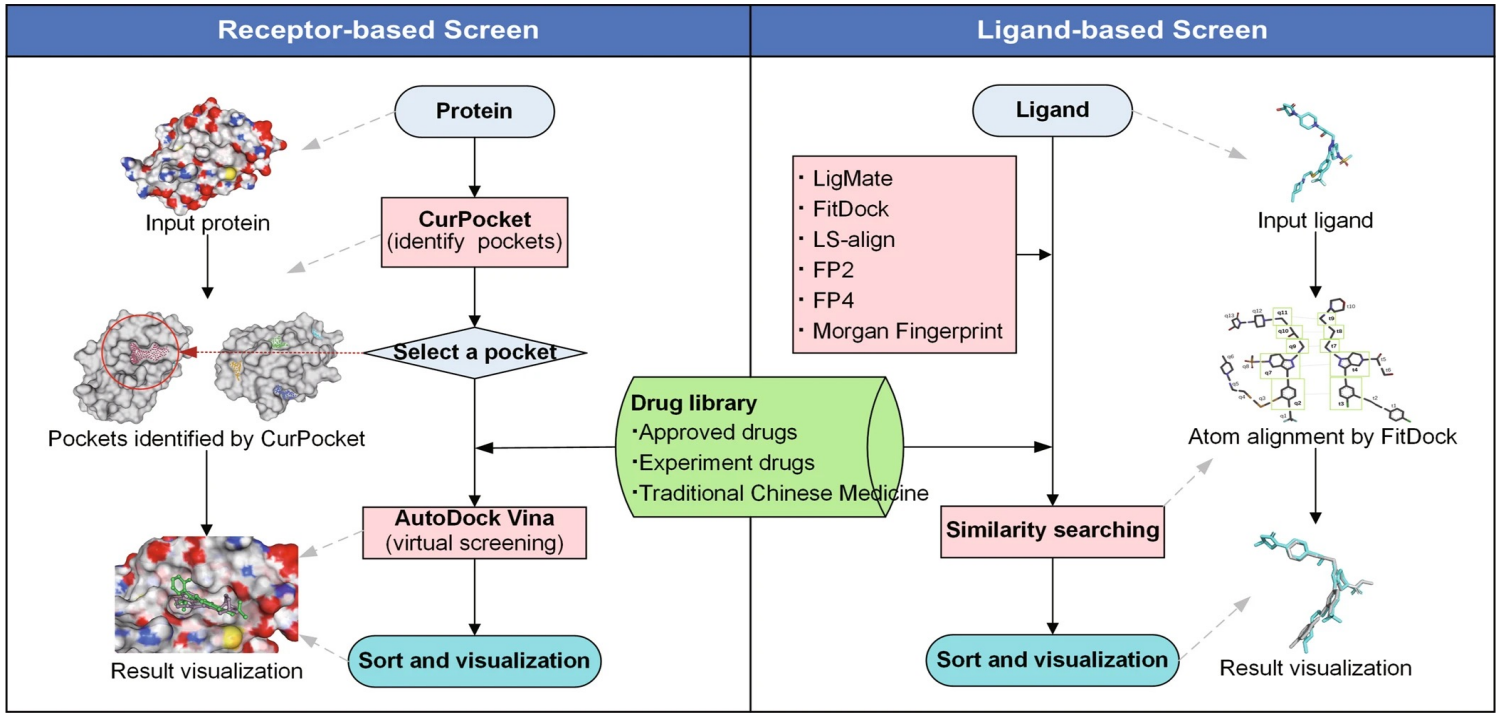
**Department of Chemical and Biomedical Engineering,**
**University of Wyoming**

# Virtual Screening

1. Virtual screening is the computational analogue of biological screening.

2. The aim is to score, rank or filter a set of chemical structures using one or more computational procedures. Docking is one of the procedure to do the virtual screening.

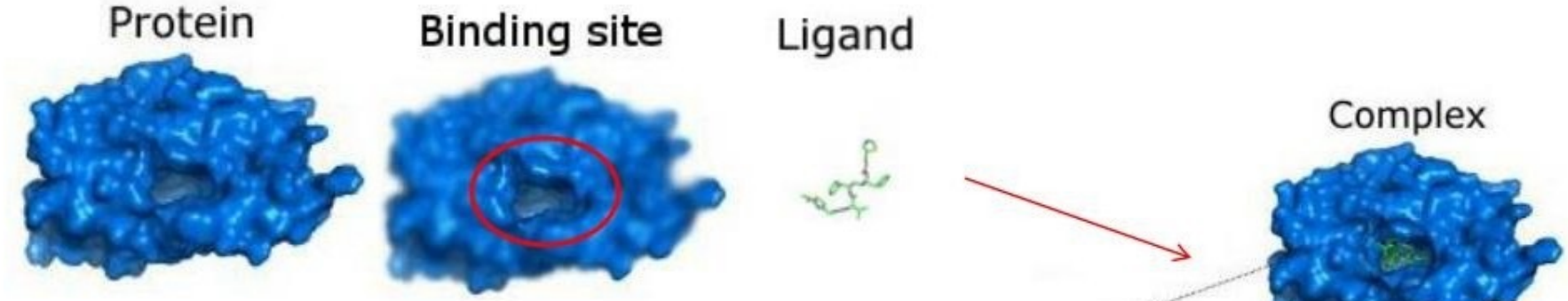
3. This used in drug discovery to search libraries of small molecules to identify those structures which are most likely to bind to a drug target, typically a protein receptor or enzyme.



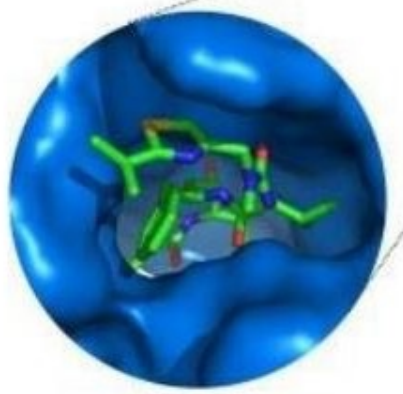

*Gan, Jian-hong, et al. "DrugRep: an automatic virtual screening server for drug repurposing." Acta Pharmacologica Sinica 44.4 (2023): 888-896.*

# Protein-ligand docking

- A Structure-Based Drug Design (SBDD) method
  - "structure" means "using protein structure"
- Computational method that mimics the binding of a ligand to a protein
- **Given...**



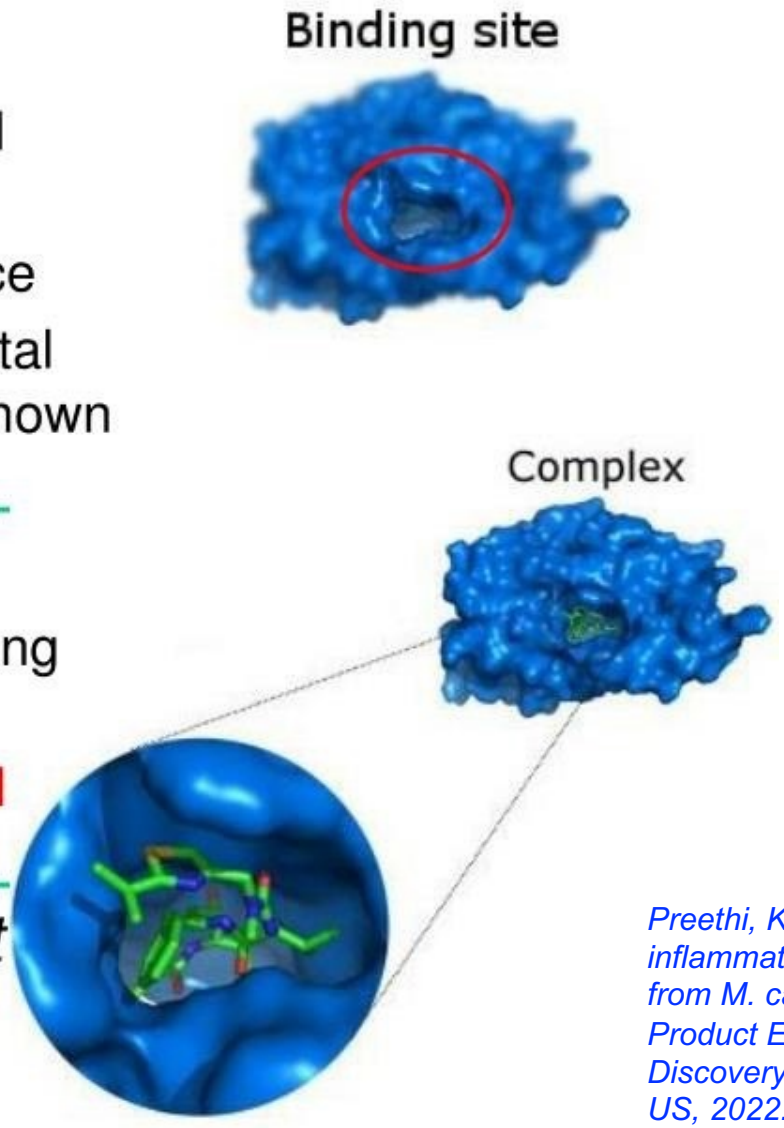

- **Predicts...**
  - The **pose** of the molecule in the binding site
  - The binding affinity or a **score** representing the strength of binding



*Ref: Preethi, K. "In Silico Analysis of Anti-inflammatory Activity of Quercetin from M. calabura Fruit." Natural Product Experiments in Drug Discovery. New York, NY: Springer US, 2022. 489-504.*

# Pose vs. Binding Site

- **Binding site** (or "active site")
  - the part of the protein where the ligand binds
  - generally a cavity on the protein surface
  - can be identified by looking at the crystal structure of the protein bound with a known inhibitor
- **Pose** (or "binding mode")
  - The *geometry* of the ligand in the binding site
  - Geometry = **location, orientation and conformation**
- *Protein-ligand docking is **not** about identifying the binding site*




*Preethi, K. "In Silico Analysis of Anti-inflammatory Activity of Quercetin from M. calabura Fruit." Natural Product Experiments in Drug Discovery. New York, NY: Springer US, 2022. 489-504.*

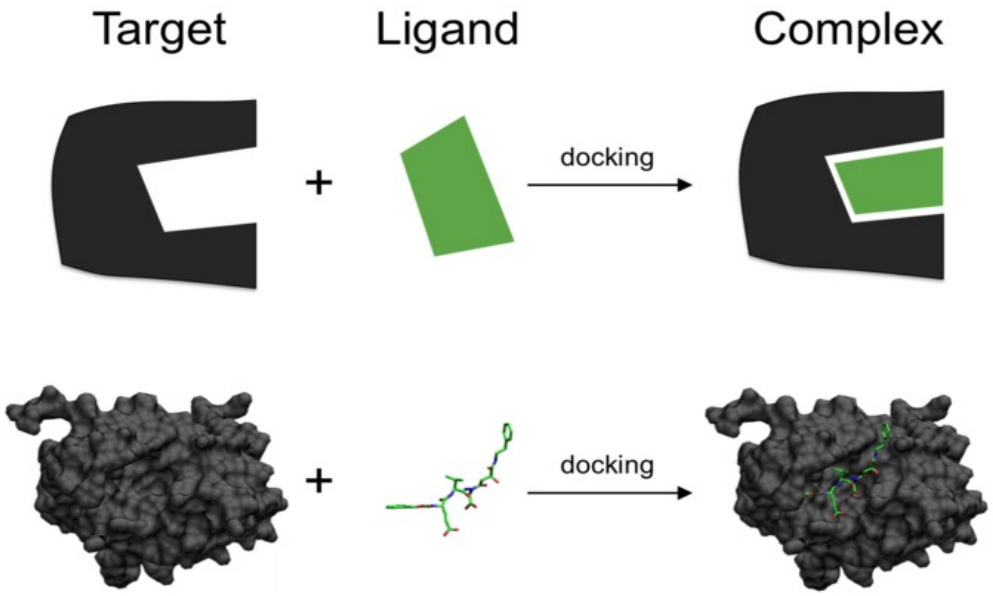
# Uses of docking

The main uses of protein-ligand docking are :

**Virtual screening** - to identify potential lead compounds from a large dataset

**Pose prediction** - If we know exactly where and how a known ligand binds

– We can see which parts are important for binding
– We can suggest changes to improve affinity
– Avoid changes that will ‘clash’ with the protein




https://en.wikipedia.org/wiki/Docking_%28molecular%29

# Types of Docking Software Available

| Name | Brief Description |
|---|---|
| Dock [62, 63] | Anchor-and-Grow based docking program, for flexible ligand and flexible protein. (http://dock.compbio.ucsf.edu/). |
| Autodock [64] | For Flexible ligand, Flexible protein side chains. Compatible for Linux, Window and Mac OS. (http://autodock.scripps.edu/). |
| Hex [65] | Mainly for protein-protein and protein-DNA docking. (http://hex.loria.fr/) |
| FTDock [66] | For rigid-body docking, based on based on Fourier correlation algorithm. (http://www.sbg.bio.ic.ac.uk/docking/ftdock.html) |
| AutoDock Vina [131] | Improved version of AutoDock4, fast, betters binding energy. (http://vina.scripps.edu/) |
| HADDock [132] | It is use for protein-protein/protein-ligand docking. (http://www.nmr.chem.uu.nl/haddock/) |

*Singla, Deepak, et al. "Open-source software and web services for designing therapeutic molecules." Current topics in medicinal chemistry 13.10 (2013): 1172-1191.*

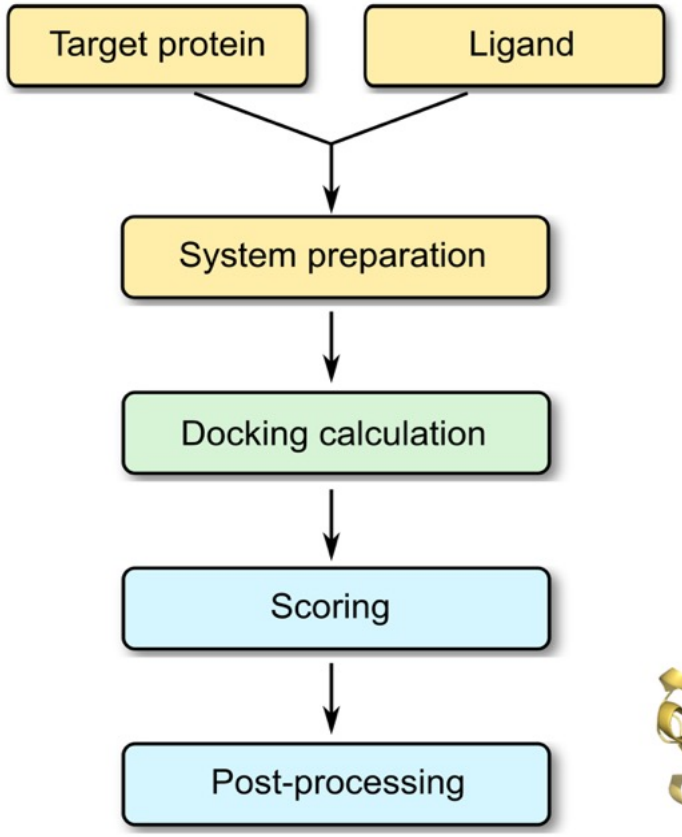
# Components of docking software

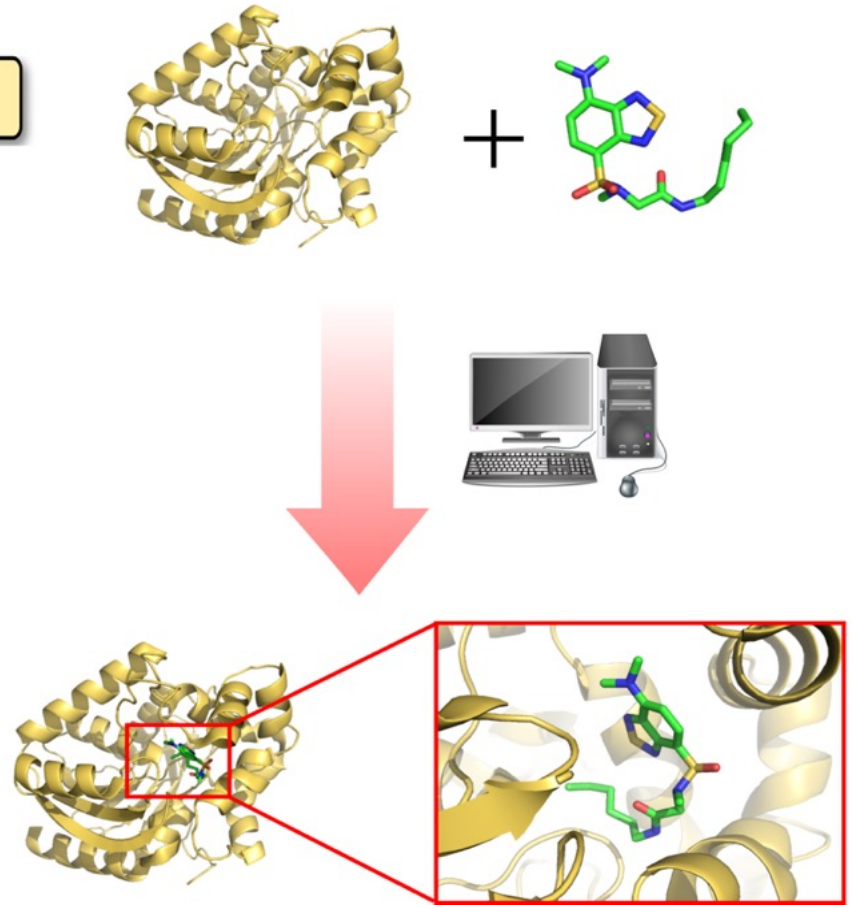
Typically, protein – ligand docking software consists of two main components which work together :

- Search algorithm
- Scoring function

Search algorithm generates large number of poses of a molecule in the binding site.

Scoring function calculates the score or binding affinity for a particular pose.



https://www.profacgen.com/protein-ligand-docking.htm

# Advantages of AutoDock-Vina

Auto Dock Vina is a free, open-source program for molecular docking that's part of the AutoDock Suite. It's commonly used in academia and has a well-accepted basic scoring function. Vina is designed to be easy to use, with no need to understand implementation details, tweak search parameters, or know advanced algebra.

It also has the following features:

- Vina improves the average accuracy of binding mode predictions compared to Auto Dock 4
- Vina's results does not have a statistical bias related to the conformation of the input structure
- Vina can take advantage of multiple CPUs or CPU cores on your system to significantly shorten its running time and robust with the multiple ligands docking.

# How to get AutoDock-Vina

**Download and install Auto Dock Vina (Standalone)**
https://vina.scripps.edu/downloads/

**OR**

**Install UCSF Chimera and use the inbuild AutoDock Vina**

# Installation of Chimera

**Step 1: Download Chimera**

1.Go to the UCSF Chimera website: Chimera Download. https://www.cgl.ucsf.edu/chimera/download.html
2.Select the appropriate version for your operating system (Windows, macOS, or Linux).
3.Download the installer file.

**Step 2: Install Chimera**

*Windows*

1.Run the downloaded installer (e.g., chimera-1.15-win64.exe).
2.Follow the installation prompts, choosing the default options unless you have specific preferences.
3.Once the installation is complete, launch Chimera from the Start Menu or desktop shortcut.

*macOS*

1.Open the downloaded disk image file (e.g., chimera-1.15-mac.dmg).
2.Drag the Chimera application to your Applications folder.
3.Launch Chimera from the Applications folder or Dock.

**Step 3: Verify Chimera Installation**

1.Launch Chimera.
2.The Chimera interface should open, indicating successful installation.

# Go to the Tutorial

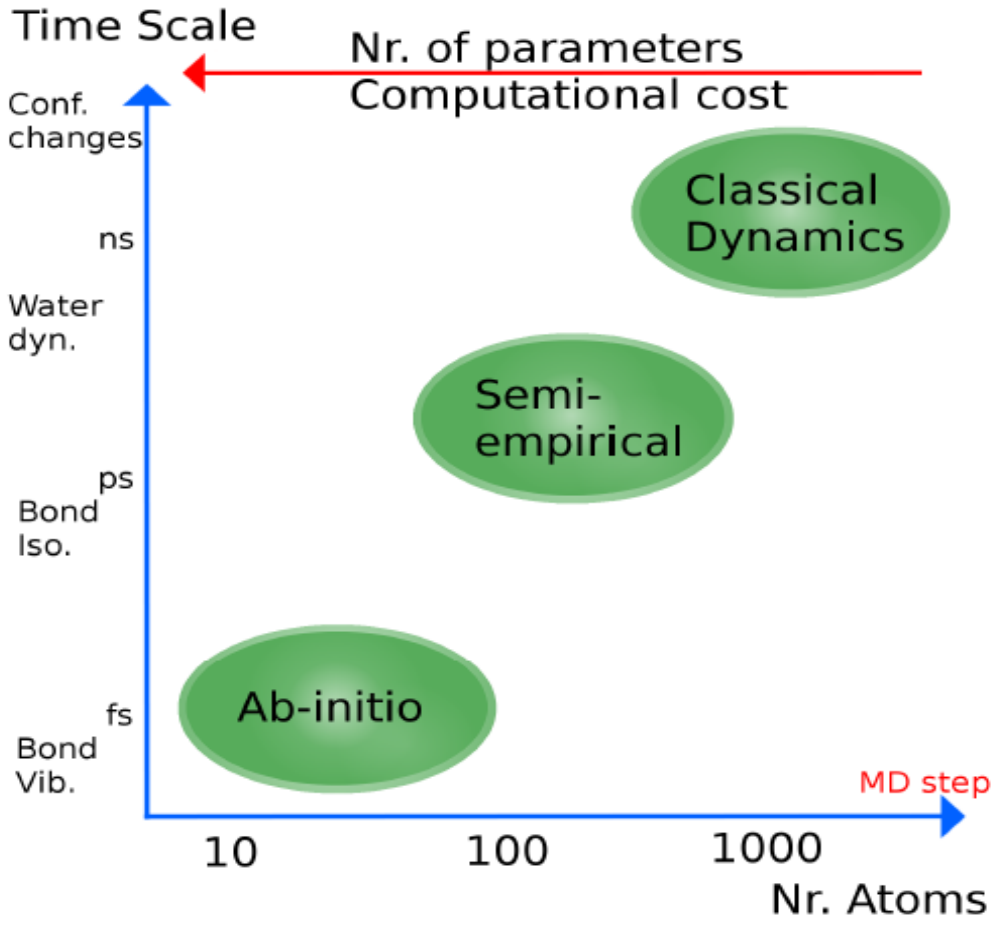
# Limitations of conventional protein-ligand docking and molecular dynamics

- Docking programs simplify protein and ligand models to expedite calculations, but this simplification may disregard crucial dynamic interactions essential for accurate binding predictions.
- Predicted ligand poses generated by docking programs often represent a single static configuration, neglecting the dynamic nature of protein-ligand interactions.
- Scoring functions utilized in docking algorithms aim to predict binding affinity by considering various factors such as interatomic interactions and solvation effects. However, these scoring functions may struggle to accurately account for all the complexities involved in protein-ligand binding, leading to potential inaccuracies in predictions.
- Conventional molecular dynamics simulations used for sampling different ligand conformations can be time-consuming due to the presence of high energy barriers that hinder conformational transitions.
- Enhanced sampling methods can accelerate sampling events by lowering energy barriers by adding bias potentials, providing a more comprehensive exploration of the ligand's conformational space, leading to a more accurate estimation of binding free energy within an affordable time.

# Why Enhanced Sampling?

The Multiple Time Scale Problem:

- Interesting molecular processes involve multiple time scales
- If there were just one long time scale, we could just do that directly
- But we need things happening at all time scales




https://www.hpc2n.umu.se/sites/default/files/pdf-files/Documentation/talk_compchemcour_20Jun2016.pdf

# Types of Enhanced Sampling Algorithms available

**Meta-dynamics** : Adds a history-dependent bias to overcome energy barriers and sample rare events.

**Umbrella Sampling** : Uses a biasing potential to enhance sampling in specific regions of the phase space, typically to calculate free energy profiles.

**Replica Exchange** : Also known as Parallel Tempering, it involves running multiple simulations at different temperatures and periodically exchanging configurations to enhance sampling.

**Accelerated Molecular Dynamics (aMD)** : Modifies the potential energy surface to reduce barriers, allowing faster transitions between states.

**Steered Molecular Dynamics (SMD)** : Applies an external force to drive the system along a specific reaction coordinate, facilitating the exploration of rare events.

**Adaptive Biasing Force (ABF)** : Computes and applies the mean force along a reaction coordinate to flatten the free energy landscape and improve sampling efficiency.

**Enhanced Sampling of Nonequilibrium Dynamics (END)** : Applies nonequilibrium work theorems to enhance sampling of systems driven out of equilibrium.

**Gaussian Accelerated Molecular Dynamics (GaMD)** : Introduces a harmonic boost potential to enhance the sampling of conformational space without knowing the reaction coordinates beforehand.

**Weighted Ensemble (WE)** : Divides the phase space into bins and redistributes trajectories to focus computational effort on under-sampled regions.

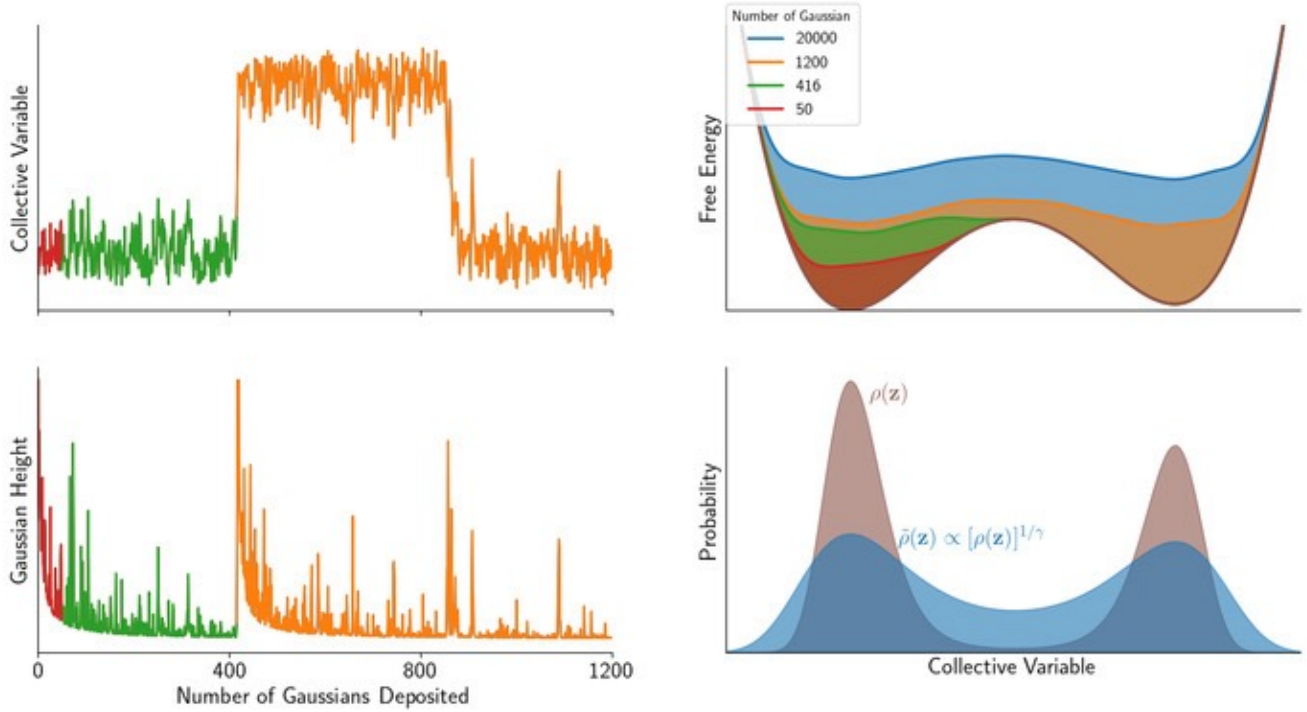
**Hamiltonian Replica Exchange** : Similar to the temperature replica exchange but exchanges different Hamiltonians (e.g., different potentials) to enhance sampling.

**Temperature Accelerated Molecular Dynamics (TAMD)** : Involves periodically scaling the temperature to escape local minima and then rescaling to the original temperature to enhance sampling.

# Metadynamics

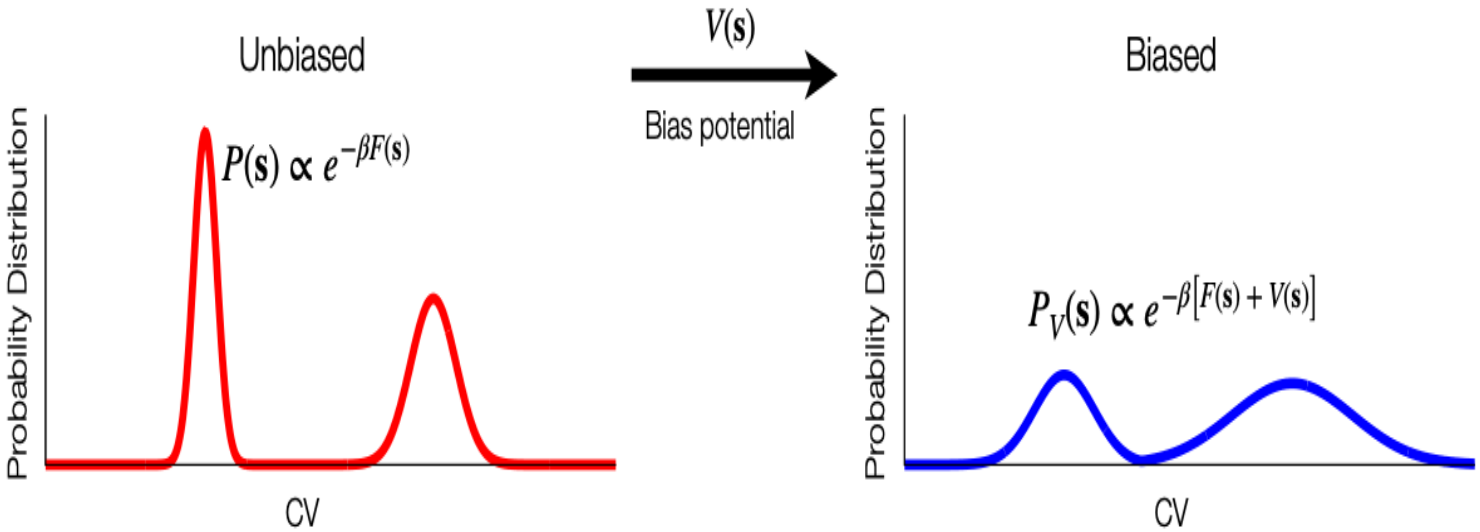
- The bare bones of metadynamics
  - Bias away from previously visited configurations
  - In a reduced space of collective variables
  - At a sequentially decreasing rate of bias

- Key thought experiments to build intuition
  - Bias size, shape, and rate
  - Good and bad collective variables

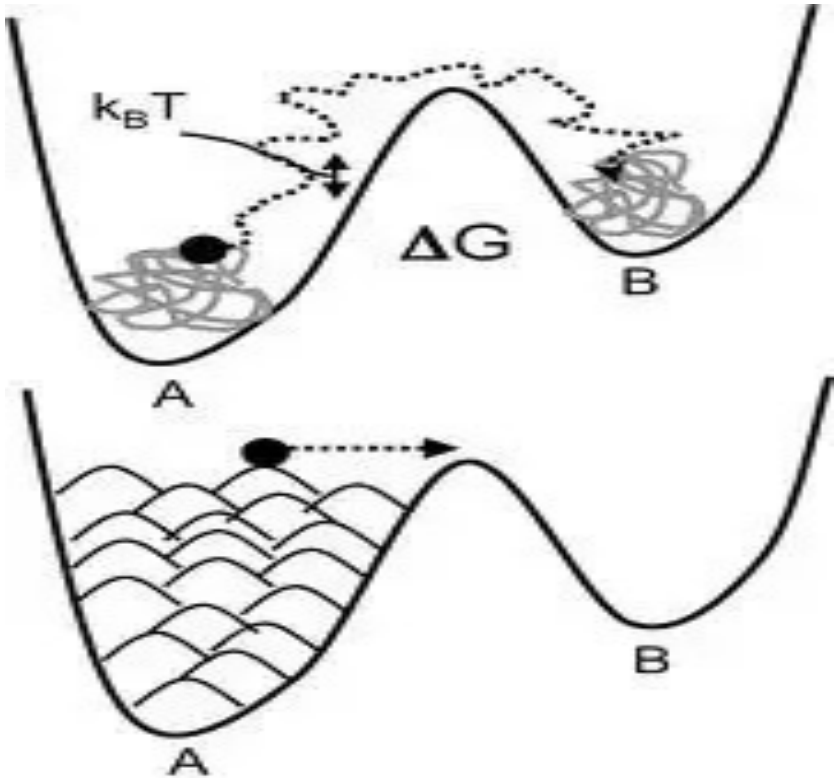



*Hénin, Jérôme, Tony Lelièvre, Michael R. Shirts, Omar Valsson, and Lucie Delemotte. "Enhanced sampling methods for molecular dynamics simulations." arXiv preprint arXiv:2202.04164 (2022).*

# Hills and Bias




*Valsson, Omar, Pratyush Tiwary, and Michele Parrinello. "Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint." Annual review of physical chemistry 67 (2016): 159-184.*



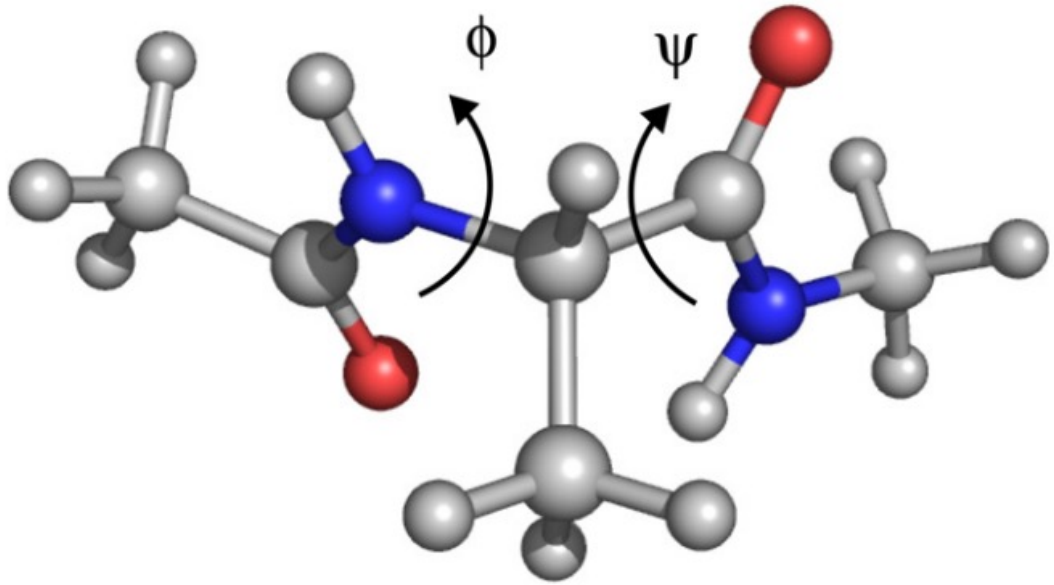

*Kellett, Kathryn Emily. "Development of chemical sensors for rapid identification of amphetamine-related new psychoactive substances." (2017).*

# Collective Variables

Any function of any number of fine gained variables:

- Position, distance, angle, dihedral
- Coordination number, density, crystalline order
- Helicity, contact map, NMR spectrum
- Strings of configurations in another CV space

**Dihedral angle as the collective variable**



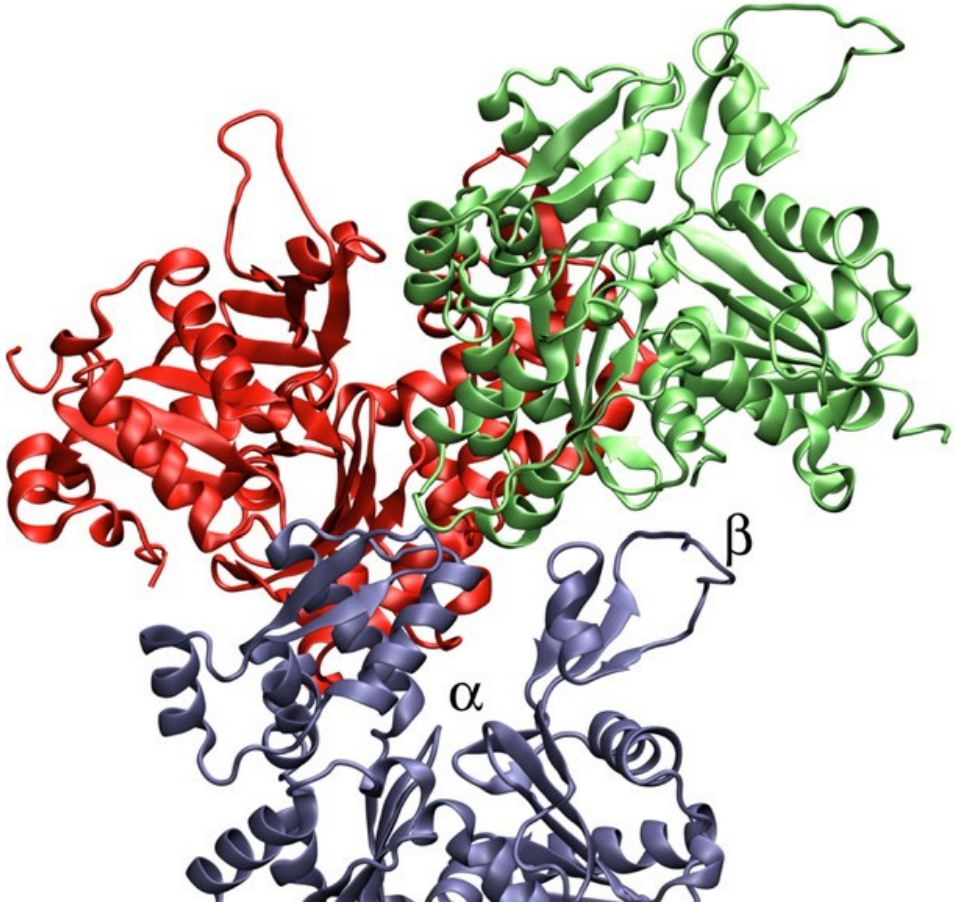

*Laio, Alessandro, and Francesco L. Gervasio. "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science." Reports on Progress in Physics 71.12 (2008): 126601.*

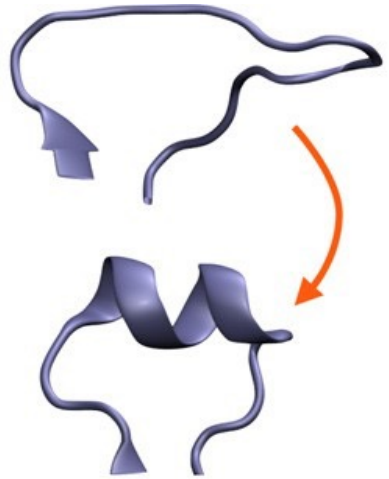
# Applications and Examples from the Literature

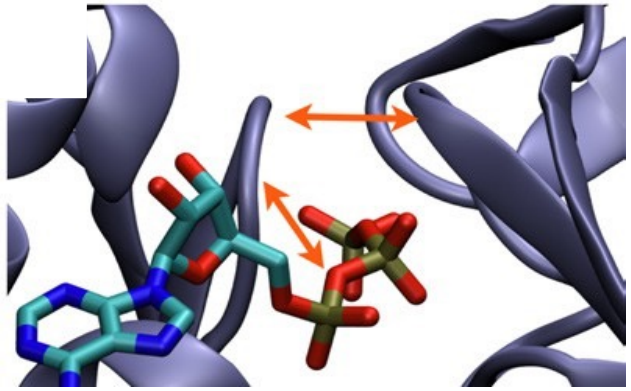
- Chemical reaction
- Phase transition
- Interfacial chemistry
- Protein function
- Protein folding

# Cytoskeleton Protein Function

*Pfaendtner, Barducci, Parrinello, Pollard, and Voth. Nucleotide--dependent conformational states of actin. 2009*





# Protein Folding

*Granata, Camilloni, Vendruscolo, and Laio. Characterization of the free--energy landscapes of proteins by NMR guided meta-dynamics. 2013*

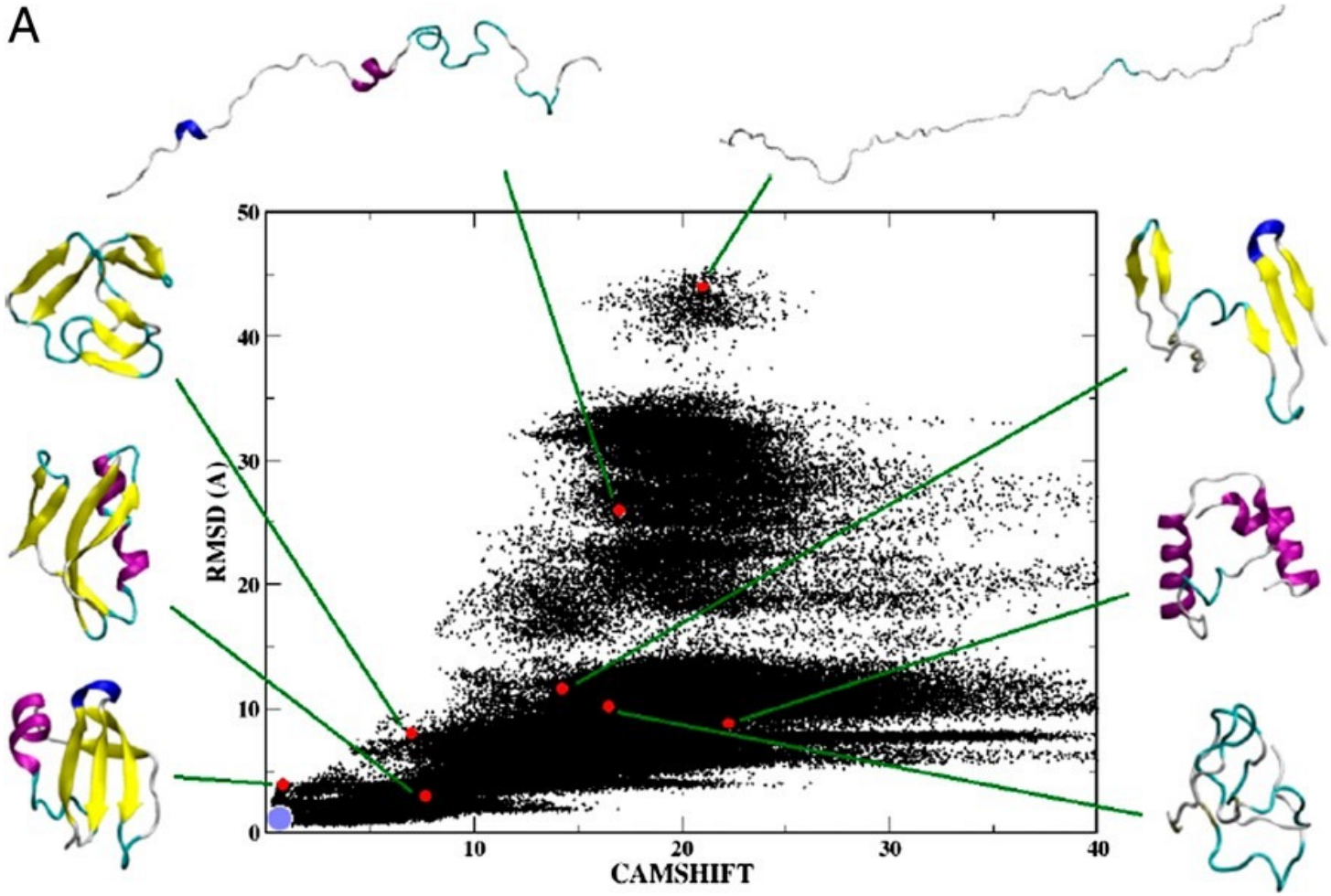
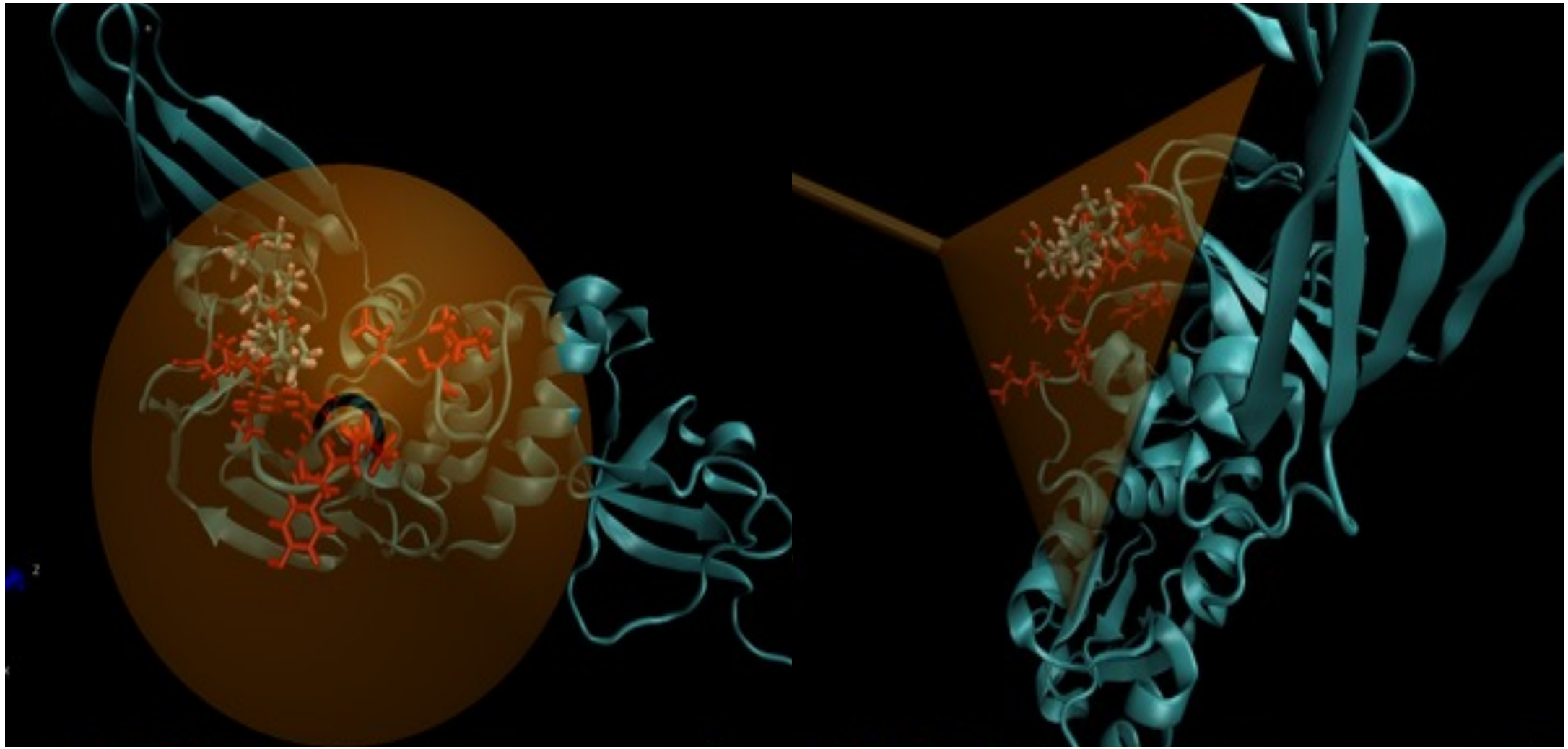



Fig. 2

# Funnel Metadynamics



# Comparison with other methods

- Molecular docking protocols have played a leading role in finding binding role in shorter period.
- However, docking calculations suffer from  poor sensitivity and specificity, inaccurate scoring function, and lack of explicit solvent.
- Umbrella Sampling (US) and steered molecular dynamics (SMD) takes more time to converge.
- Potential mean force (PMF) of US depends on sampled path, so same system might converge to diverse results.
- Simulation other enhanced sampling methods (eg metadynamics simulation) spends a lot of time exploring the unbound state. Therefore, it takes long time to converge.

# Advantages of Funnel Metadynamics

- No a priori knowledge on the ligand binding mode is required.
- The ligand, the protein and the solvent molecules are explicit – helps determine the effects of solvent molecules.
- The entire ligand binding mechanism can be constructed - binding pathway of the ligand from its solvated state to its final binding mode in the target structure can be elucidated
- Multiple back and forth binding and unbinding events can be observed
- FM is self-diagnostic of nonconvergent results – free energy value does not converge unless appropriate CV is used.

# Funnel Metadynamics

- From the BFES, one can identify the ligand binding mode as the lowest free-energy minimum and compute the absolute protein–ligand binding free energy using the following formula.

$$\Delta G_{\mathrm{b}}^{0} = -k_{\mathrm{B}}\, T \ln\left(C^{0} K_{\mathrm{b}}\right).$$

- $C^0$ is the standard concentration of 1 M for all reacting molecules, $k_B$ is the Boltzmann constant; T is the temperature of the system; and $K_b$ is the equilibrium binding constant obtained from the free-energy difference between the bound and unbound state

- The binding free energy depends exclusively on the free-energy values of these two states, independently of the path that connects one to the other.

# Go to the Tutorial

# Limitations of Funnel Metadynamics

- Positioning of the funnel in the system.
- There is no methods for selecting appropriate CVs.
- Cost of free energy construction grows exponentially with the number of CVs.
- In the case of a bad CV choice simulation does not converge.