Molecular Dynamics Simulations: Theory and Applications (in Drug Design)

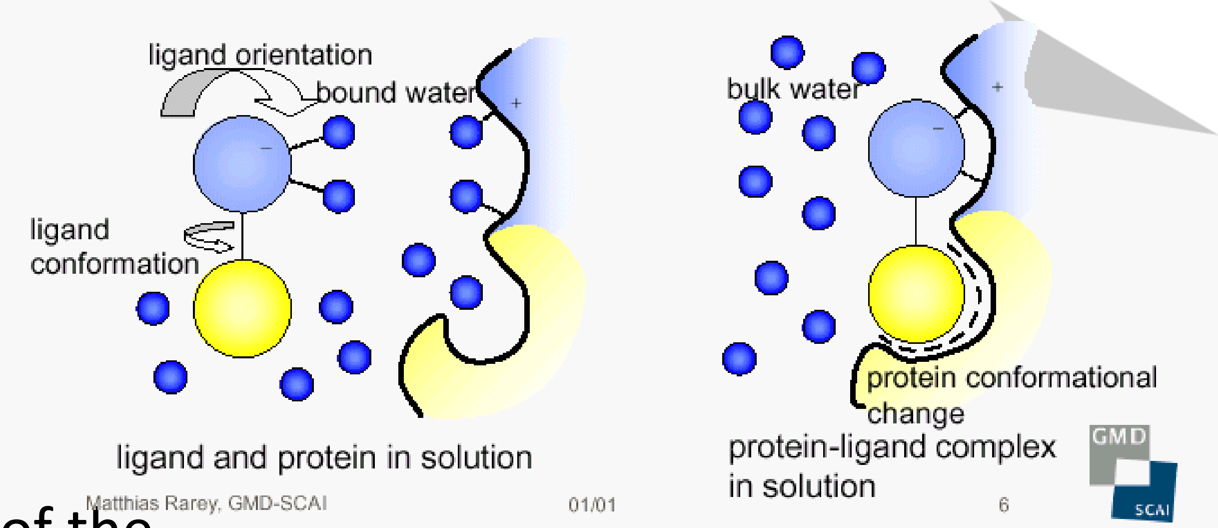
Hanoch Senderowitz
Department of Chemistry, Bar-Ilan University, Israel

6th Strasbourg Summer School in Cheminformatics, June 2018,
Strasbourg France

Binding Free Energy

- Docking and Scoring
- Pharmacophore
- QSAR

↓



- Incomplete treatment of the binding process
- Static, single molecule view

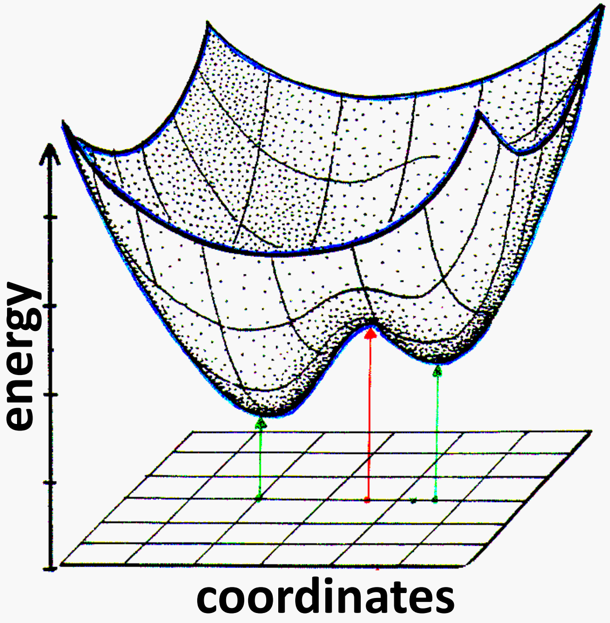
$$A = kT \ln\left(\iint d\mathbf{p}^N d\mathbf{r}^N \exp\left(+ \frac{E\left(\mathbf{p}^N, \mathbf{r}^N\right)}{kT} \right) \rho\left(\mathbf{p}^N, \mathbf{r}^N\right) \right)$$

The Potential Energy Surface

Building the PES

Force Field

HF



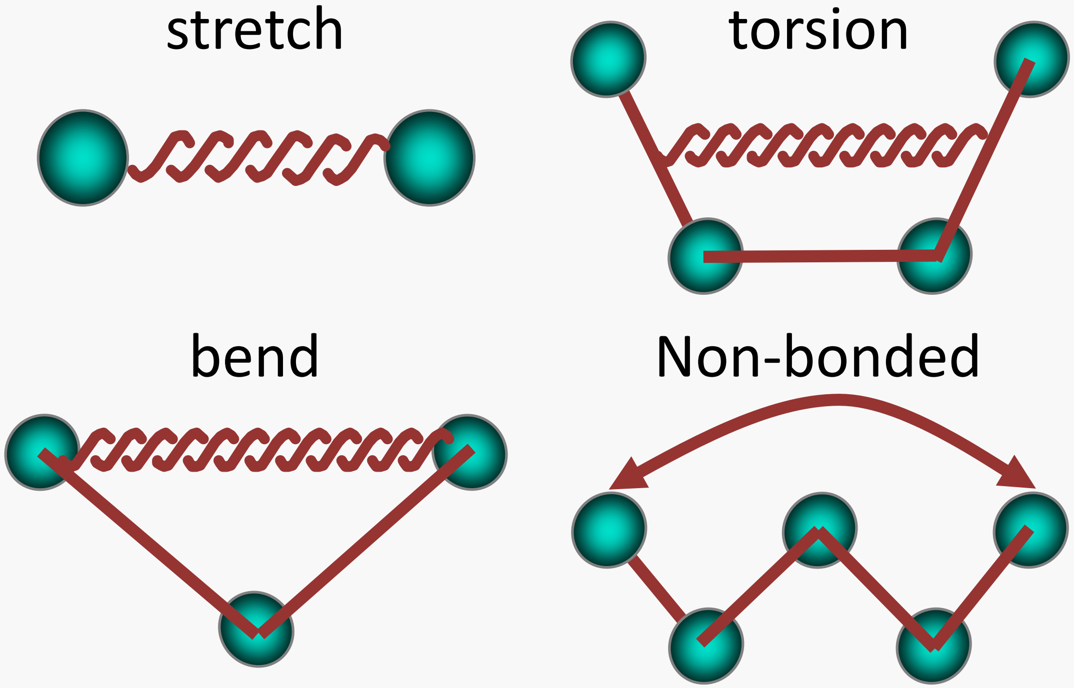
DFT

QM/MM

Sampling the PES

- Energy minimization
- Conformational search
- Molecular dynamics

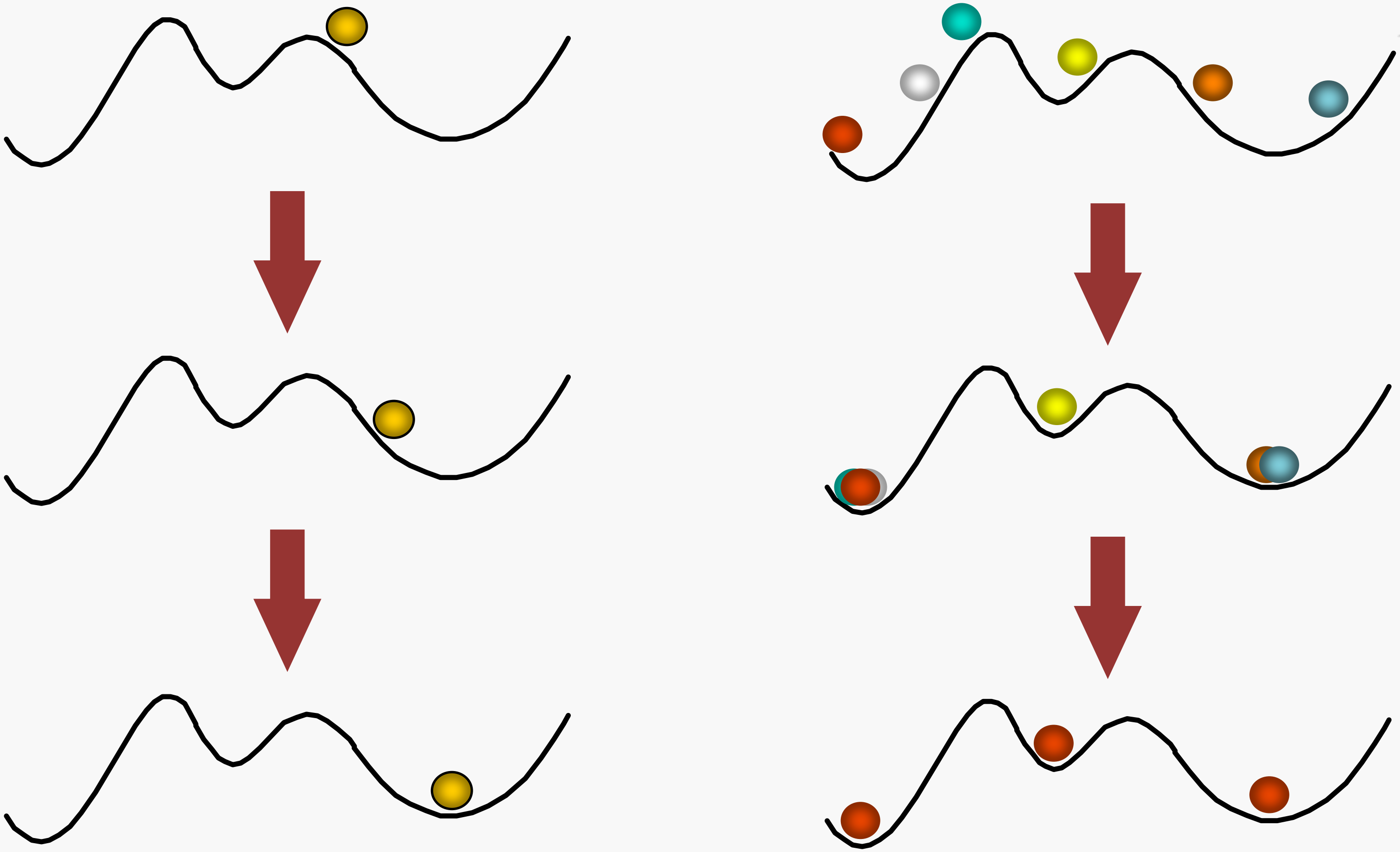
Force Fields



- Force Fields
  - Defined by equations and parameters
  - Empirical
  - Are not correct / incorrect but rather useful / not useful
  - Build unique PESs

$$V(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1+\cos(n\omega-\gamma)) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) + \text{cross terms}$$

Energy Minimization and Conformational Search



MD: PES → Phase Space

- Any experimentally measurable property is obtained by a weighted average of that property over all phase space:

$$< A >= \int_{\Omega} A(\boldsymbol{p}^N, \boldsymbol{r}^N)\rho(\boldsymbol{p}^N, \boldsymbol{r}^N) d\Omega$$

- The weighting $\rho(\mathbf{P}^N, \mathbf{r}^N)$ function gives the probability of finding this particular system state in the phase space
- For the NVT (canonical) ensemble, the probability function is given by the Boltzmann function:

$$\rho(\boldsymbol{p}^N, \boldsymbol{r}^N) = \exp(-E(\boldsymbol{p}^N, \boldsymbol{r}^N)/k_B T)/Q$$

$$Q_{NVT} = C\int d\boldsymbol{p}^N d\boldsymbol{r}^N \exp\left[-\frac{E(\boldsymbol{p}^N, \boldsymbol{r}^N)}{k_B T}\right]$$

The Ergodic Hypothesis

- In order for us to obtain the ensemble average <A>, we need to prepare many systems, each in a different state. This can’t be done.
- Thus instead of averaging over many systems, we can propagate a single system through the phase space and average over time:

$$A_{ave} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\boldsymbol{p}^N(t)\boldsymbol{r}^N(t))dt$$

- The ergodic hypothesis states that given enough time, a trajectory in phase space passes near every point in the space and spends in its region an amount of time proportional to its ensemble weight. Thus:

$$< A >= A_{ave}$$

- However since we can’t simulate a process on a computer to infinity:

$$< A > \approx A_{ave}$$

Molecular Dynamics: The Basis

- Solve Newton’s equations of motion
  - 1st law

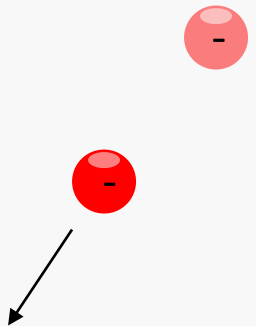
$$v(t) = \frac{dr(t)}{dt}$$

  - 2nd law

$$F = m \cdot a(t) = m \cdot \frac{dv(t)}{dt}$$

  - From the PES

$$F_q = -\frac{dU}{dq}$$

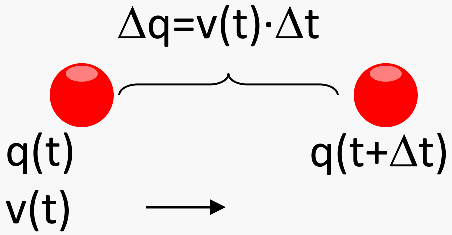


$$F = -\frac{dU}{dr} = -\frac{d}{dr}\left(\frac{Q_1Q_2}{r}\right) = \frac{Q_1Q_2}{r^2}$$

Molecular Dynamics: The Method

- Assuming we have the positions, q, and velocities, v, at time *t*:
- The position at time t+Δt is then given by:

$$\mathbf{q}(t+\Delta t)=\mathbf{q}(t)+\mathbf{v}(t)\Delta t$$



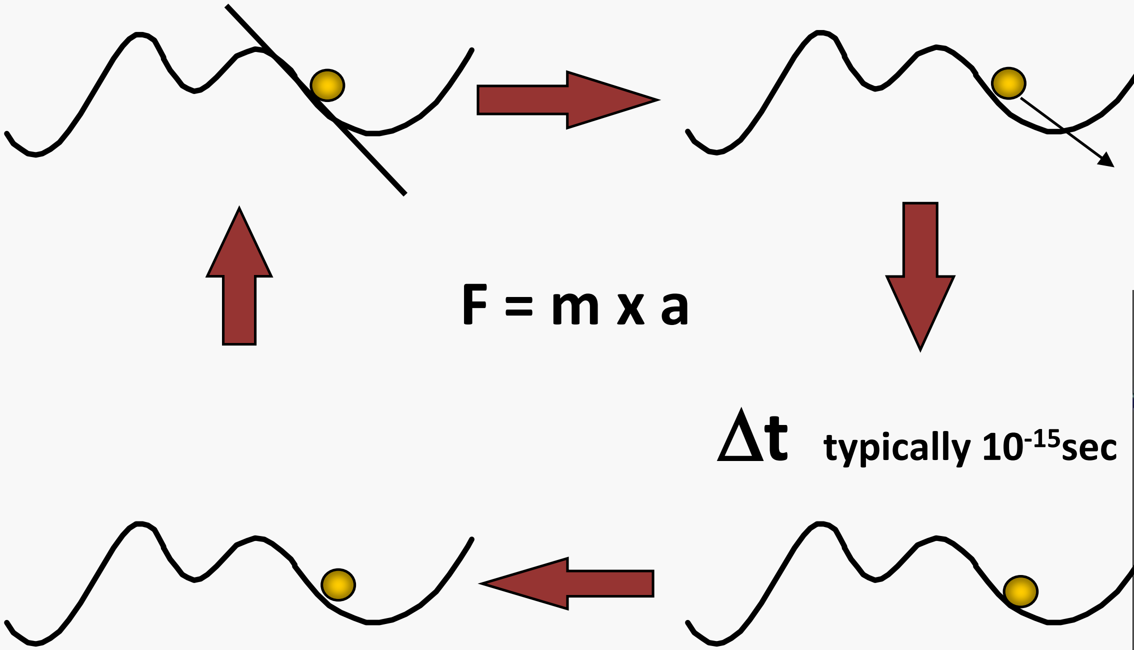
- The velocity at time t+Δt is then given by:

$$\mathbf{v}(t+\Delta t)=\mathbf{v}(t)+\mathbf{a}(t)\Delta t \qquad a=\frac{\boxed{F}}{m}$$

From energy function

- If we can compute the acceleration from the forces acting on each particle at any instant, we can simulate trajectories

Molecular Dynamics: The Trajectory



Calculate force based on surface curvature

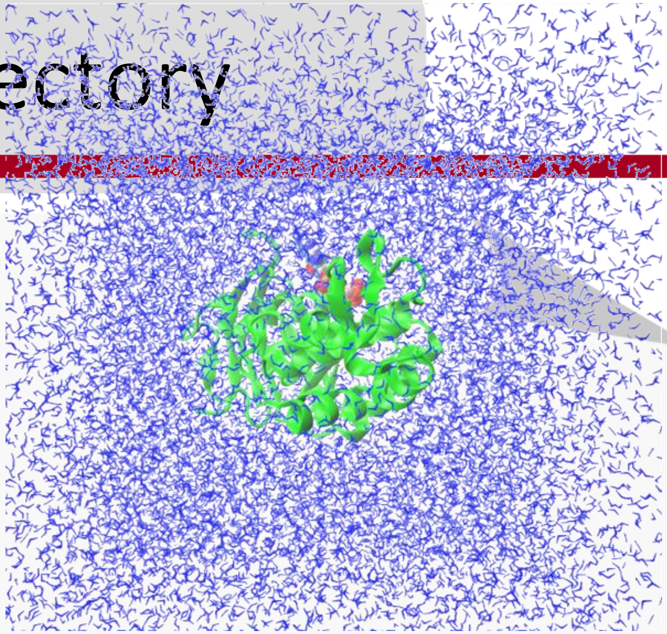
Calculate velocities based on force

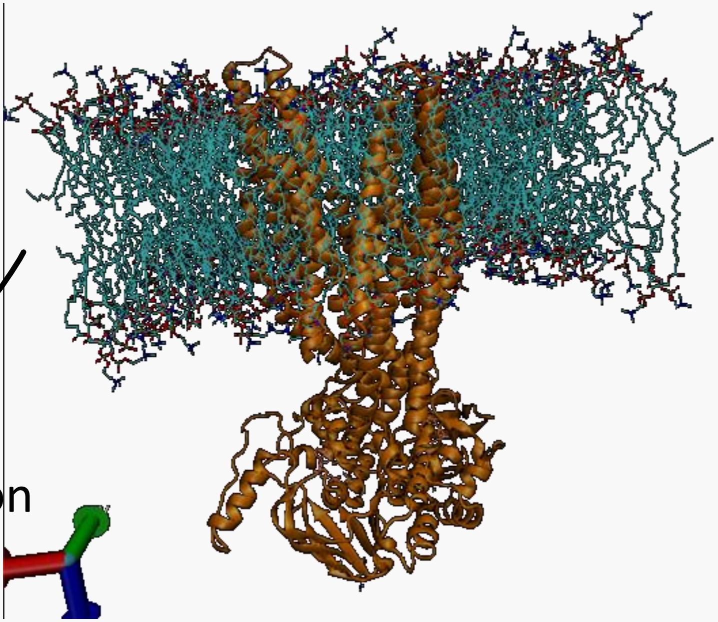
F = m x a

Δt typically 10^{-15}sec

Calculate new position

Save energy and geometry for average





Setting Up an MD Simulation

Initial coordinates

Minimize structure

Initial velocities

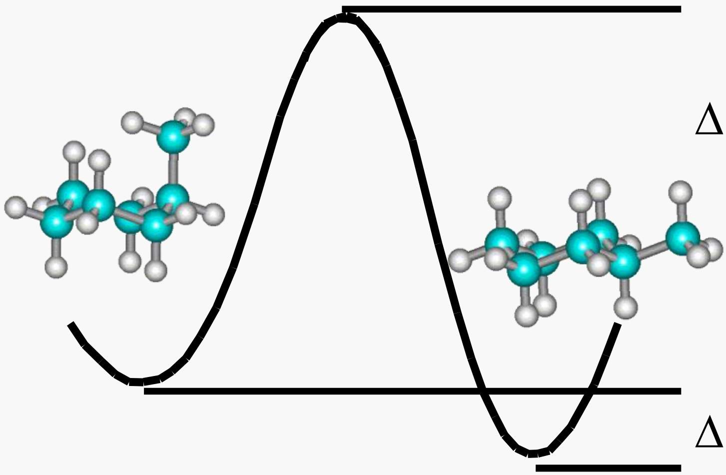
Heating dynamics

Equilibration dynamics

Production dynamics

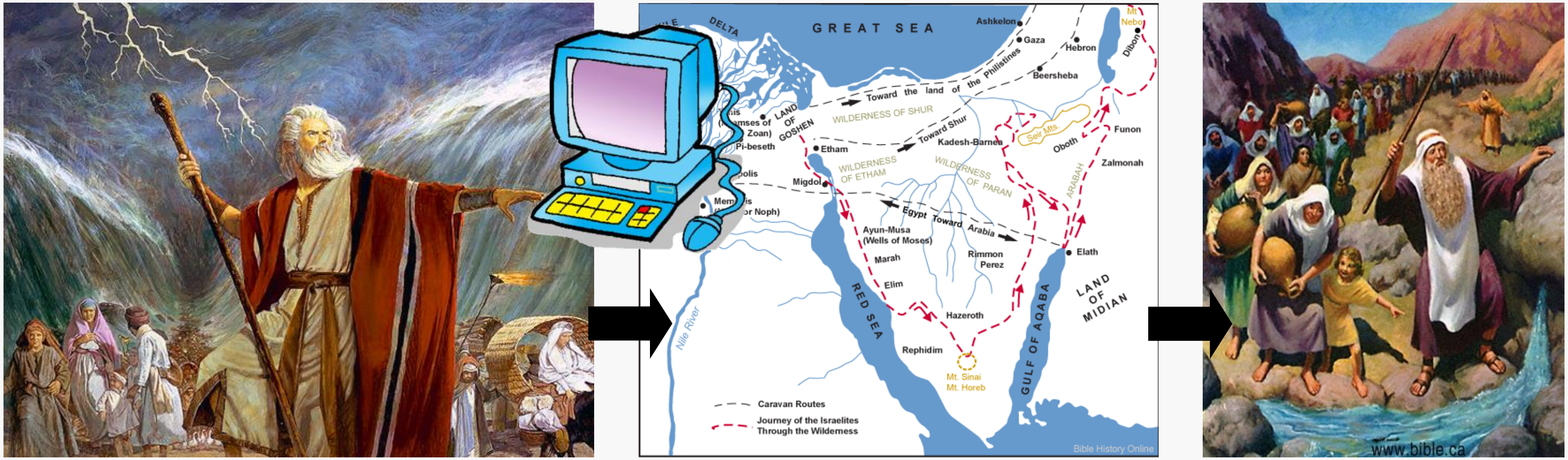
Trajectory analysis

Convergence Problems in Molecular Dynamics (a 20 Years Old Example…)



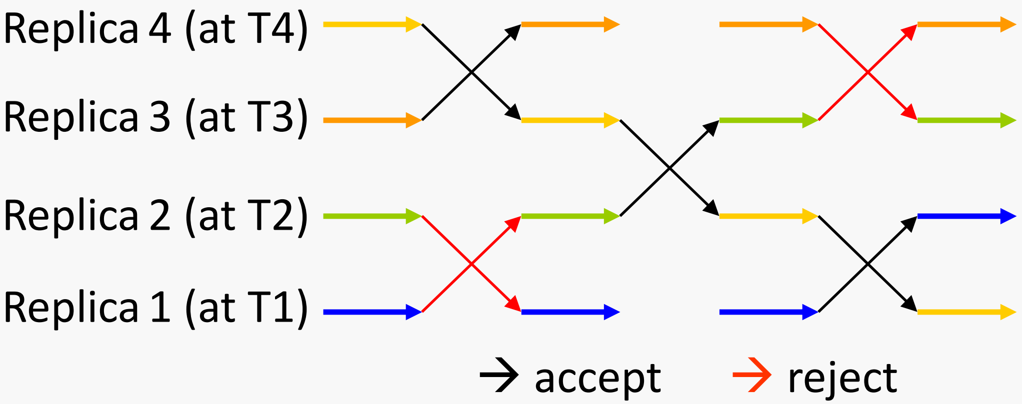
$\Delta G \sim 10$ kcal/mol

$\Delta G \sim 1.7$ kcal/mol



Enhanced Sampling Methods

Replica exchange



Replica 4 (at T4)

Replica 3 (at T3)

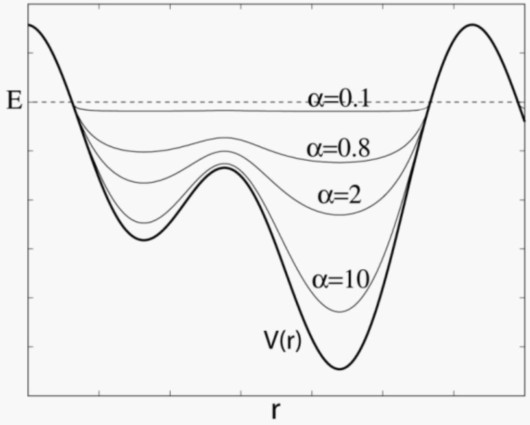
Replica 2 (at T2)

Replica 1 (at T1)

→ accept → reject

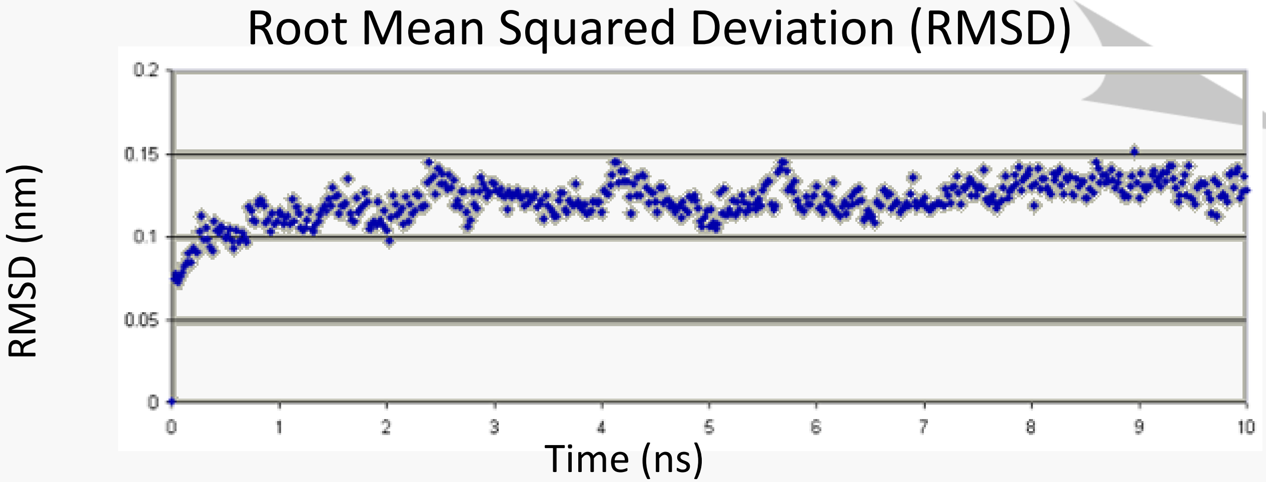
Accelerated MD

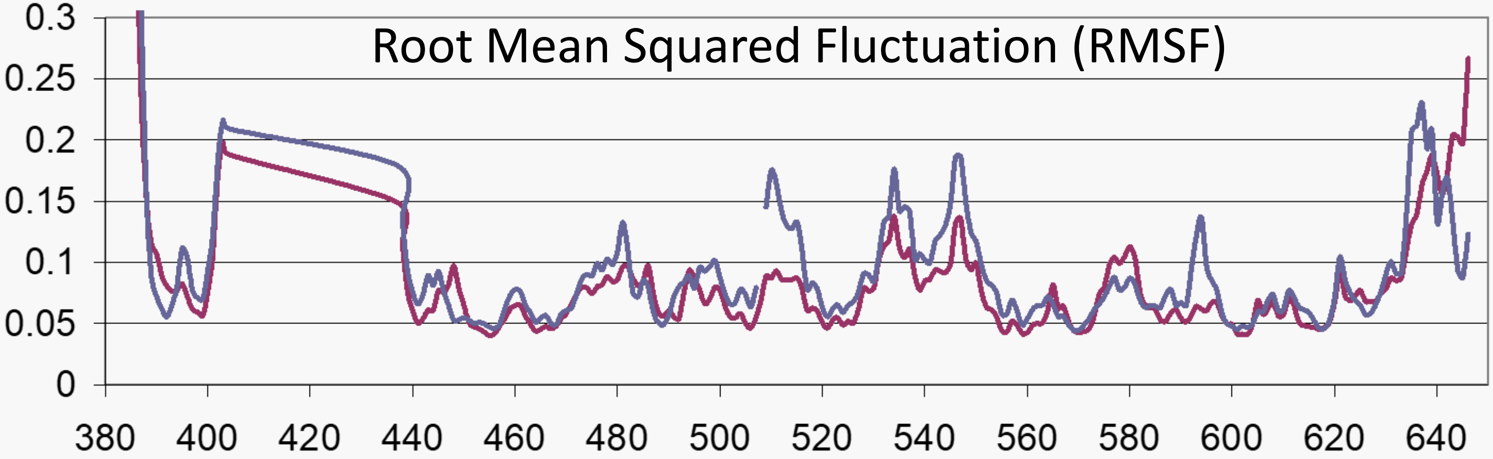
$$\Delta V(\mathbf{r}) = \begin{cases} 0 & V(\mathbf{r}) \geq E \\ \frac{(E-V(\mathbf{r}))^2}{\alpha+E-V(\mathbf{r})} & V(\mathbf{r}) < E. \end{cases}$$



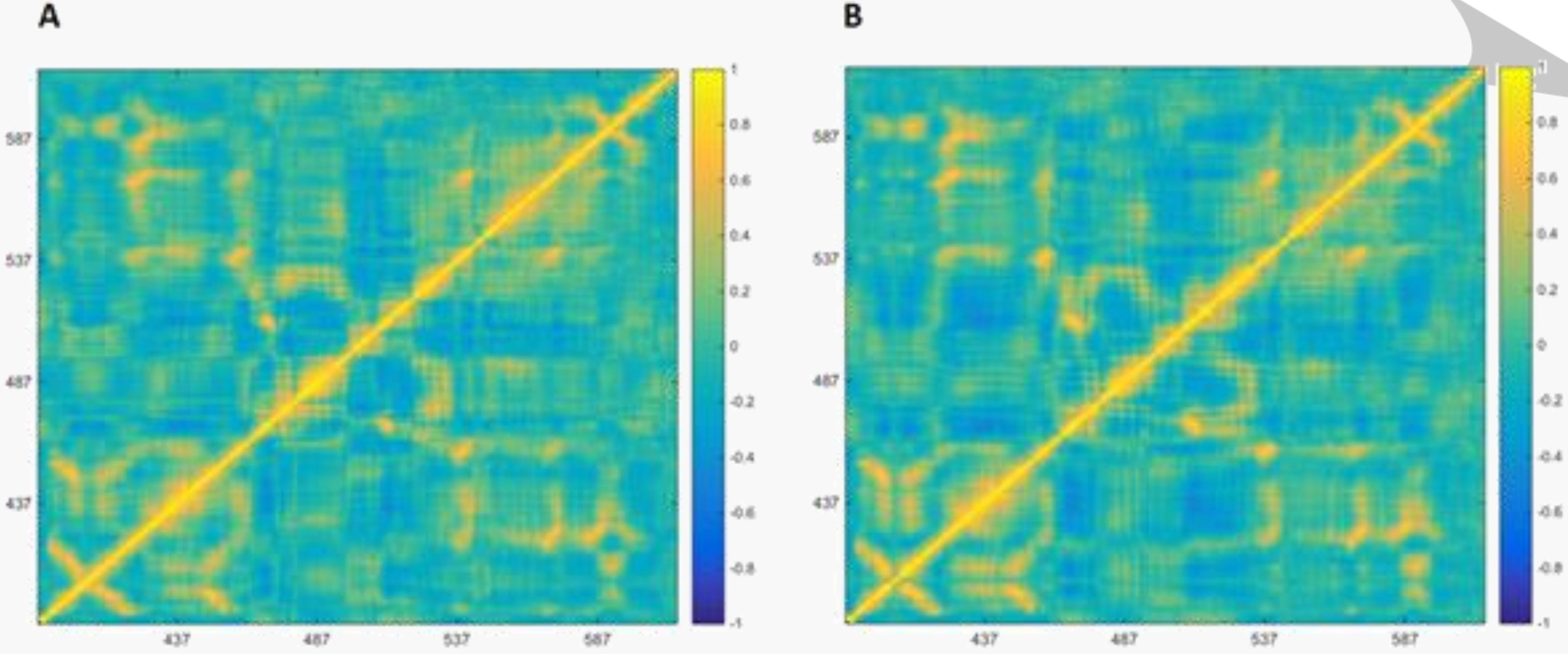
$$\langle A \rangle = \frac{\langle A(\mathbf{r}) \exp(\beta \Delta V(\mathbf{r})) \rangle^*}{\langle \exp(\beta \Delta V(\mathbf{r})) \rangle^*}$$ ’13

Analysis


Root Mean Squared Deviation (RMSD)

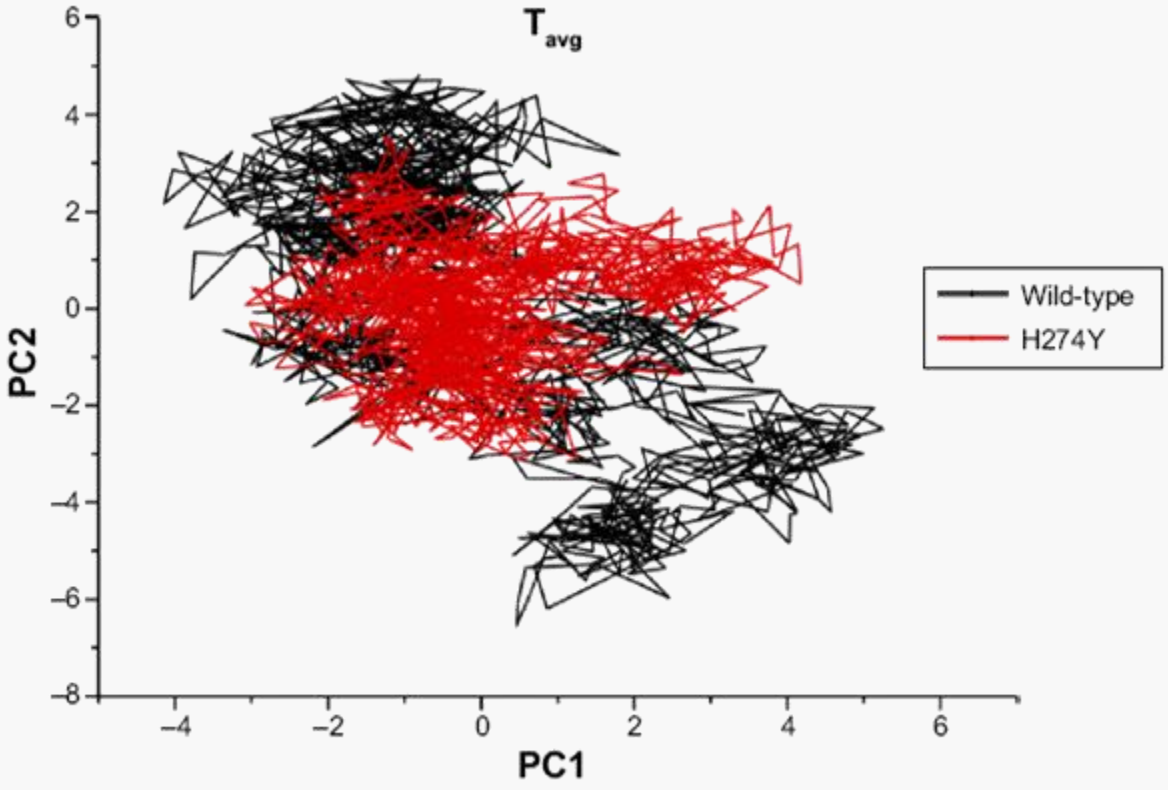

Root Mean Squared Fluctuation (RMSF)

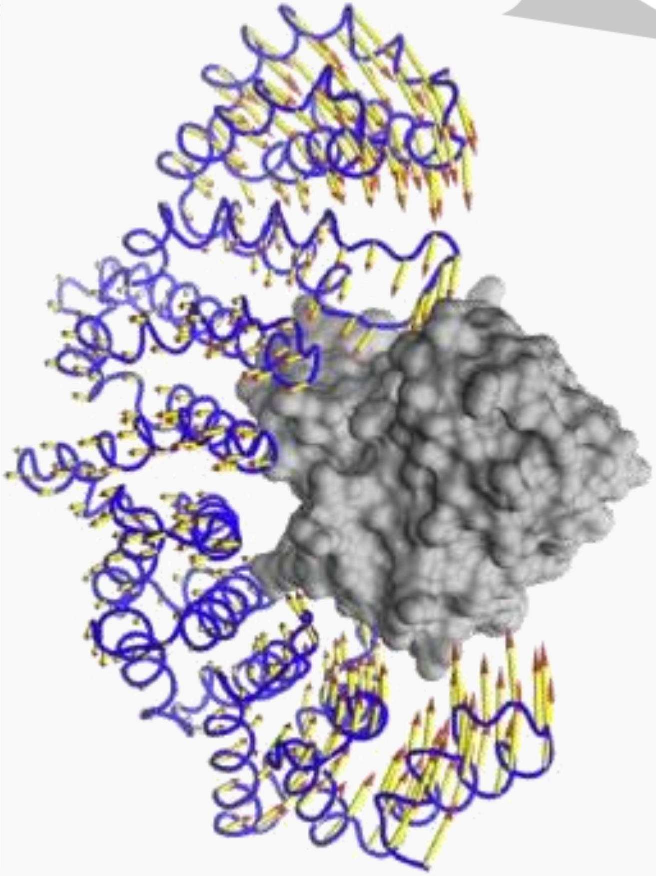
Covariance Matrix



$$N_{\mathrm{cov}(X,Y)} \frac{\sum((X_i - < X >)\cdot(Y_i - < Y >))}{\sqrt{\sum(X_i - < X >)^2 \cdot \sum(Y_i - < Y >)^2}}$$

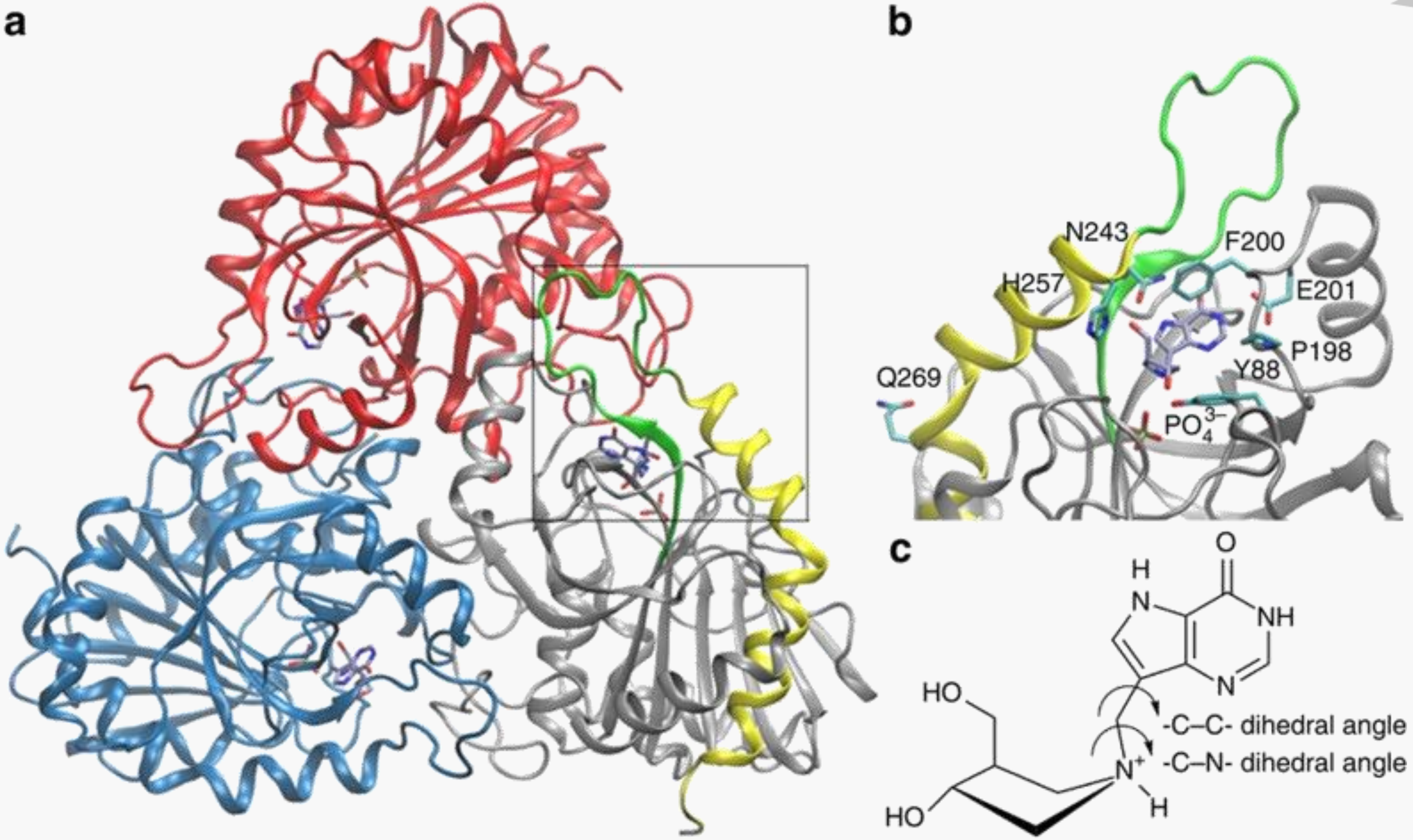
Principle Component Analysis (PCA)





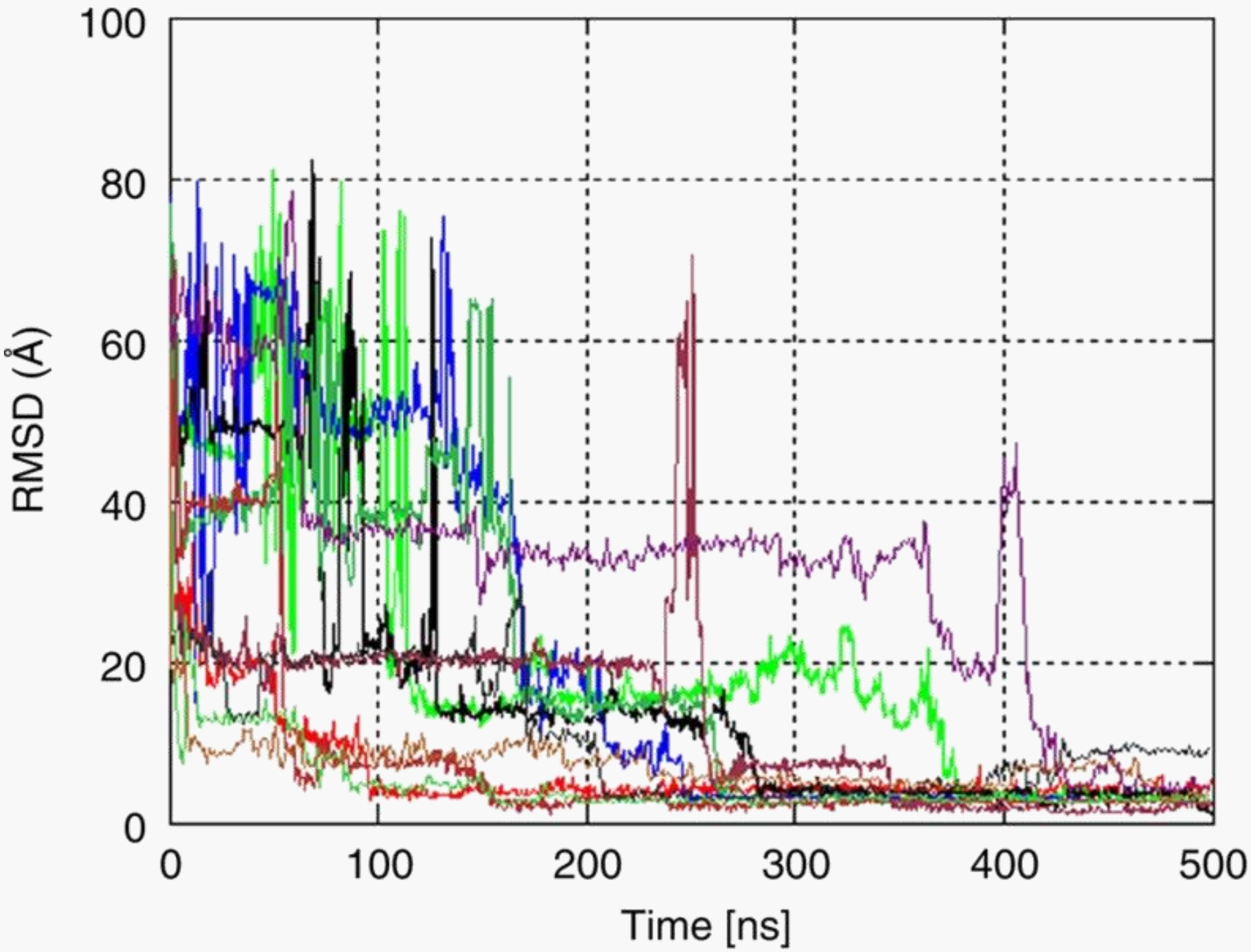
Direct Binding Simulations

purine nucleoside phosphorylase



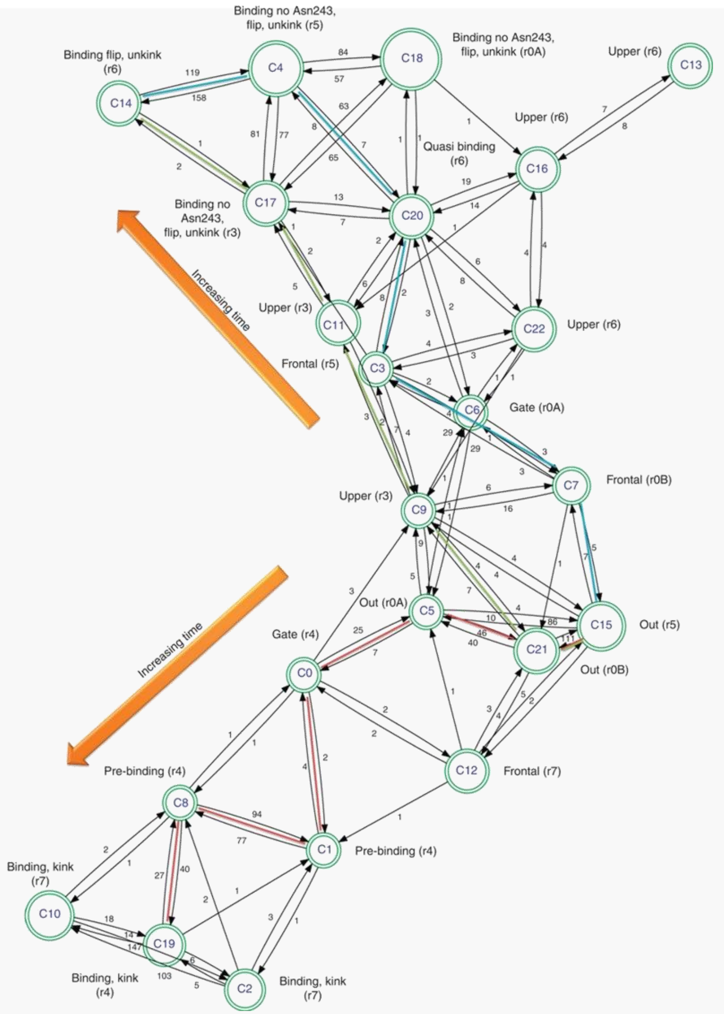
Decherchi et al., Nature Comm. 2015, 6, 6155

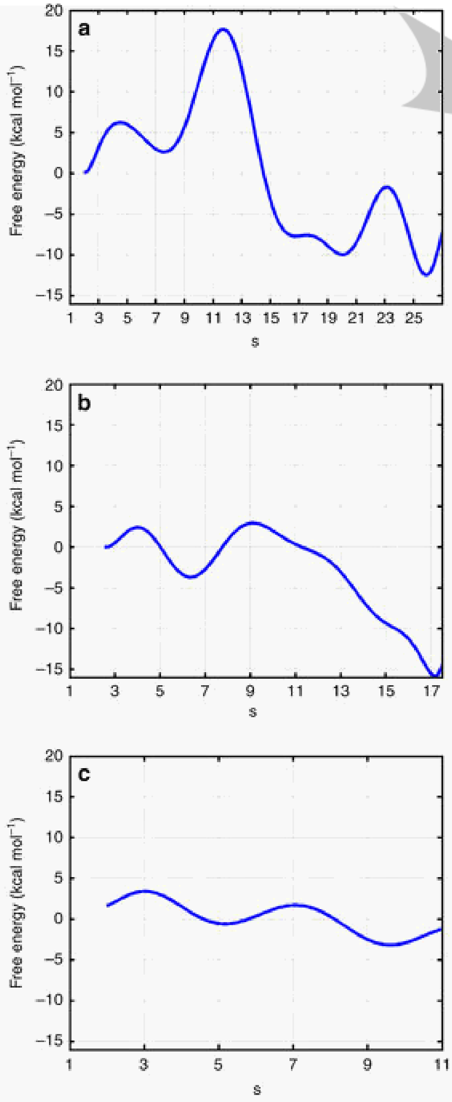
Direct Binding Simulations



Decherchi et al., Nature Comm. 2015, 6, 6155

Direct Binding Simulations



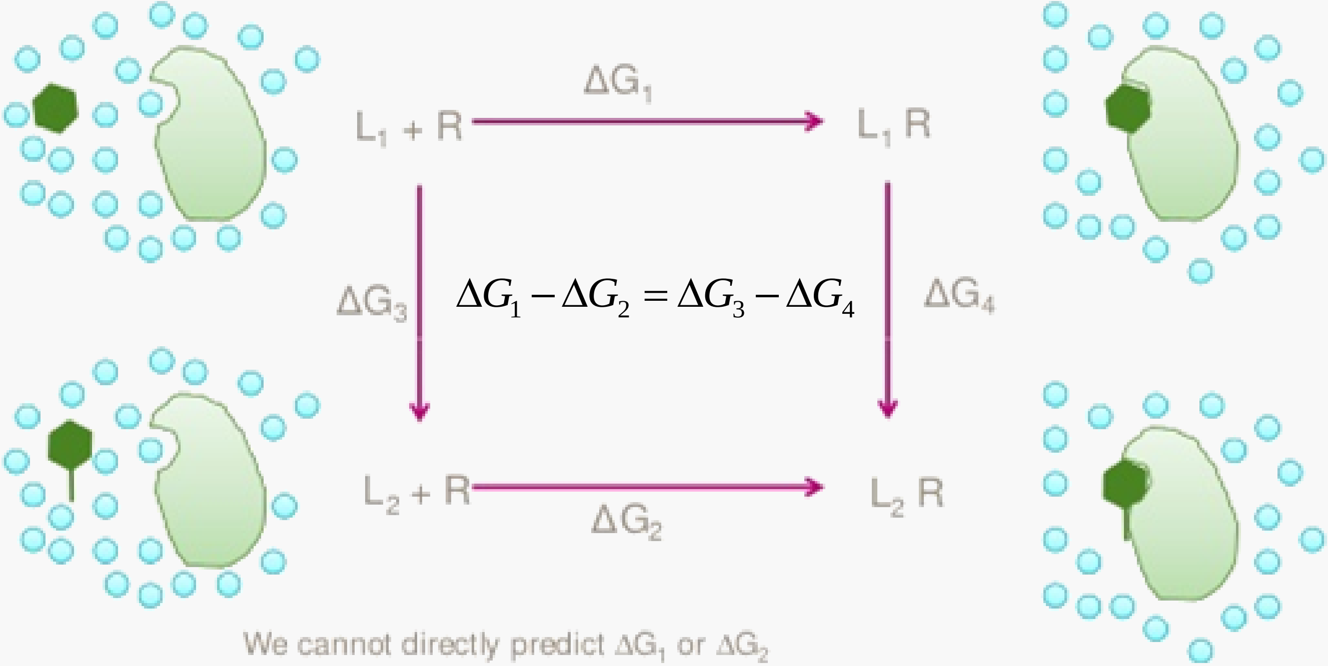


Decherchi et al., Nature Comm. 2015, 6, 6155

Direct Binding Simulations

Decherchi et al., Nature Comm. 2015, 6, 6155

Free Energy Perturbations



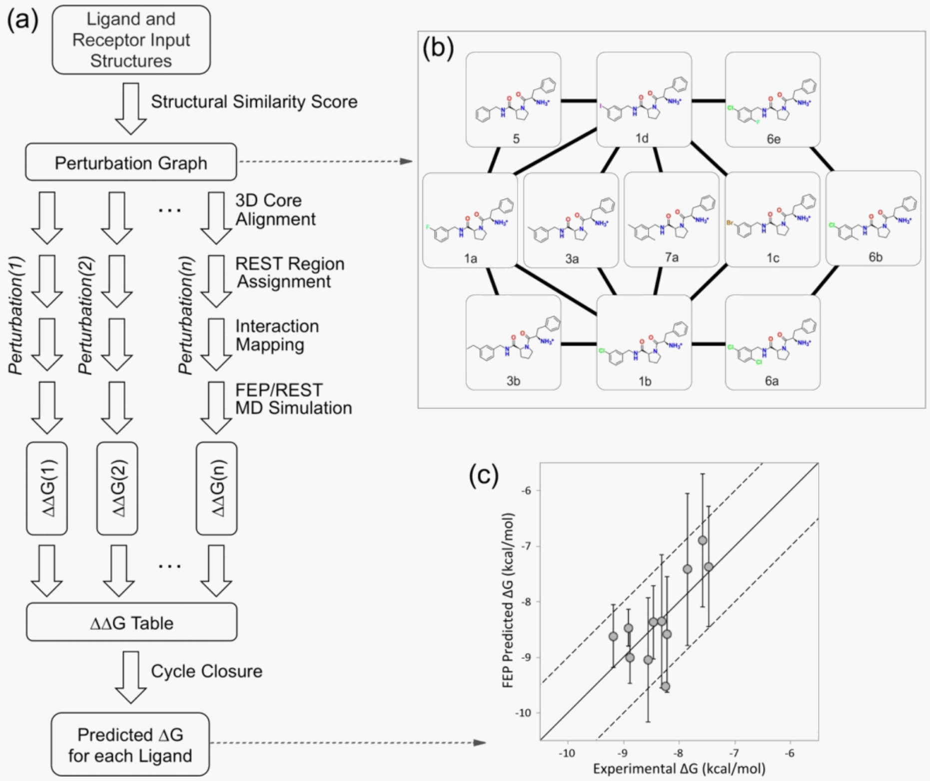
$$\Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$$

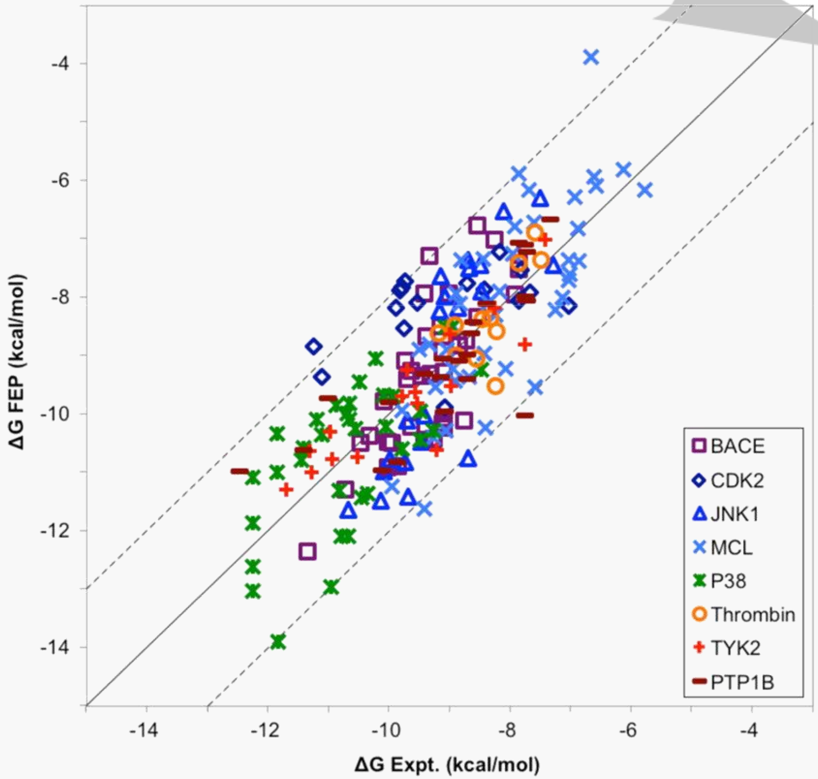
We cannot directly predict ΔG_1 or ΔG_2

Zwanzig equation (1954)

$$\Delta F(A \rightarrow B) = F_B - F_A = -k_B T \ln \left\langle \exp\left(-\frac{E_B - E_A}{k_B T} \right) \right\rangle_A$$

Free Energy Perturbations

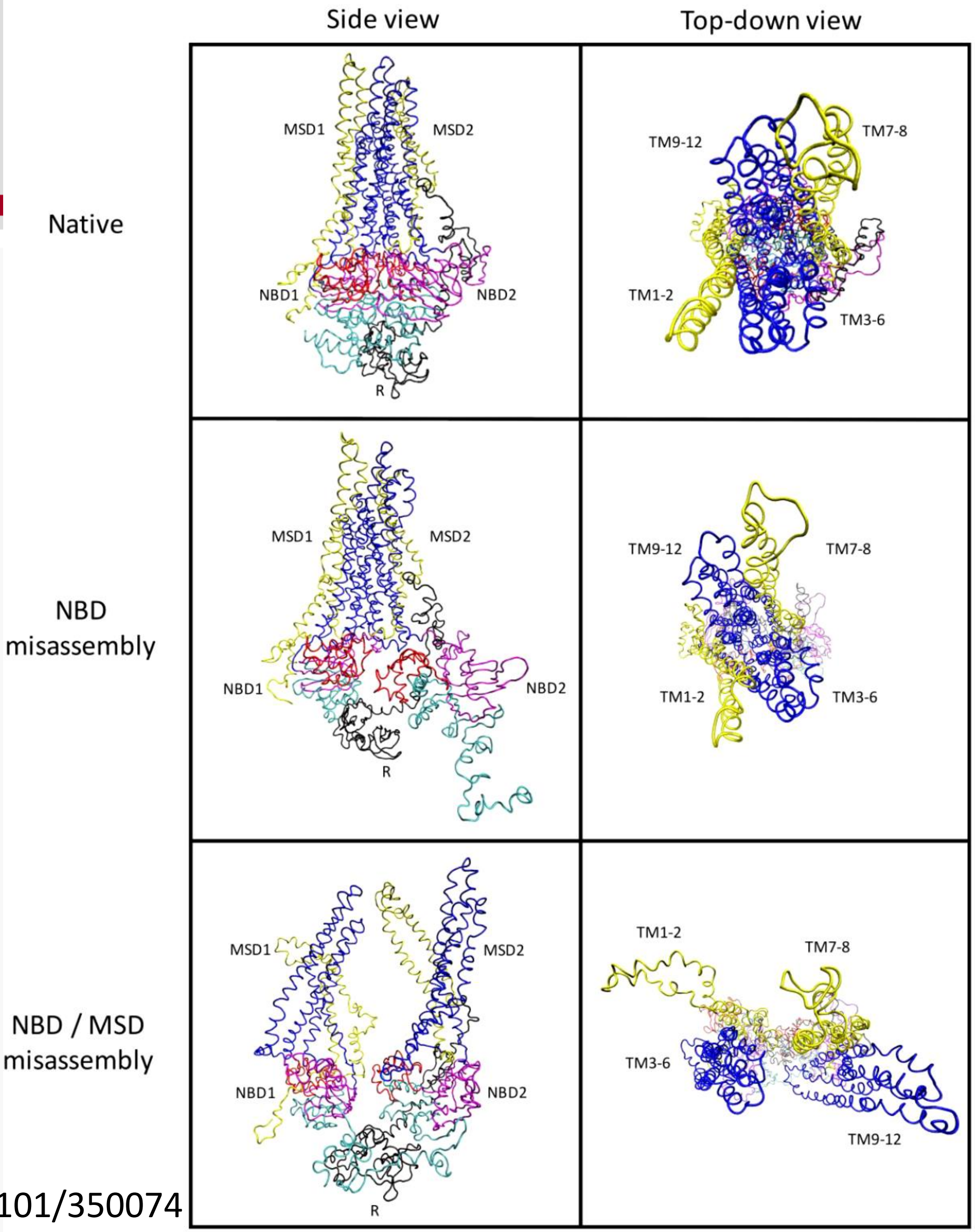




Wang et al., JACS. 2015, 13:2695-703

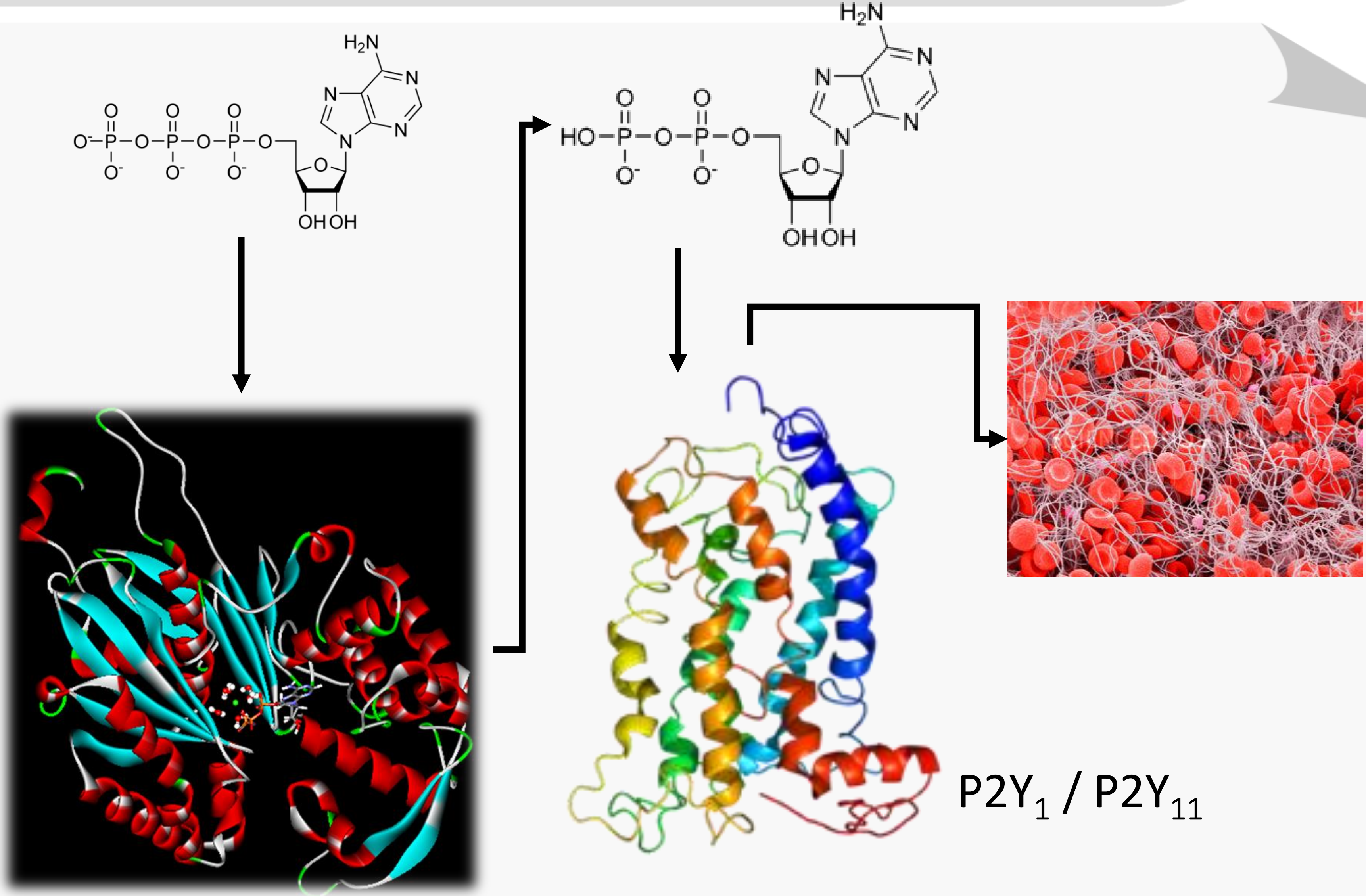
Protein Folding

- Native-centric force field
- Multiple repeats
- Not all trajectories successful
- Environment is important
- Information about folding efficiency



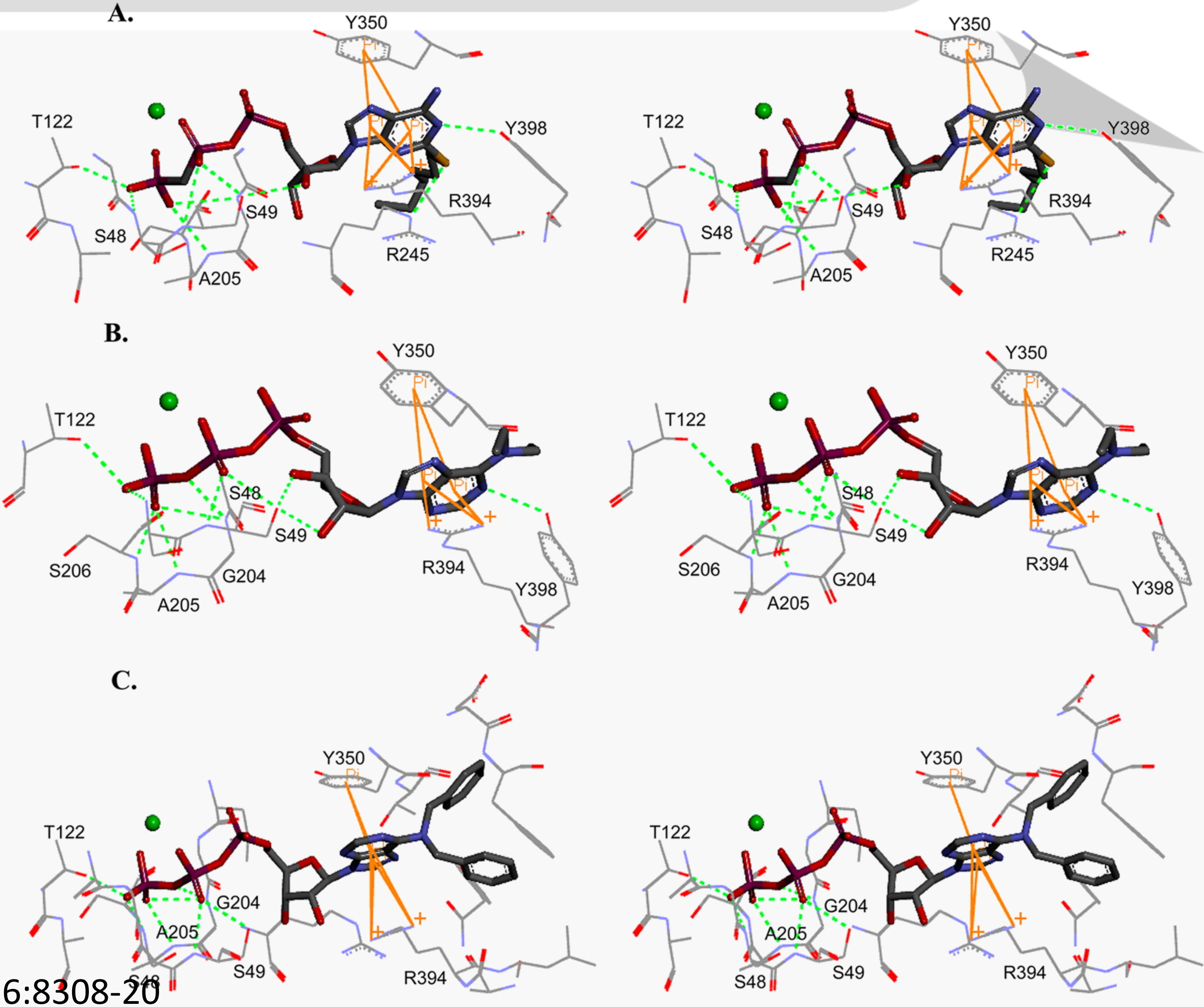
Shi et al., bioRxiv , doi: . http://dx.doi.org/10.1101/350074

NTPDase2 Inhibitors: Potential Anti-Coagulates



$P2Y_1$ / $P2Y_{11}$

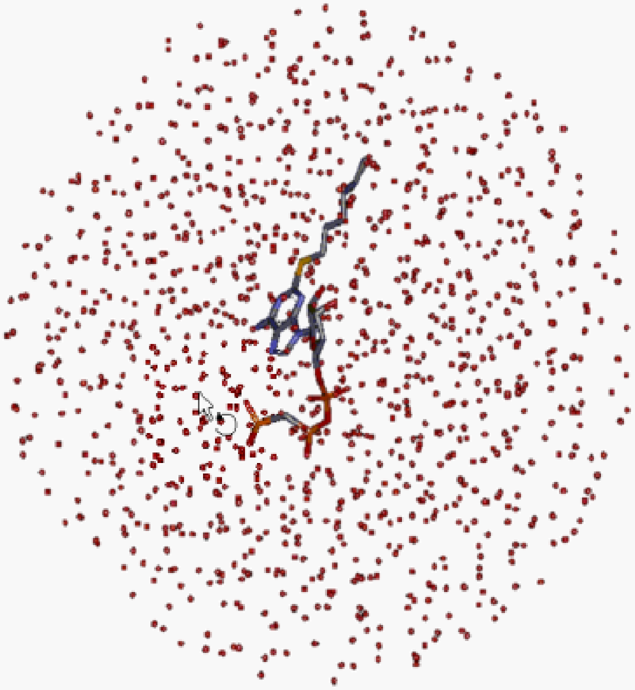
Homology Modeling and Docking



Lecka et al., J Med Chem. 2013 56:8308-20

Linear Interaction Energy (LIE) (Because Docking Didn't Work…)

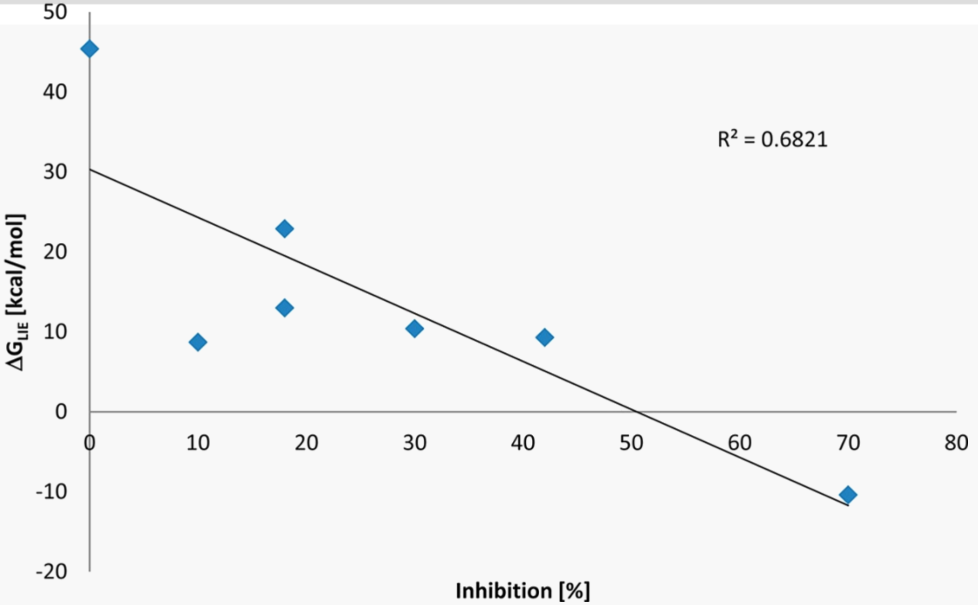
$$\Delta G_{bind} = \alpha(<V_{l-s}^{vdw}>_{bound} - <V_{l-s}^{vdw}>_{free}) + \beta(<V_{l-s}^{el}>_{bound} - <V_{l-s}^{el}>_{free})$$



Lecka et al., J Med Chem. 2013 56:8308-20

Results

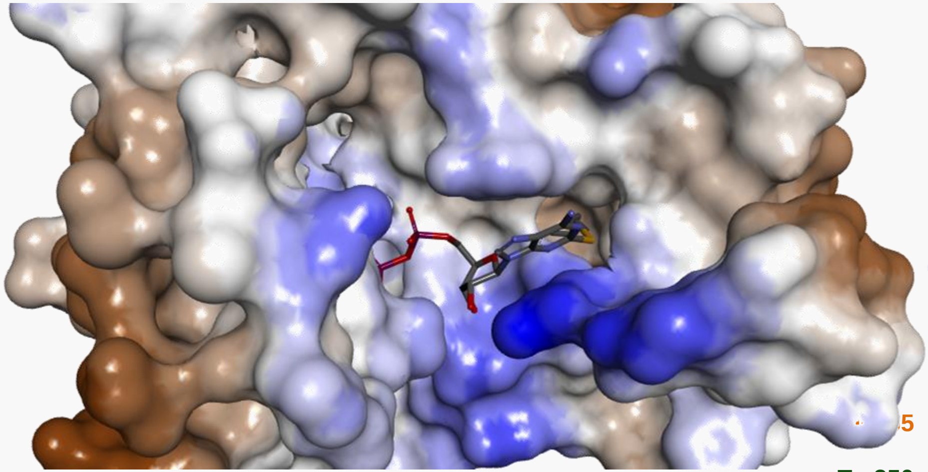
c.



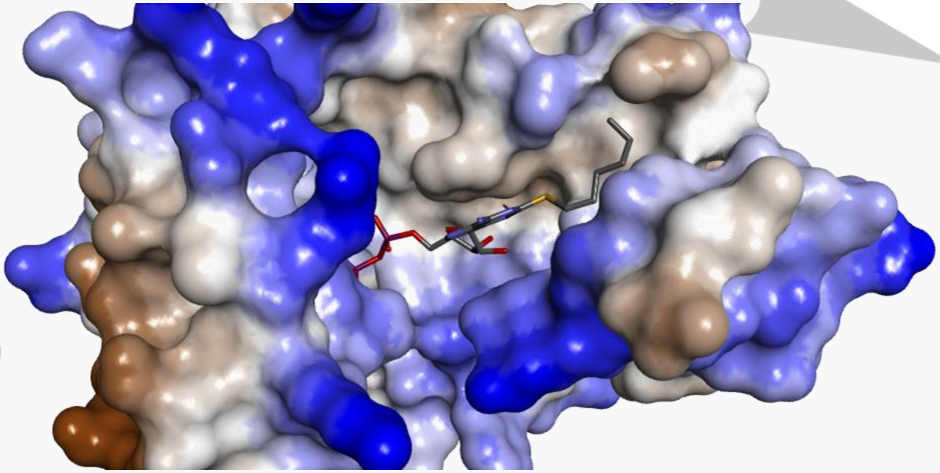
Compound	Percent of inhibition	$\langle V_{l\text{-}s}^{vdw}\rangle_w$	$\langle V_{l\text{-}s}^{el}\rangle_w$	$\langle V_{l\text{-}s}^{vdw}\rangle_p$	$\langle V_{l\text{-}s}^{el}\rangle_p$	ΔG_{LIE}
2	70	10.0±0.6	-1640±10	-29.6±0.6	-1647±7	-10.4
6	42	11.3±0.9	-1653±15	-25.6±1.0	-1621±11	9.3
7	0	0.9±0.5	-1686±13	-31.9±4.1	-1583±3	45.4
9	10	-6.7±0.4	-588.6±6.3	-29.5±0.9	-563±8	8.7
18	30	13.1±0.4	-1668±14	-31.1±1.6	-1631±6	10.4
19	18	2.2±0.4	-1097±5.5	-36.1±0.4	-1057±4	13.0
20	18	14.7±0.8	-1710±21	-30.6±1.9	-1648±11	22.9

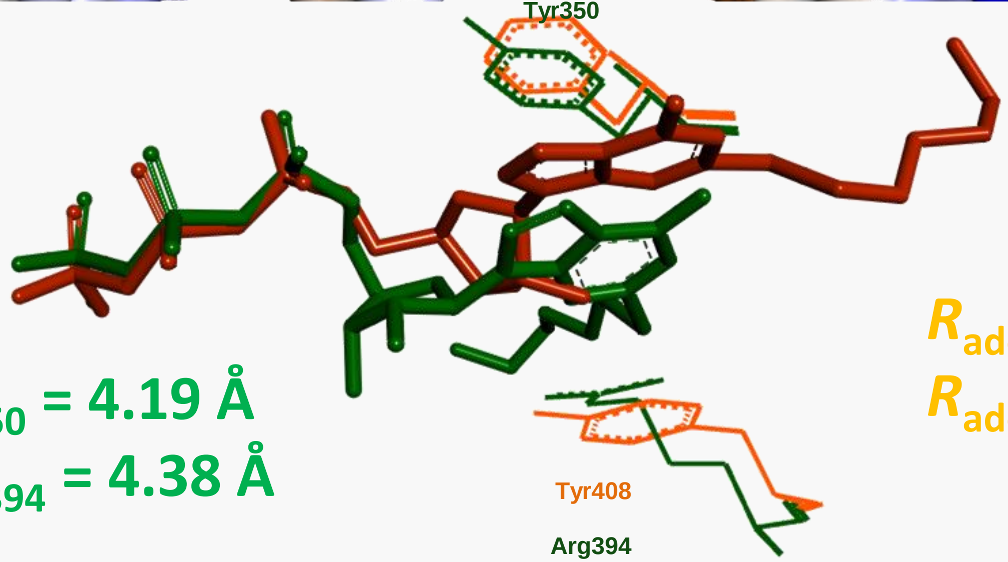
Selectivity of Most Potent Compound

NTPDase2 (20 μM)



NTPDase1 (inactive)





$R_{\text{adenine-Tyr350}} = 4.19$ Å

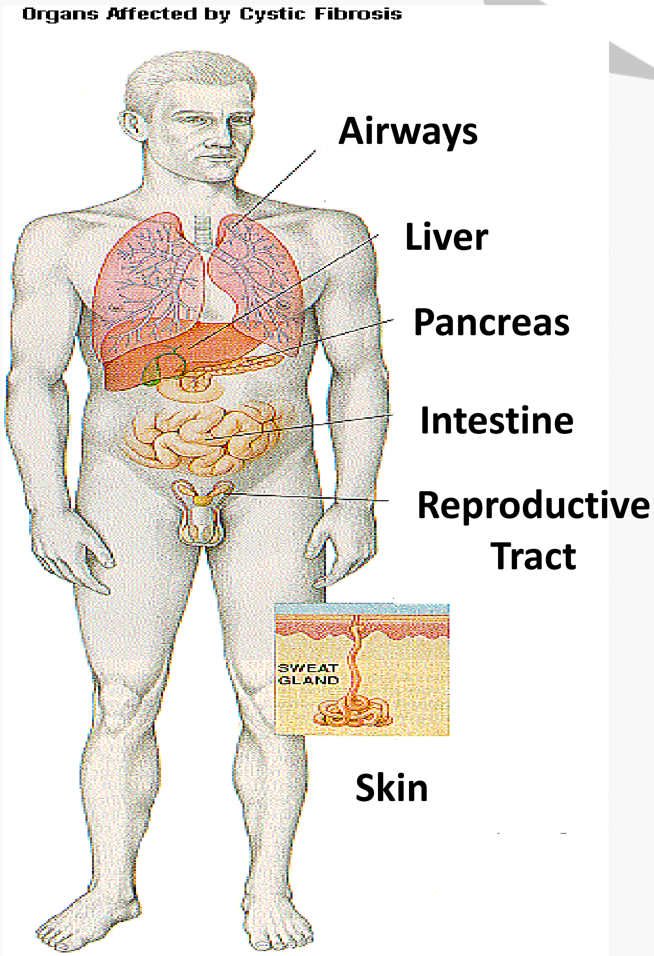
$R_{\text{adenine-Arg394}} = 4.38$ Å

$R_{\text{adenine-Tyr350}} \approx 4.19$ Å

$R_{\text{adenine-Tyr408}} = 5.52$ Å

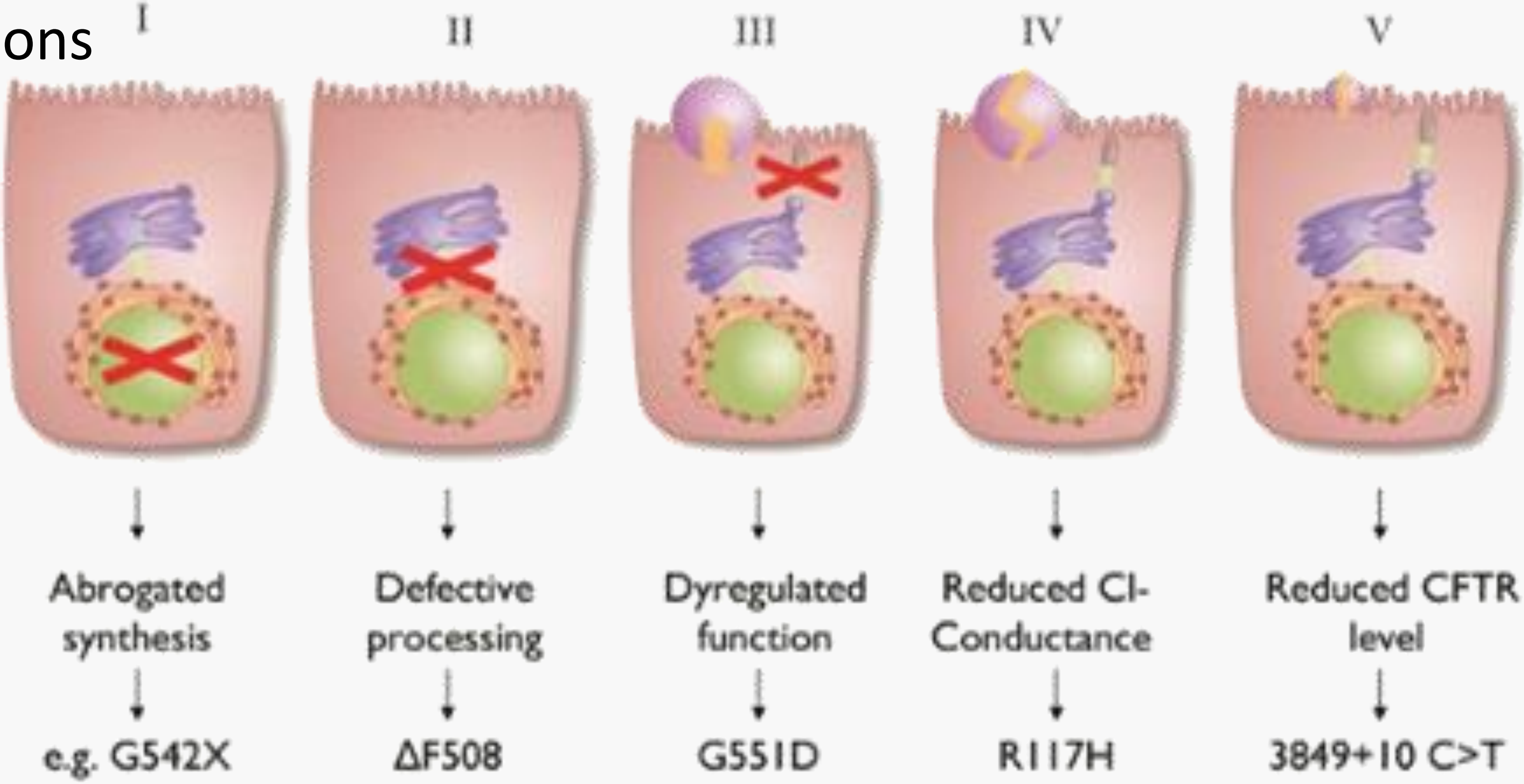
The Cystic Fibrosis Disease

- CF is the most common lethal, inherited disease among people of European descent
- The number of CF patients is estimated at 90,000 worldwide, about 30,000 of which are in the US (~700 in Israel, **~7000 in France**)
- Median survival age is ~40 years
- CF results in pathologies in multiple organs
  - ❖ Depressed lung function, lung infection, inflammation, and advanced lung disease
- **Currently, there is no general cure for CF and most of the treatments are symptomatic**
- **CF is caused by mutations to the CFTR chloride channel**



CFTR Mutations

- ~2000 CFTR mutations
- > 300 mutations confirmed as CF-causing
- 12 mutations confirmed as non CF-causing
- All CF-causing mutations compromise the ability of CFTR to conduct Cl^- ions

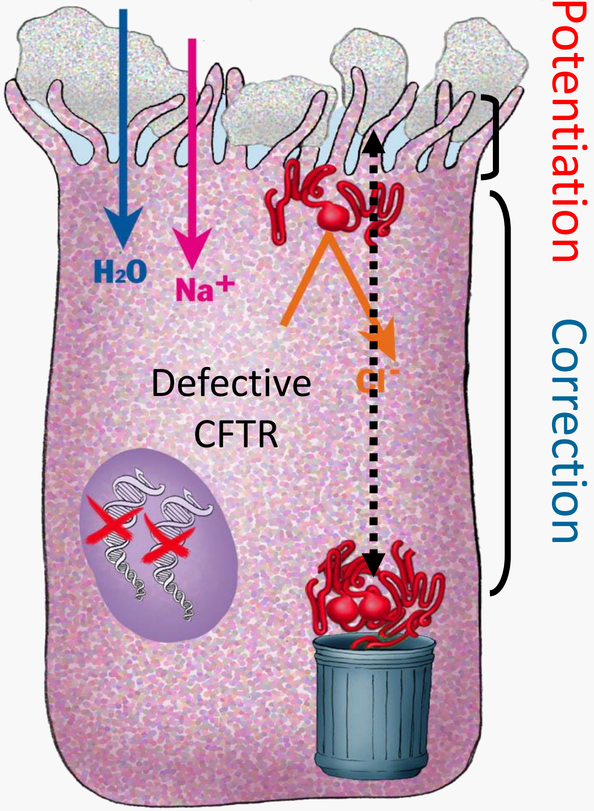


CF Treatment Hypothesis

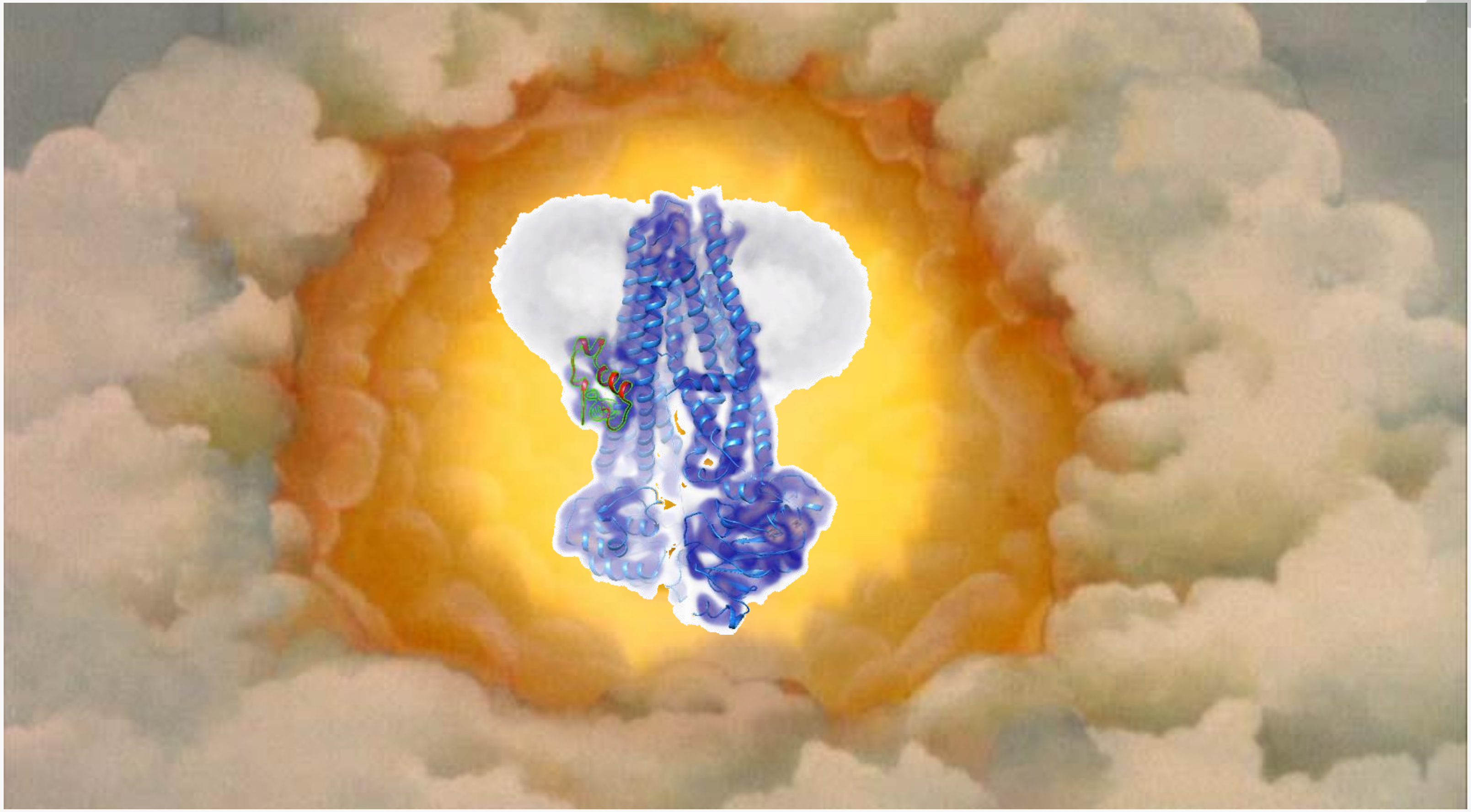
- Impaired Cl^- conductance disrupts the salt-water balance across epithelial cells leading to accumulation of a viscous mucus layer which is colonized by bacteria
- Restoring Cl^- conductance to “normal” levels will ameliorate CF pathologies

Current ~ [# channels] * [open probability]

- **CFTR corrector**: Corrects the folding defect and increases the number of CFTR channels at the cell membrane
- **CFTR potentiator**: Increases the open probability of CFTR channels at the membrane
- **Combo therapy:** Does both



The Holy Grail



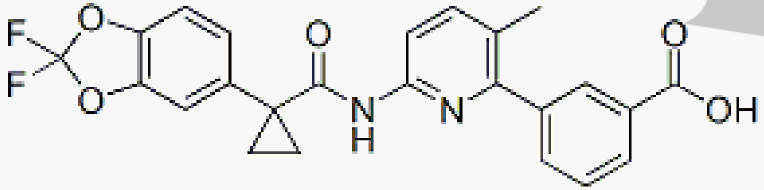
CF Therapeutics

- Mucociliary clearance agents
  - 2 available, plus 5 in pipeline
- Anti-inflammatories
  - 1 available, plus 4 in pipeline
- Antimicrobials
  - 4 available, plus 8 in pipeline
- Agents to restore CFTR function
  - 3 available, 11 in clinical trials, 5 preclinical stage

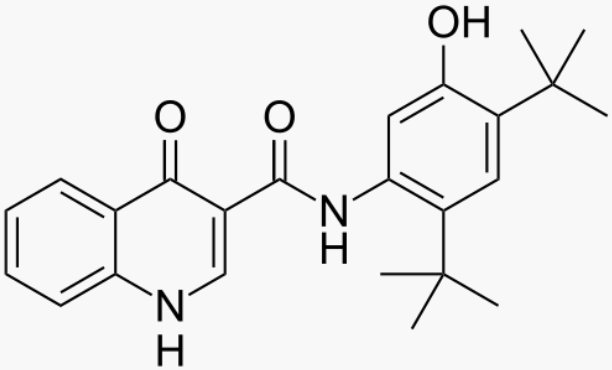
Available CFTR Modulators

Orkambi

Corrector: Lumacaftor (VX-809)

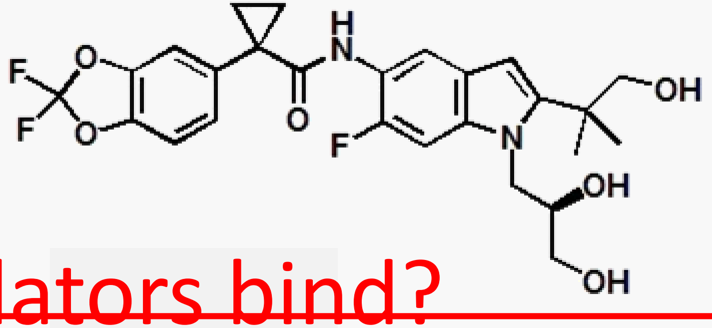


Potentiator: Ivacaftor (VX-770) approved for CF patients with the G551D (~4% of CF patient population), and other 22 gating mutations



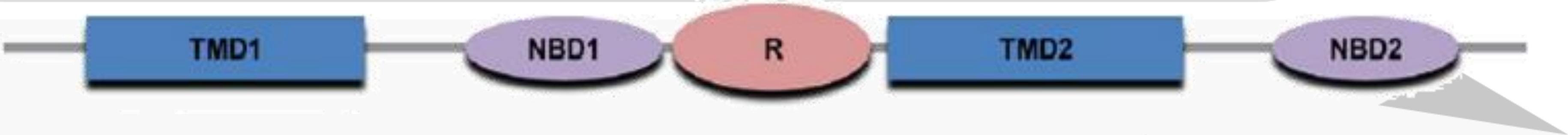
Symdeko

Corrector: Tezacaftor (VX-661)

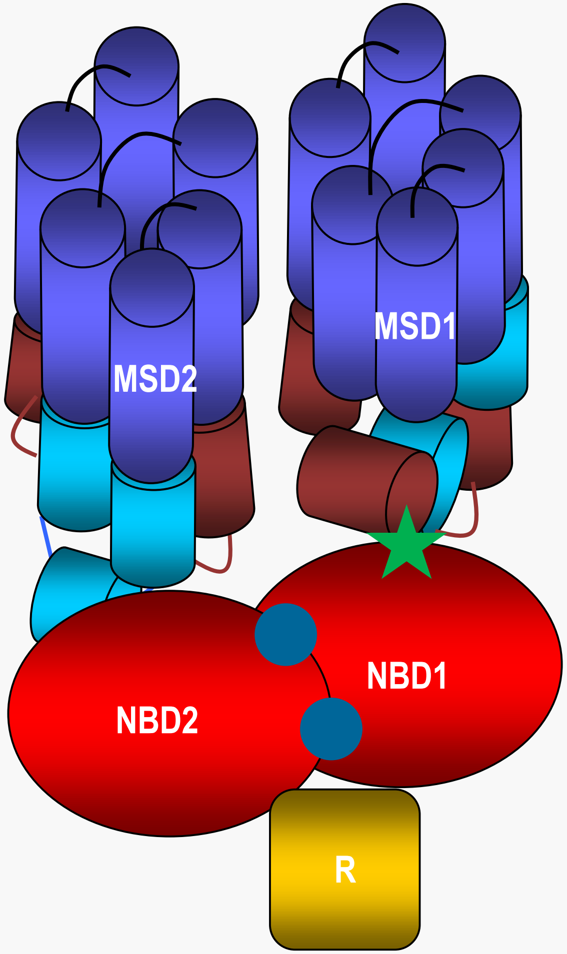


Where do modulators bind?

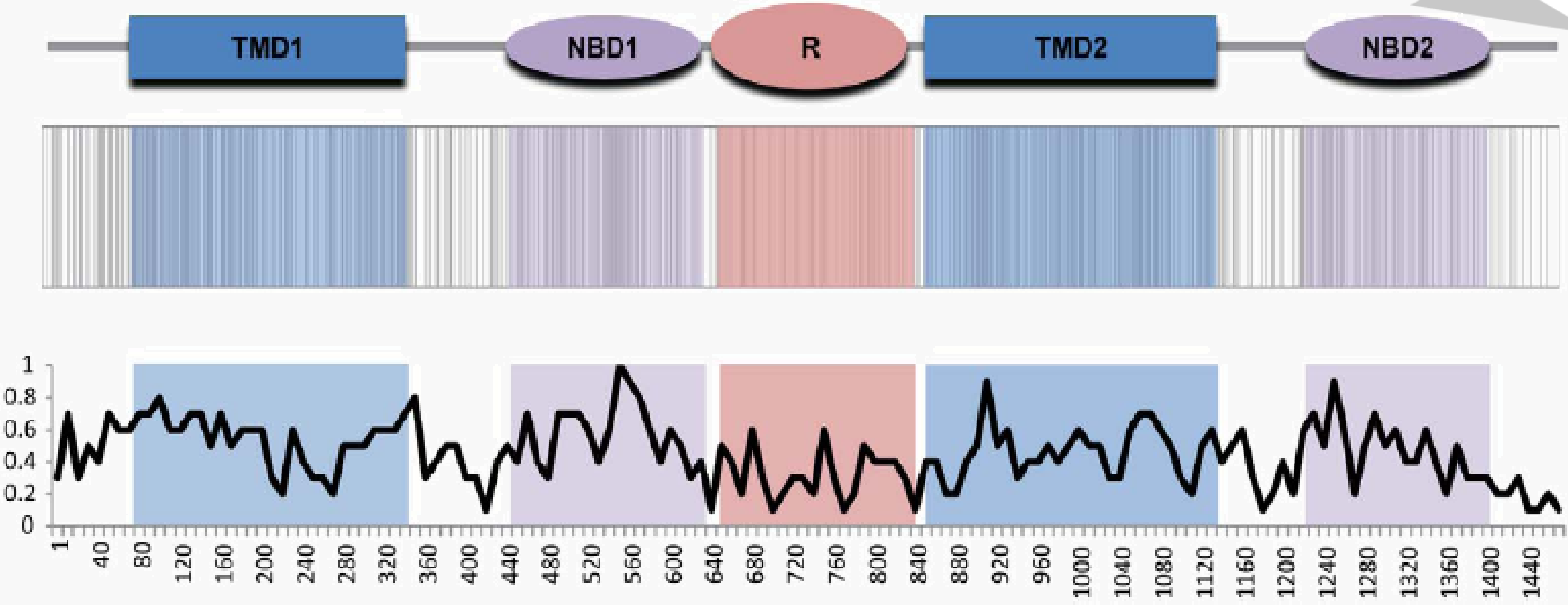
The Structure of CFTR



- CFTR is an ABC transporter
  - Membrane proteins
  - Found in prokaryotes and eukaryotes
  - Harness the energy of ATP hydrolysis for substrate transport across cell membranes

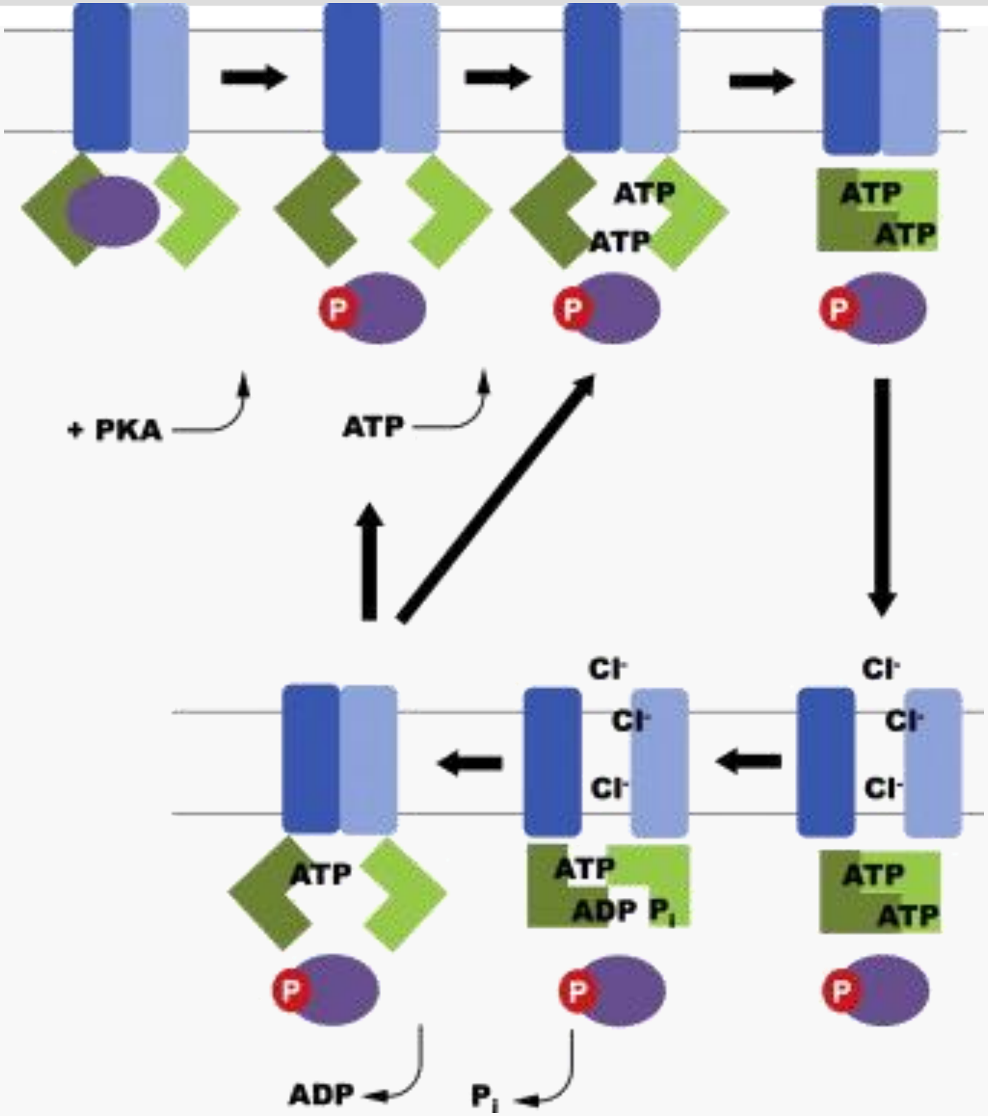


CFTR Sequence and Mutations

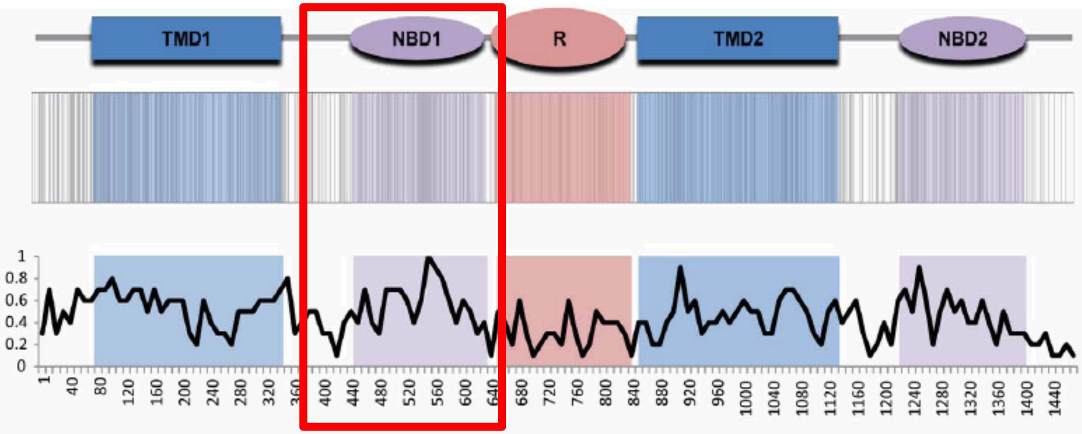


Rishishwar L et al. PLOS One 2012;7(8):e42336.

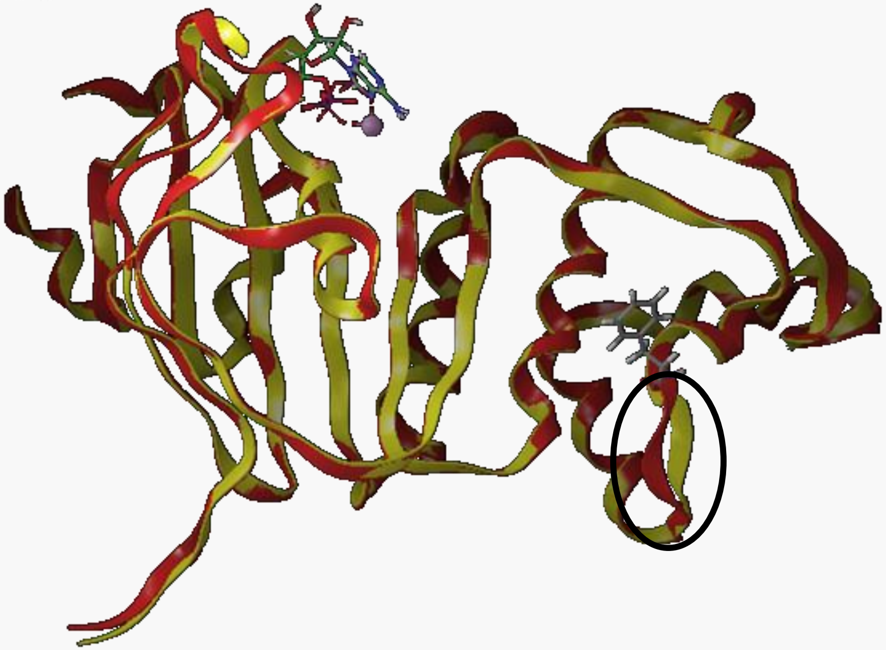
The Gating Cycle of CFTR



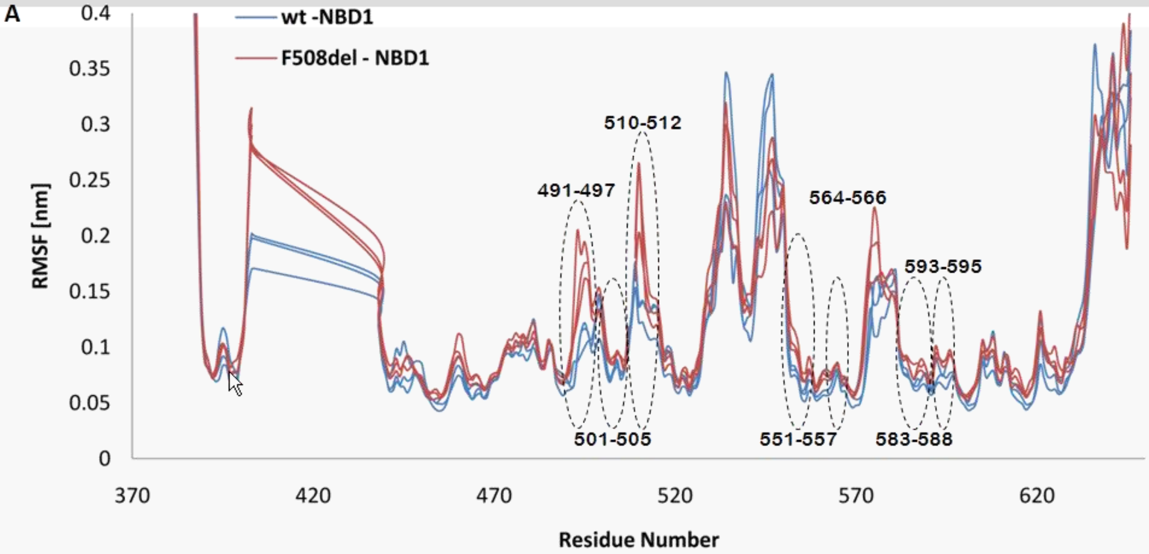
NBD1

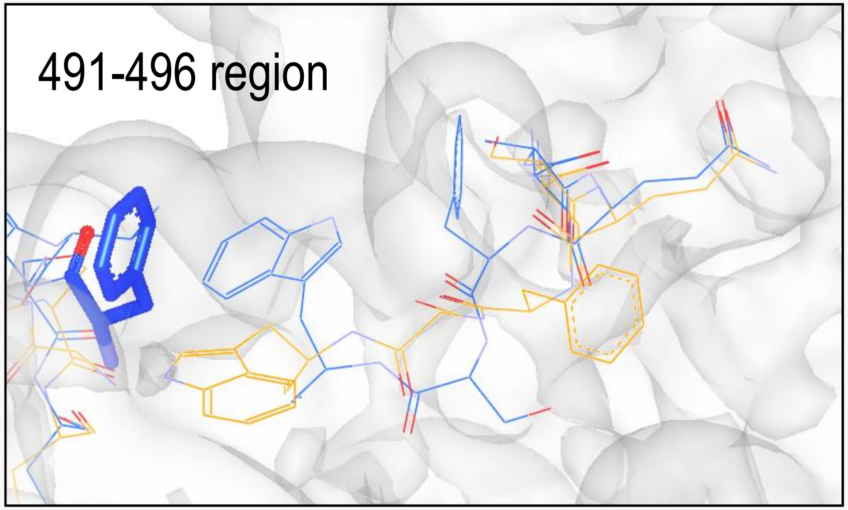


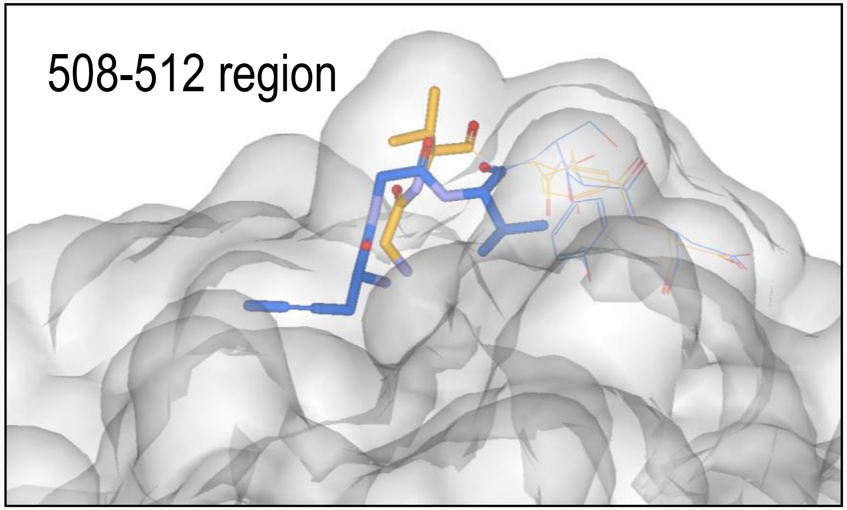
- >20 CFTR NBD1 crystal and NMR structures are currently available
- Studied mostly with respect to F508del



NBD1

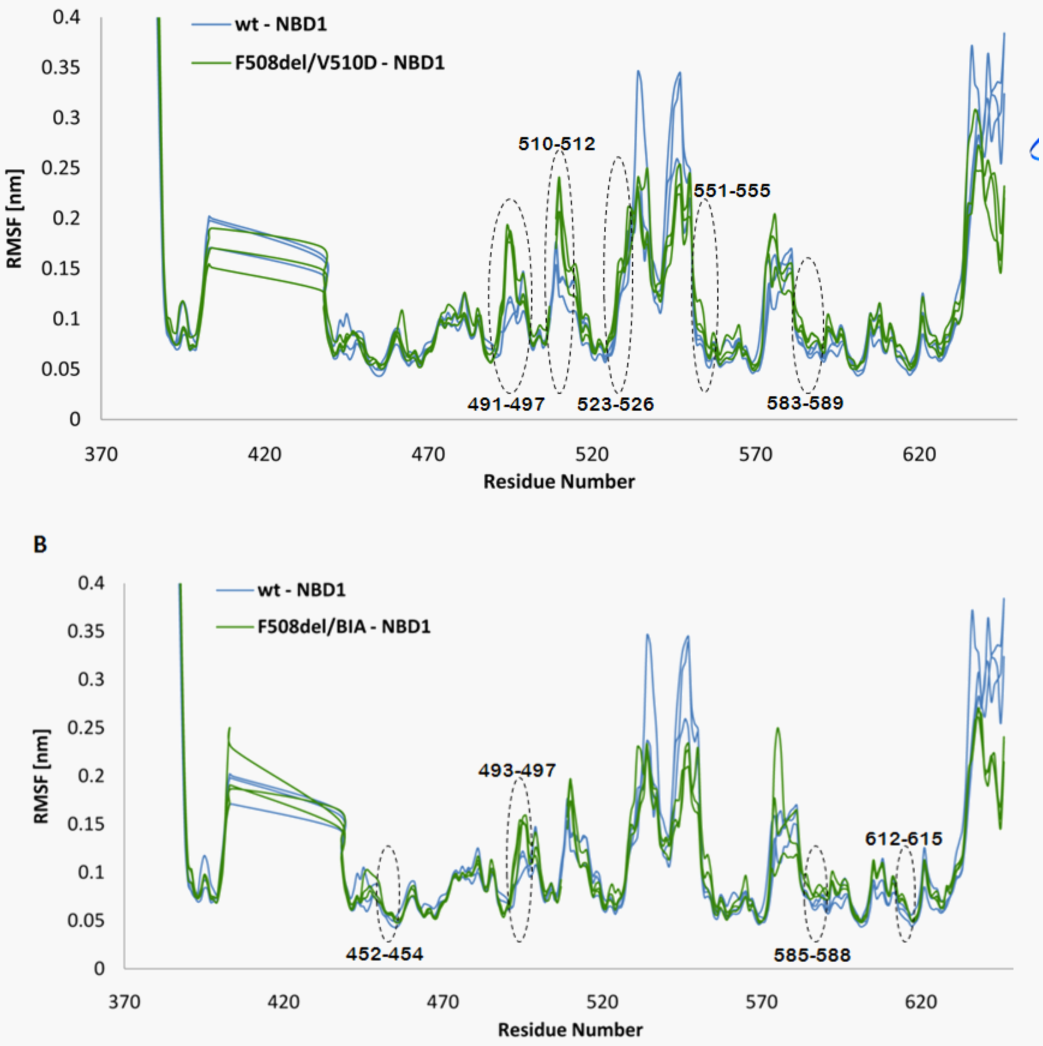


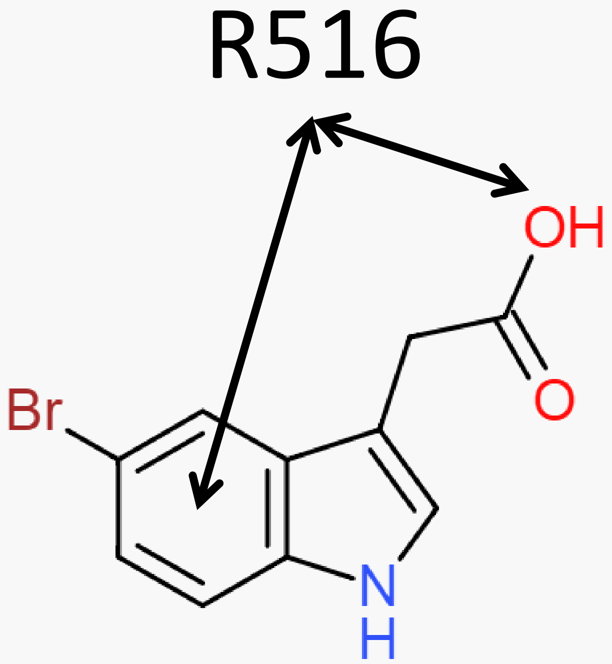




Zhenin et al., JCIM 2015, 55, 2349-64

NBD1

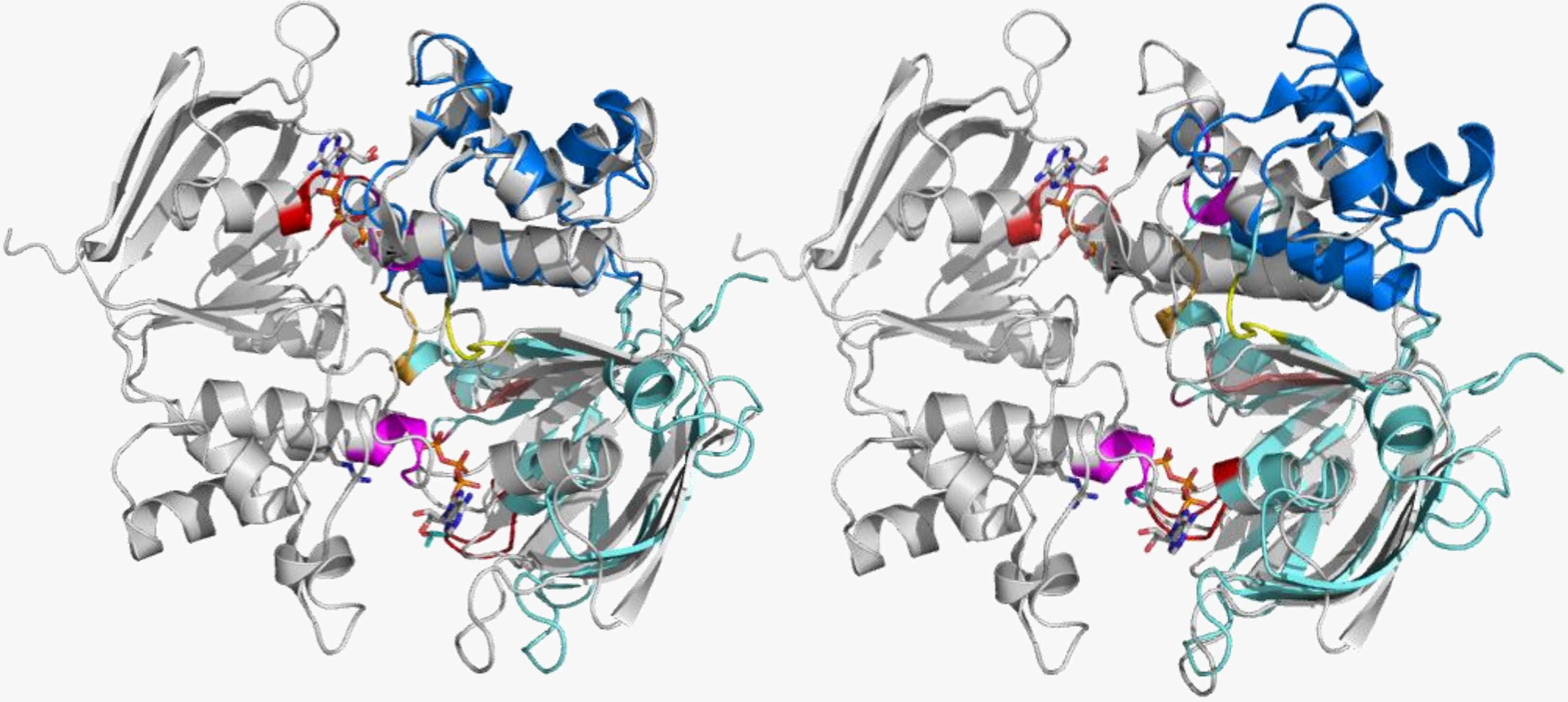




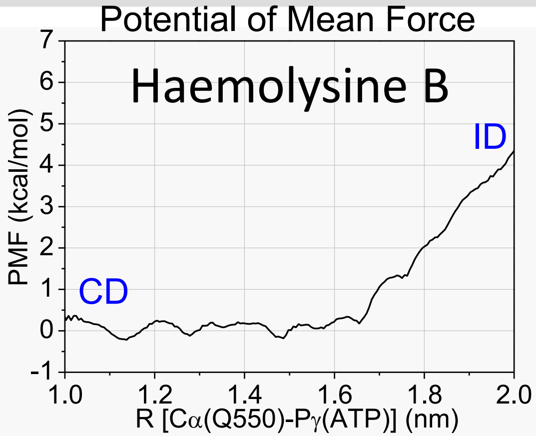
Zhenin et al., JCIM 2015, 55, 2349-64

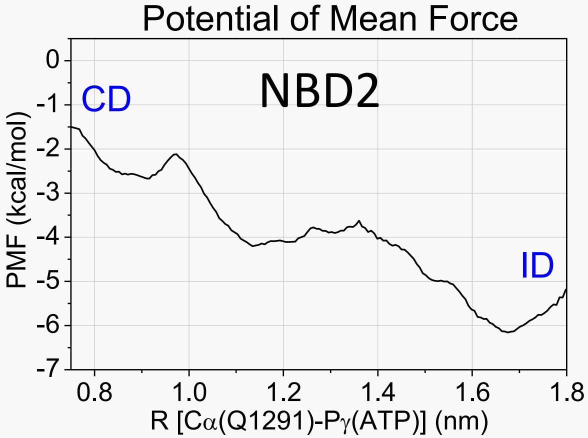
NBD2

- Poorly behaved!
- A single structure of NBD2 is available
  - Dimer incompatible conformation
  - Catalytically inactive(H1402A)
- All attempts to demonstrate NBD1:NBD2 coupling have failed.

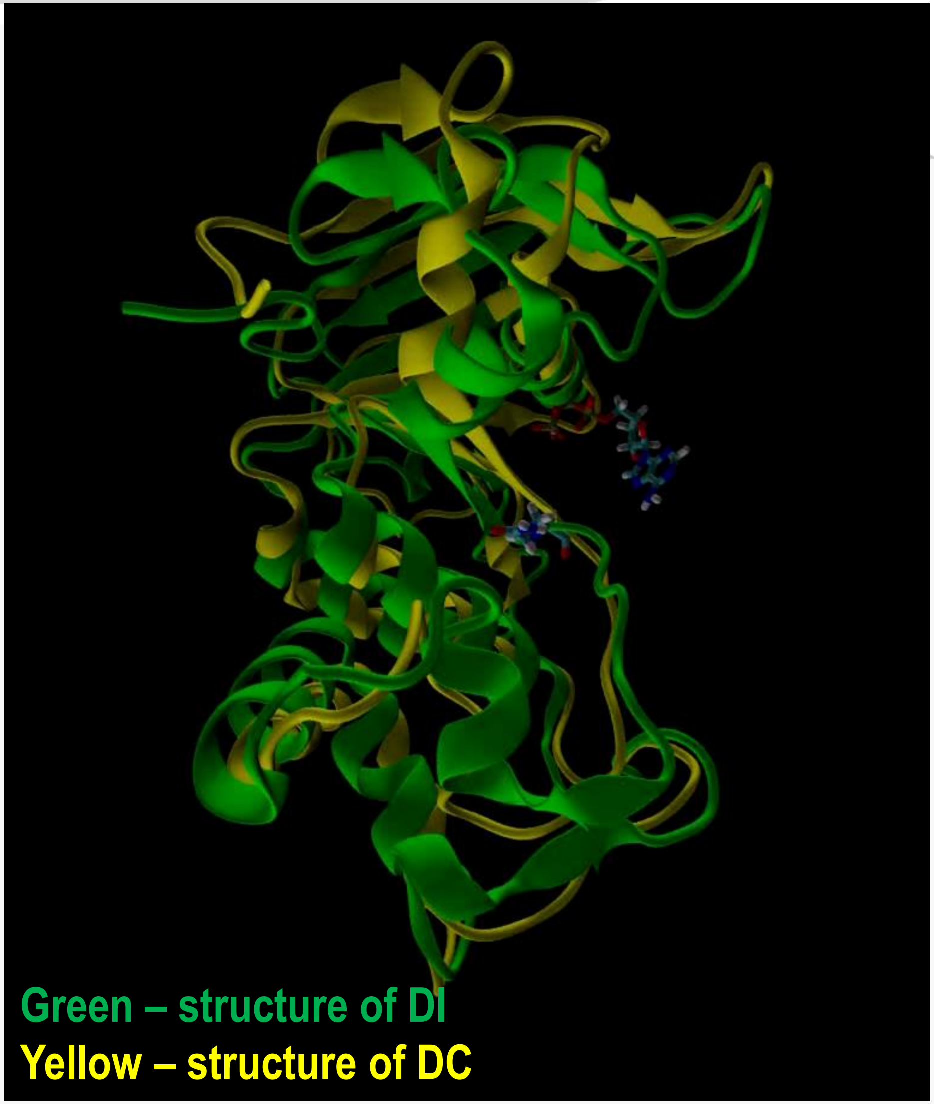


NBD2


Potential of Mean Force


Potential of Mean Force

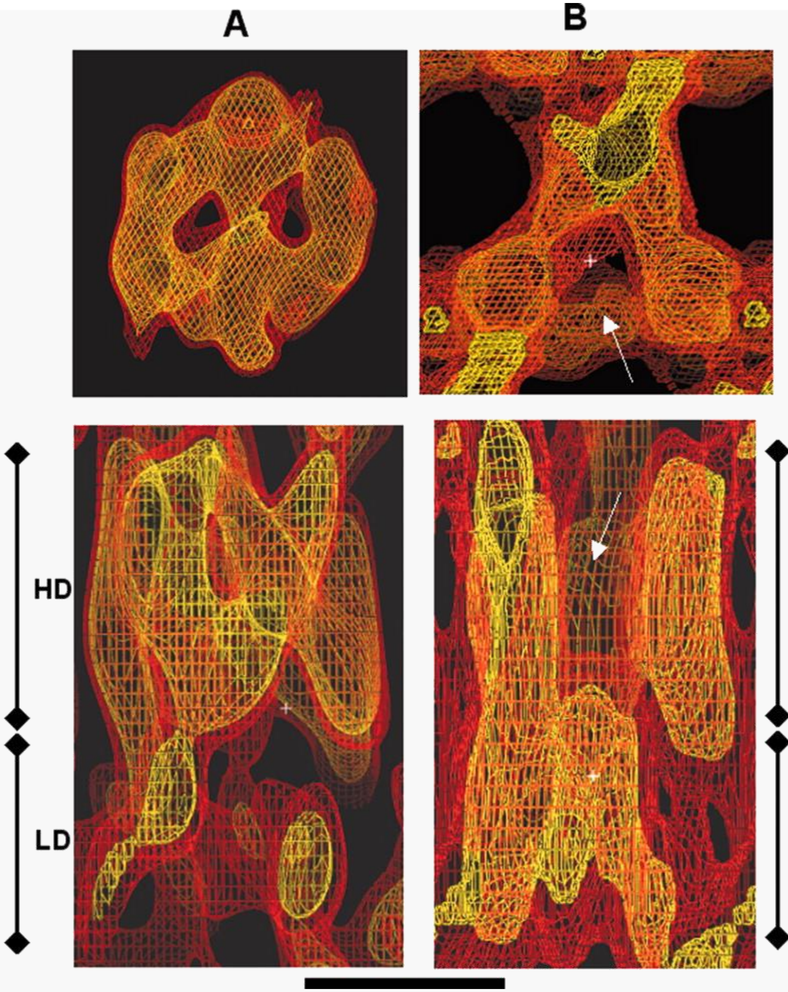
ΔG	Δ (CD-ID) kcal/mol
NBD2	4.5
HylBB	-4



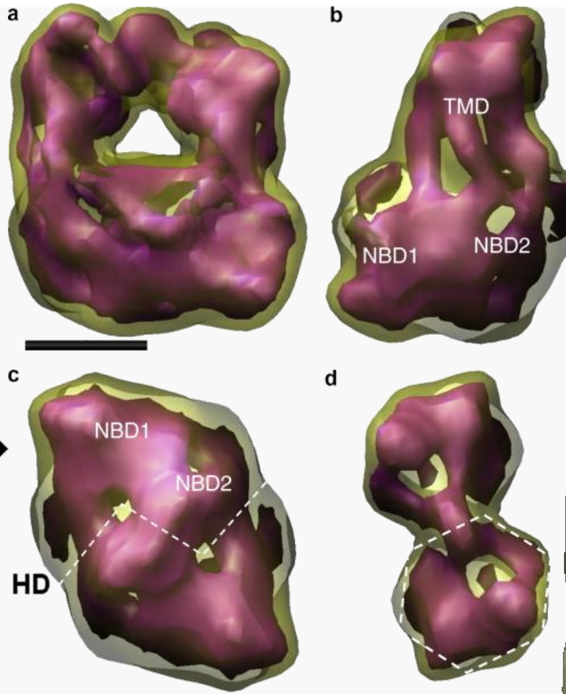
Green – structure of DI

Yellow – structure of DC

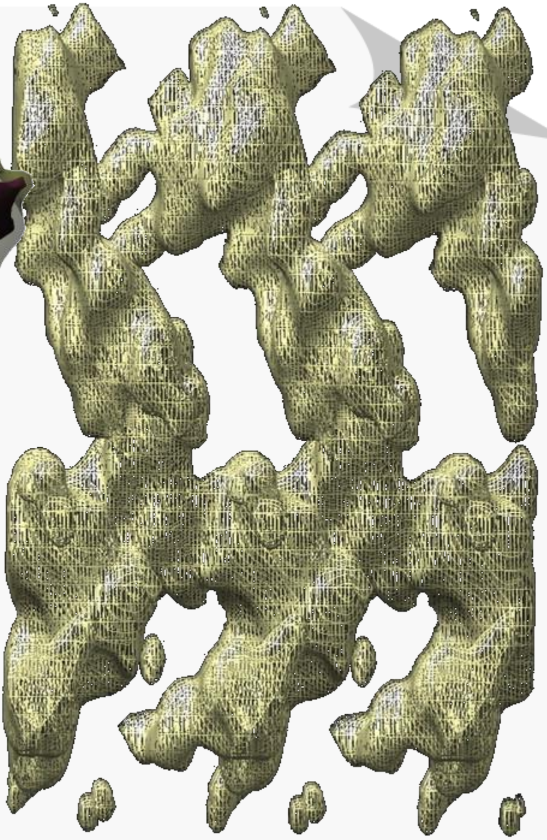
EM Maps of CFTR



Rosenberg et al., JBC 2004, 273, 39051-7



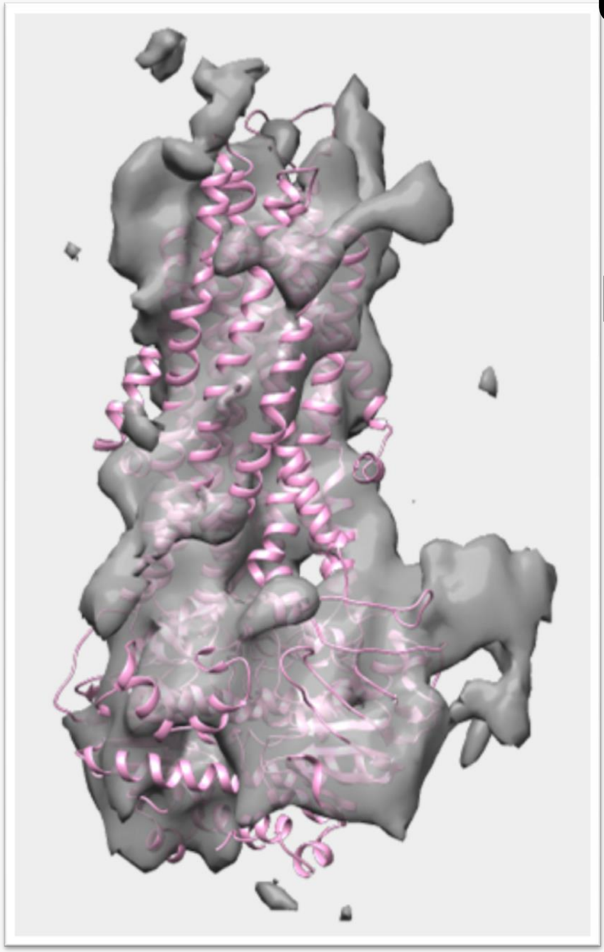
Zhang et al., JSB 2009, 167, 242-251

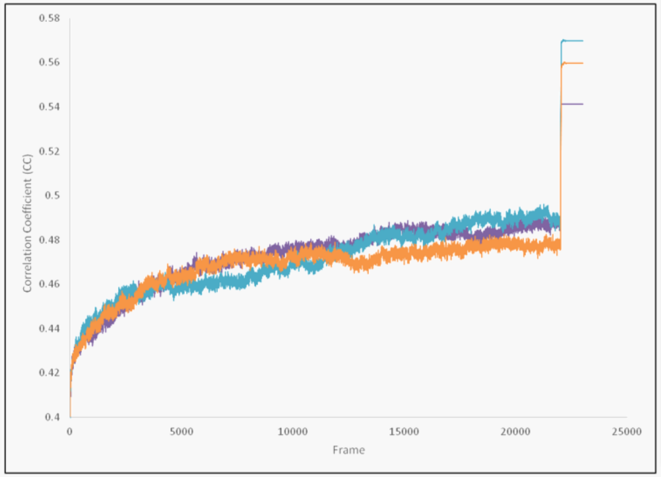


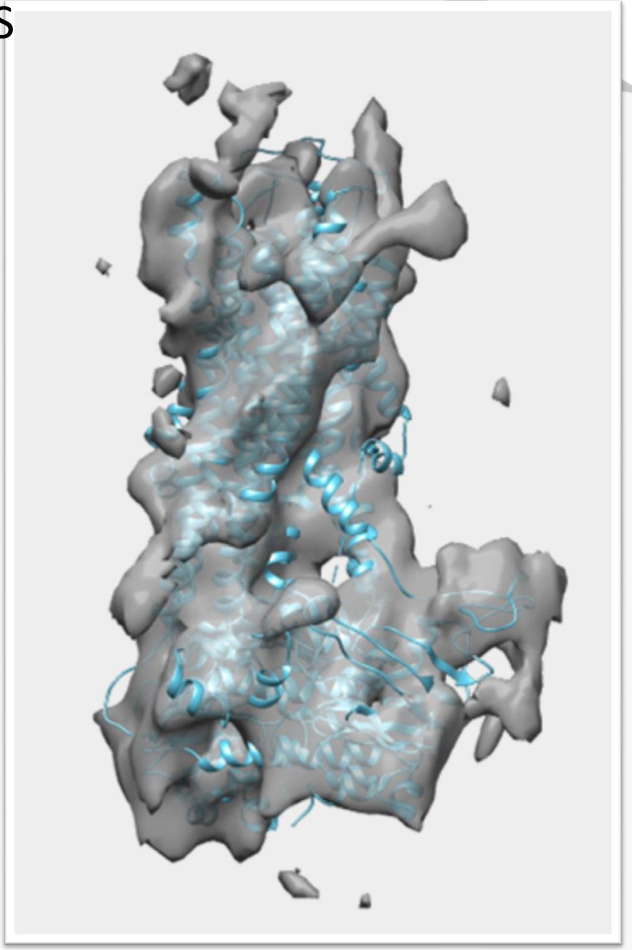
Rosenberg et al., JBC 2011, 286, 42647-54

Map Fitting

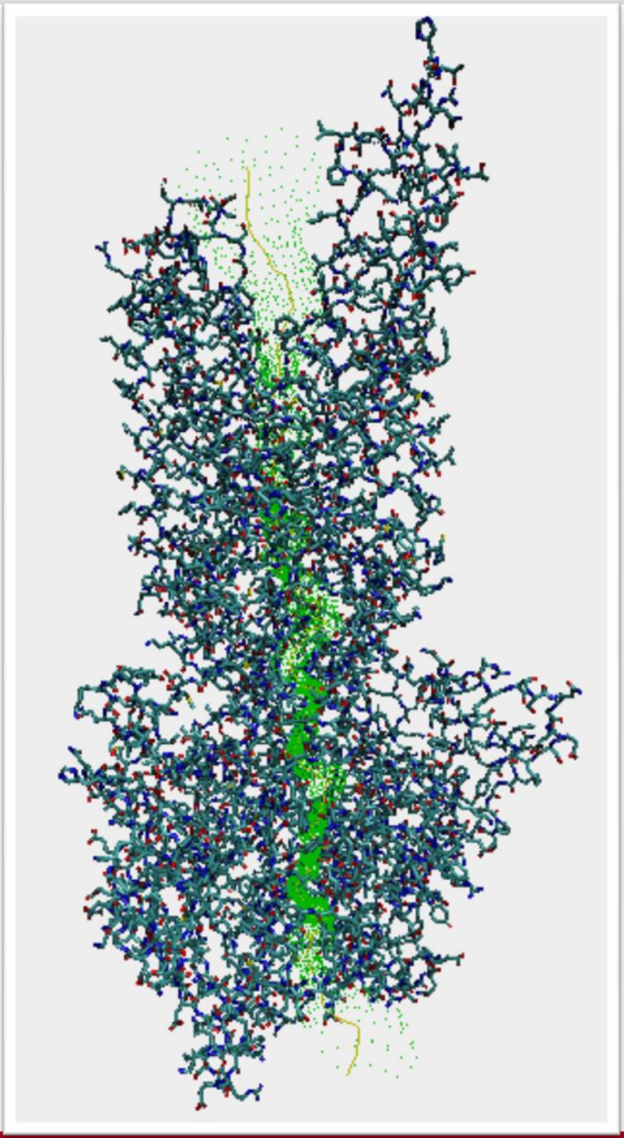
$$U_{total} = U_{MD} + U_{EM} + U_{SS}$$

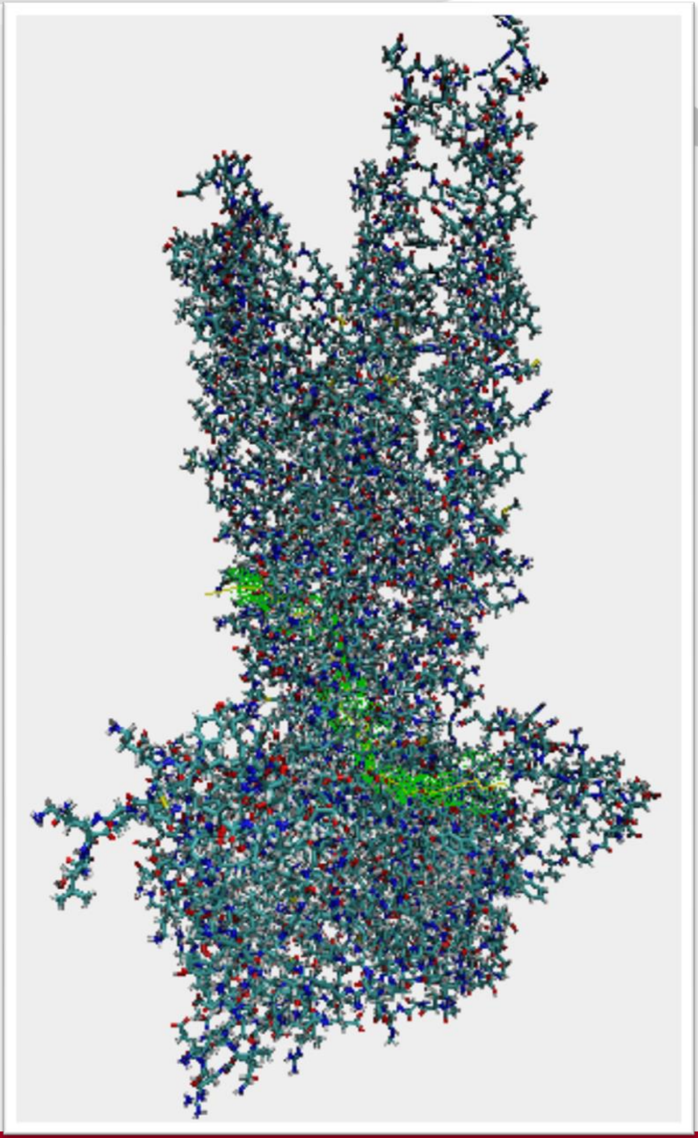




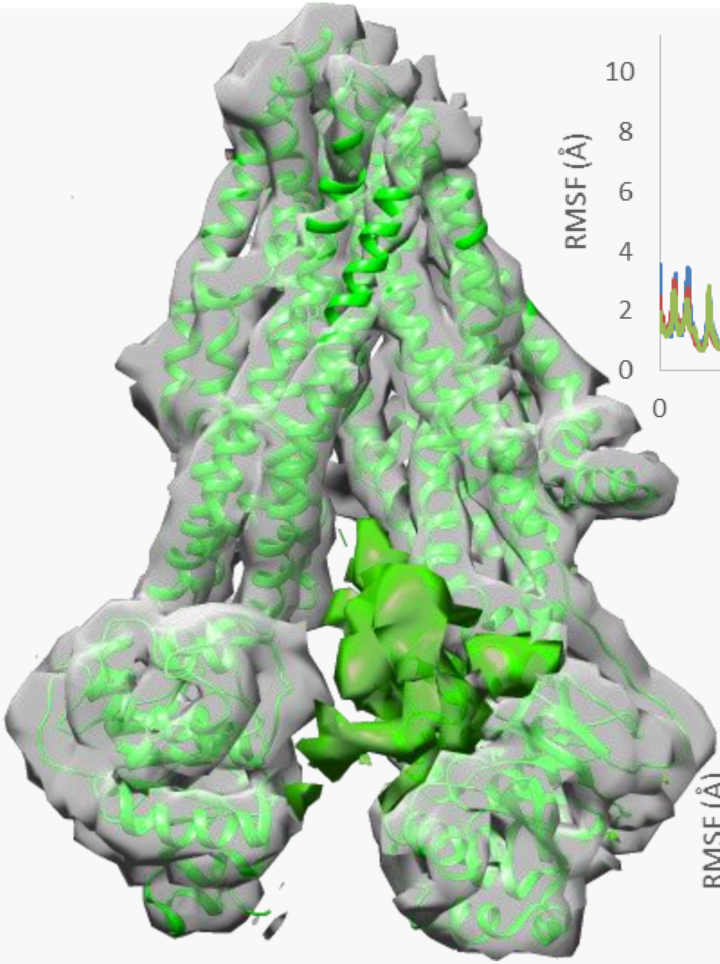


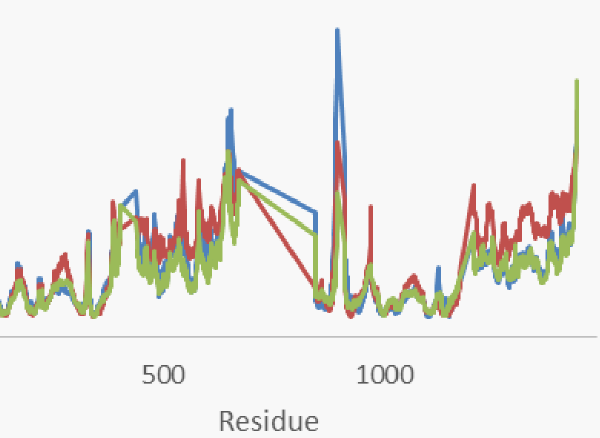
Open vs. Closed Channel

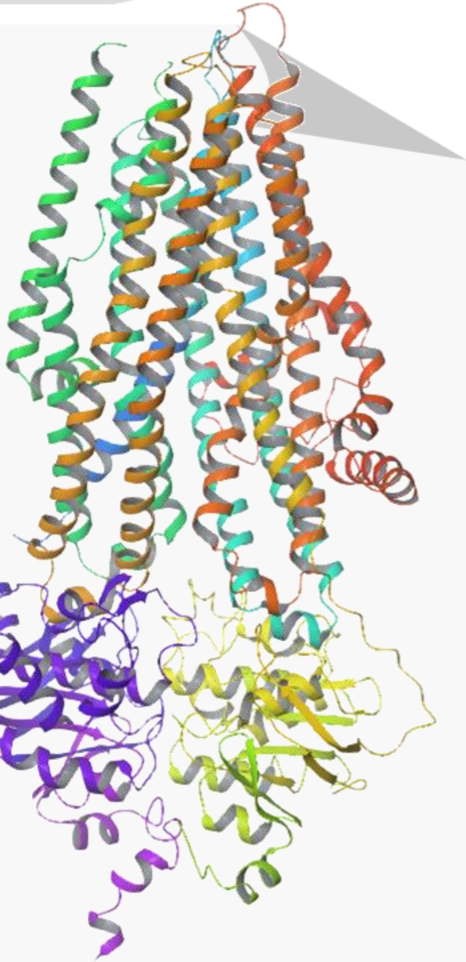


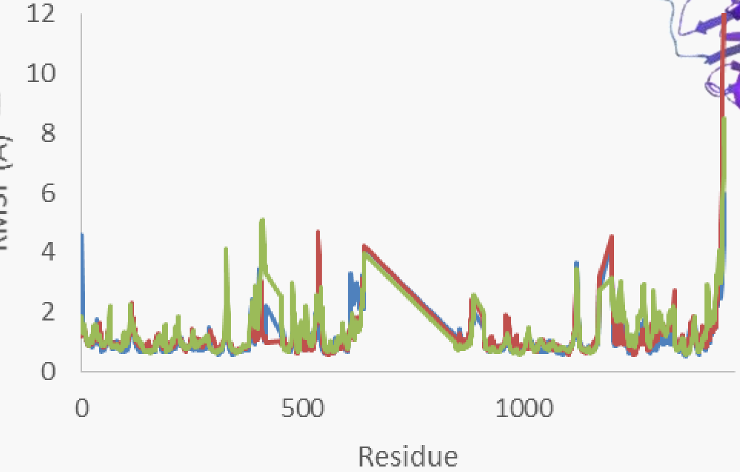


Structure**S** Are Now Available



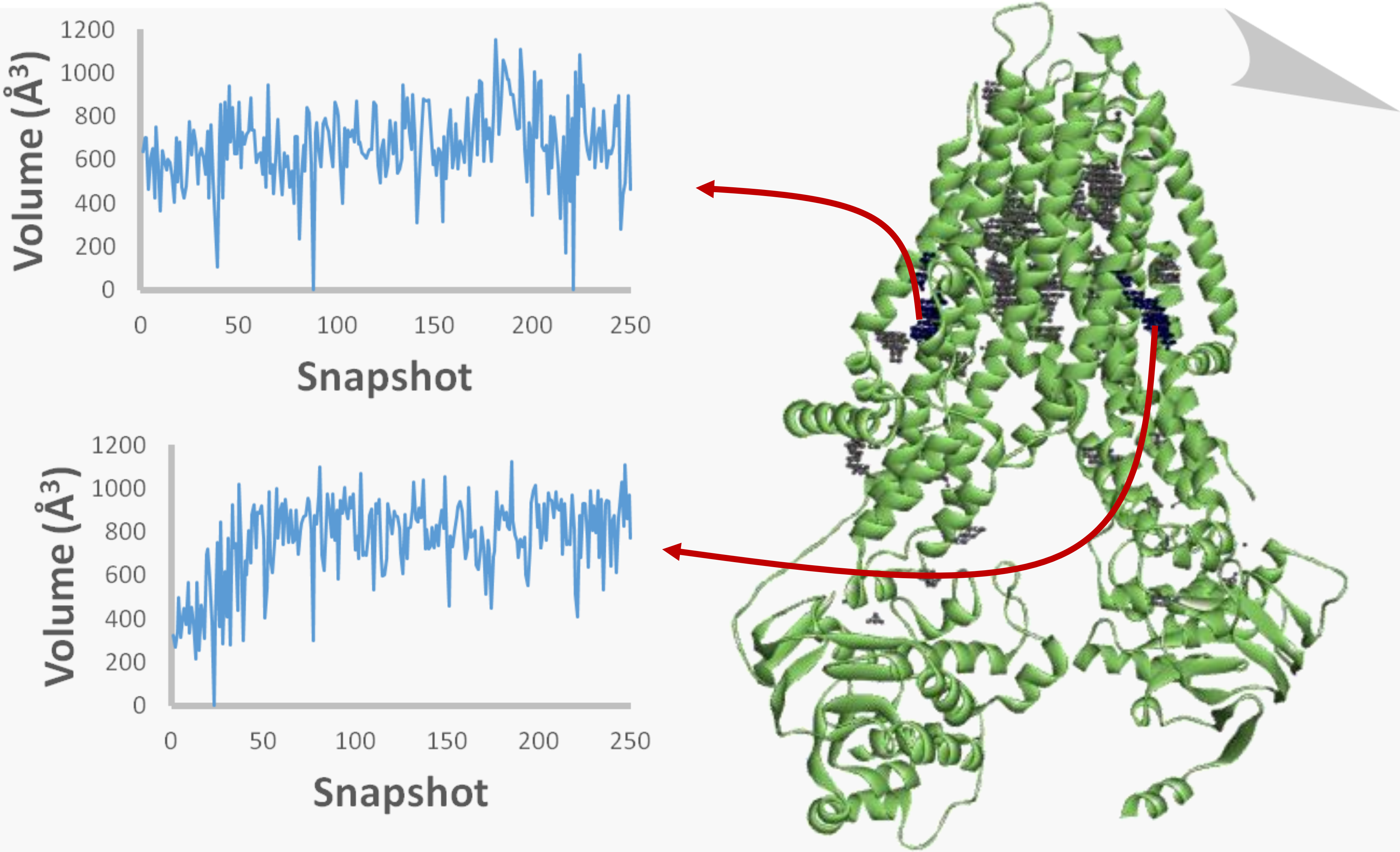




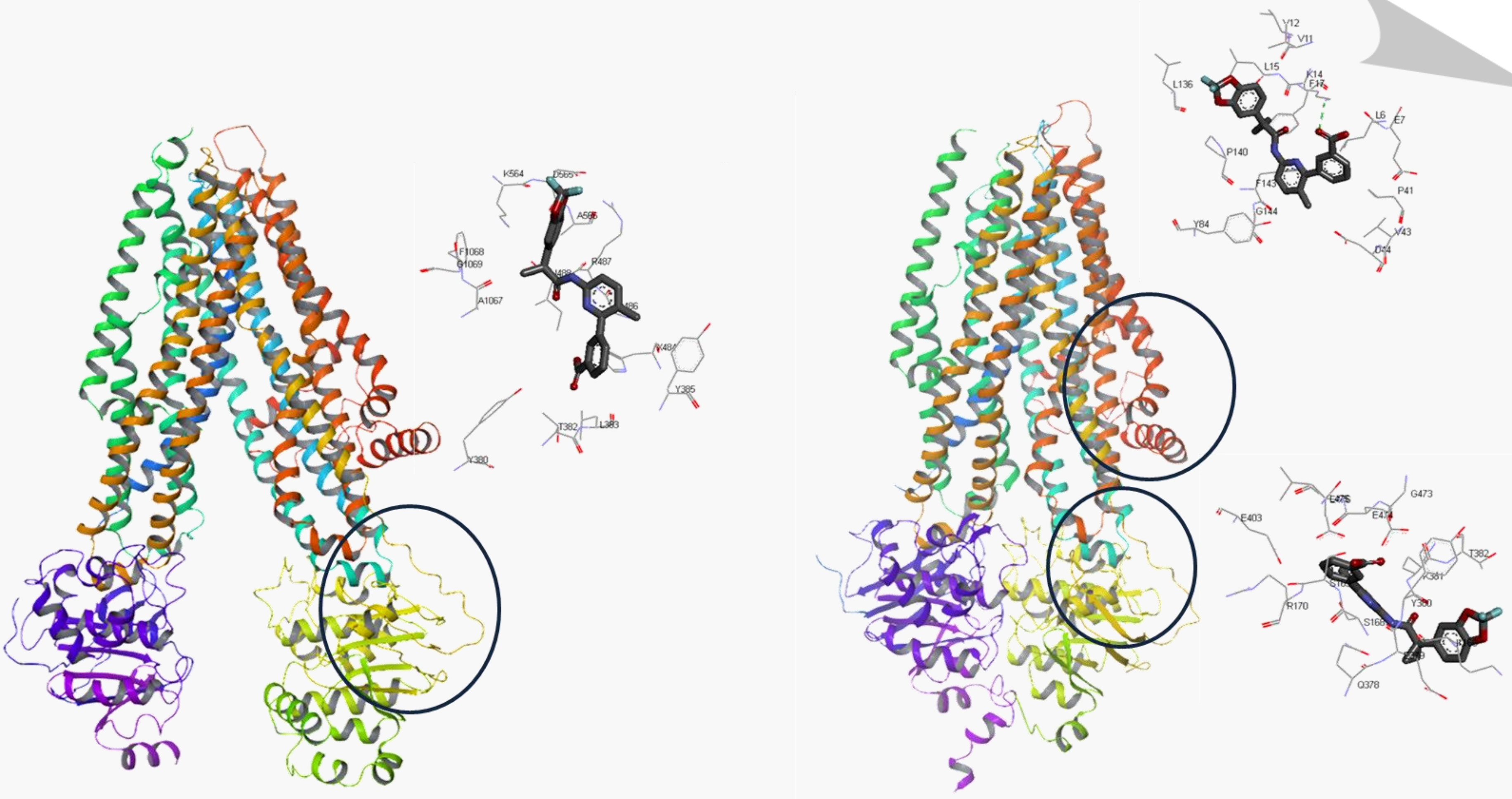


Cell 2016, 167, 1586–1597;
Cell 2017, 169, 85-95;
Cell 2017; 170, 483-491

“Dynamic” Sites



P67L-CFTR: Where VX-809 Binds?



Take Home Messages

- Always look at your data
    - Don’t just rely on numbers
- Its bad practice to deduce anything from a single simulation
    - Results vary and also depend on the simulations setup
- Hold yourself to the same standards you require from experimentalists
    - Multiple repeat, positive and negative controls

Acknowledgments

Group members

- Luba Simachev
- Netaly Khazanov
- Michael Zhenin
- Kobi Spiegel
- Malkeet Singh
- Hadar Binyamin
- Omer Kaspi

Former Group members

- Oren Nahum
- Efrat Noy
- Hannah Avgy
- Gal Fradin
- Tamar Getter
- Shirin Kahremani
- Reut Gigi
- Seema Dhail
- Avi Yosipof
- Yocheved Gilad

Collaborators

- Bilha Fischer (BIU)
- CFFT consortium

arimnet
Coordination of Agricultural Research
in the Mediterranean Area

BARD

ELA
ASSOCIATION EUROPÉENNE
CONTRE LES LEUCODYSTROPHIES

BSF
United States – Israel
Binational Science Foundation

INNI Israel National Nanotechnology Initiative

CYSTIC FIBROSIS FOUNDATION
THERAPEUTICS