USI Lugano

AIDD Summer School 2022

13 May 2022

University of Naples
"Federico II"

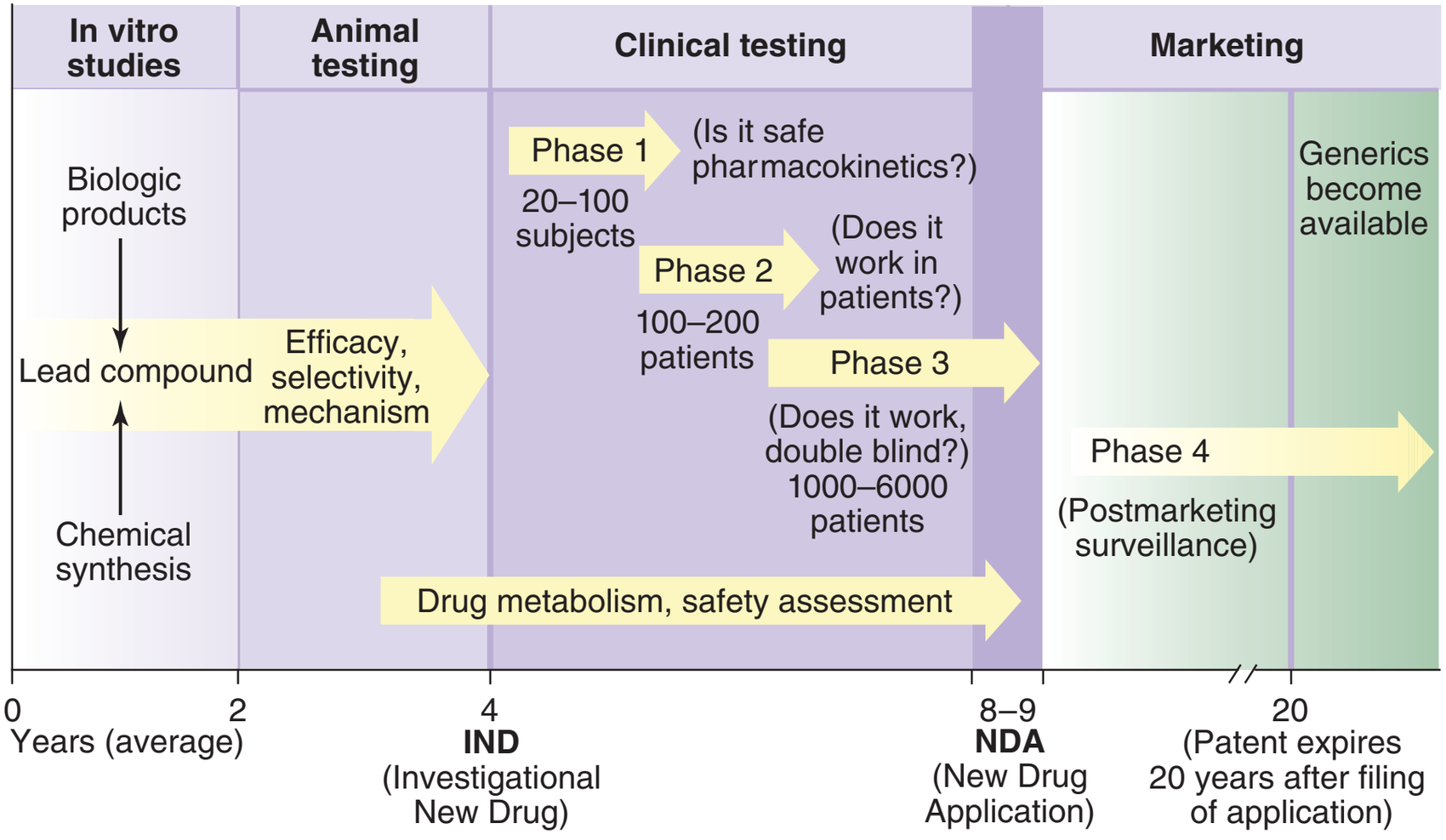
Can we learn chemical and energetic properties from molecular simulations?

Vittorio Limongelli
vittoriolimongelli@gmail.com

Outline of the Talk

- Introduction to drug discovery and drug/target binding (DTB)
- Theory and simulation methods of DTB
- Machine learning augmented sampling

Drug Discovery - a (very) long process



ML at the service of DD

ML Methods

- Neural Network
- Graph Theory
- Deep Learning
- Active Learning
- …

DD Applications

- Molecules Generation
- Molecule Synthesis
- Drug/Target affinity
- Molecule Pharmacokinetics
- Molecule Pharmacodynamics

Literature:
Vamathevan et al. *Nat Rev Drug Discov* (2019)
Chen et al. Drug *Discovery Today* (2018)
Bickerton et al. *Nat Chem* (2012)

ML at the service of DD

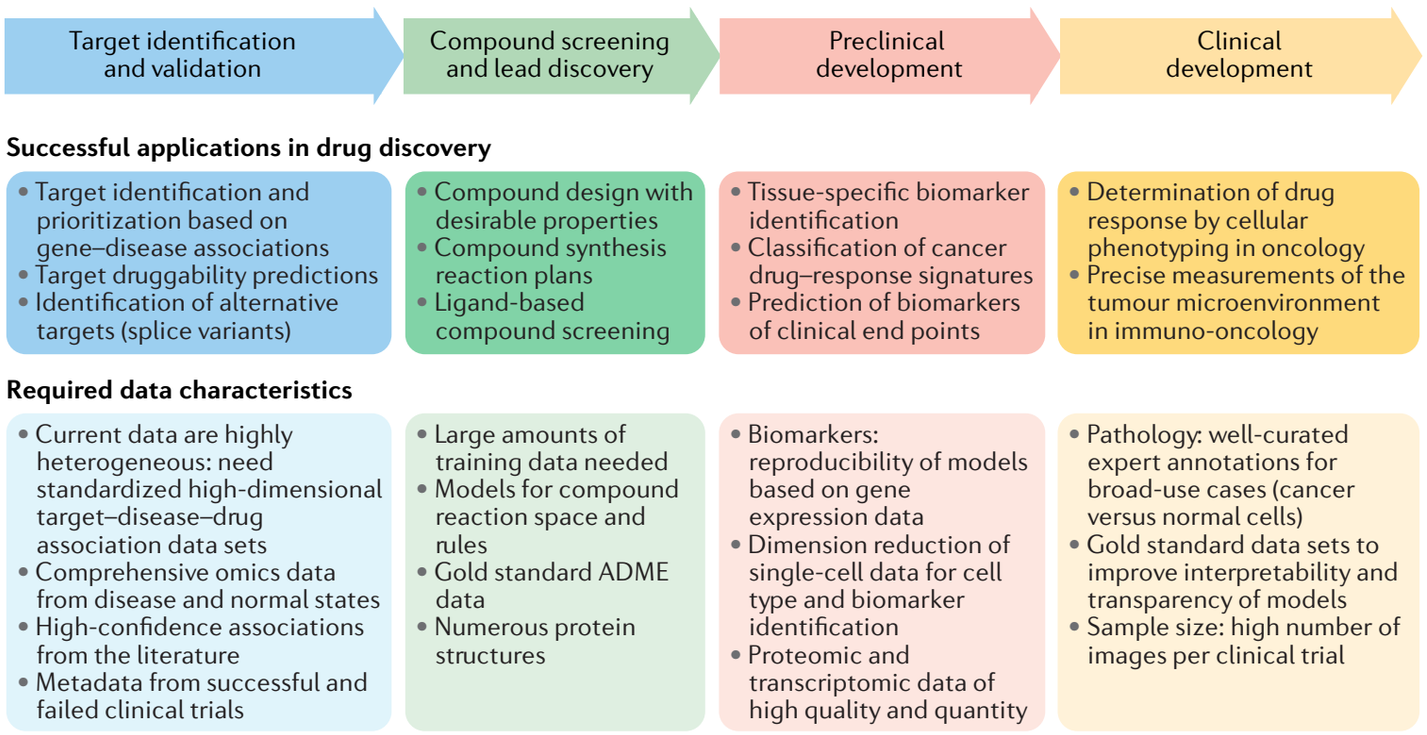
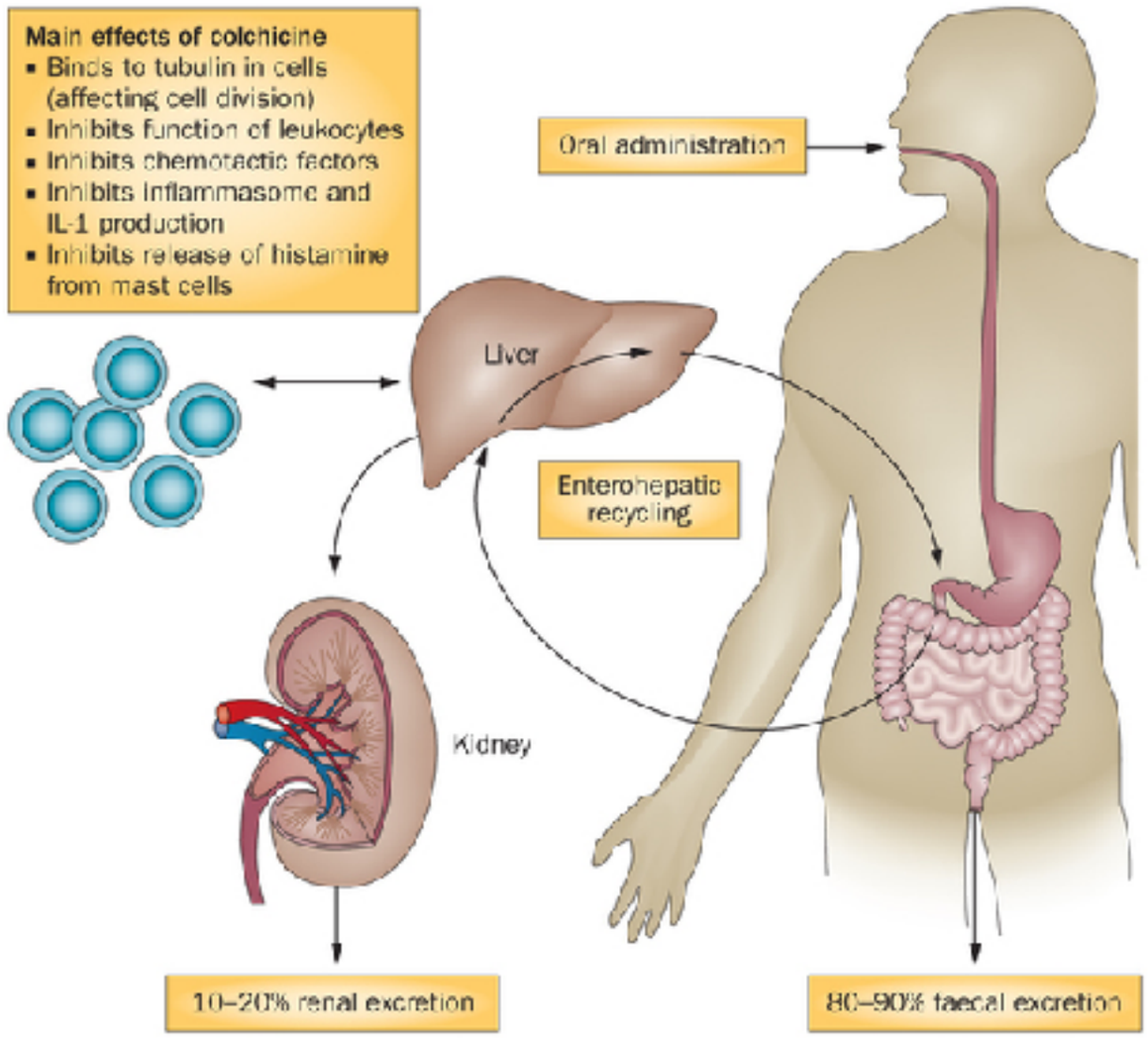


Fig. 1 | **Machine learning applications in the drug discovery pipeline and their required data characteristics.** Several successful applications of machine learning in various stages of the drug development pipeline in pharmaceutical companies have been published. However, within each data domain, there are still challenges related to the standard of data quality and data quantity needed to capitalize on the full potential of these methods for discovery. ADME, absorption, distribution, metabolism and excretion.

Vamathevan et al. *Nat. Rev. Drug Discov.* (2019)

From Outside to Inside

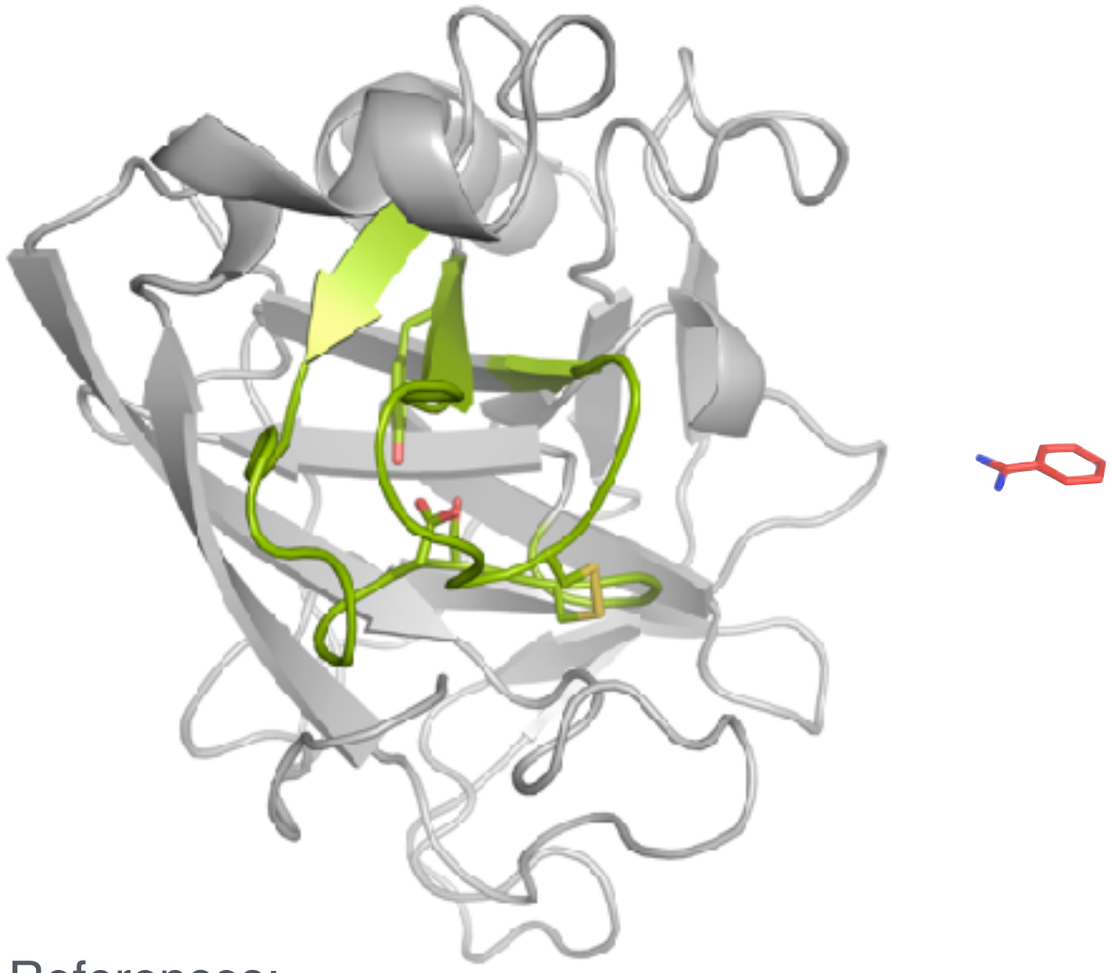


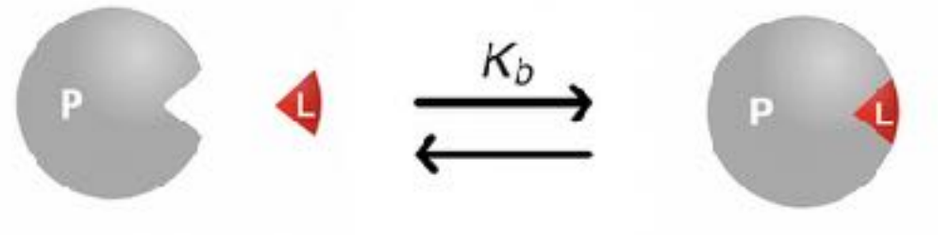
From Outside to Inside

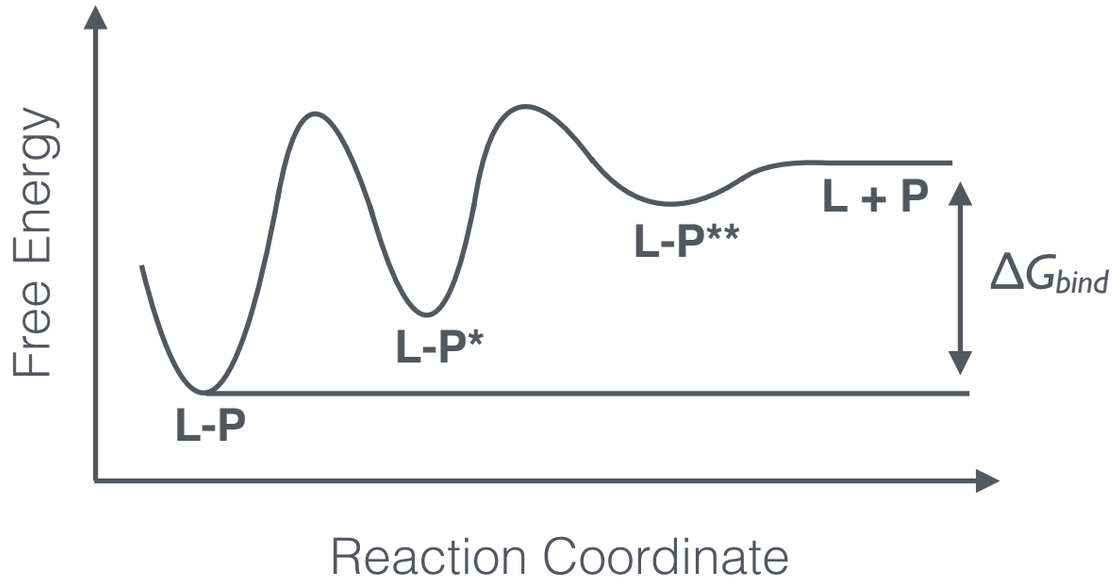
Pharmacokinetics: The actions of the body on the drug (ADME)

Pharmacodynamics: The actions of a drug on the body

Drug/Target Binding







References:

Limongelli et al. PNAS (2010)

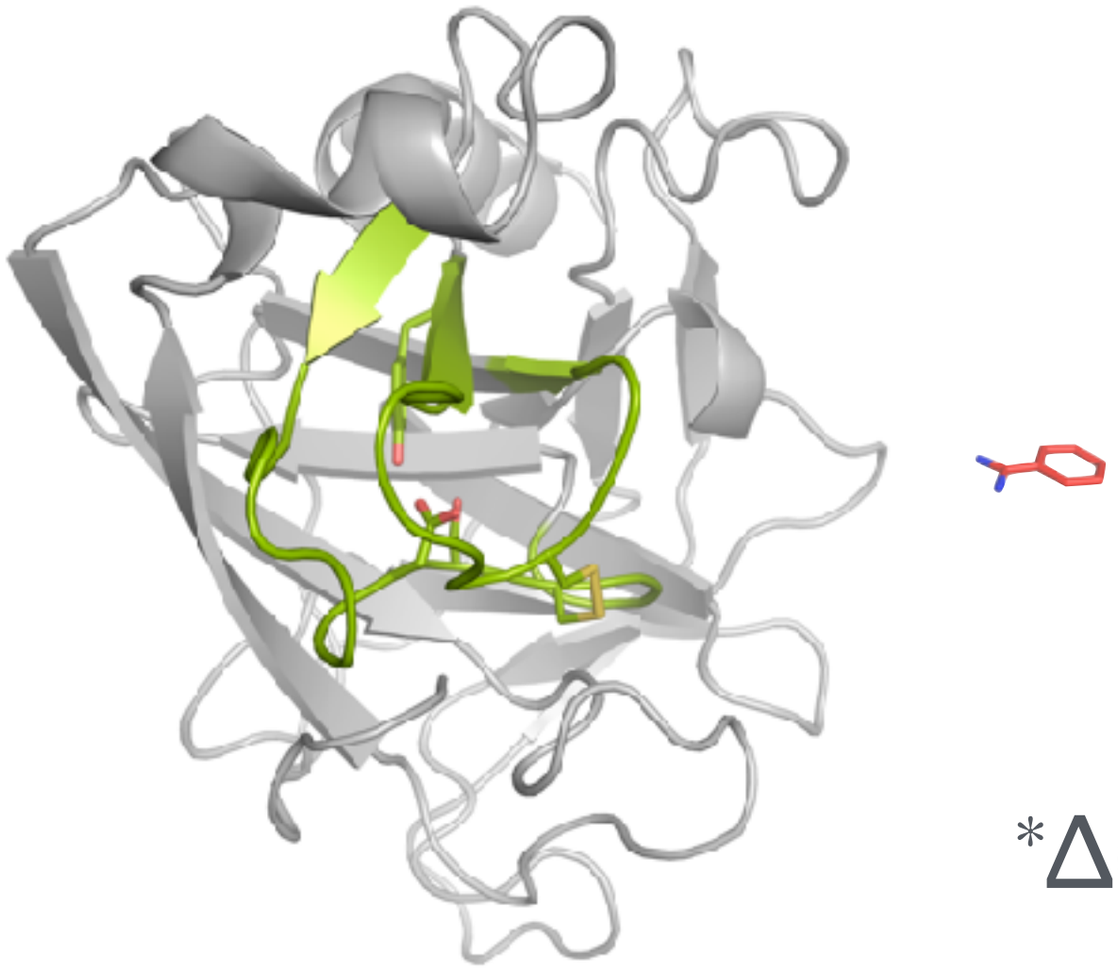
Limongelli et al. PNAS (2013)

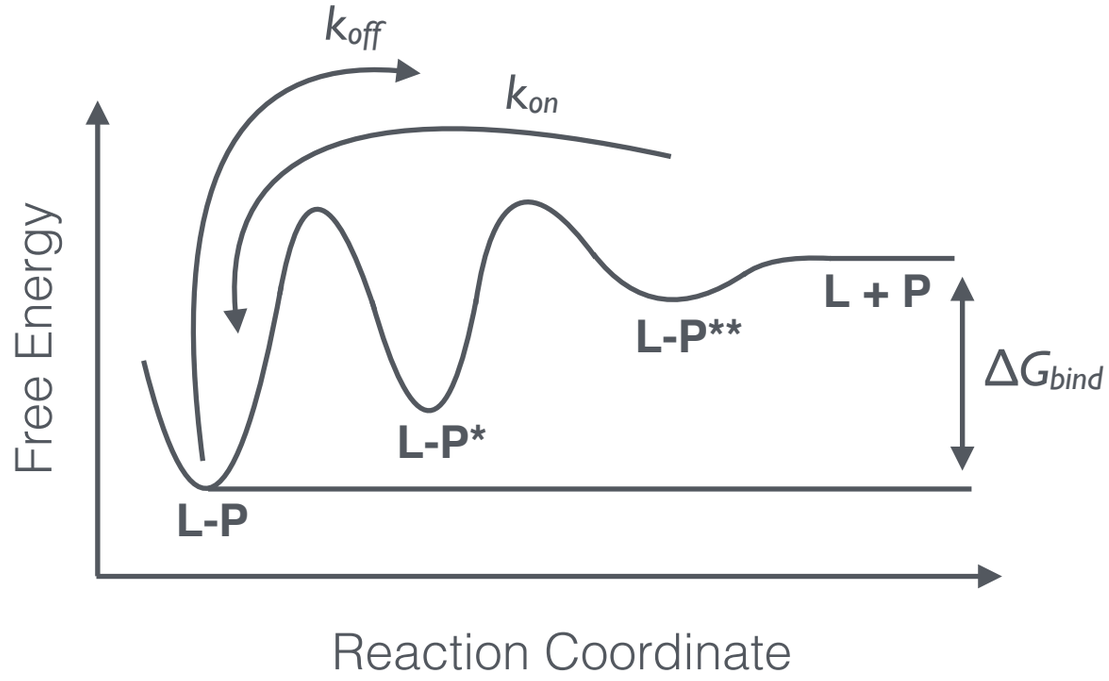
Tiwary, **Limongelli** et al. PNAS (2015)

Limongelli WIREs Comp Mol Sci (2020)

$$[L] + [P] \underset{}{\overset{k_b}{\rightleftharpoons}} [LP] \qquad K_b = \frac{[LP]}{[L]\,[P]}$$

Drug/Target Binding





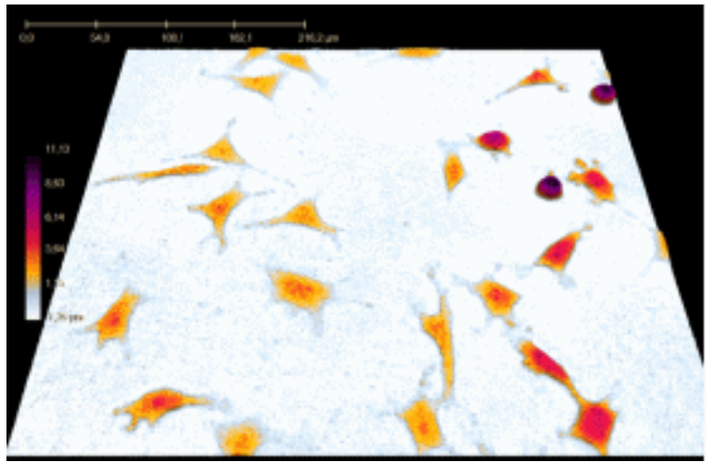
$$^{*}\Delta G_{bind} = G_{[LP]} - G_{[L][P]} = -k_b T \ln K_b$$

* in case of competitive inhibitors

$$\Delta G^0 = \mu_{LP} - \mu_L - \mu_P = -k_b T ln \left(\frac{C^0}{8\pi^2} \frac{\int e^{-(U_{(rLP)} + S_{(rLP)}/k_b T)} dr_{(LP)}}{(\int e^{-(U_{(rL)} + S_{(rL)}/k_b T)} dr_{(L)})(\int e^{-(U_{(rP)} + S_{(rP)}/k_b T)} dr_{(P)})} \right)$$

Limongelli *WIREs Comput. Mol. Sci.* (2020)
Gilson, Zhou. *Annu Rev Biophys Biomol Struct* (2007)

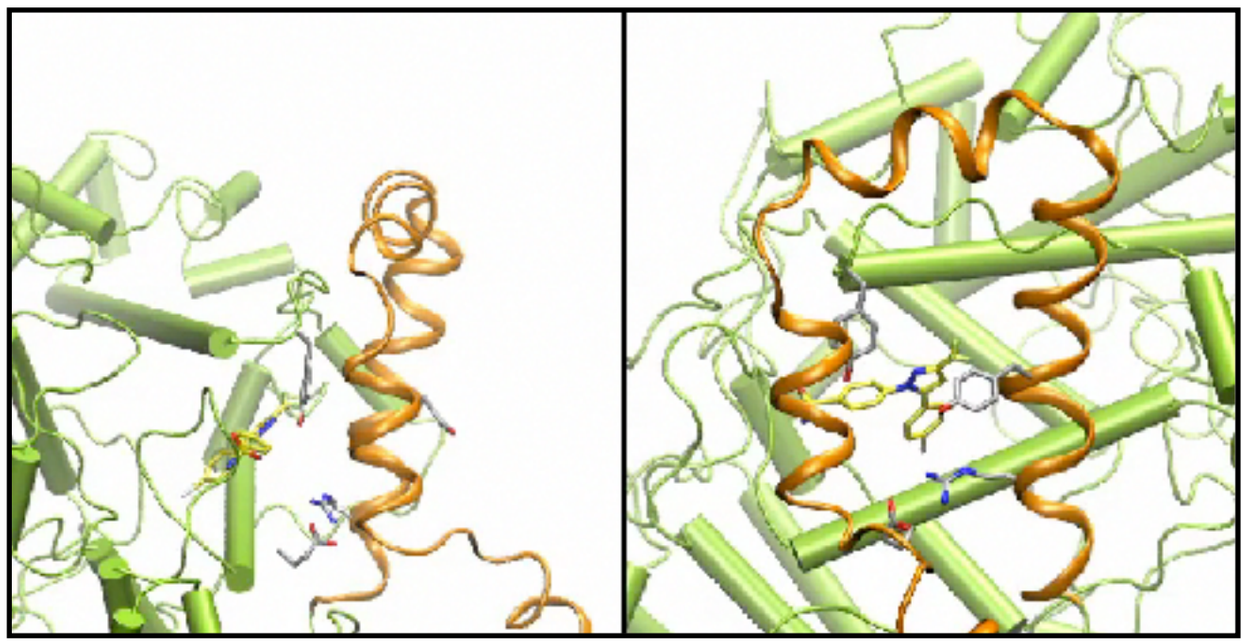
Life is Dynamics



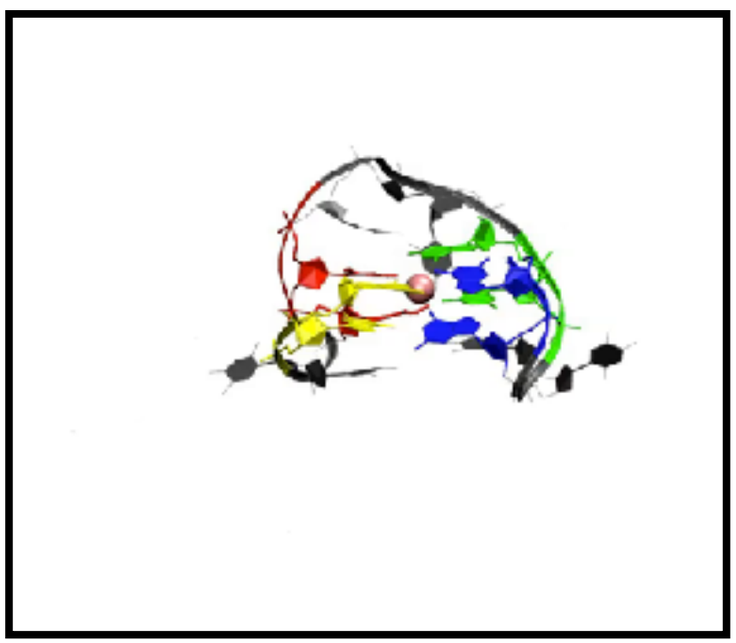
Source: wikimedia commons

…including proteins and DNA

Limongelli et al., *PNAS* (2010)



Limongelli et al., *Angew. Chem. Int. Ed.* (2013)

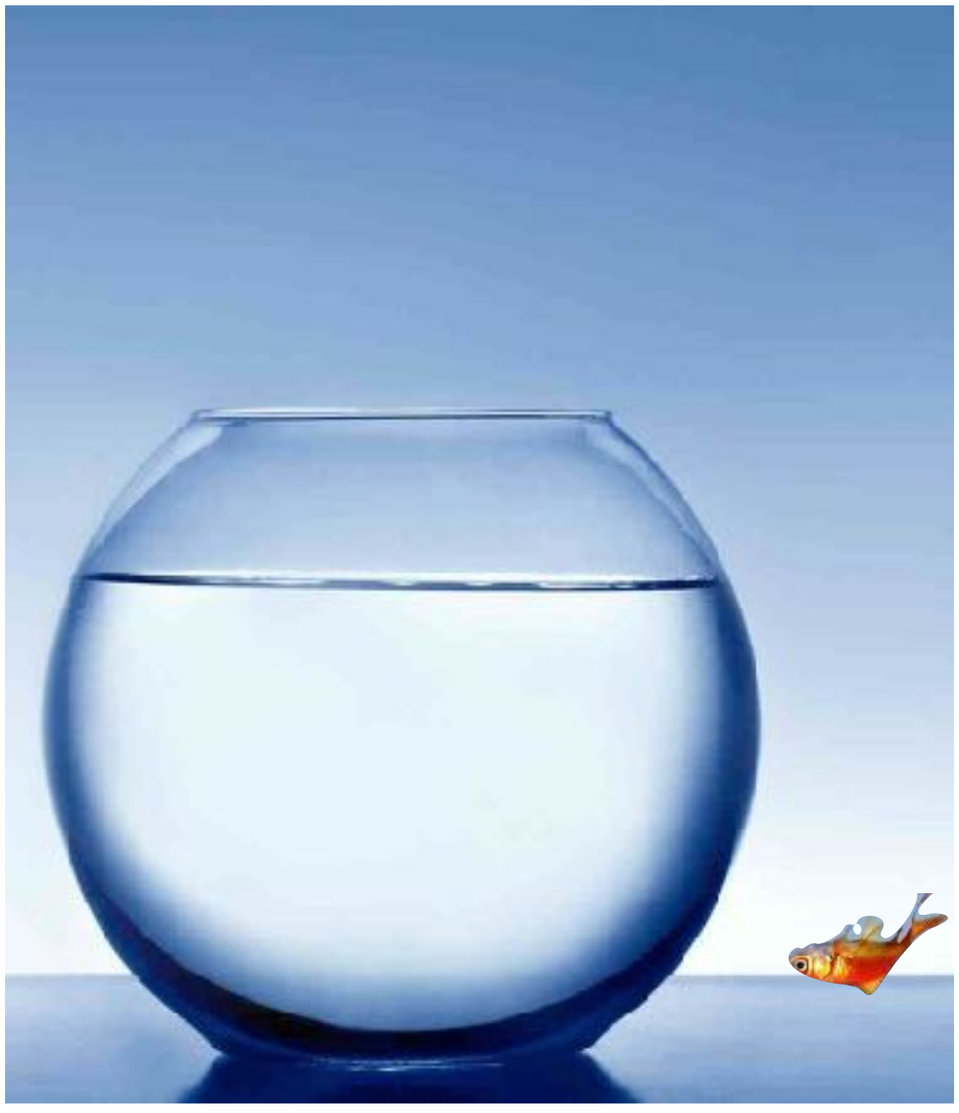


Life is Dynamics

What if we neglect dynamics?

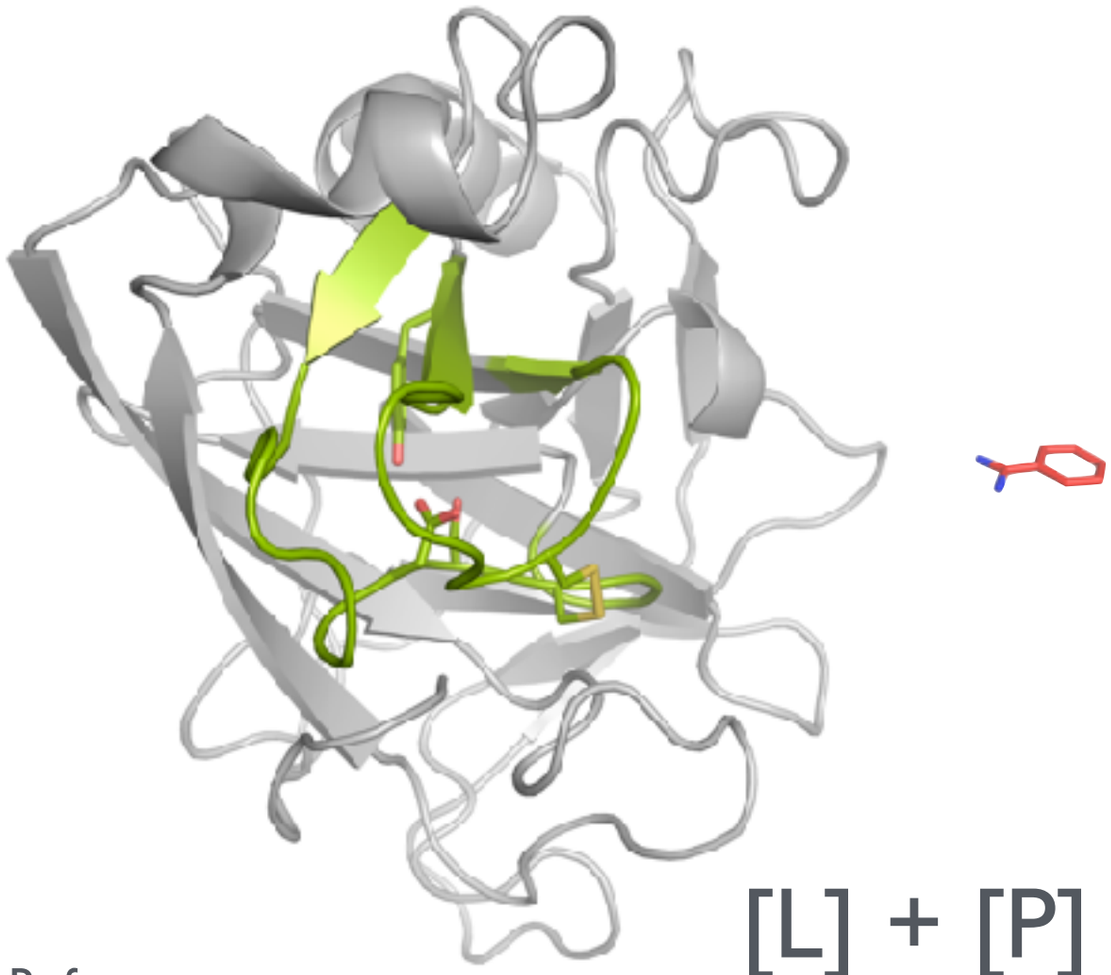
:)

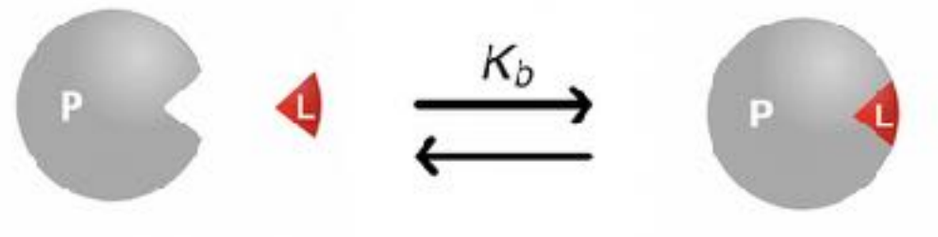
Life is Dynamics

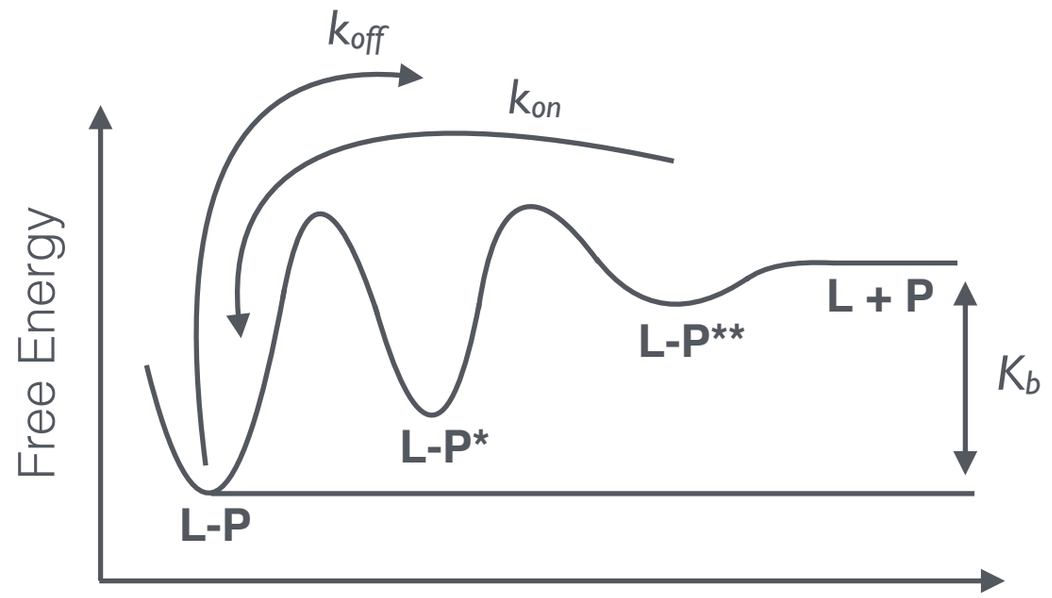


:(

Drug/Target Binding







$$[L] + [P] \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} [LP]$$

References:

Limongelli et al. PNAS (2010)

Limongelli et al. PNAS (2013)

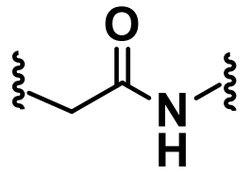
Tiwary, **Limongelli** et al. PNAS (2015)

Limongelli WIREs Comp Mol Sci (2020)

$$K_b = \frac{[LP]}{[L]\,[P]} \qquad K_b = \frac{k_{on}}{k_{off}}$$

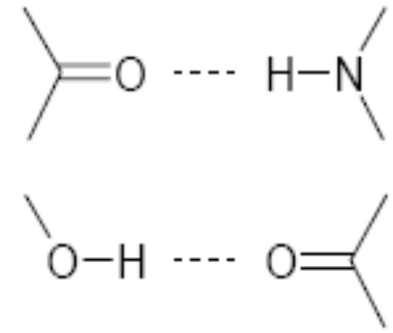
Drug/Target Interaction

protein ligand



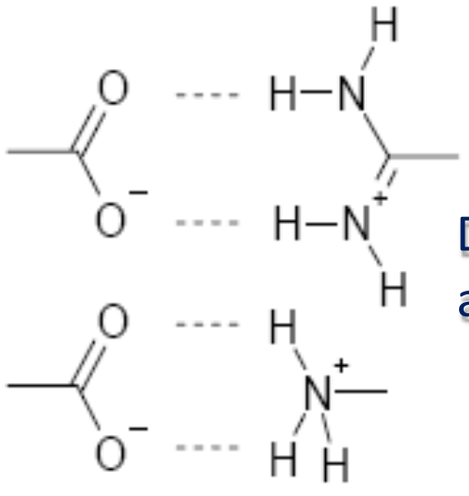
Covalent Binding
irreversible

protein ligand



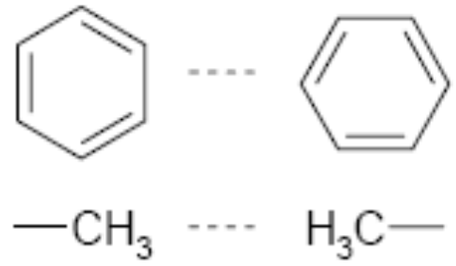
Hydrogen Bond
Depends on the distance and orientation of H-bond donor and acceptor groups

protein ligand



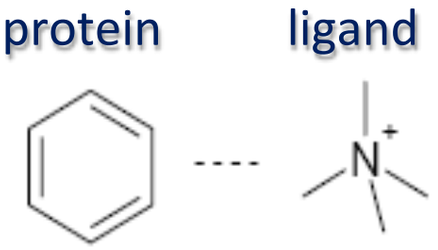
Salt Bridge Interaction
Depends on the distance and orientation of donor and acceptor groups

protein ligand



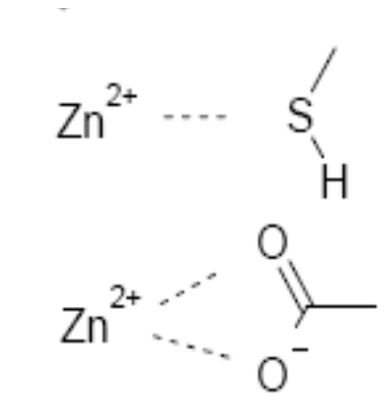
Van der Waals interactions (electron transfer (π–stacking and T-shaped) and hydrophobic)

protein ligand



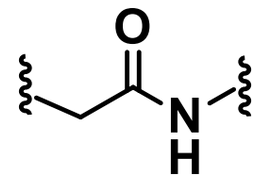
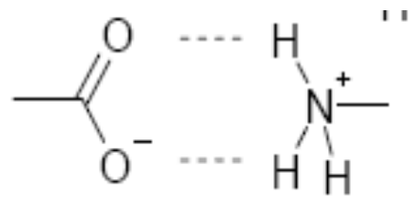
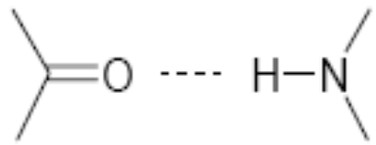
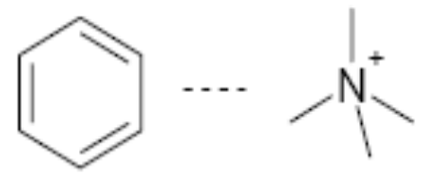
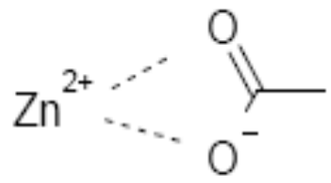
Cation-π interactions
Between potively charged group and π systems

protein ligand

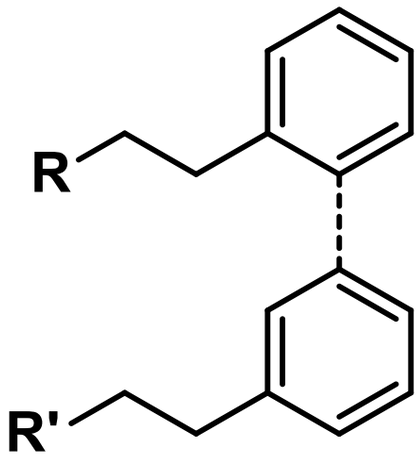
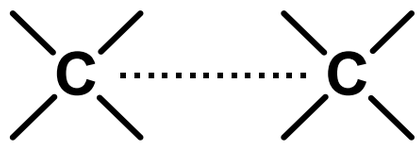


Metal ions coordination

Drug/Target Interaction

Interaction type	Energy (kcal/mol)	Example
Covalent	40-140	
Ionic	5-10	
H-bond	1-7	
Cation-π	5-10	
Chelation	5-10	

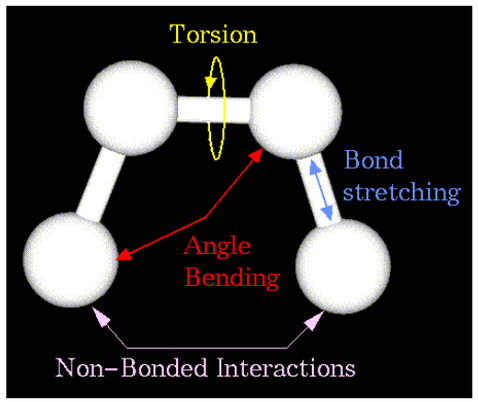
Drug/Target Interaction

Interaction type	Energy (kcal/mol)	Example
Ion-dipole	1-7	R_4N^+ ... $:NR_3$
Dipole-dipole	1-7	δ^- δ^+ O=C ... $:NR_3$
π–stacking	1	
van der Waals	0,5-1	

Atomistic Model

A force field estimates the potential energy of the system and how it changes related to different conformations (states)

$$E_p = E_{\text{bond}} + E_{\text{non-bond}} + E_{\text{angle}} + E_{\text{dihedral}}$$



The **force constants** and **reference values** (bond distance, angle, torsion, etc.) are based on similar data measured (computed) on small systems

The main approximation of FF is that from the equilibrium states of small systems we can derive information on equilibrium of large system

Molecular Dynamics (MD)

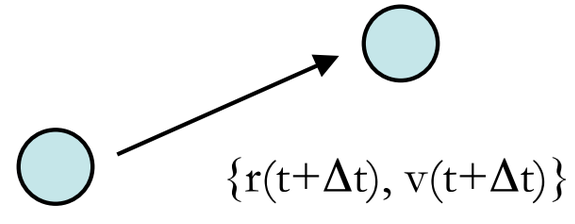
- We can induce motion to explore the phase space
- Molecular dynamics methods (MD) use the Newton's equation of motion to accelerate atoms

$$m_i \ddot{\mathbf{R}}_i = -\nabla_{\mathbf{R}_i} V$$

- These are deterministic approaches that provide a trajectory of the system evolution
- MD can be used for a wide range of applications (system stability, calculation of the thermodynamics and kinetics, protein flexibility, protein/DNA folding, ligand/protein binding, material science etc.)

MD algorithm

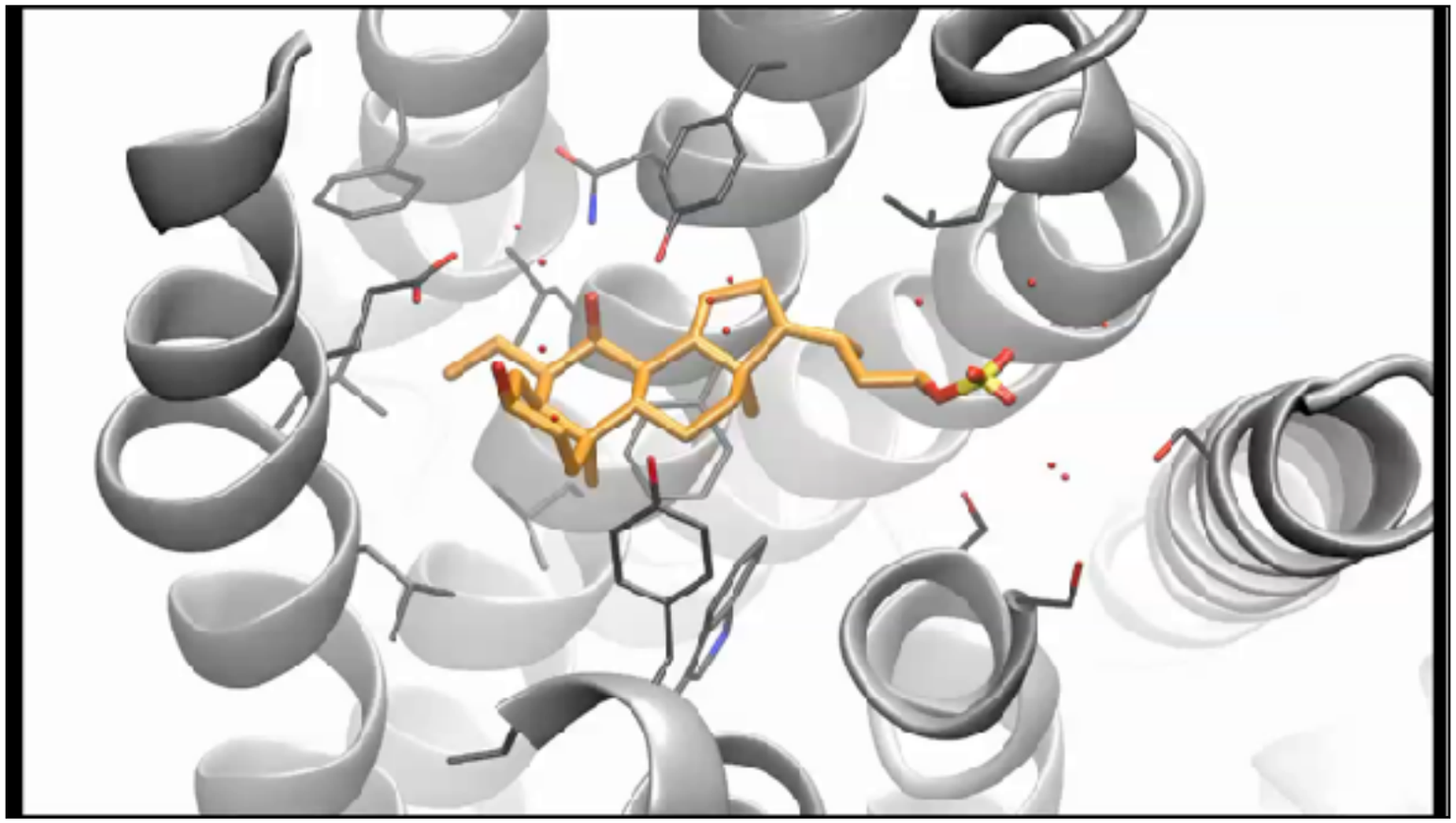
- Initialise system
  - Ensure particles do not overlap in initial positions (can use lattice)
  - Randomly assign velocities
- Move and integrate



Leapfrog algorithm

1	solve for a_i at t using:	$-\frac{dE}{dr_i} = F_i = m_i\, a_i(t)$
2	update v_i at $t + \Delta t/2$ using:	$v_i(t + \Delta t/2) = v_i(t - \Delta t/2) + a_i(t)\, \Delta t$
3	update r_i at $t + \Delta t$ using:	$r_i(t + \Delta t) = r_i(t) + v_i(t + \Delta t/2)\, \Delta t$

MD algorithm



Limongelli and co. *J Med Chem* (2014)

Limitations:

- time scale accessible (sampling issue)
- accuracy of atomistic force fields

Dimensionality Reduction

It is possible to describe a physical/chemical process in terms of a small number of coarse descriptors of the system:

$$\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{R}) = (S_1(\boldsymbol{R}), \ldots, S_d(\boldsymbol{R}))$$

Key quantity of thermodynamics is the free energy as a function of these variables:

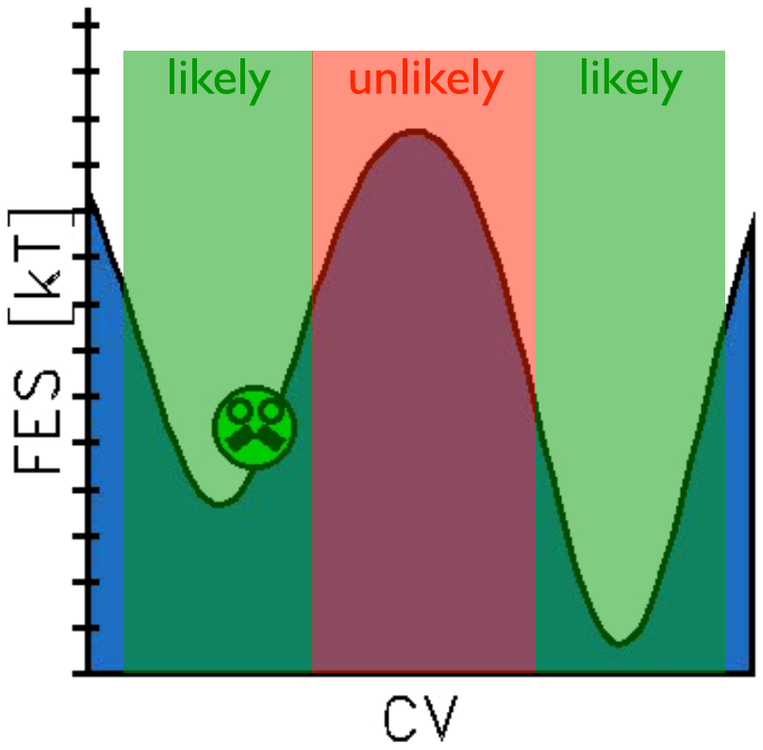
$$F(\boldsymbol{S}) = -\frac{1}{\beta} \ln P(\boldsymbol{S})$$

Canonical ensemble

$$P(\boldsymbol{S}) = \frac{\int d\boldsymbol{R}\, \delta(\boldsymbol{S} - \boldsymbol{S}(\boldsymbol{R}))\, e^{-\beta U(\boldsymbol{R})}}{\int d\boldsymbol{R}\, e^{-\beta U(\boldsymbol{R})}}$$

Zwanzig *Phys Rev* (1961)

Dimensionality Reduction



$$F(A) - F(B) = -k_B T \ln \frac{N_A}{N_B}$$

How to estimate a free energy difference if we never see a transition?

Biased Sampling

The idea is to add a bias potential that acts on the collective variables:

$$U(\mathbf{R}) \rightarrow U(\mathbf{R}) + V(\mathbf{S}(\mathbf{R}))$$

In this biased ensemble the free energy becomes:

$$F'(\mathbf{S}) = -\frac{1}{\beta} \ln P'(\mathbf{S}) + C$$

where $$P'(\mathbf{S}) = \frac{\int d\mathbf{R}\, \delta(\mathbf{S} - \mathbf{S}(\mathbf{R}))\, e^{-\beta[U(\mathbf{R}) + V(\mathbf{S}(\mathbf{R}))]}}{\int d\mathbf{R}\, e^{-\beta[U(\mathbf{R}) + V(\mathbf{S}(\mathbf{R}))]}}$$

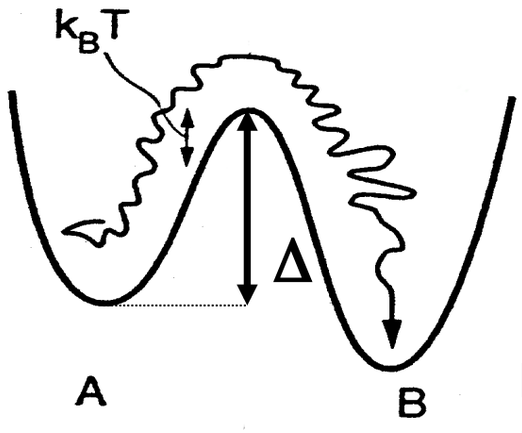
leading to:

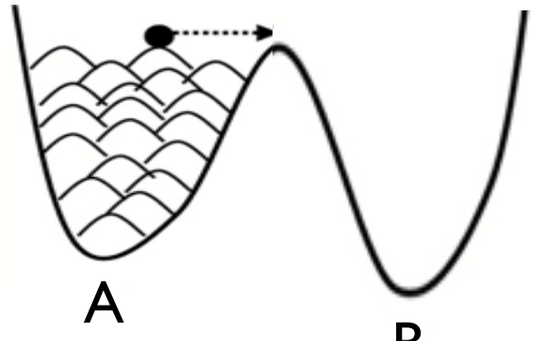
$$F'(\mathbf{S}) = F(\mathbf{S}) + V(\mathbf{S})$$

The Sampling Issue

$$F(s,t) = -\frac{T+\Delta T}{\Delta T}\, V(s,t)$$

Classical MD





Metadynamics

MetaD Potential

$$\dot{V}(s;t) = \omega e^{-\frac{V(s;t)}{\Delta T}}\, \delta_{s,s(t)}$$

initial rate

rate decreases as exp(-V/ΔT)

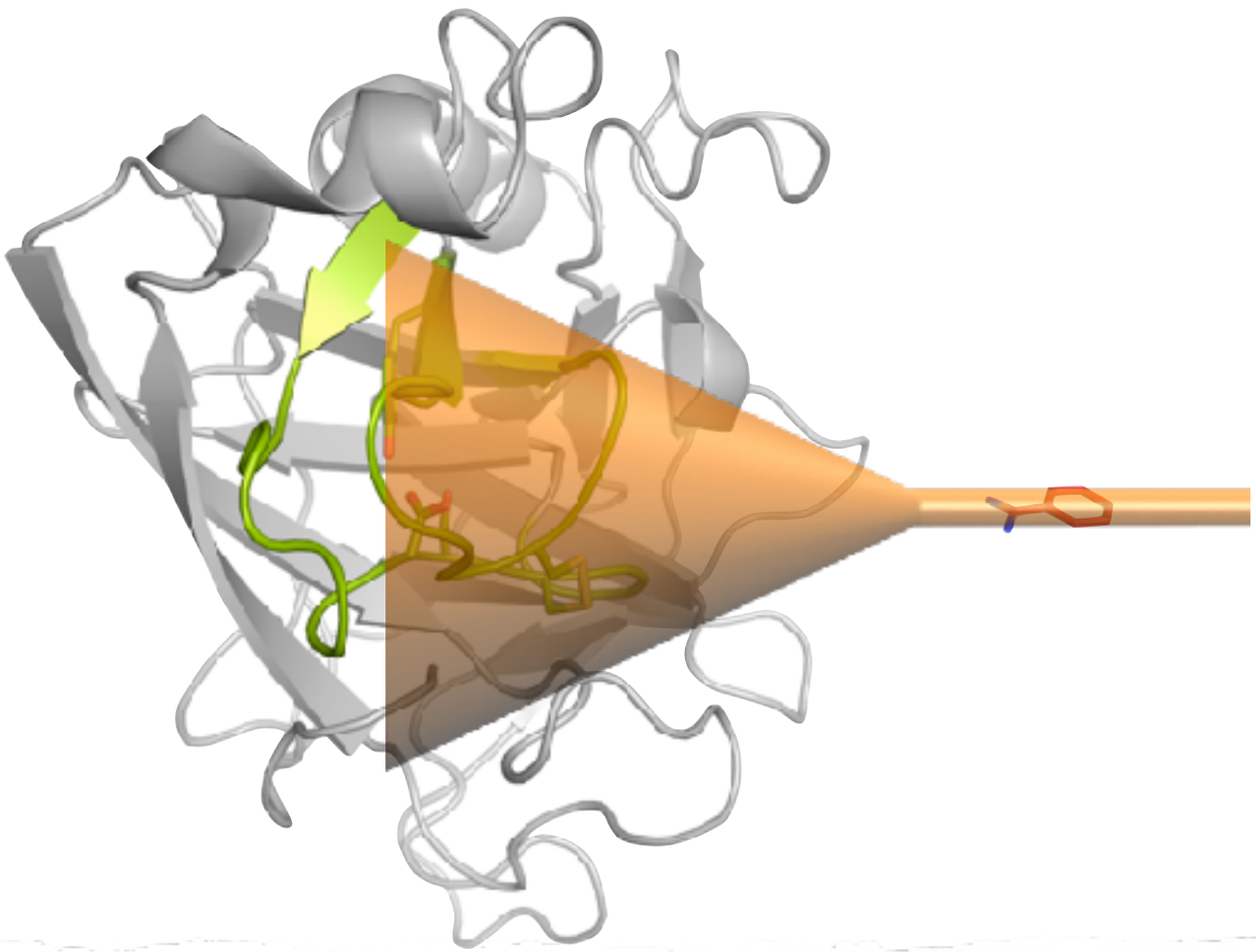
References:

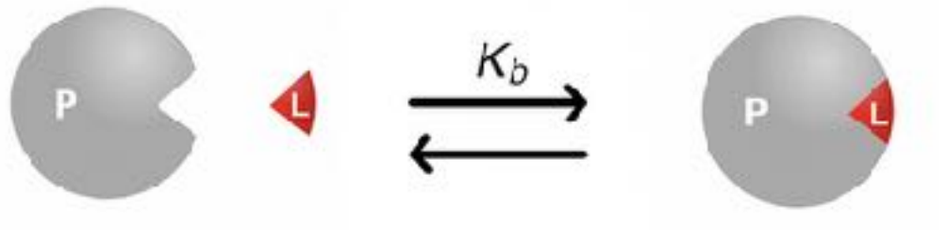
Laio and Parrinello, *PNAS* (2002)

Barducci, Bussi, Parrinello, *PRL* (2008)

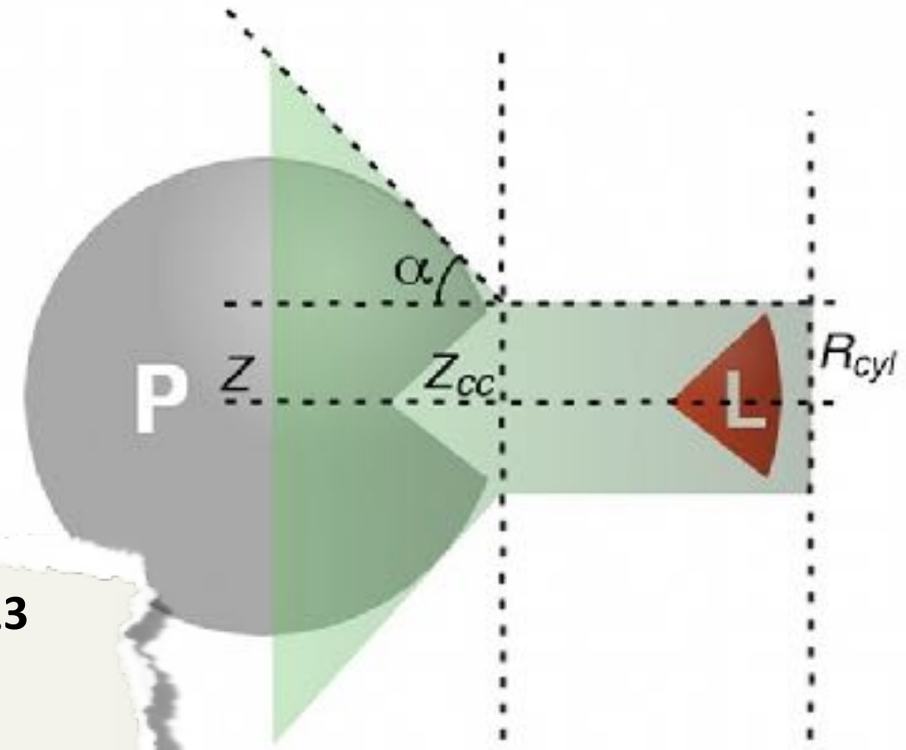
Limongelli *WIREs Comput. Mol. Sci.* (2020)

Protein/Ligand Binding





$$\Delta G_b^0 = -\frac{1}{\beta} ln(C^0 \ K_b)$$

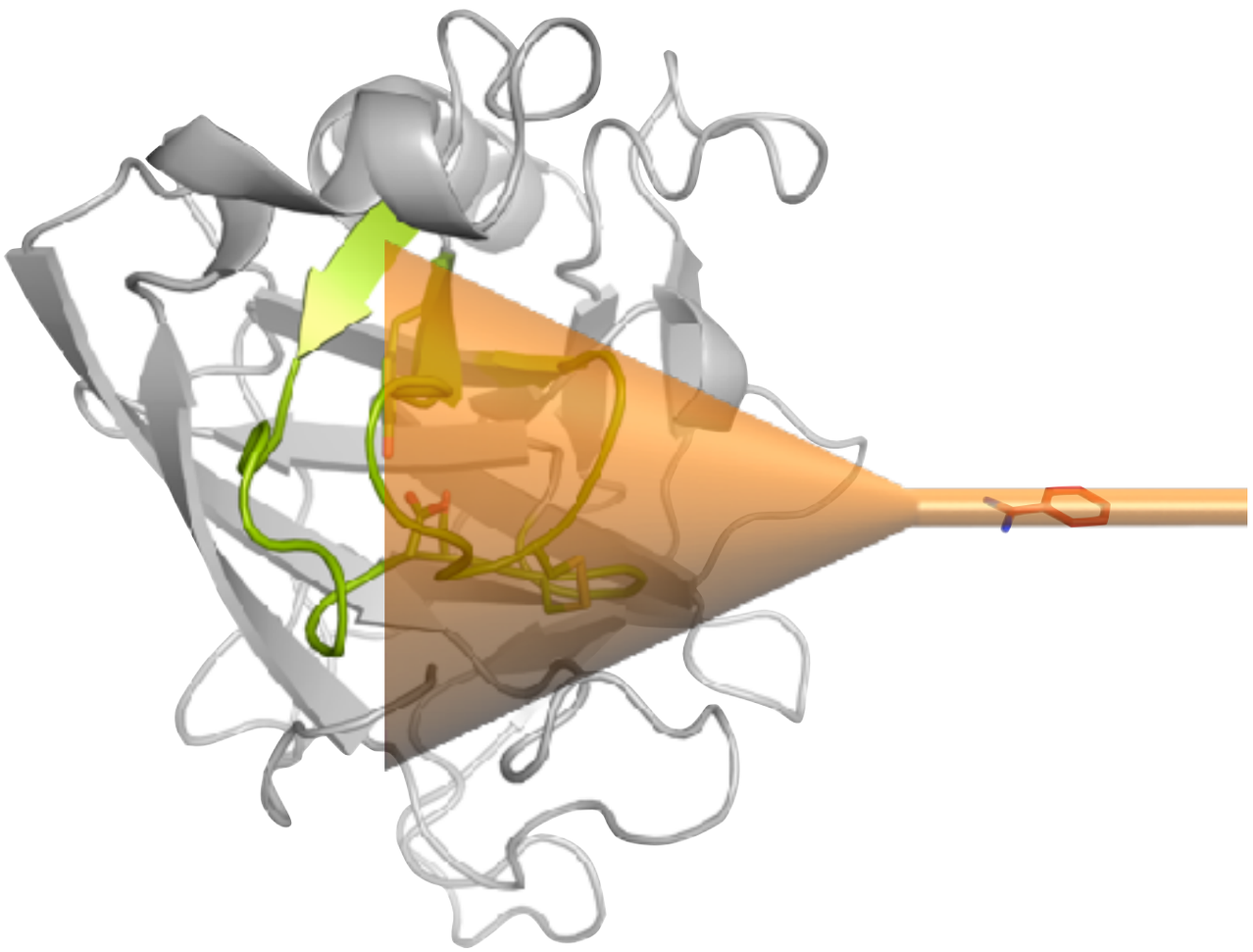


Funnel metadynamics as accurate binding free-energy method **2013**

Vittorio Limongelli[a,1], Massimiliano Bonomi[b], and Michele Parrinello[c,d,1]

[a]Department of Pharmacy, University of Naples Federico II, I-80131 Naples, Italy; [b]Department of Bioengineering and Therapeutic Sciences, and California Institute of Quantitative Biosciences, University of California, San Francisco, CA 94158; [c]Department of Chemistry and Applied Biosciences, Eidgenössiche Technische Hochschule (ETH), 8006 Zürich, Switzerland; and [d]Facoltà di Informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, CH-6900 Lugano, Switzerland

Funnel-Metadynamics (FM)



$$\Delta G_b^0 = -\frac{1}{\beta} ln(C^0 \ K_b)$$

ΔG_{site} is equal to 0

potential of the unbound state

$$^*K_b = e^{\beta \Delta G_{site}} \int_{site} dz \ e^{-\beta[W(z) - W_{ref}]} \ S_u$$

potential of the bound state

Su is equal to πR_{cyl}^2

$$^{**}\Delta G_b^0 = \Delta G - \frac{1}{\beta} ln(\pi R_{cyl}^2 C^0)$$

Free Energy difference between the bound and unbound state

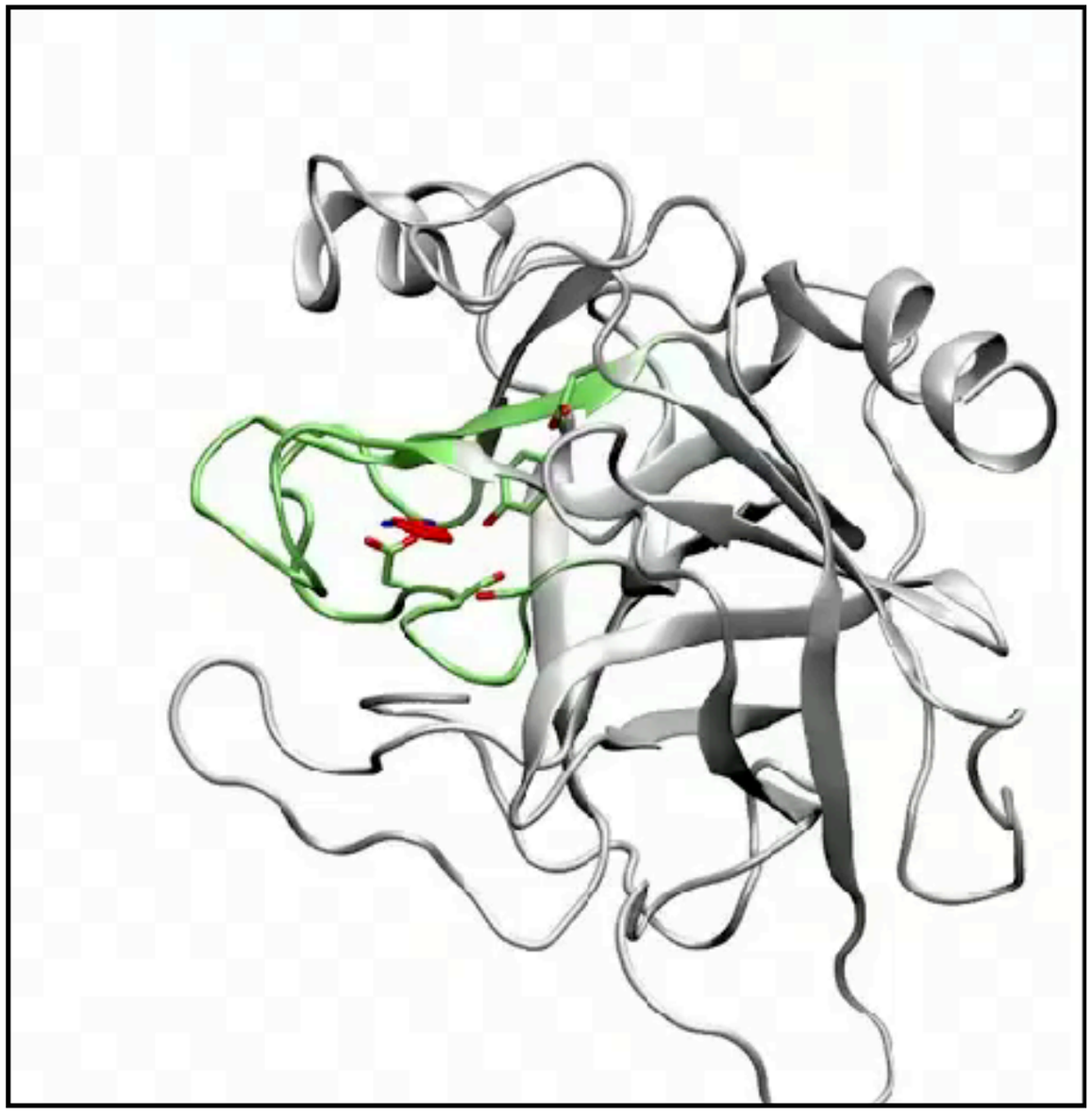
Analytical correction

References:

* Allen et al., *PNAS* (2004); Roux et al., *J. Chem. Phys.* (2008)

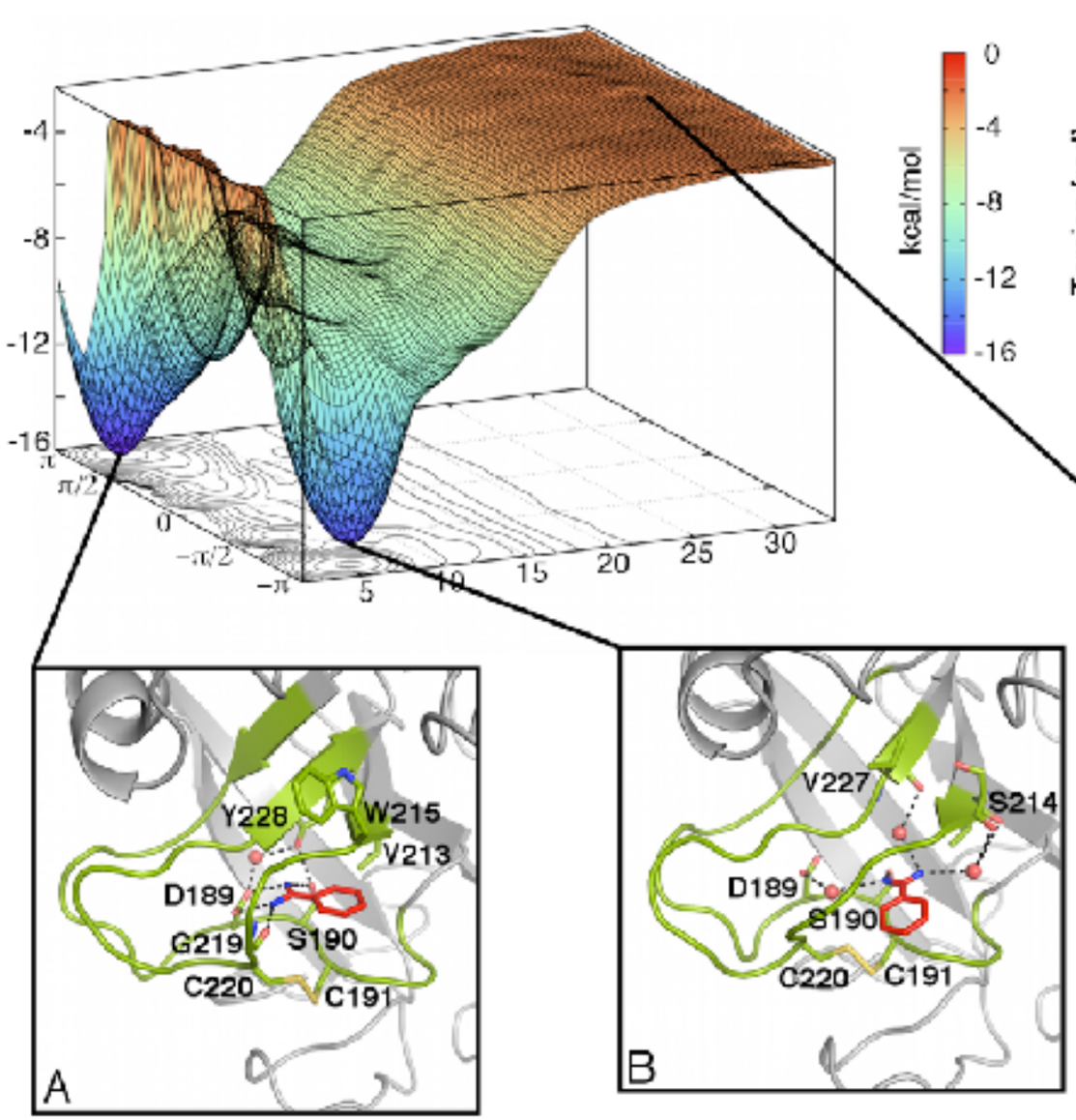
** **Limongelli**, Bonomi & Parrinello, *PNAS* (2013)

Trypsin/Benzamidine Binding with FM



Funnel-Metadynamics code available on my website and GitHub:
https://sites.google.com/site/vittoriolimongelli/home

Ligand Binding Free Energy



$G_{(A)}$-$G_{(C)}$ = - 12.3 kcal/mol

3.8 kcal/mol

$$\Delta G_b^0 = \Delta G - \frac{1}{\beta} ln(\pi R_{cyl}^2 C^0)$$

ΔG_b^0 = −8.5 ± 0.7 kcal/mol

Limongelli, Bonomi & Parrinello, *PNAS* (2013)

Previous estimates −5.5 to −9.0 kcal/mol*

*Talhout et al., *Eur. J. Bochem.* (2001)
Doudou et al., *JCTC* (2009)

Further Reading:
Raniolo & **Limongelli** *Nature Protolos* (2020)
Limongelli and co. *PNAS* (2017)

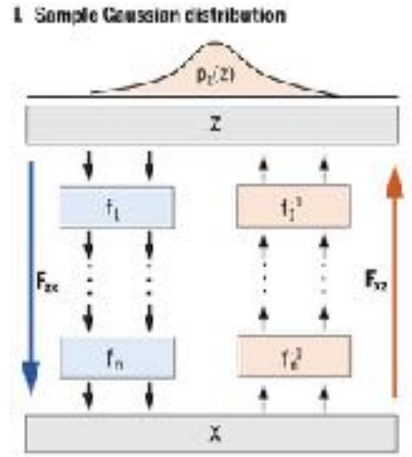
Limongelli et al. *PNAS* (2010)
Limongelli et al. *PNAS* (2012)
Grazioso, **Limongelli** et al. *JACS* (2012)

Is it possible to overcome the sampling issue with ML?

Pioneering Studies

- Generative neural networks to produce samples from a given target distribution

  Noé et al. *Science* (2019)



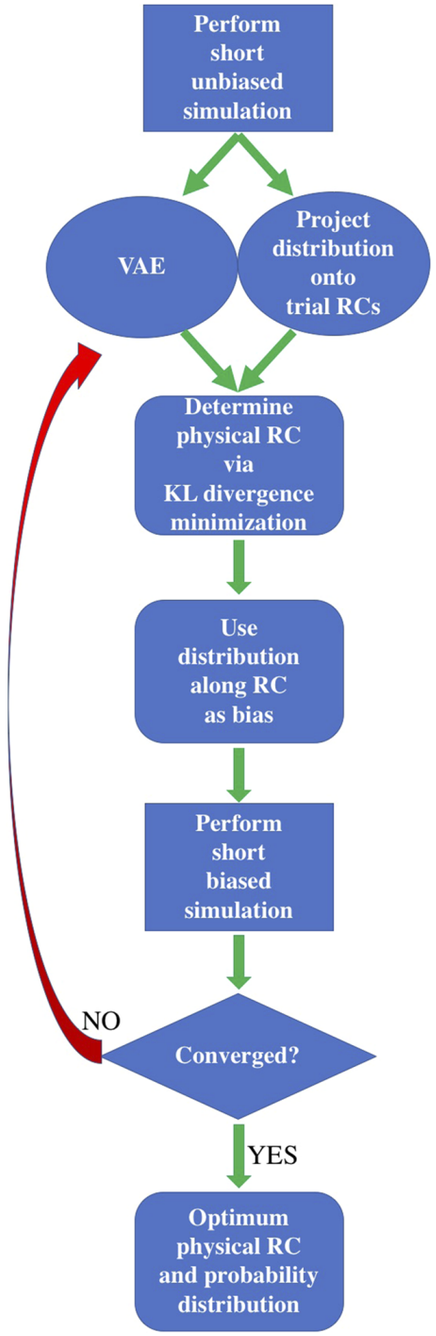
- Reweighted autoencoded variational Bayes for enhanced sampling (RAVE)

  Ribeiro et al. *JCP* (2018)

Pioneering Studies

- Generative neural networks to produce samples from a given target distribution

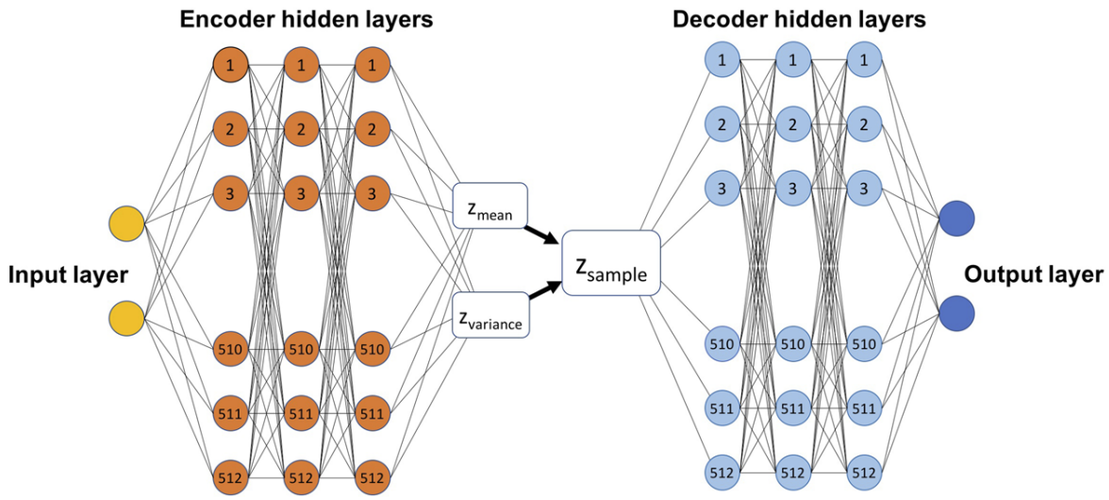
  Noé et al. *Science* (2019)



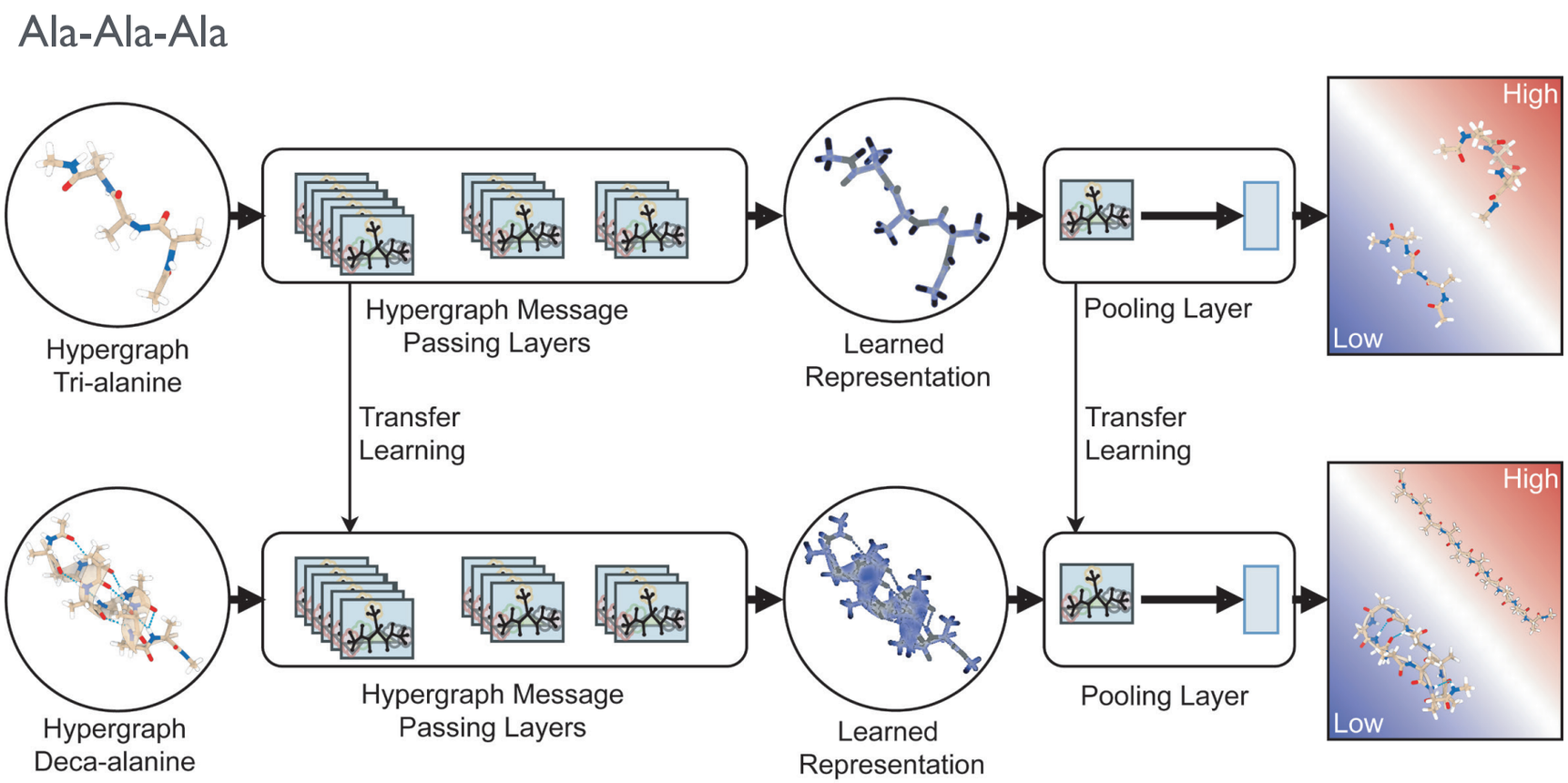
- Reweighted autoencoded variational Bayes for enhanced sampling (RAVE)

  Ribeiro et al. *JCP* (2018)

$$p(\mathbf{x}) = p(\mathbf{x}|z)p(z).$$



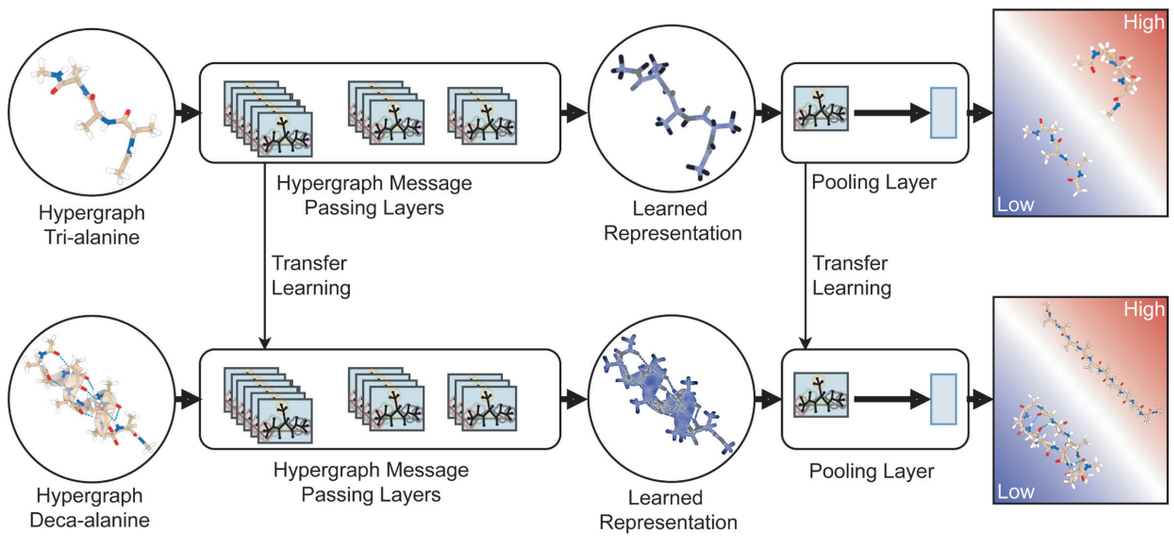
Transfert Learning



Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala

Limongelli and co. *Nat. Comput. Sci.* (under review)

The Model



Hypergraph *H(V,E,X,W)*

$$W_i = \begin{bmatrix} \text{1 if } e_i \text{ is a bond, 0 otherwise} \\ \text{Coulomb force if } e_i \text{ denotes a Coulomb interaction, 0 otherwise} \\ \text{Van der Waals force if } e_i \text{ denotes Van der Waals interaction, 0 otherwise} \\ \text{angle between three atoms if } |e| = 3\text{, 0 otherwise} \\ \text{dihedral between four atoms if } |e| = 4\text{, 0 otherwise} \end{bmatrix}$$

Hypergraph message passing NN

$$M_v = f_v(X_v^{(t)})$$

$$W_e^{(t+1)} = g_w(W_e^{(t)}, \sum_{v \in e} M_v)$$

$$M_e = f_w(W_e^{(t)}, \sum_{v \in e} M_v)$$

$$X_v^{(t+1)} = g_v(X_v^{(t)}, \sum_{e,v \in e} M_e)$$

Loss function

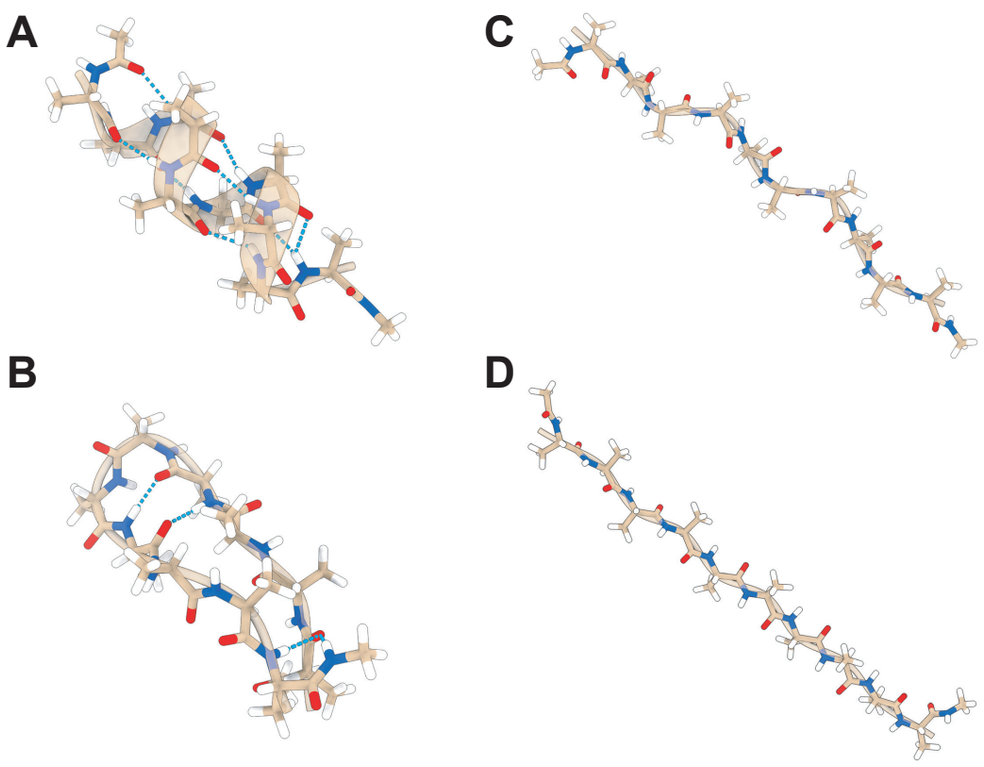
$$\text{BinaryCrossEntropyLoss} + l_2 + \text{TargetLoss}$$

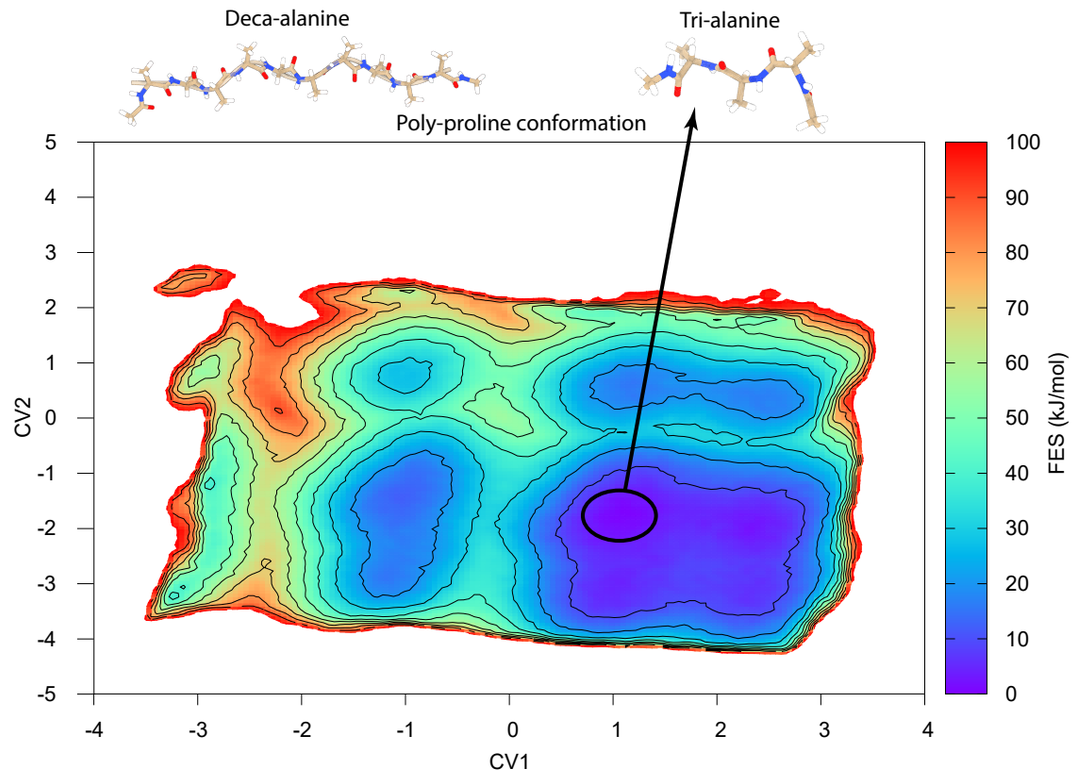
$$\frac{1}{|N|} \sum_{i=1}^{N} y_i log(p(y_i)) + (1 - y_i) log(1 - p(y_i))$$

$$g(x, m) = L(x) + m$$

Limongelli and co. *Nat. Comput. Sci.* (under review)

The challange





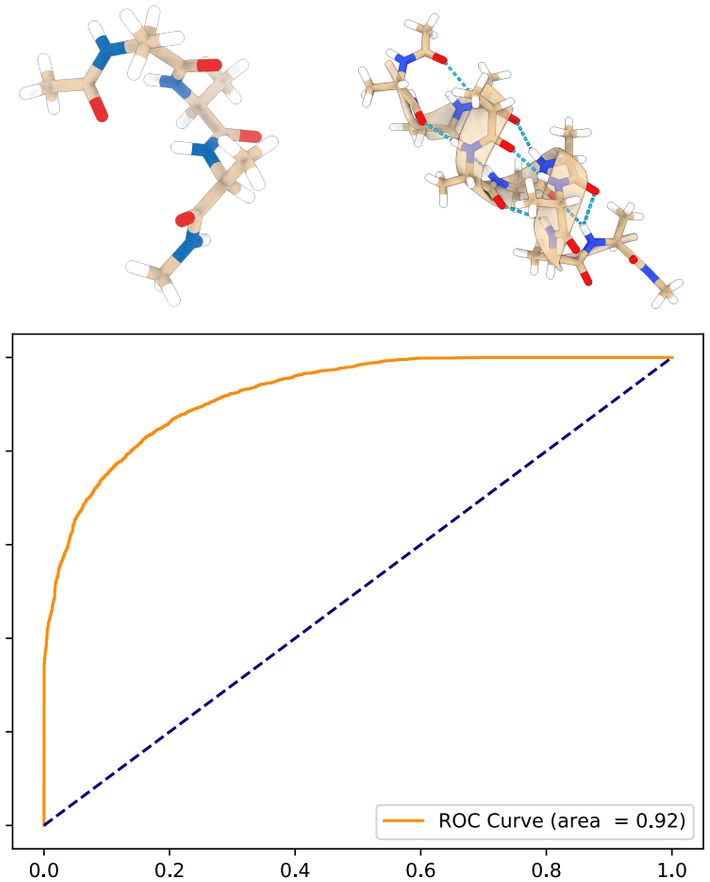
Low and high energy states of deca-alanine

Limongelli and co. *Nat. Comput. Sci.* (under review)

Facoltà di scienze biomediche

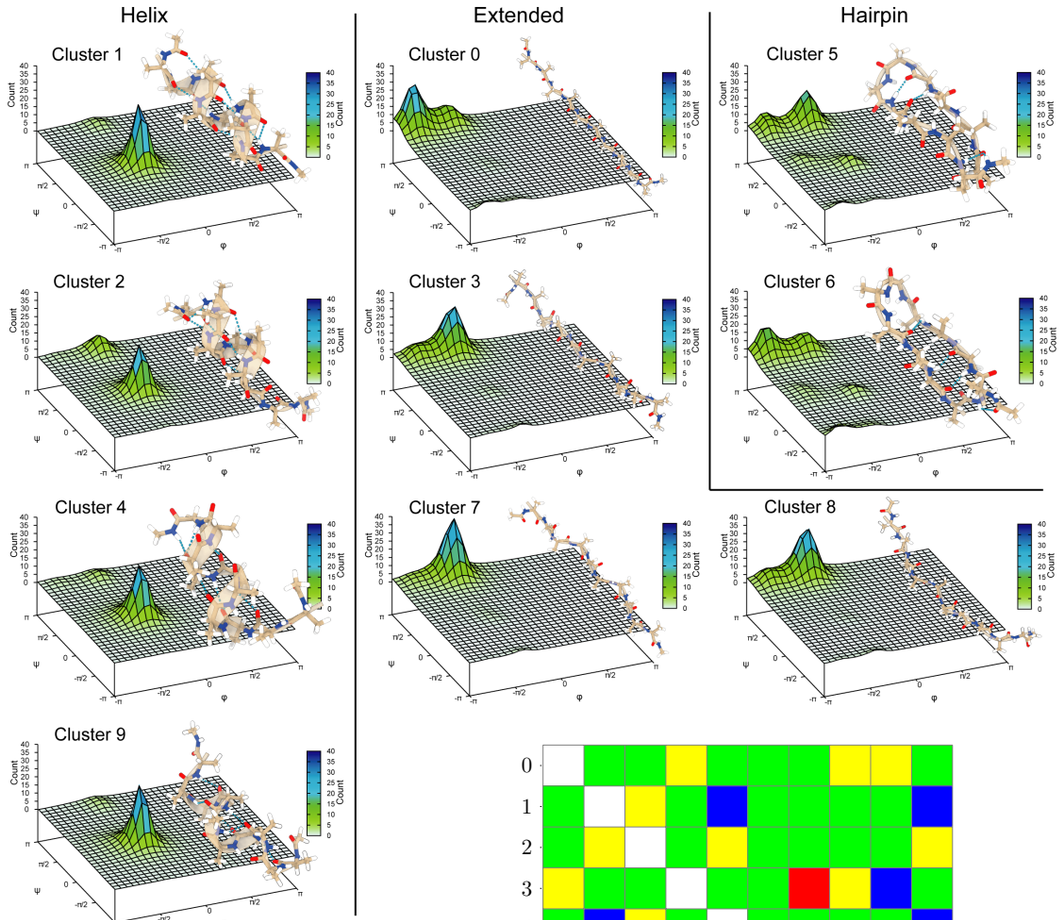
The Results

Classification of low/high energy states

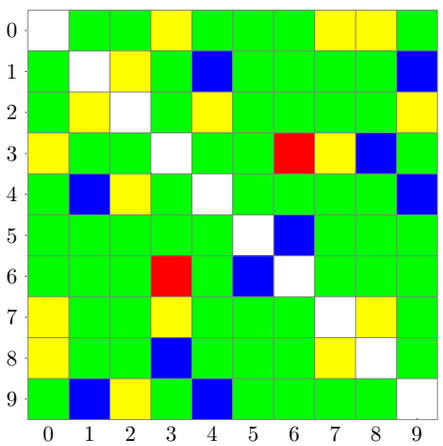


Threshold	Precision ($\frac{tp}{tp+fp}$)	Recall ($\frac{tp}{tp+fn}$)
Low $\leq$ 0.45	0.956	0.836
High>0.45	0.979	0.624
Low $\leq$ 0.5	0.923	0.950
High>0.5	0.921	0.880
Low $\leq$ 0.55	0.907	0.999
High > 0.55	0.877	0.894

Unsupervised secondary structure prediction



p-value < 0.01



Limongelli and co. *Nat. Comput. Sci.* (under review)

Next Steps

- Free energy estimate of low/high states (thermodynamics and kinetics)

- Sampling of unvisited states

- Sampling high-dimensional and multimodal distributions typical of pharmacologically relevant systems (drug/protein complexes)

Hands-on Session

- Can we build a ML model that works as molecular simulator?
- How do we ensure these methods scale to very high-dimensional, pharmacologically relevant systems?

- Funnel-Metadynamics (FM) + FMAP

ligand binding free-energy calculation

Limongelli, Bonomi, Parrinello *PNAS* (2013); Raniolo, **Limongelli** *Nat. Protoc.* (2020)

- Coarse-Grained Metadynamics (CG-MetaD)

protein/protein binding free-energy calculation

Limongelli and co. *JACS* (2016)

PLUMED-NEST

www.plumed-nest.org

Bonomi, Bussi, Camilloni, Tribello…**Limongelli** et al. *Nature Methods* (2019)

Limongelli Research Group (LRG)

Stefano Raniolo

Daniele Angioletti

@Gothenburg
Antonio Carlesso

Paolo Conflitti

Vincenzo D'Amore

@Naples
Francesco Di Leva
Federica Moraca
Catalanotti Bruno

?

PhD Student
&
PostDoc

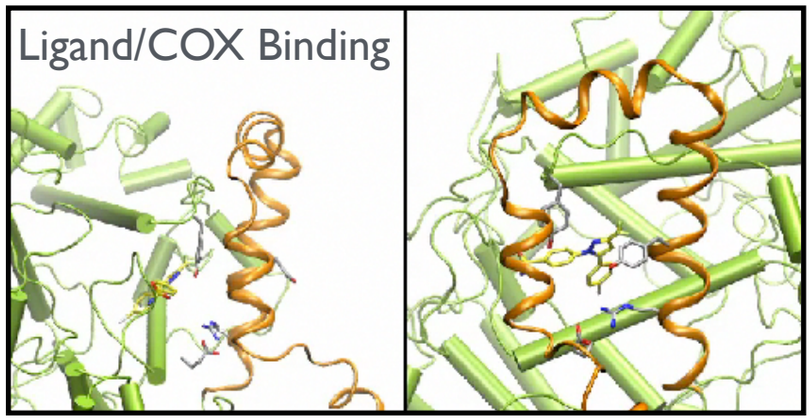
Vins Cardenas

https://www.facebook.com/LimResGrp/

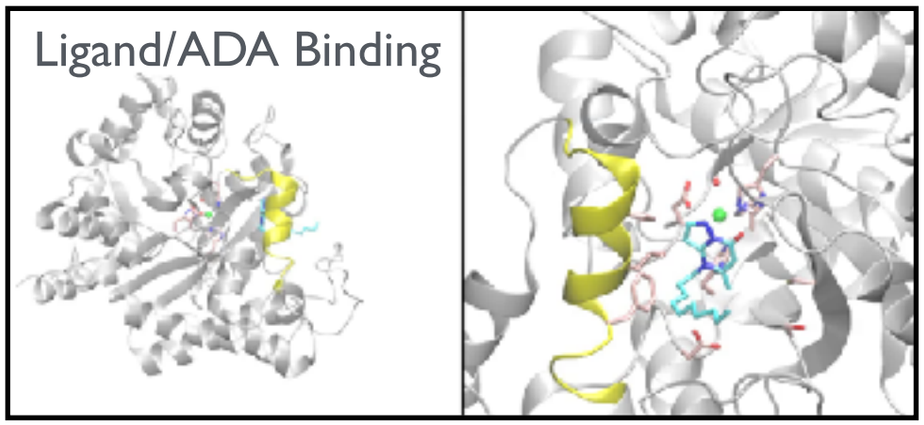
https://twitter.com/limresgrp

vittoriolimongelli@gmail.com

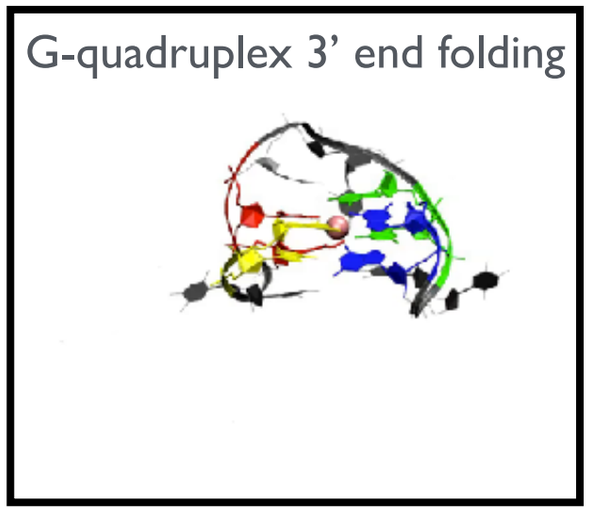
LRG Media Library



Limongelli et al., *PNAS* (2010)

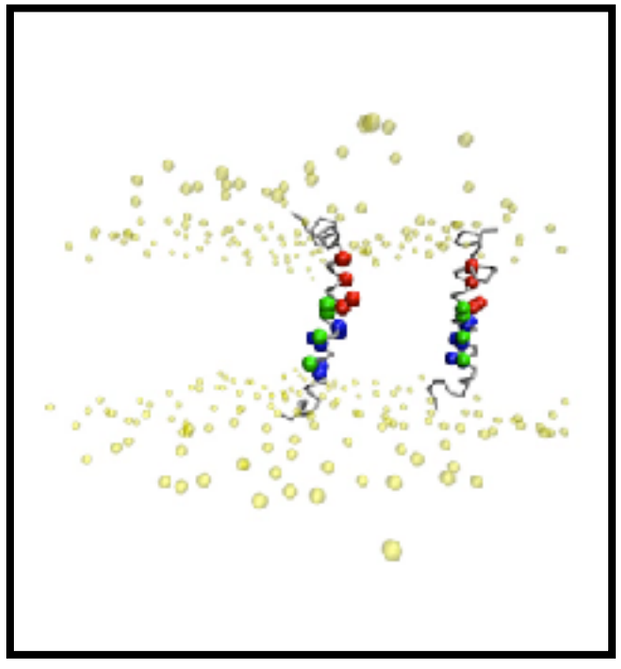


Limongelli et al., *PNAS* (2012)



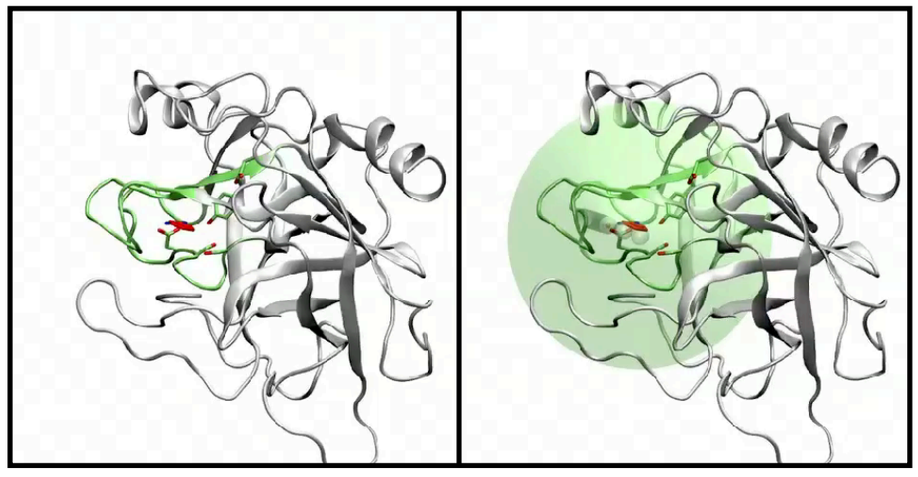
Limongelli et al., *Angew. Chem. Int. Ed.* (2013)

Coarse-Grained MetaDynamics (CG-MetaD)

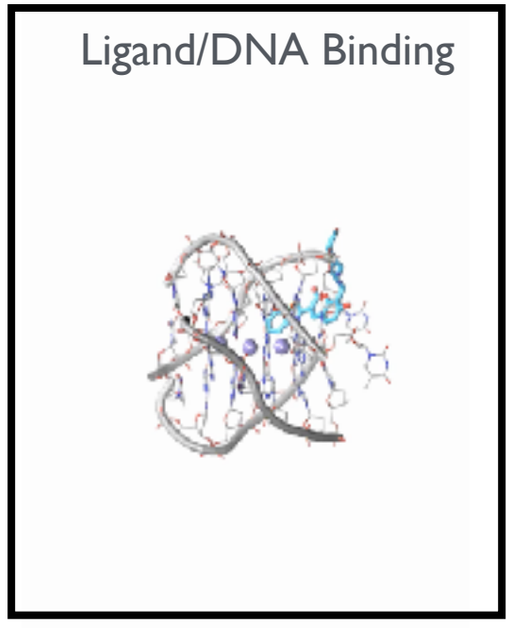


Limongelli and co. *JACS* (2016)

Funnel-Metadynamics (FM)



Limongelli et al., *PNAS* (2013)



Limongelli and co., *NAR* (2014)
Limongelli and co., *PNAS* (*2017*)

YouTube Limongelli Research Group channel

vittoriolimongelli@gmail.com