# Drug Discovery, Docking, and GPCRs

Magnus Andersson

*magnus.andersson@scilifelab.se*

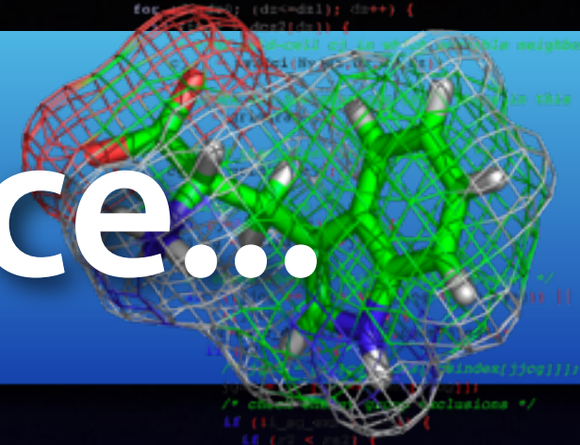
Theoretical & Computational Biophysics

SciLifeLab

KTH VETENSKAP OCH KONST

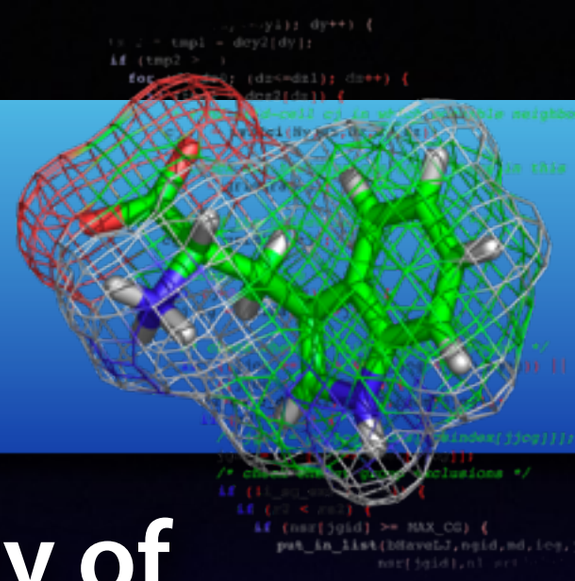
UNIVERSITET STOCKHOLMS

# The final puzzle piece...

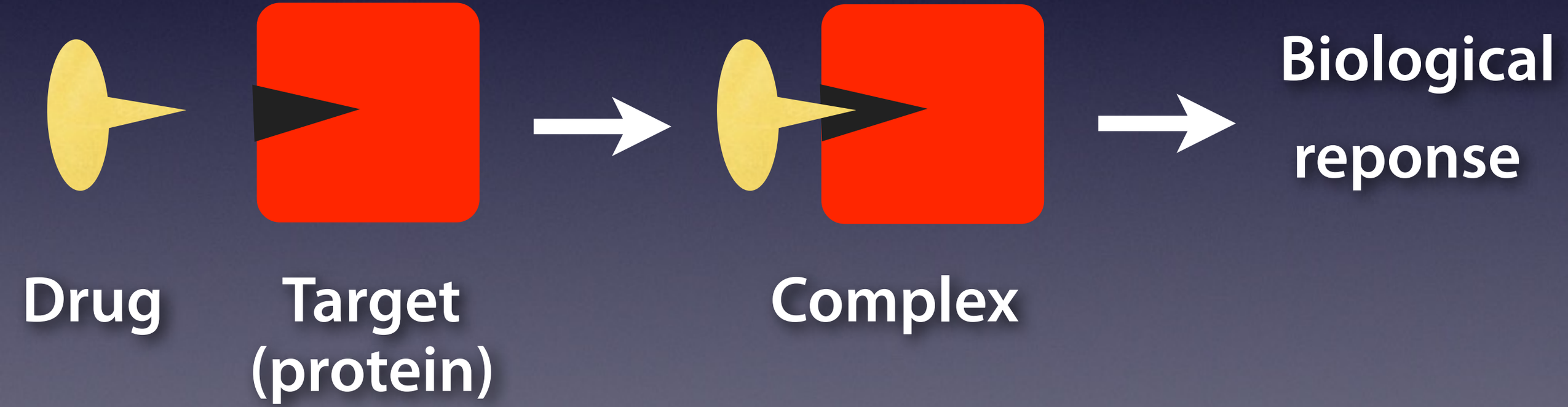


- From what we've learnt this far, we can:
  - Start from a sequence and try to predict the fold by homology
    - Use force field-based methods to build side-chains and missing structure parts
    - Energy minimize the structure
    - Simulate models of structure
- But can we predict binding of ligands/drugs?

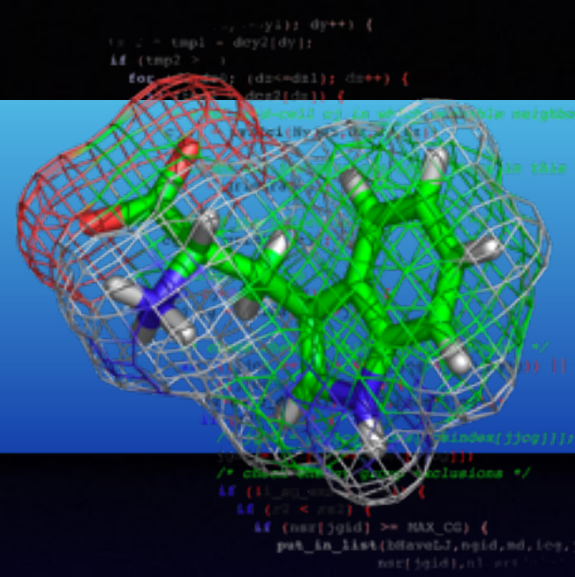
# What do drugs do?

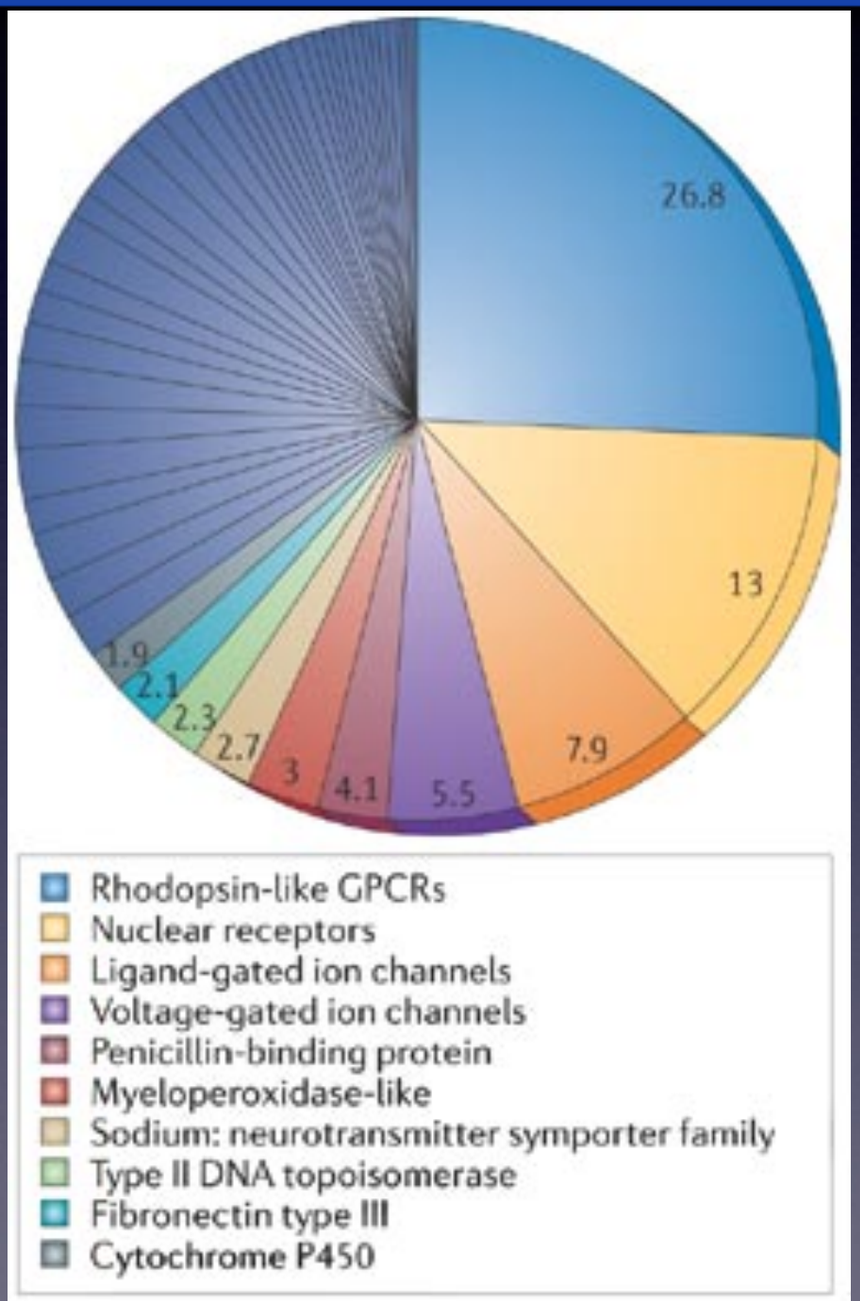


- Small molecules that modulate the activity of proteins



# What do drugs hit?

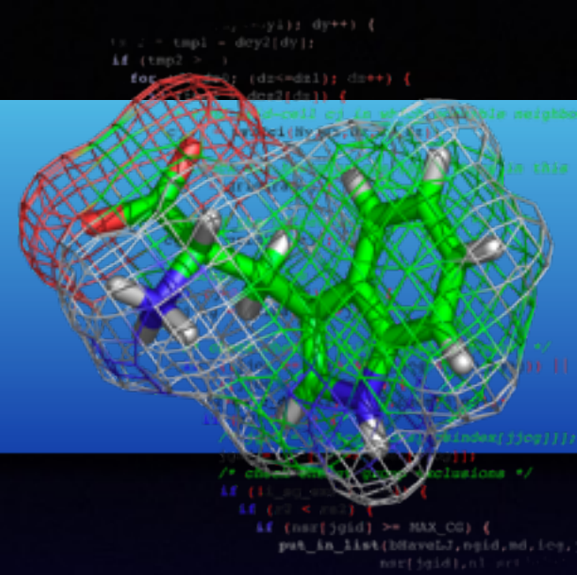




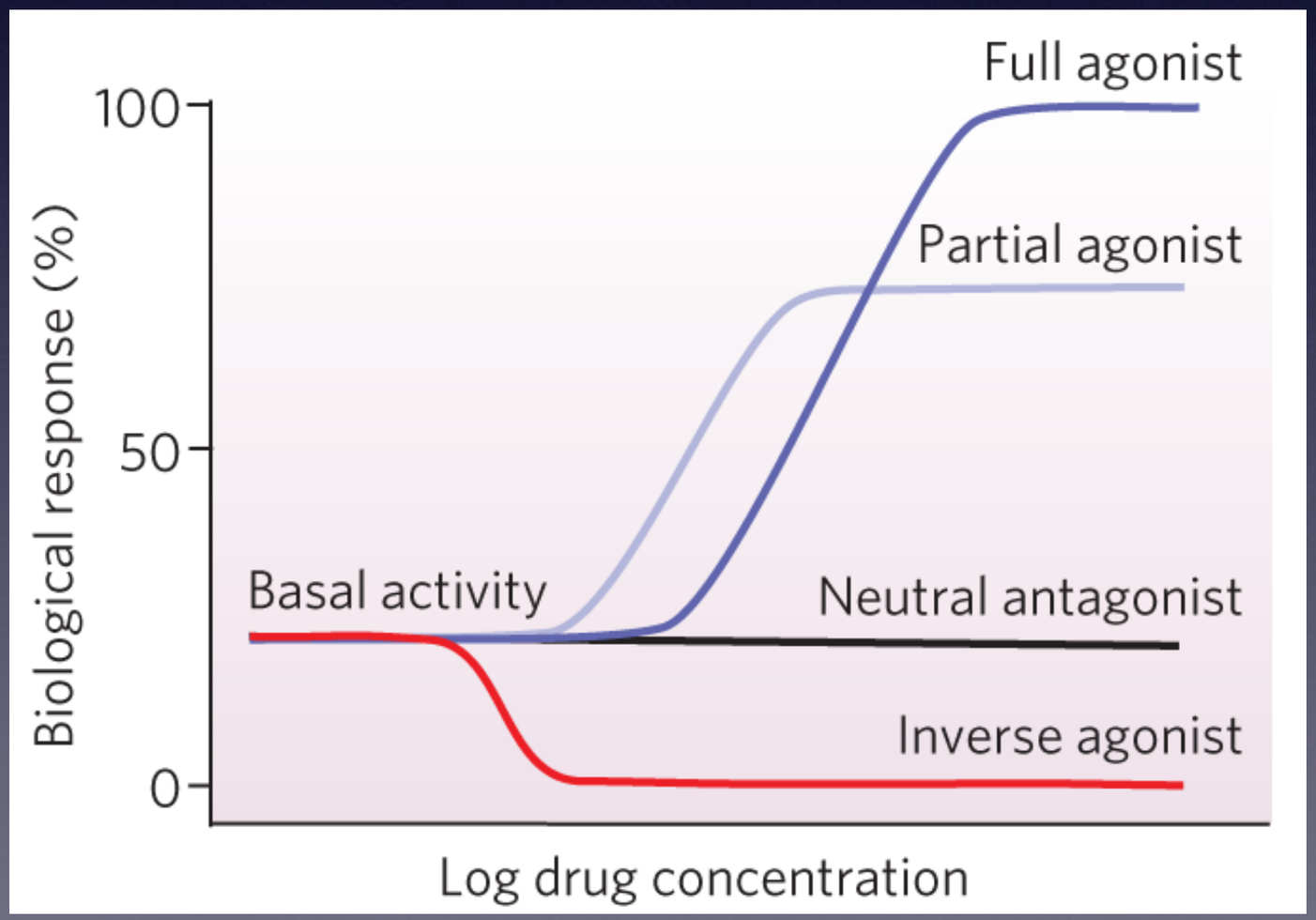



- **G-Protein Coupled Receptors: 27%**
  - Recognizes (e.g.) neurotransmitters Example: Antipsycotics/ Dopamine receptor
- **Nuclear receptors: 11%**
  - Transcripton factors Example: breast cancer drugs / Estrogen receptor
- **Ion Channels: 12%**
  - Ligand-gated/Voltage-gated. Example: Anaesthetics / GABA receptor

# What do drugs do?



- Agonists, Partial agonists, Antagonists (inhibitors), Inverse agonists

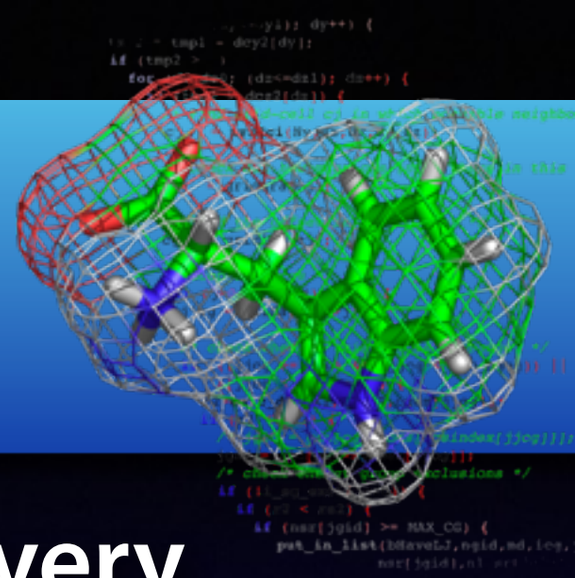


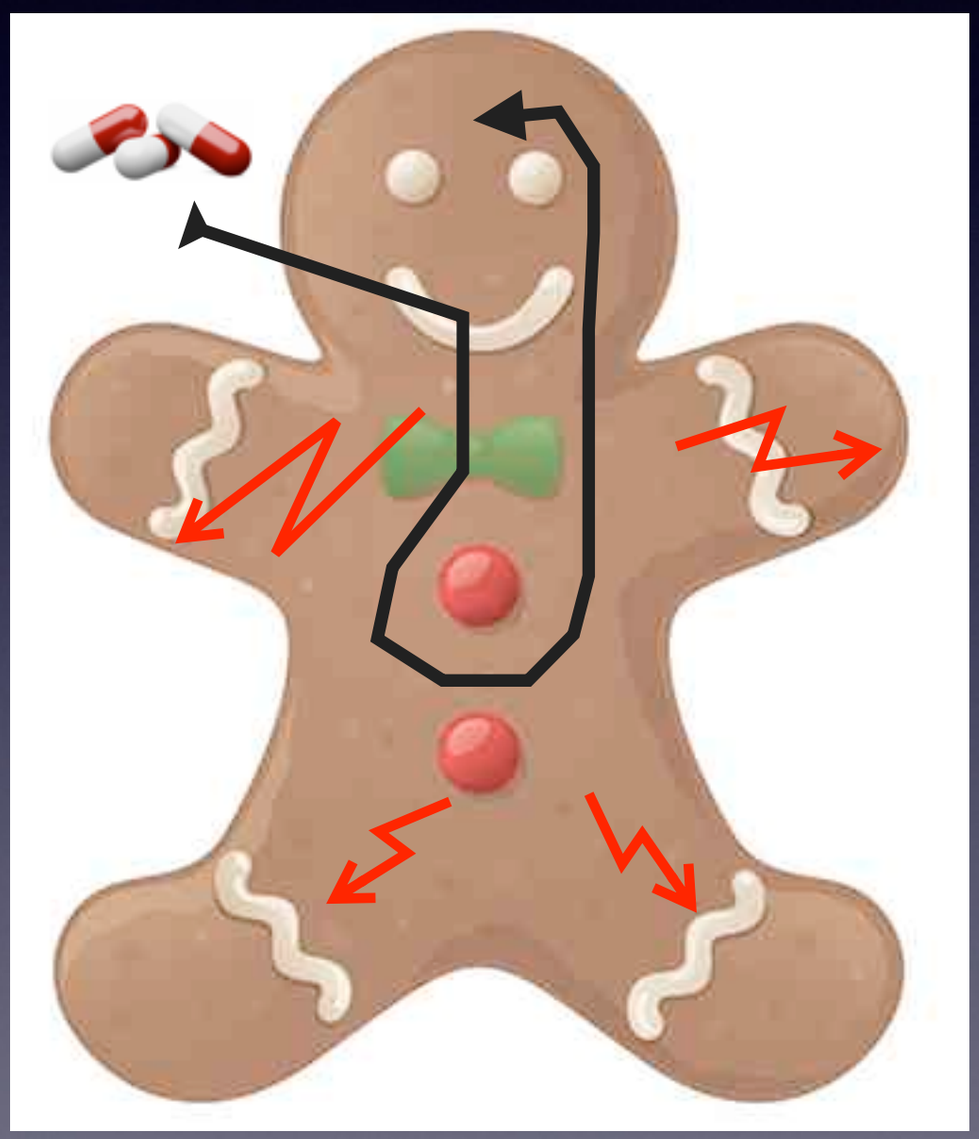

Activate

Inhibit

Deactivate

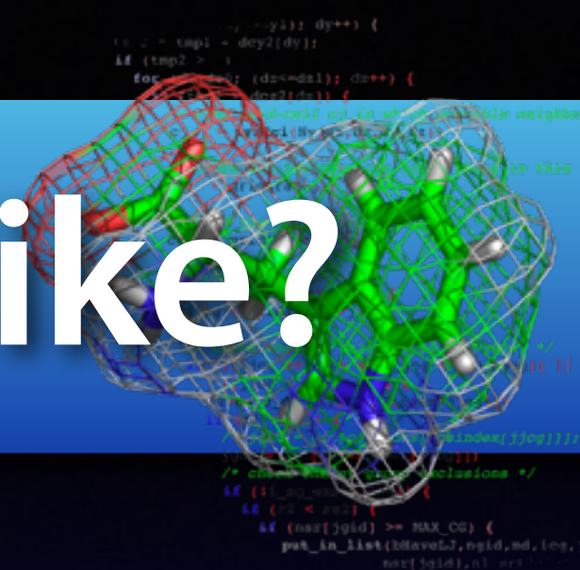
# Biology is complex...





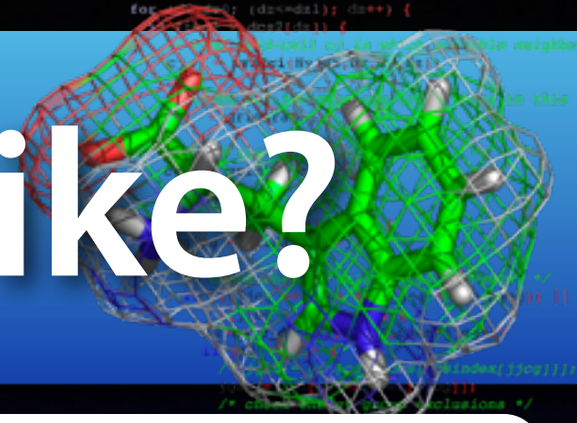
- Why is drug discovery difficult?
  - Compound must bind to the target (protein).
  - Compound must not bind to other targets (side effects).
  - Compound must survive from administration to target (e.g. brain).

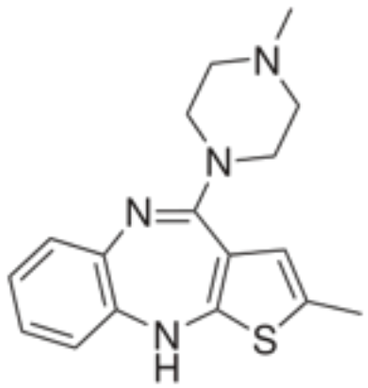
# What do drugs look like?



- Lipinski's rule of Five: (ADME)
  - MW < 500
    - Small enough to get transported
  - LogP < 5
    - Polar enough to get into blood stream
  - < 5 hydrogen bond donors
  - < 10 hydrogen bond acceptors
    - Non-polar so that it can cross membranes.

# What do drugs look like?

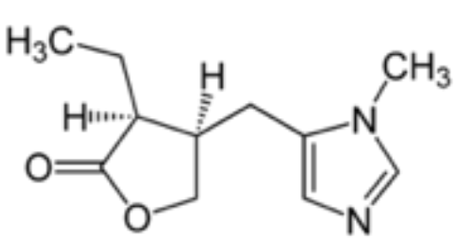




**Olanzapine:**

Indication: Schizofrenia

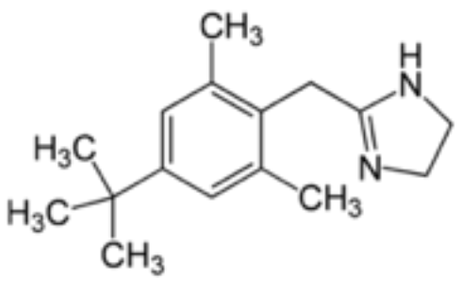
Target: Dopamine, Serotonin, Muscarinic, adrenergic, histamine receptors (antagonists, polypharmacology).



**Pilocarpine:**

Indication: Glaucoma

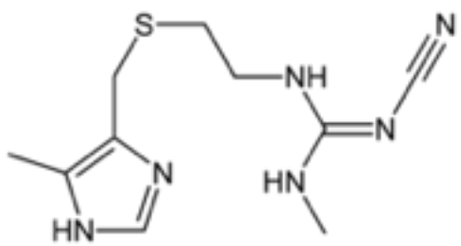
Target: Muscarinic receptors (agonist)



**Xylometazoline:**

Indication: Nasal decongestant

Target: alpha adrenergic receptors (agonist)



**Cimetidine:**

Indication: Ulcer & heartburn

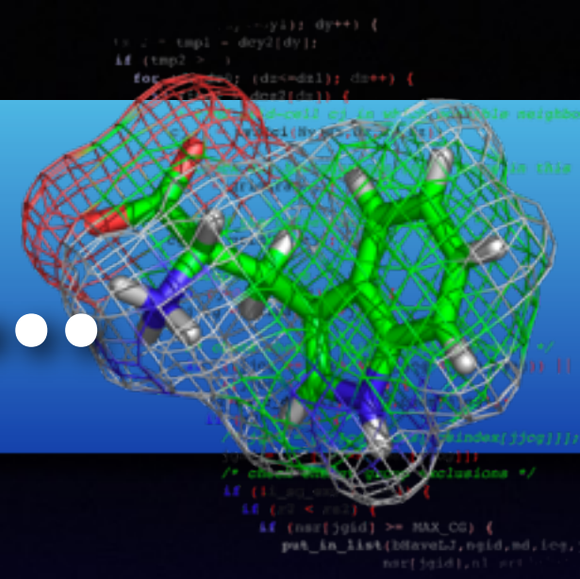
Target: H2 histamine receptor (antagonist)

# “Modern” drug discovery

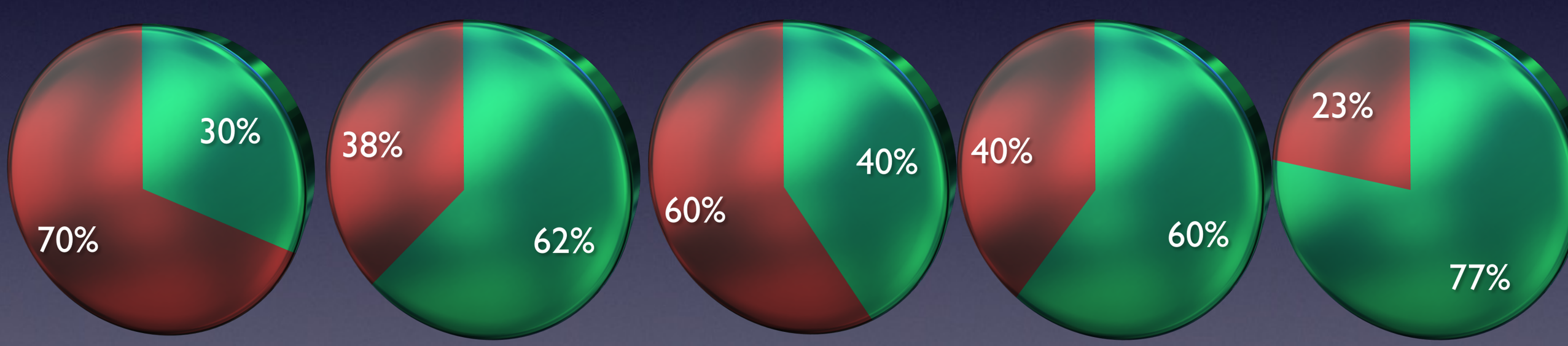
- Target identification: What do we want to hit?
- Pre-clinical:
  - Finding something to start with (hit)
  - Does it seem to have any effect?
  - Optimize affinity/efficacy (lead)
  - *in vitro* tests, animal tests
- Phase I: Is it safe in humans?
- Phase II: Is it efficient in humans?
- Phase III: Efficacy compared to alternatives

Increasing cost

# Failure is the norm…



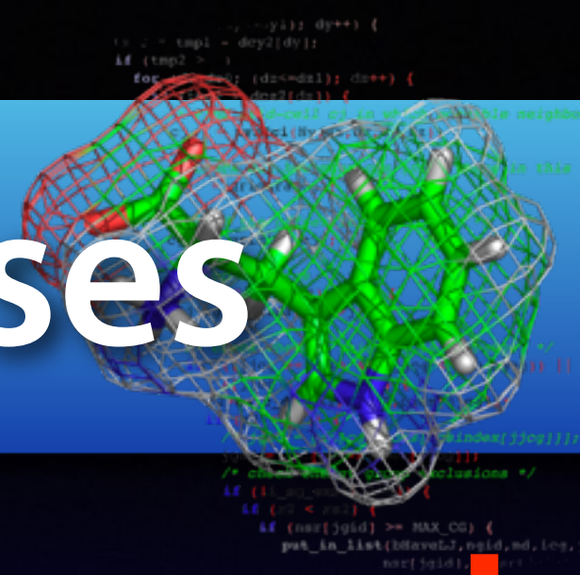
Preclinical Phase I Phase II Phase III FDA




Only ~3% of drug discovery projects make it all the way

**Fail cheap - Fail early**

# Development phases

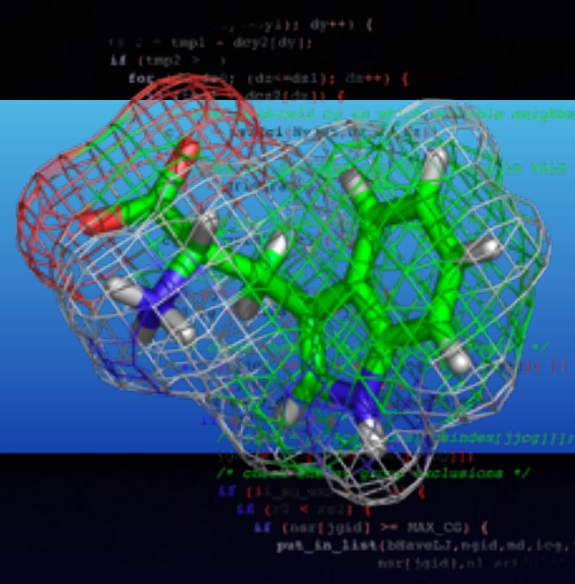


- Target identification:
  - What do we want to hit?
- Pre-clinical:
  - Finding a hit.
  - Does it seem to have any effect?
  - Optimize affinity/efficacy (hit to lead)
  - vitro tests, animal tests
- Phase I: Is it safe in humans?
- Phase II: Is it efficient in humans?
- Phase III: Efficacy compared to alternatives

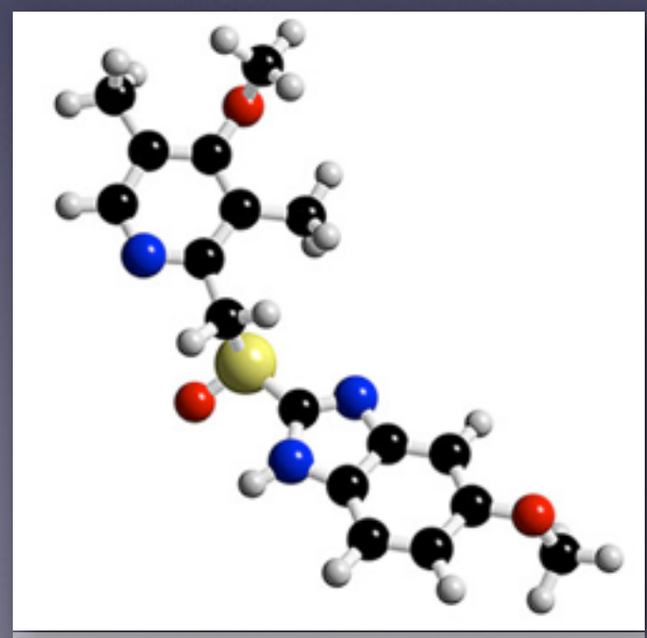
Computational tools?
Protein structure?

Increasing cost

# Lead discovery

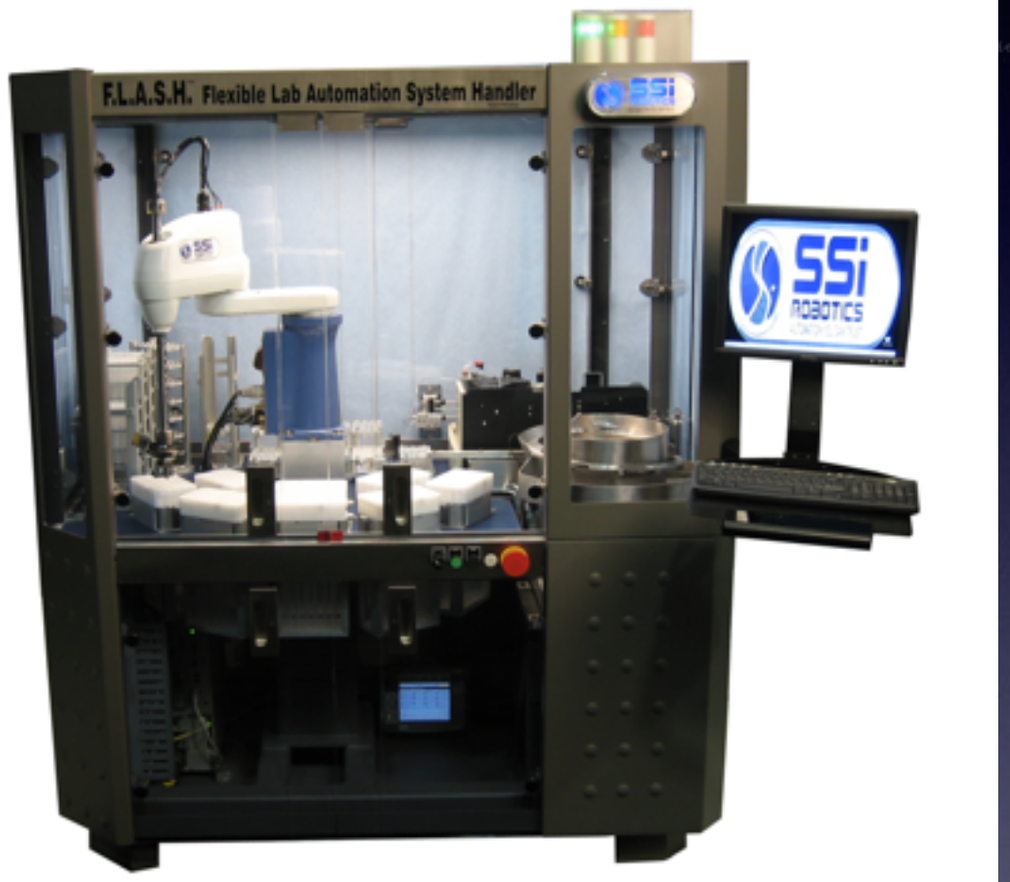


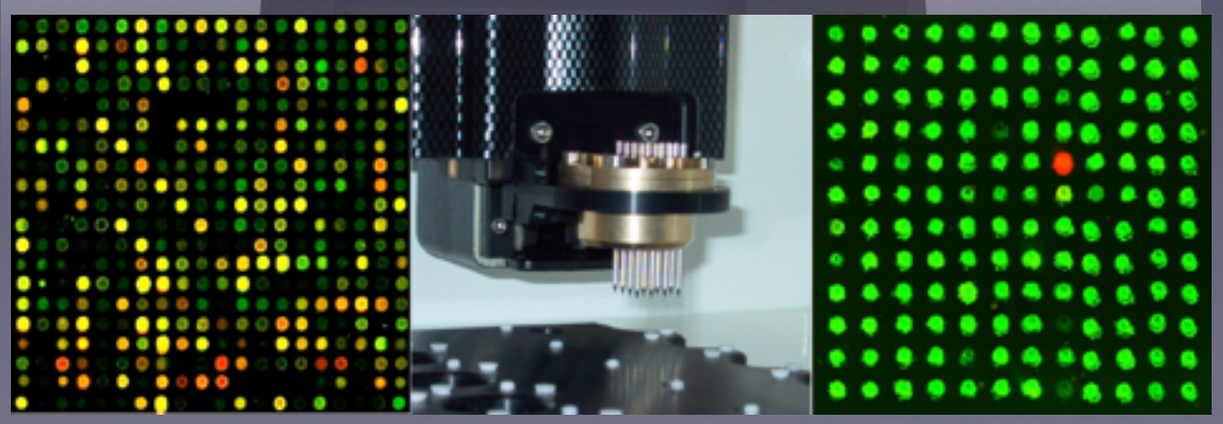
- Identify ‘hit’ molecules with *measureable* activity in a biological assay.
  - High-throughput Screening
  - QSAR/Pharmacophore modeling
  - Molecular docking screening (Virtual screening)



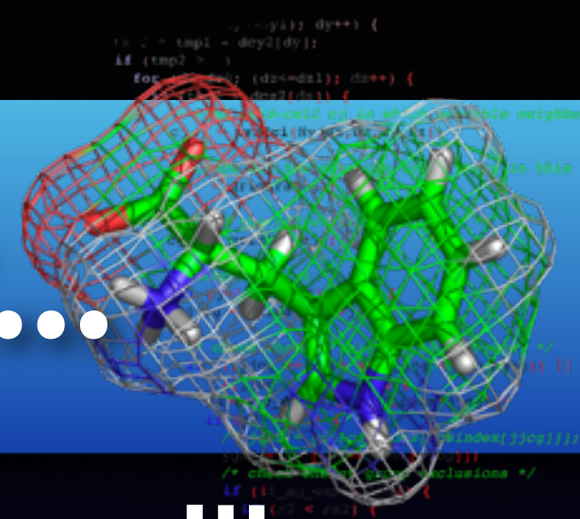
# High Throughput Screening

- ~150,000 tests/day
- Screen up to 1,000,000 different compounds
- Dozens to thousands of target molecules
- If lucky, 100 leads
- Cost: ~$1/well
- Expensive
  Good assay necessary



# HTS shouldn't work...

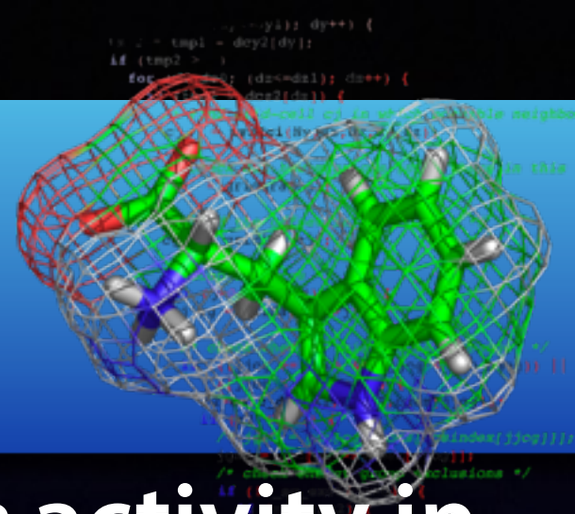


- **Chemical space contains ~$10^{60}$ drug-like molecules: The probability of finding a ligand by random screening $10^{6}$ of these is negligible!**
- **... but sometimes it does! Two examples:**

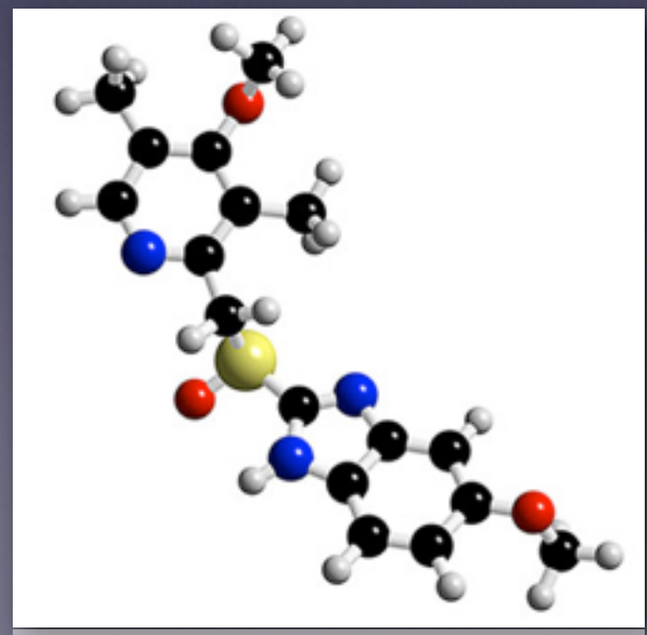
| Target | #cmpds | Exp. hits | Docking hits |
|---|---|---|---|
| Lactamase | 300,000 | 0 | ? |
| Cruzain | 200,000 | 146 | ? |

- **Conclusions: Very few hits in HTS screening!**

# Hit discovery



- Identify 'hit' molecules with *measureable* activity in a biological assay. Alternatives:
  - Serendipity (luck), e.g. Penicillin & Viagra
  - Litterature/Patents - has the target been studied before?
  - High-throughput Screening
  - QSAR/Pharmacophore modeling
  - Molecular docking screening (Virtual screening)



# Experimental vs Computational Screening

vs

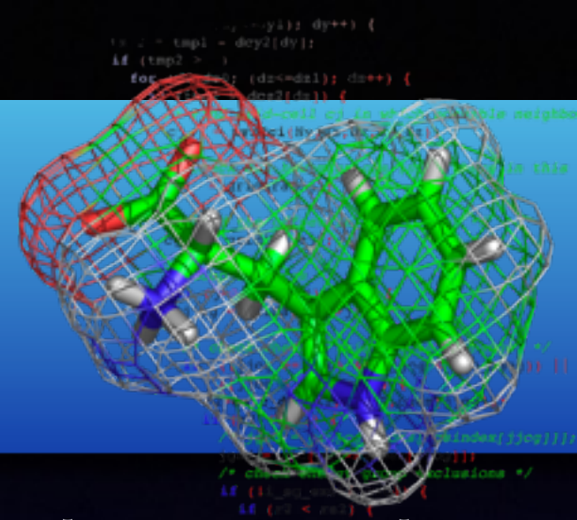

**Experimental screening:** Expensive & slow

| | |
|---|---|
| NIH screening center, NCGC | $3\times10^5$ |
| AstraZeneca | $10^6$ |

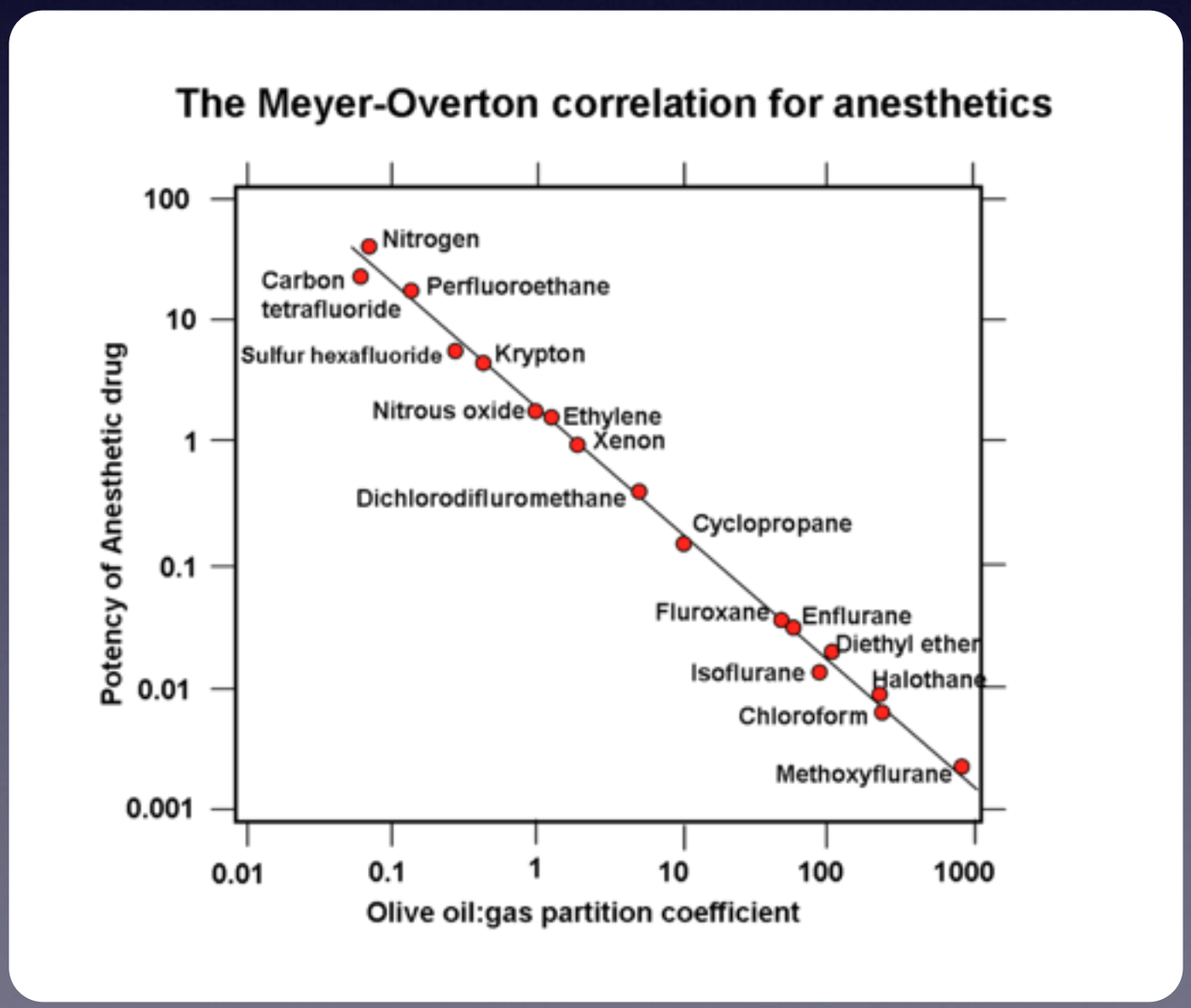
**Computational screening:** Cheap,fast,accurate?

| | |
|---|---|
| QSAR, Docking | ~ $10^6$-$10^9$ |

# QSAR

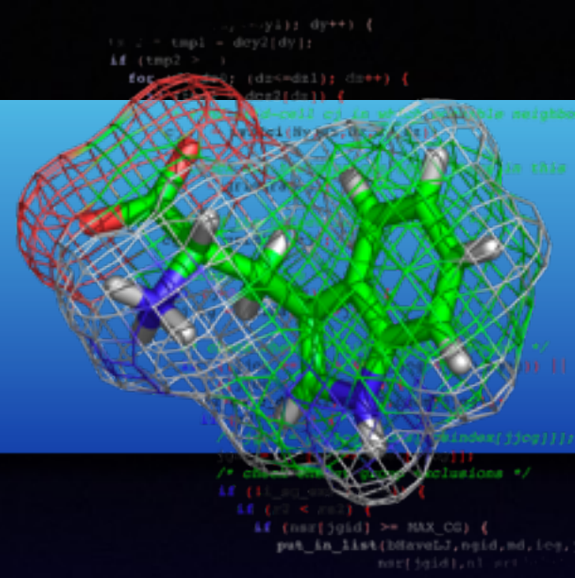


- Correlate biological activity with simple chemical properties.



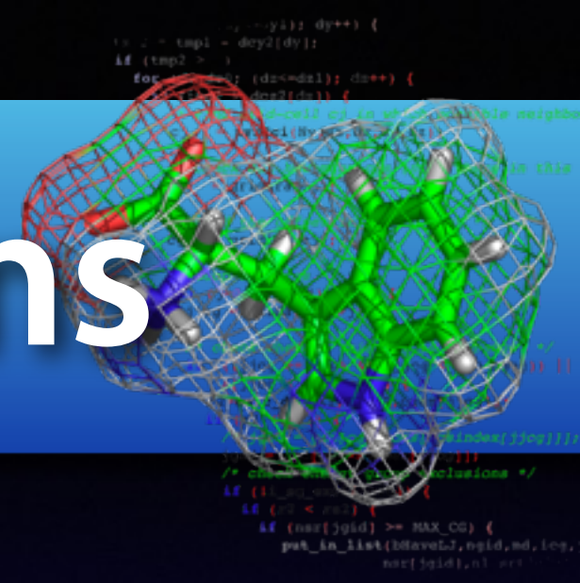

- Remarkable correlation between partition coefficient and anesthetic potency!

# QSAR



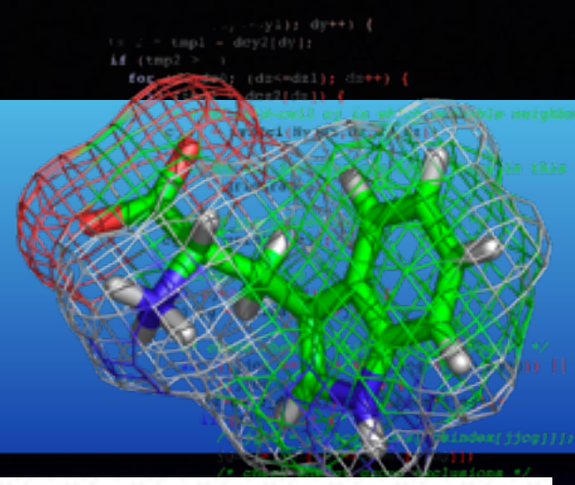
- Quantitative Structure-Activity Relationship
- Fast(!) and simple approach: Correlate affinity with “all sorts of properties”
  - Molecular weight
  - Charge, dipole moment, surface
  - Partition coefficients (water/octanol)

# QSAR: Pros and Cons

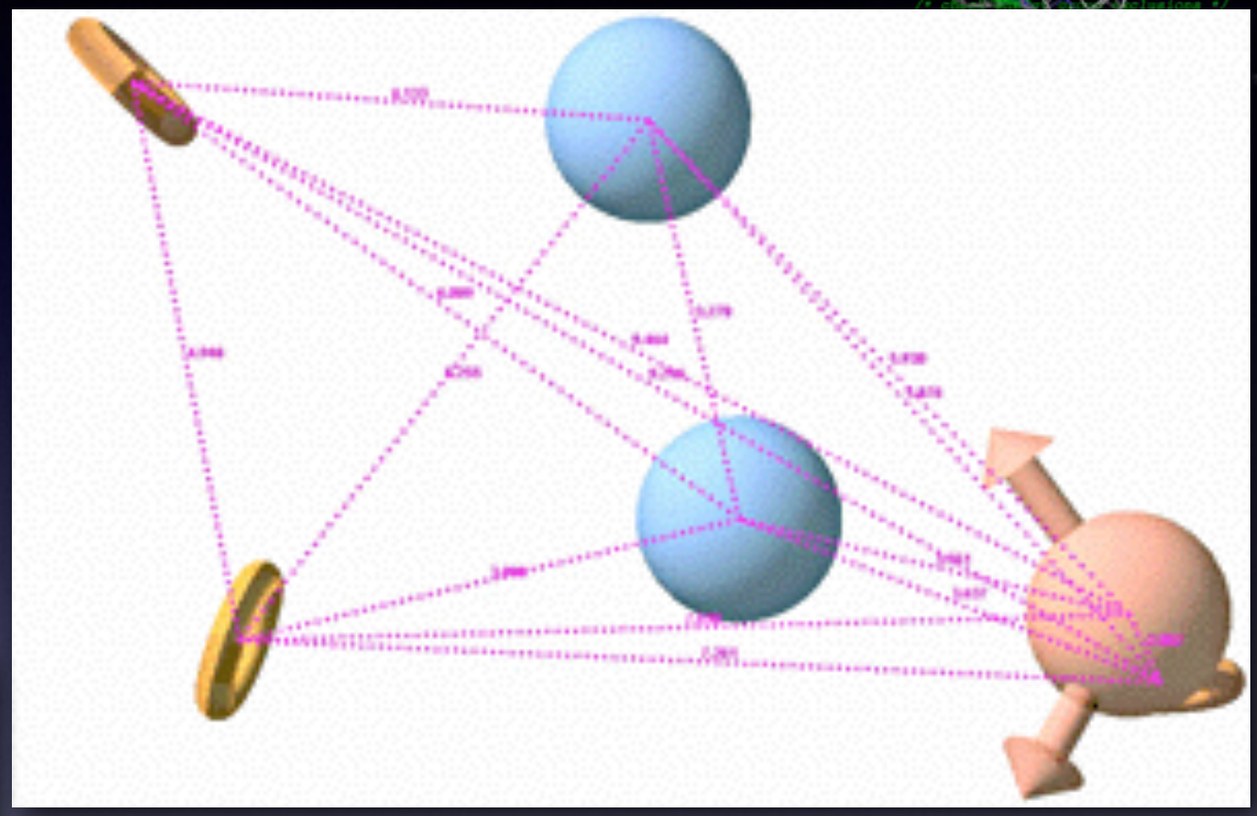


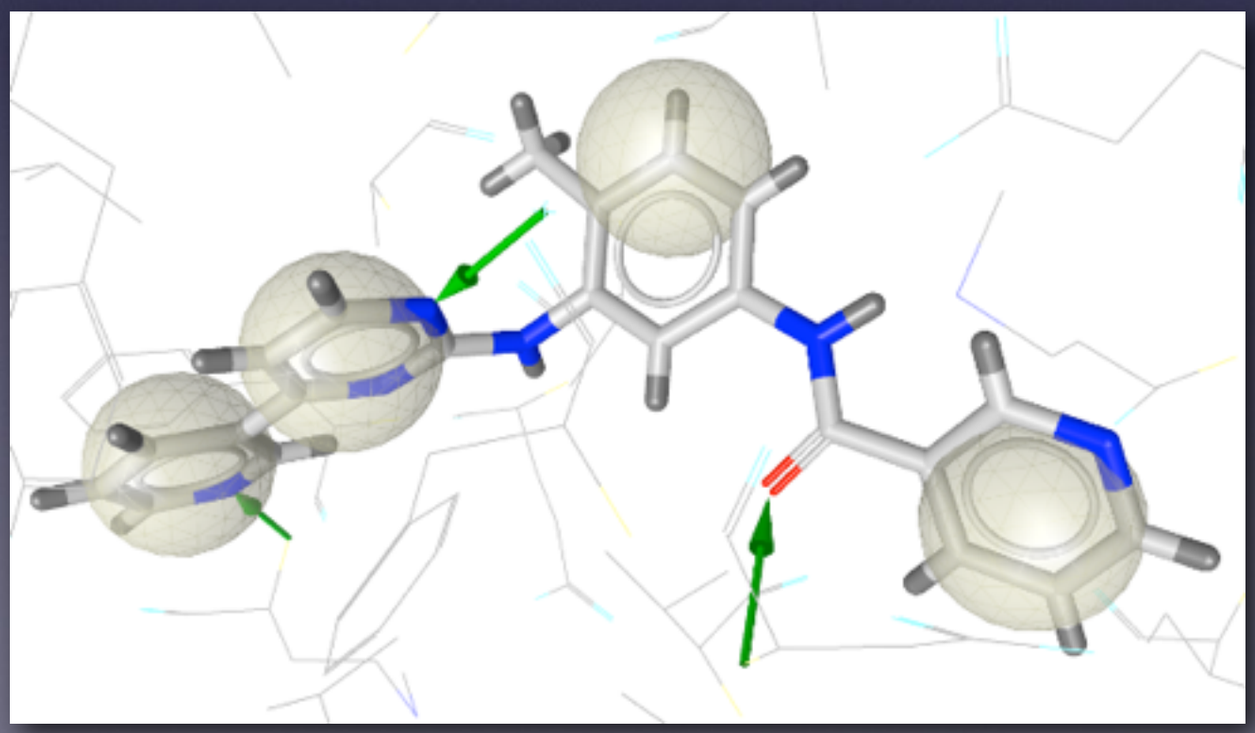
- **Advantages**
  - Very fast to screen large chemical databases.
  - The method works quite well - finds ligands!
- **Disadvantages:**
  - If you make a very specific model, you will only discover what you already know.
  - Flexible molecules require that you identify the “correct” conformation to base the model on.
  - What if the molecules bind to different sites?
  - Important to include non-binding molecules in model building too.

# Pharmacophores



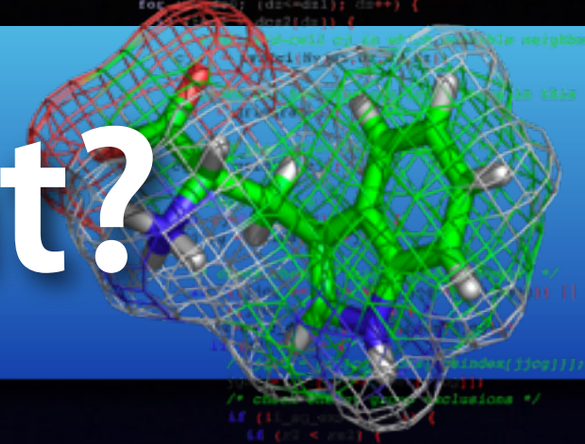
- Given a lead series, characterize properties computationally
- Distances between important groups



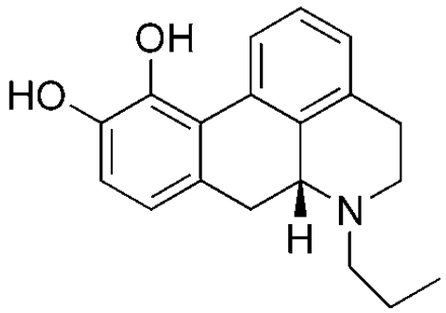


- Find patters
- Refine properties
- Screen databases

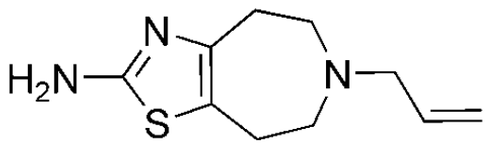
# Common element?



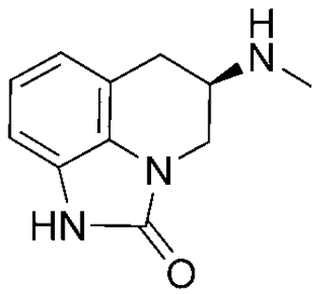
Full agonists



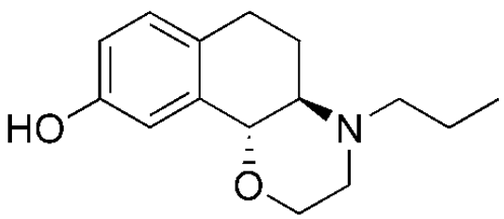
(*R*)-NPA (**1**)



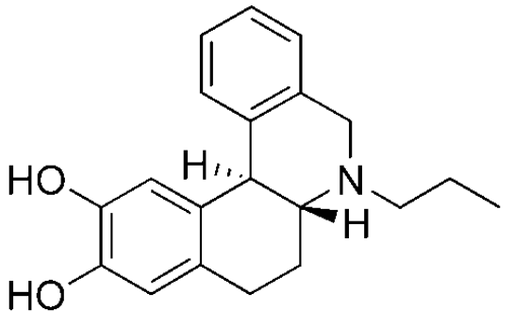
talipexole (**2**)



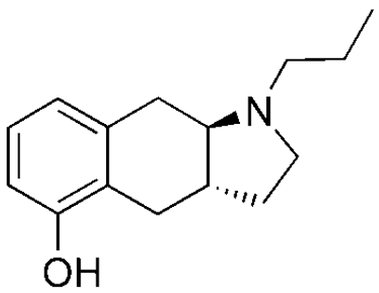
sumanirole (**3a**)



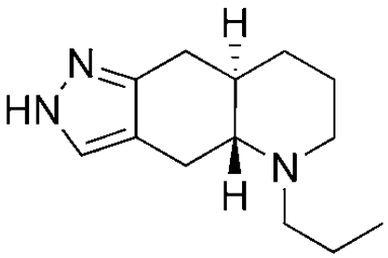
(*R*,*R*)-PHNO (**4a**)



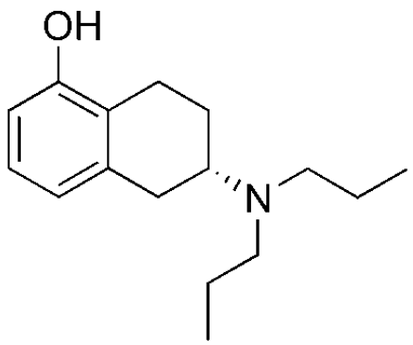
*n*Pr-DHX (**5**)



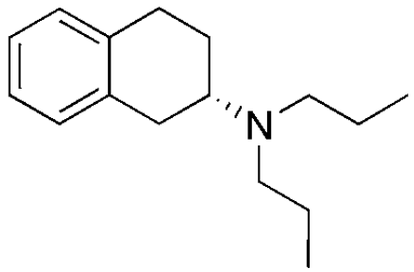
(–)-(3*S*,9*R*)-**6a**



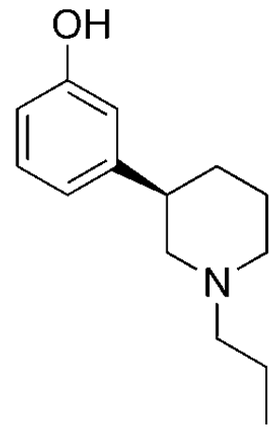
quinpirole (**7**)



(*S*)-5-OH-DPAT (**8a**)

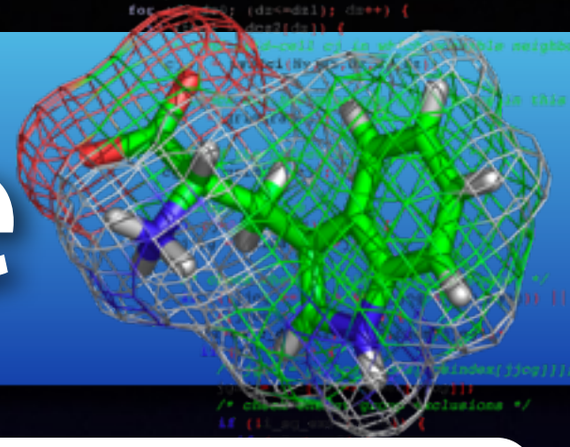


(*S*)-DPAT (**9**)

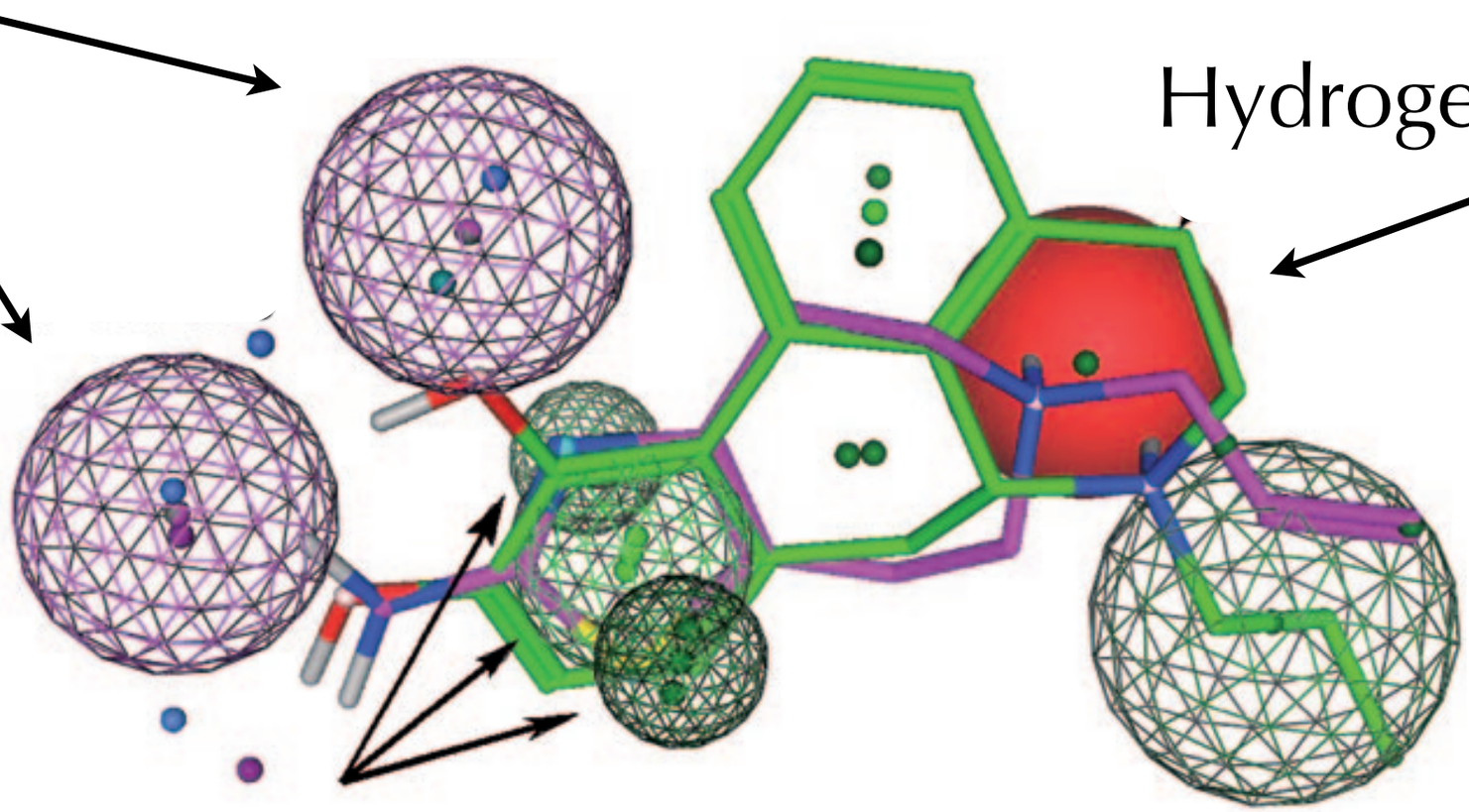


(*R*)-3-PPP (**10a**)

# Pharmacophore



Hydrogen bond donor/acceptor

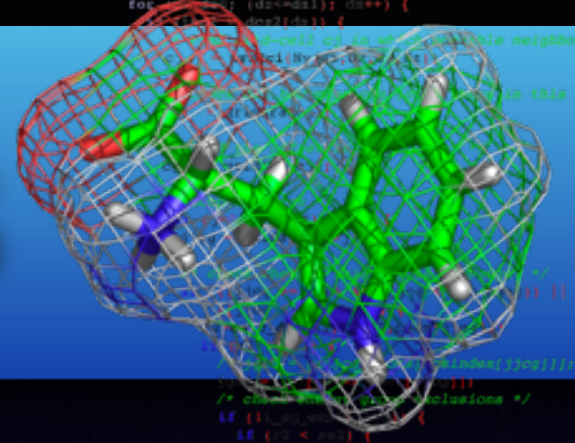


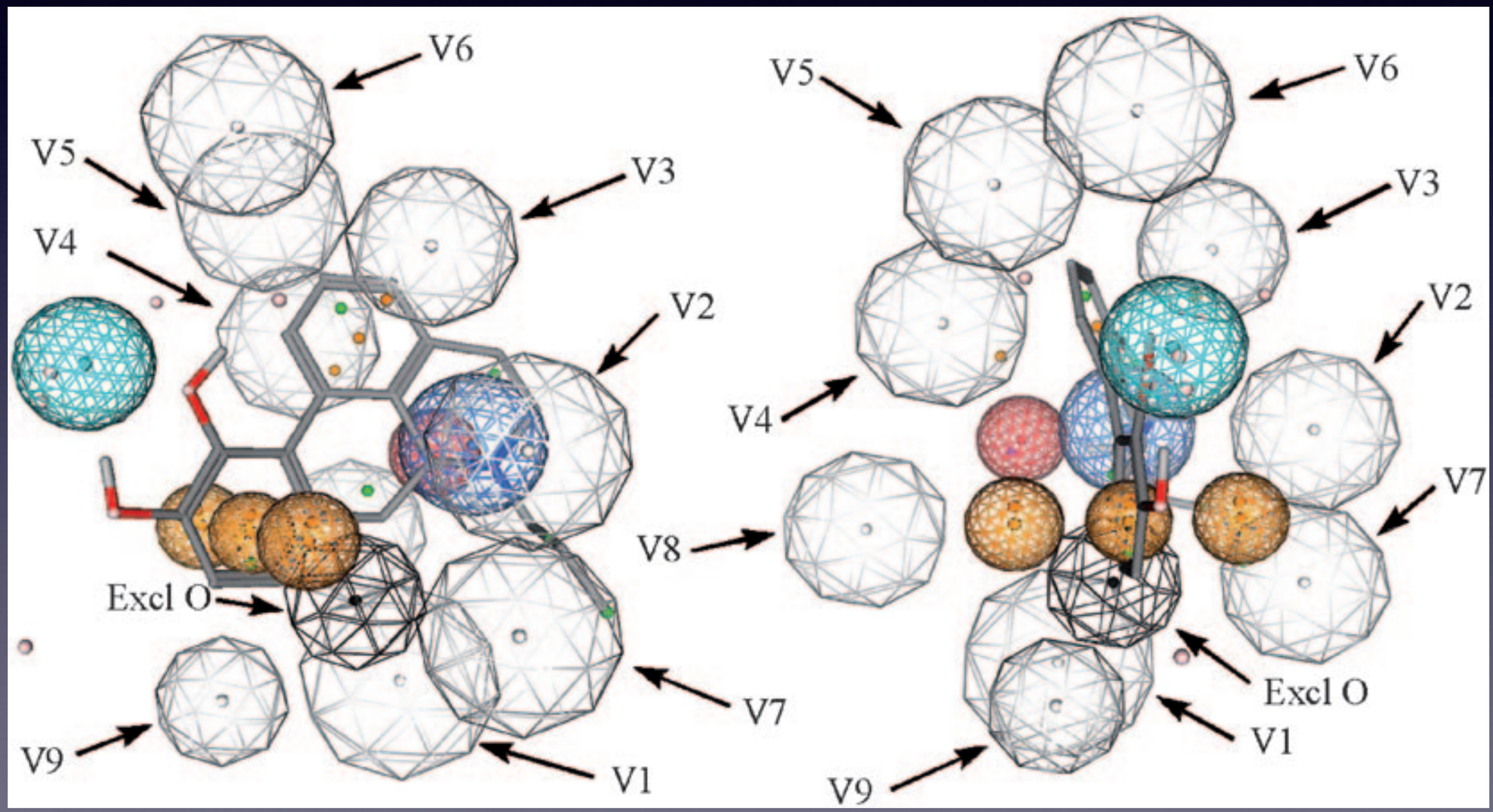
Hydrogen bond donor

Aromatic

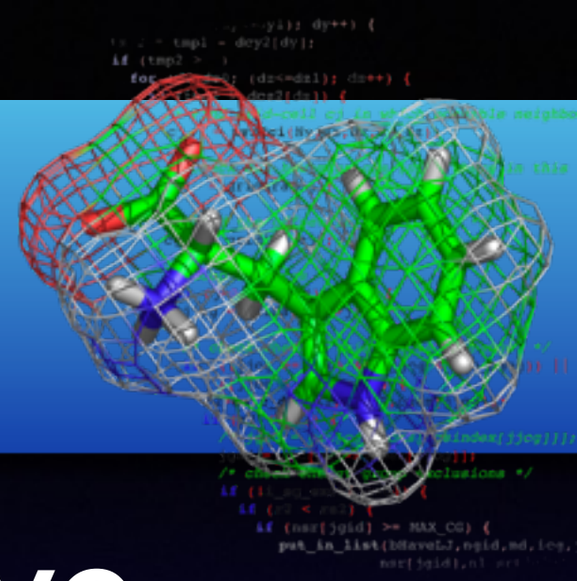
Hydrophobic

# Excluded volumes



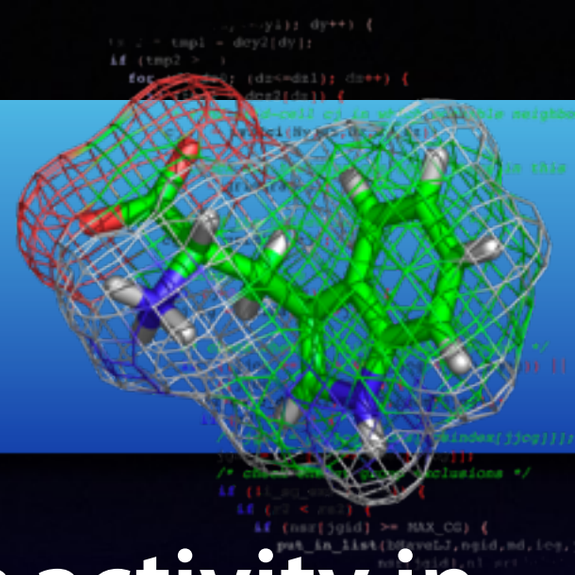


# STOP!

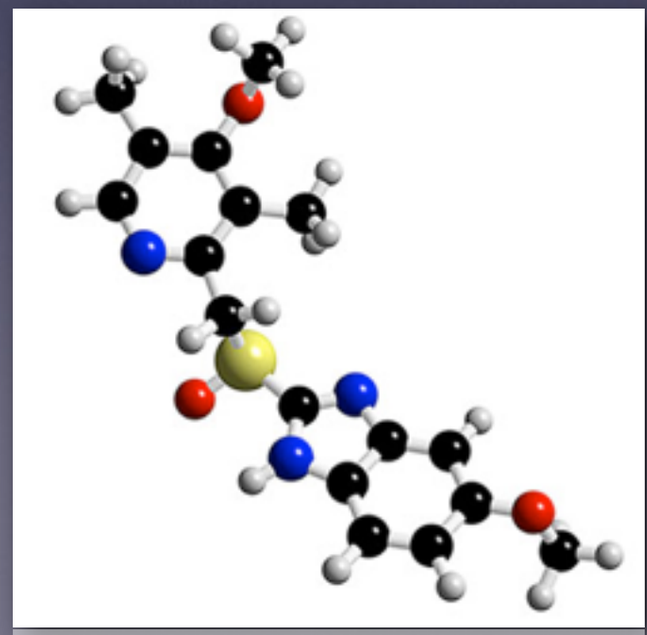


- Note that to this point we have not used any protein structures: We do not need structures to do drug discovery - but maybe it can be useful????

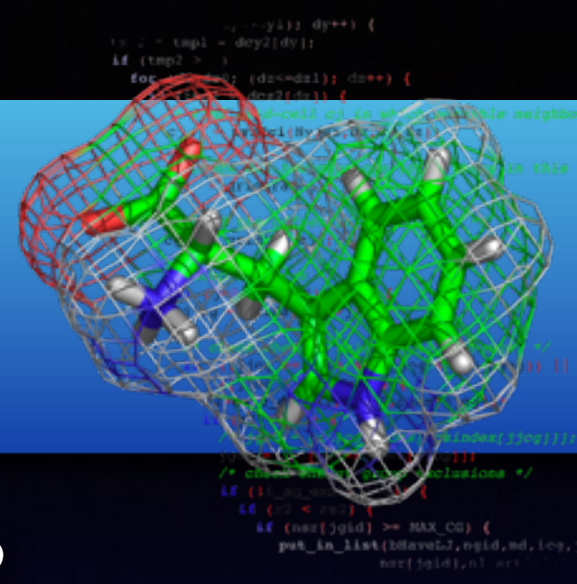
# Hit discovery



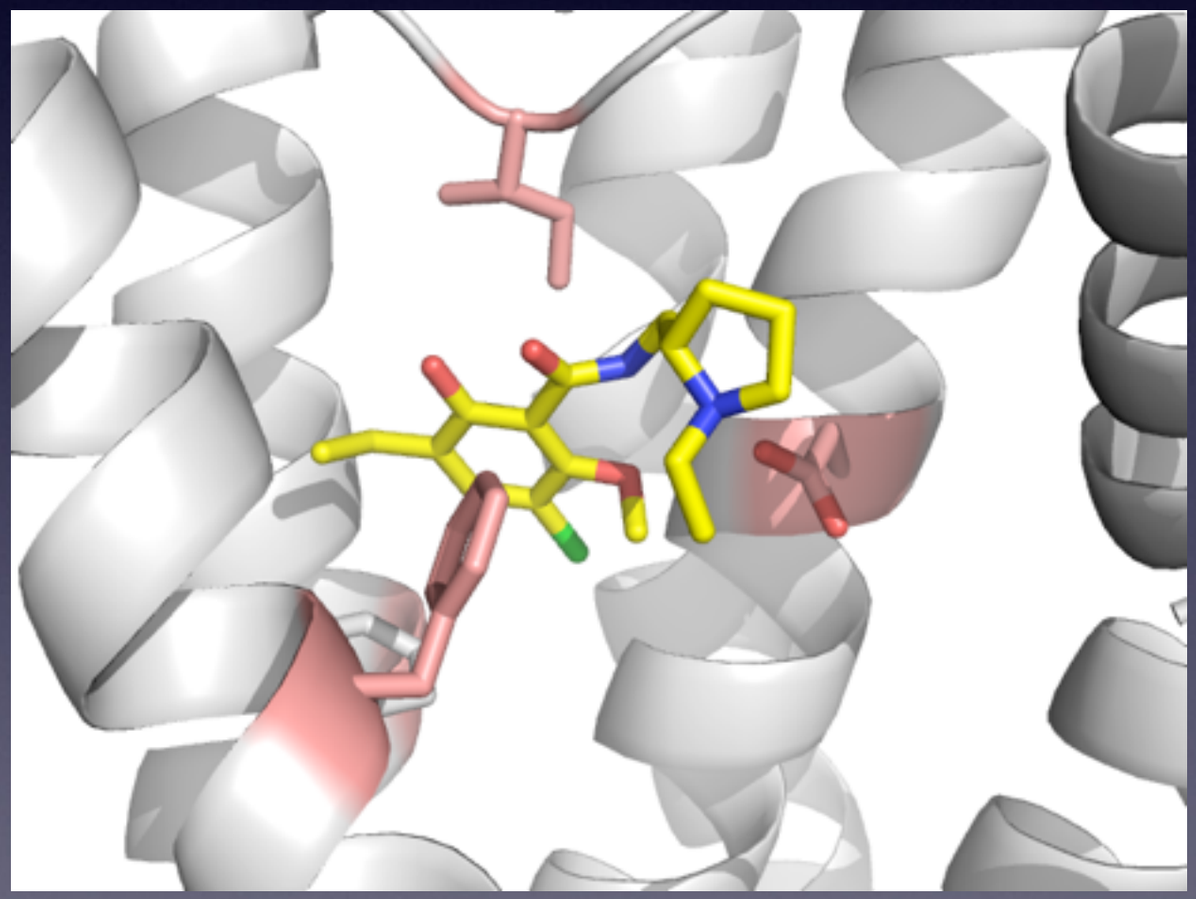
- Identify ‘hit’ molecules with *measureable* activity in a biological assay. Alternatives:
  - Serendipity (luck), e.g. Penicillin & Viagra
  - Litterature/Patents - has the target been studied before?
  - High-throughput Screening
  - QSAR/Pharmacophore modeling
  - **Molecular docking screening**

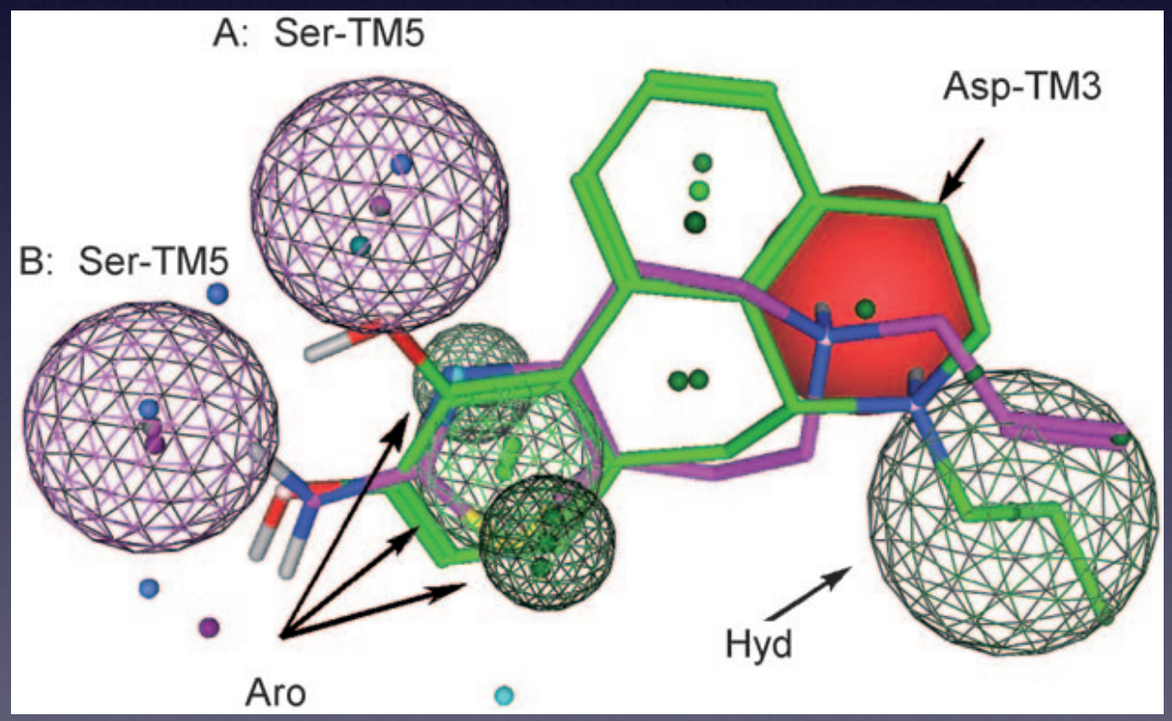


# The dopamine receptor structure



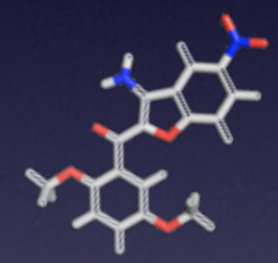
- Do we understand the pharmacophore?

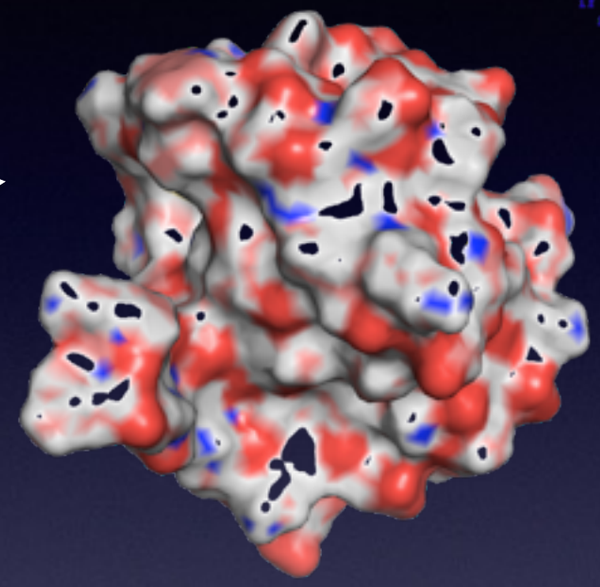




# Molecular docking

Easy



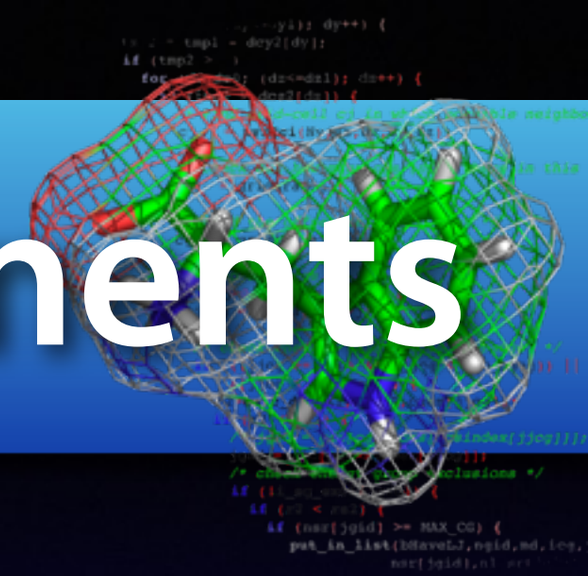


Hard

Focus on the structure of the receptor instead of the structure of the ligand(s)!

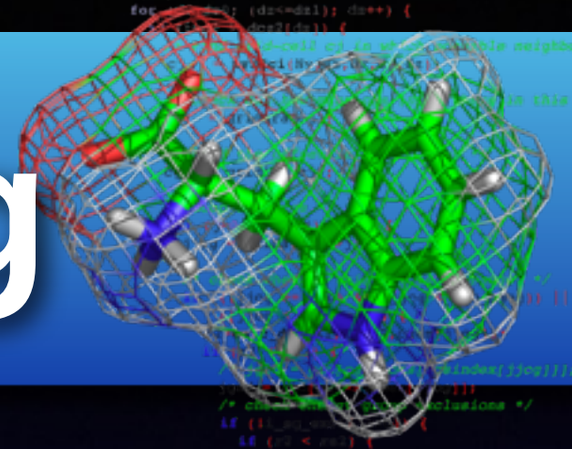
NOTE: We need a crystal structure or a homology model to do this!

# Docking: Two components

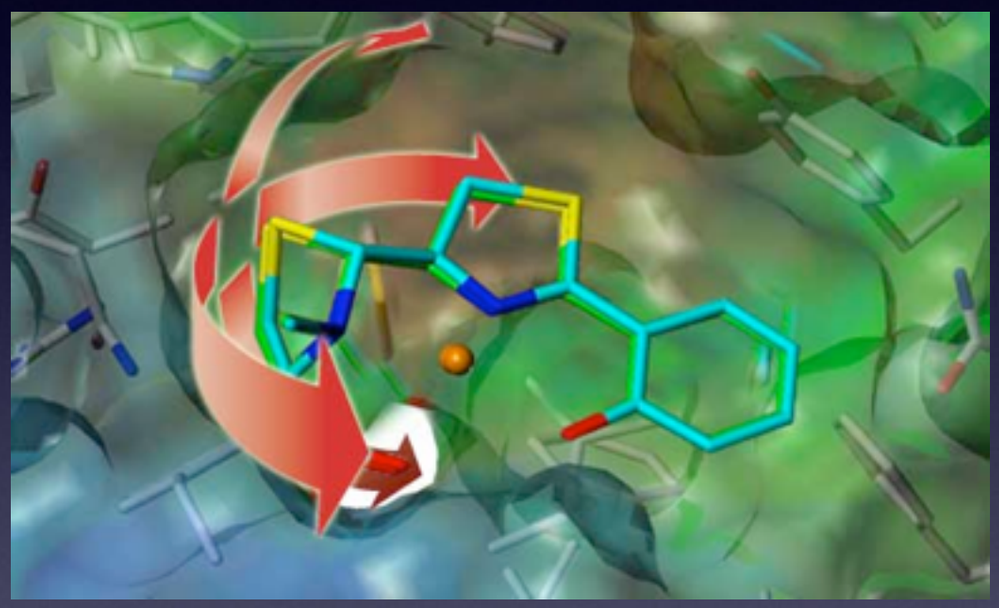


- **1. Sampling:** Generation of conformations of molecule in the receptor. Receptor is held rigid.
- **2. Scoring:** Determining which of the suggested conformations that has the lowest energy.

# Ligand sampling

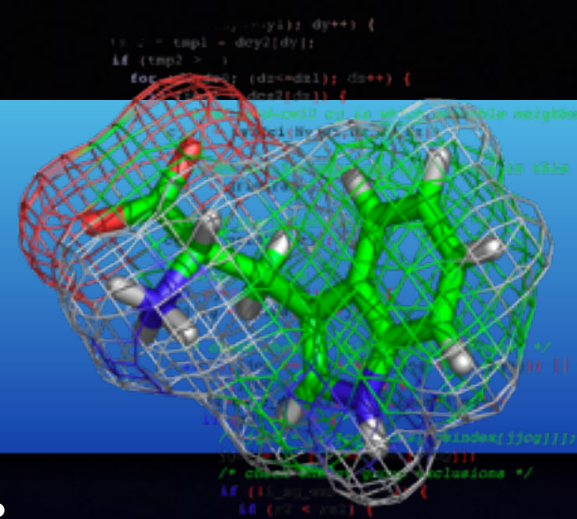


- Even ligand sampling  is expensive:
  - Ligand
    - 6 rotation/translation
    - 4 rot bonds
  - 10^3 Å box
  - Sample angles in 10 degree increments
  - Sample translation in 0.5Å steps
  - Sample 100 conformations/sec



**It would take 200 years to finish this docking!**

# Genetic algorithm



The center of mass, rigid rotation, and torsions are genes. The score is the phenotype and selection is made based on . Both Darwinian and Lamarcian

Make initial population randomly

↓

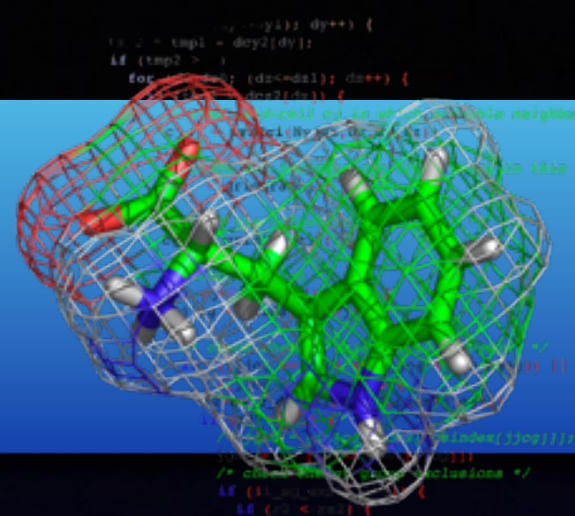
Evaluate docking scores

↓

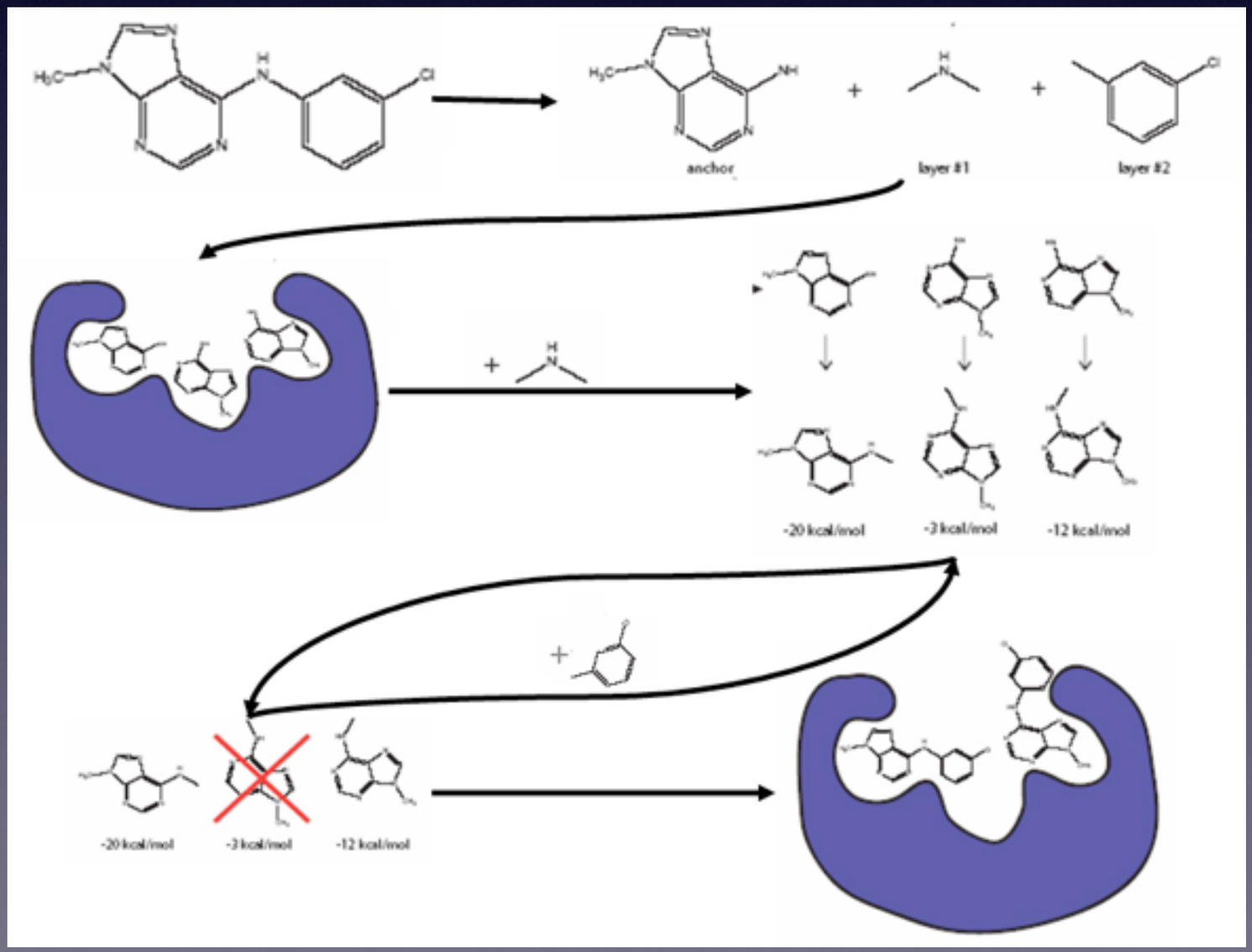
Extract best scoring conformations
(survival of the fittest)

"Mutations"

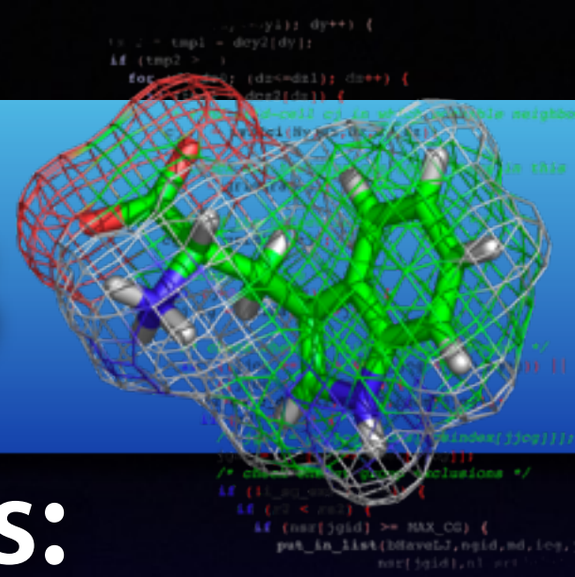
# Anchor-and-grow



- Divide molecule into rigid fragments. Dock the first fragment and then add next, etc.

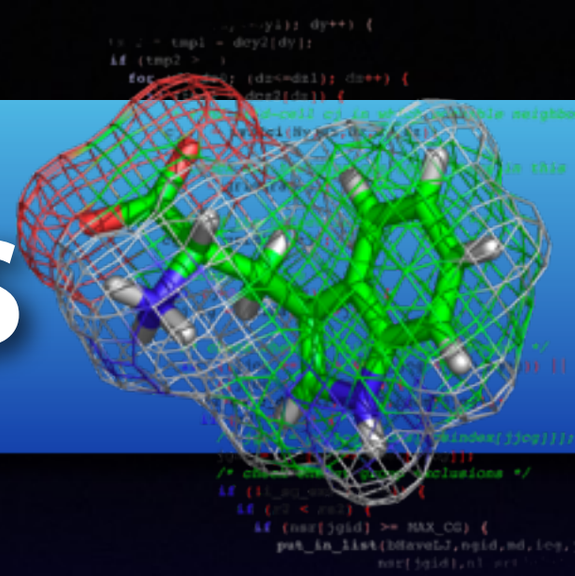


# 2. Scoring: Types

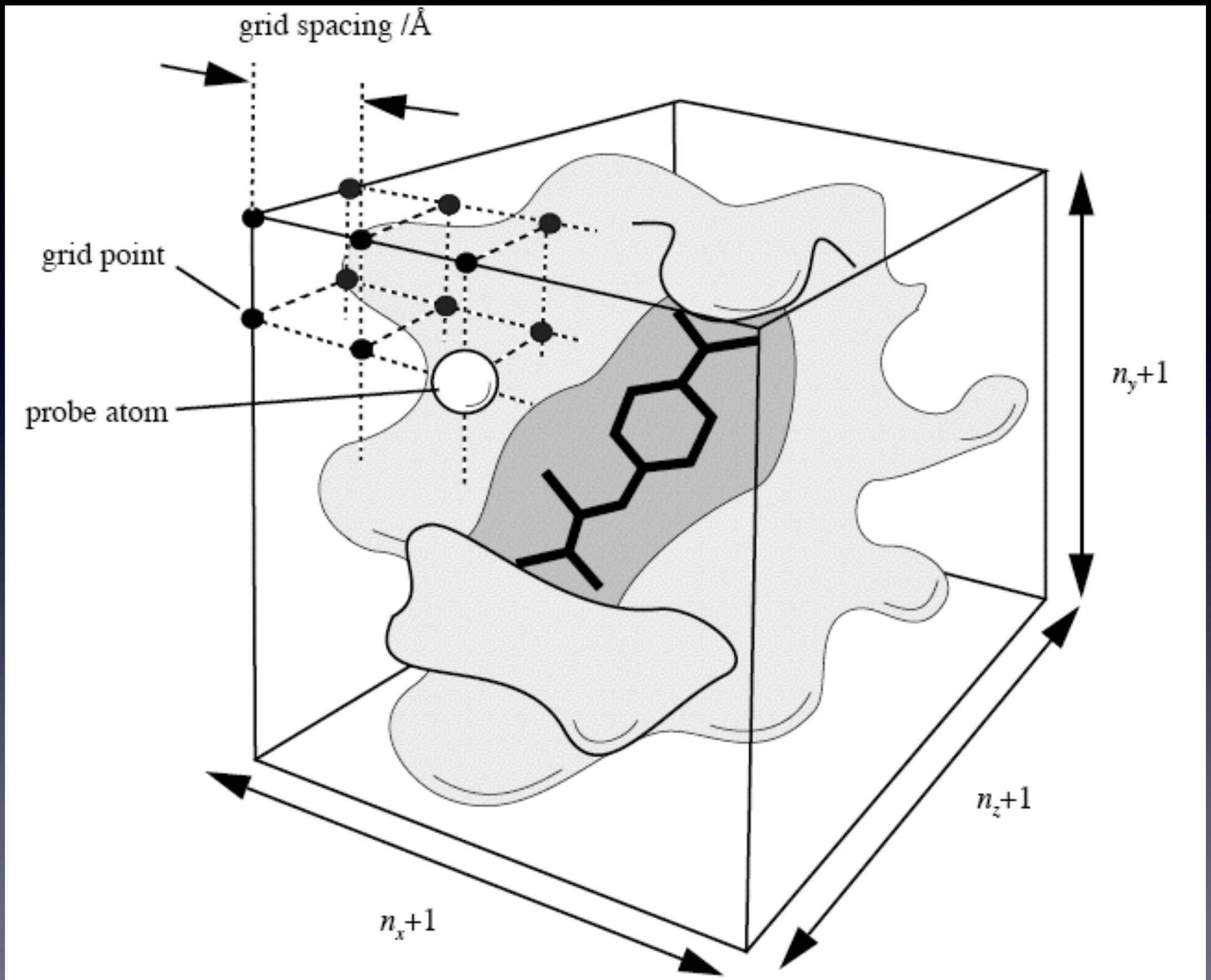


- Three types of scoring functions:
- Force field based: Physics terms for coulombic and van der Waals interactions.
- Empirical: Give “points” for formation of hydrogen bonds and hydrophobic interactions. Parameterized to reproduce experimental protein-ligand complexes.
- Knowledge-based: Statistical potential. Make statistics for which interactions that are favorable (O-O, N-O, C-C etc for all known complexes). Give score based on its probability.

# 2. Scoring: Grids



- Expensive to re-calculate all interactions with the protein/receptor for every new position of the ligand
  - Simplify the problem by assigning the protein properties (Lennard Jones, charges) to a grid instead
  - Calculate interactions of “typical” atoms such as C,N,O,H on each grid point
- Find the best position of each ligand on this grid


grid spacing /Å
grid point
probe atom
$n_y$+1
$n_x$+1
$n_z$+1

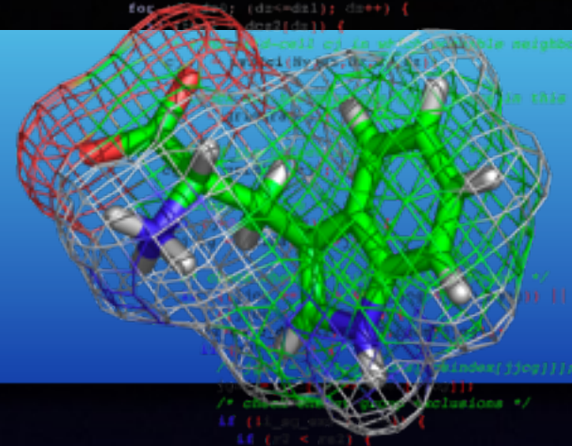
# Docking / screening

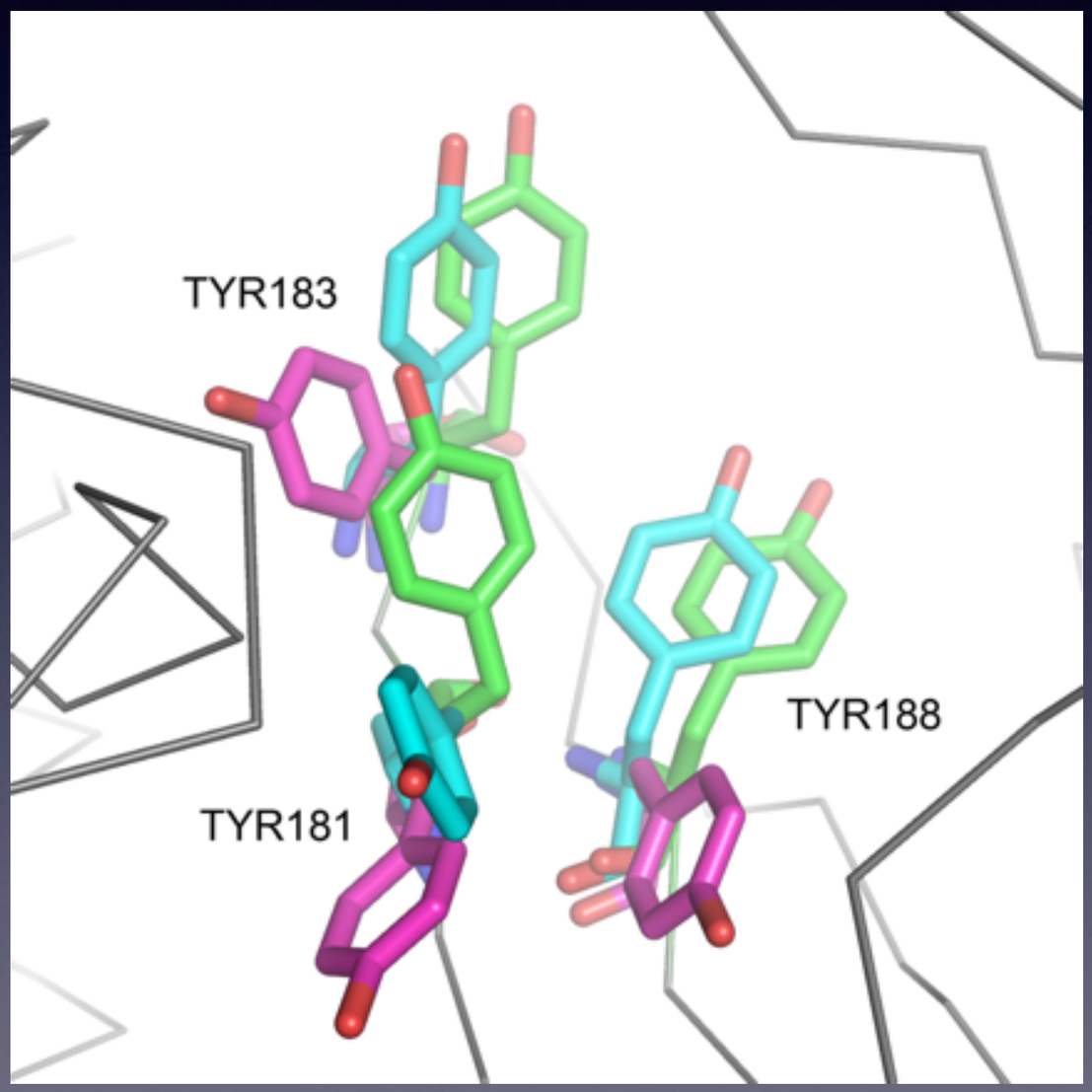
- **Does it work?**

| Target | #cmpds | Exp. hits | Docking hits |
|---|---|---|---|
| Lactamase | 300,000 | 0 | 2 |
| Cruzain | 200,000 | 146 | 5 |

- Docking can't find all hits, but that's ok - even a hit rate of 5% would save money!

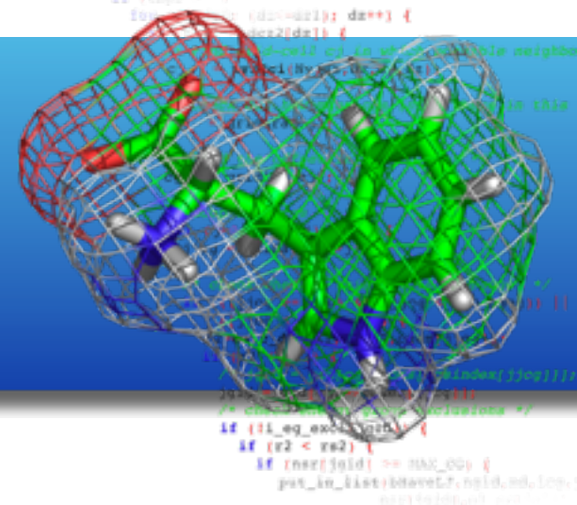
# Receptor flexibility

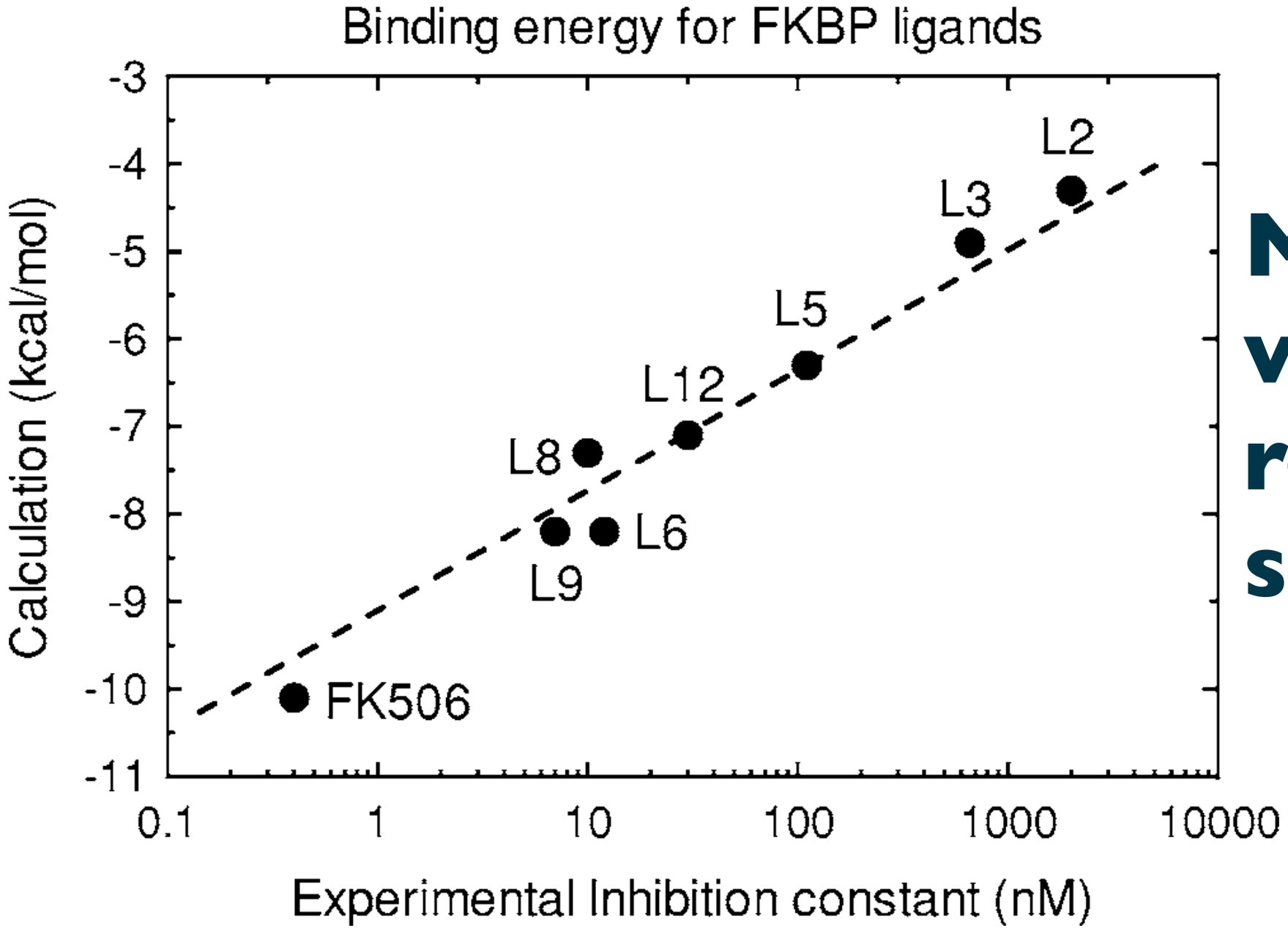





- Including receptor flexibility should be crucial for improving the accuracy of molecular docking. Many proteins change conformation when they bind ligands.

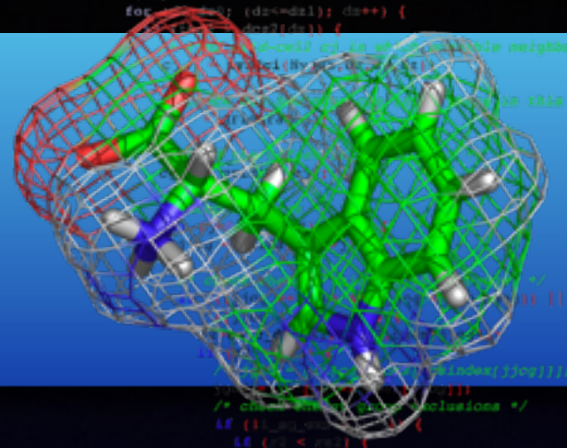
# Predictive Power

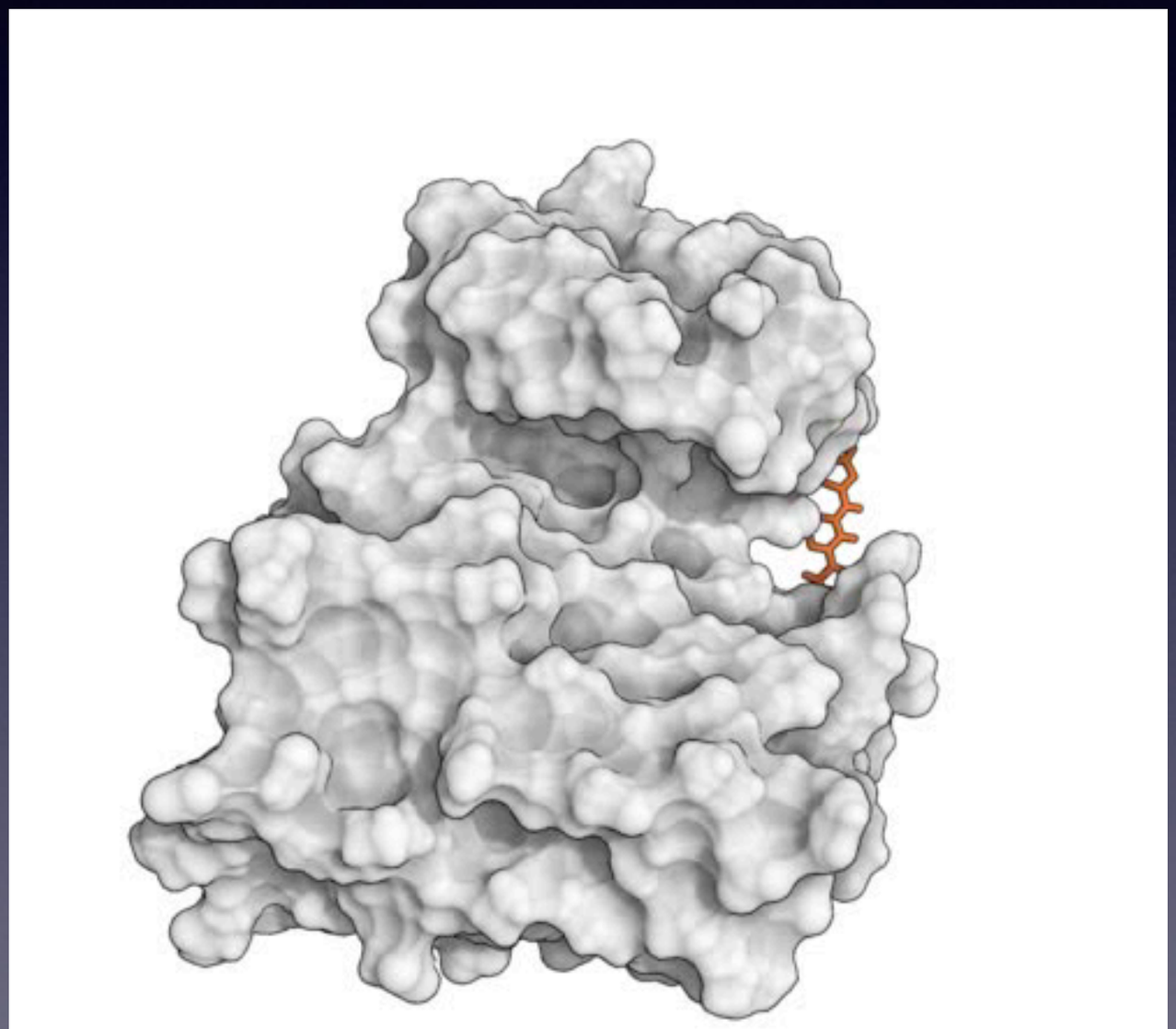





**MD Can give very accurate results in some cases!**

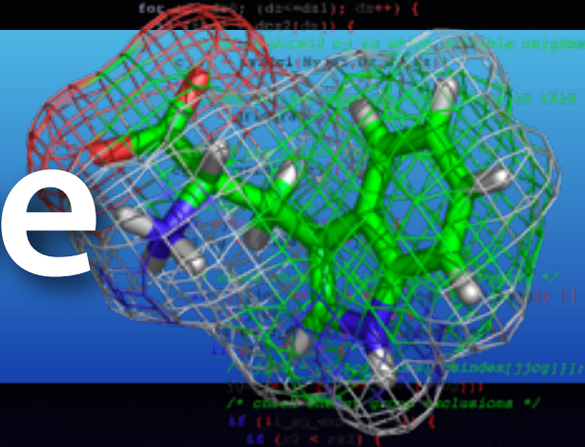
# Brute force MD

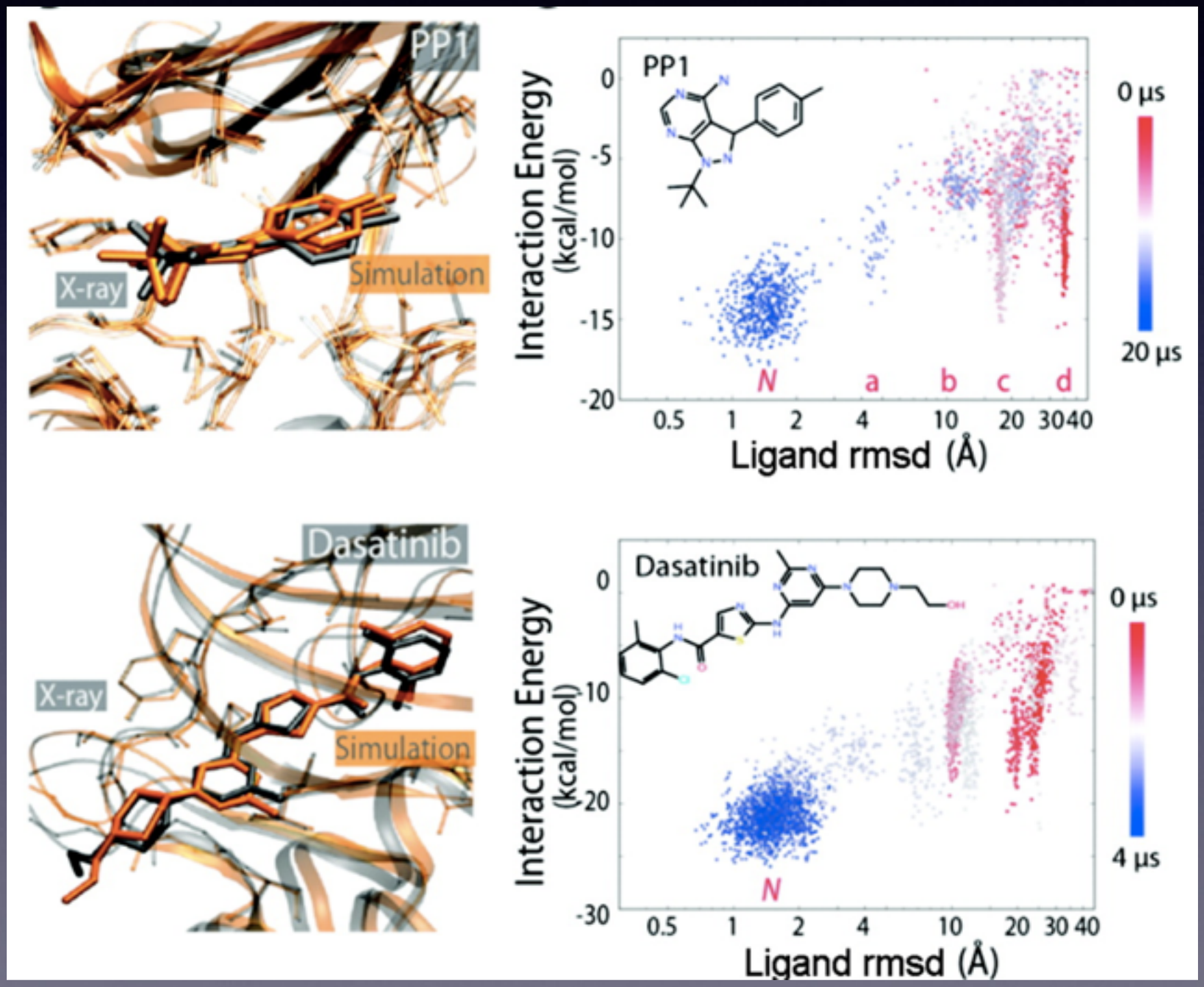




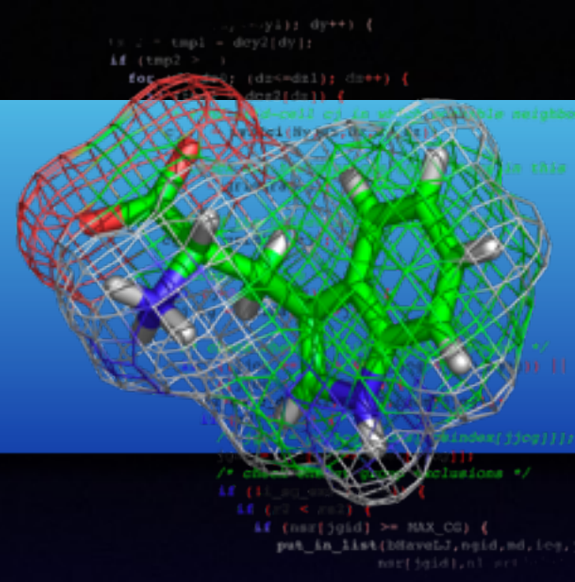
David Shaw
2011

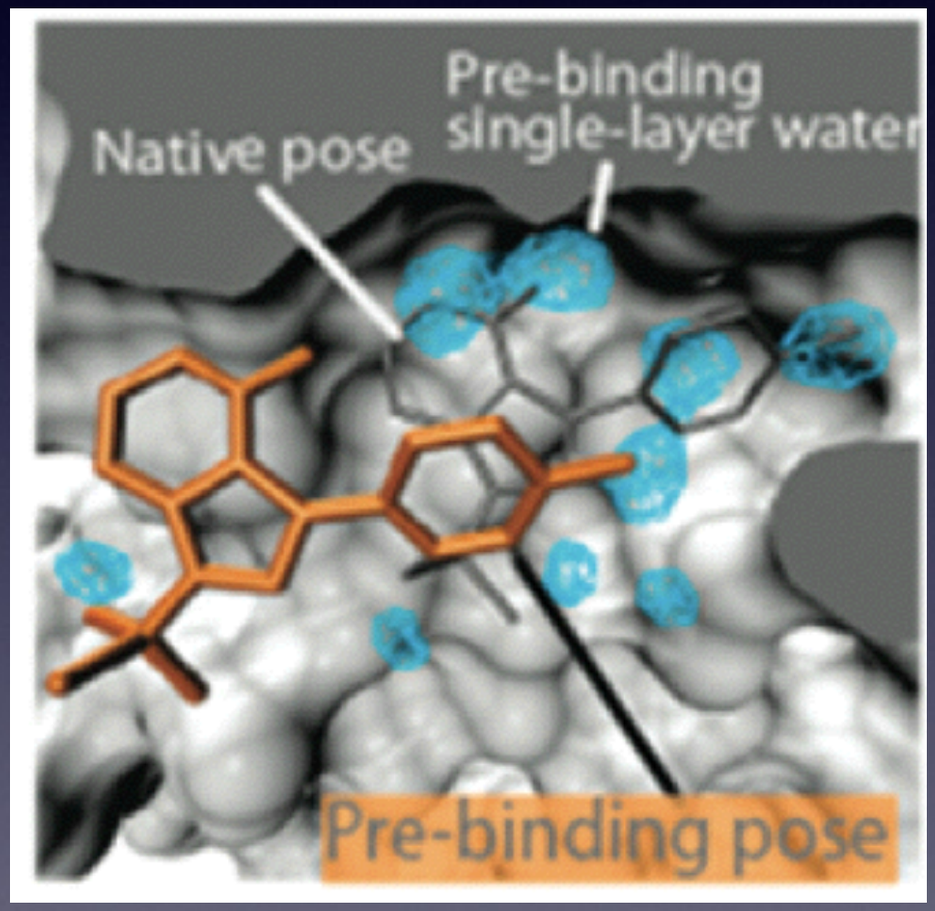
# Energy landscape

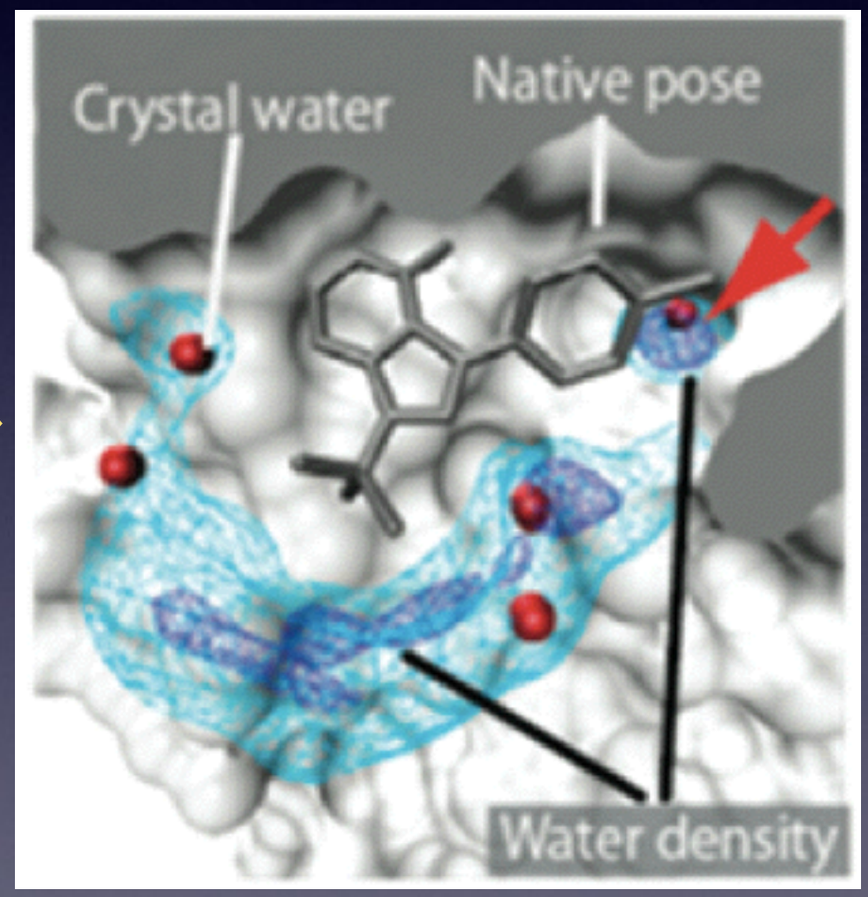




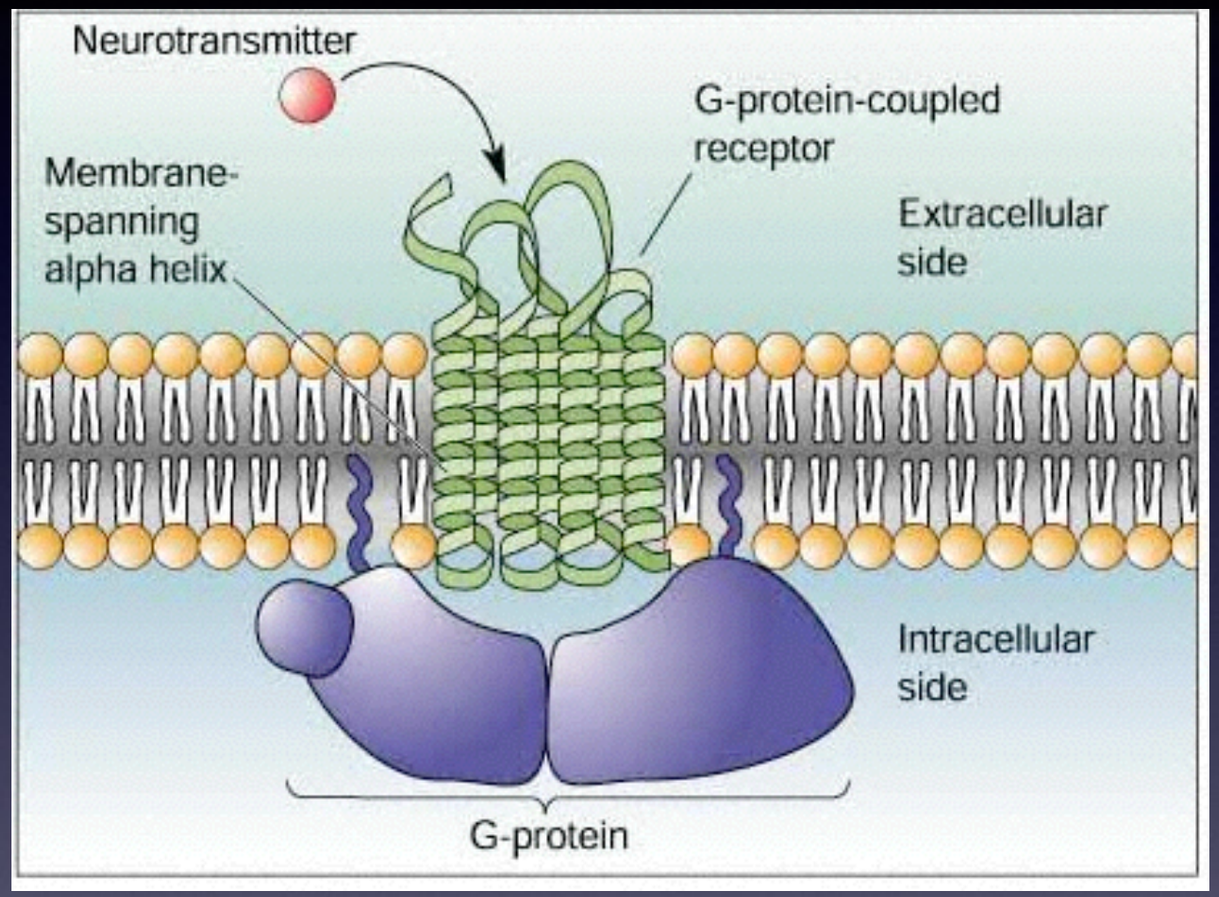
# Kinetic barrier

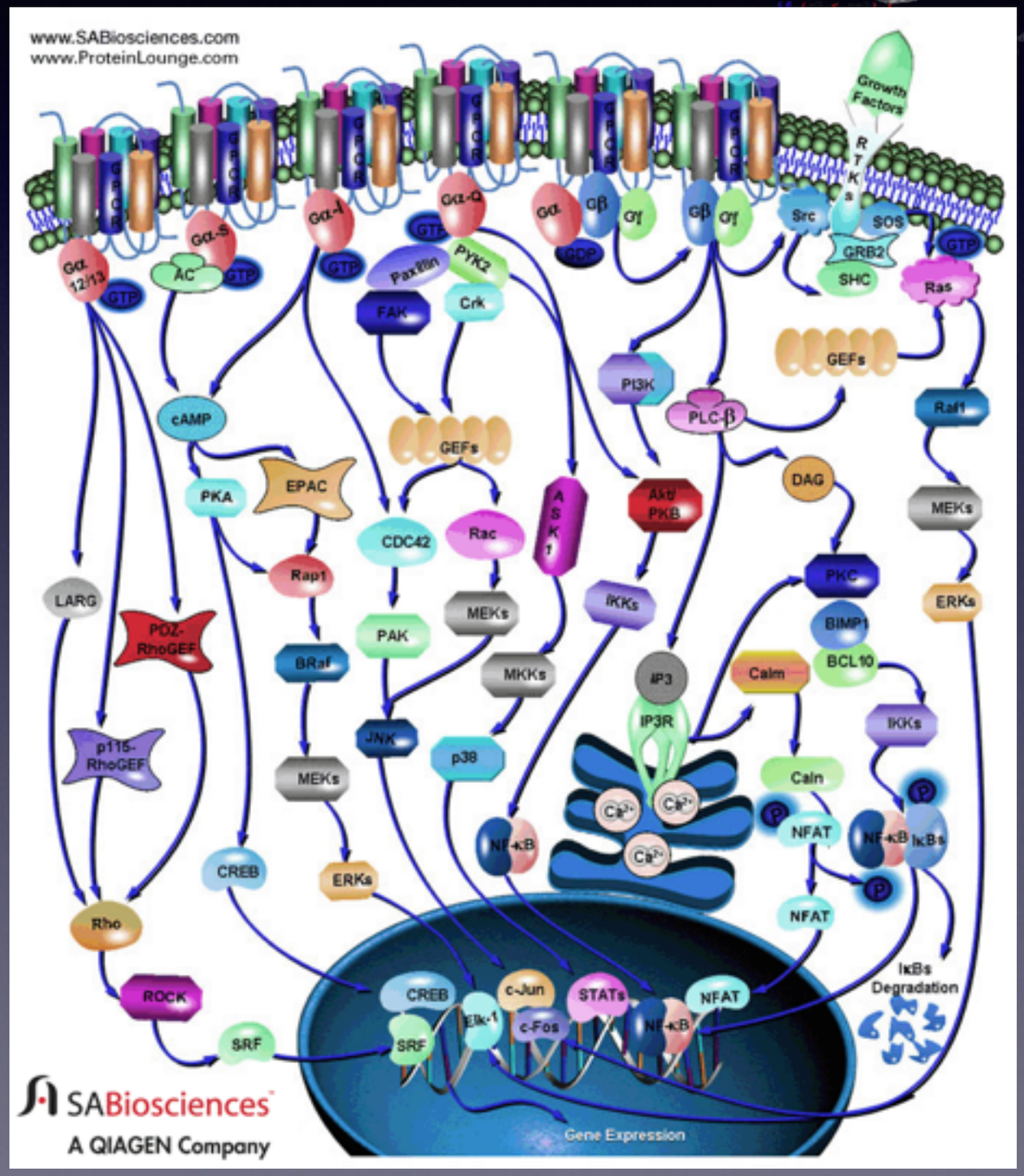




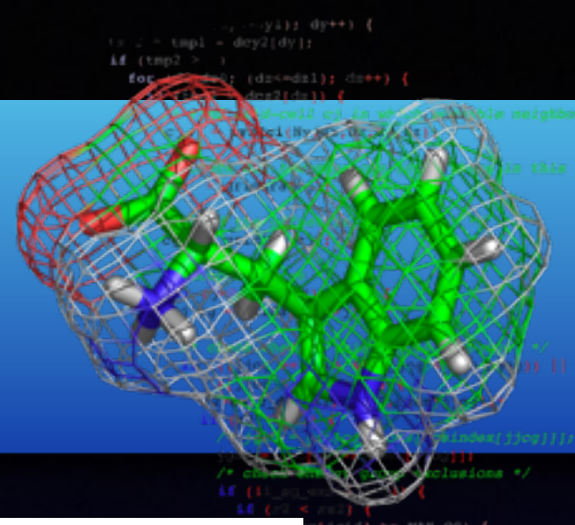
# GPCRs

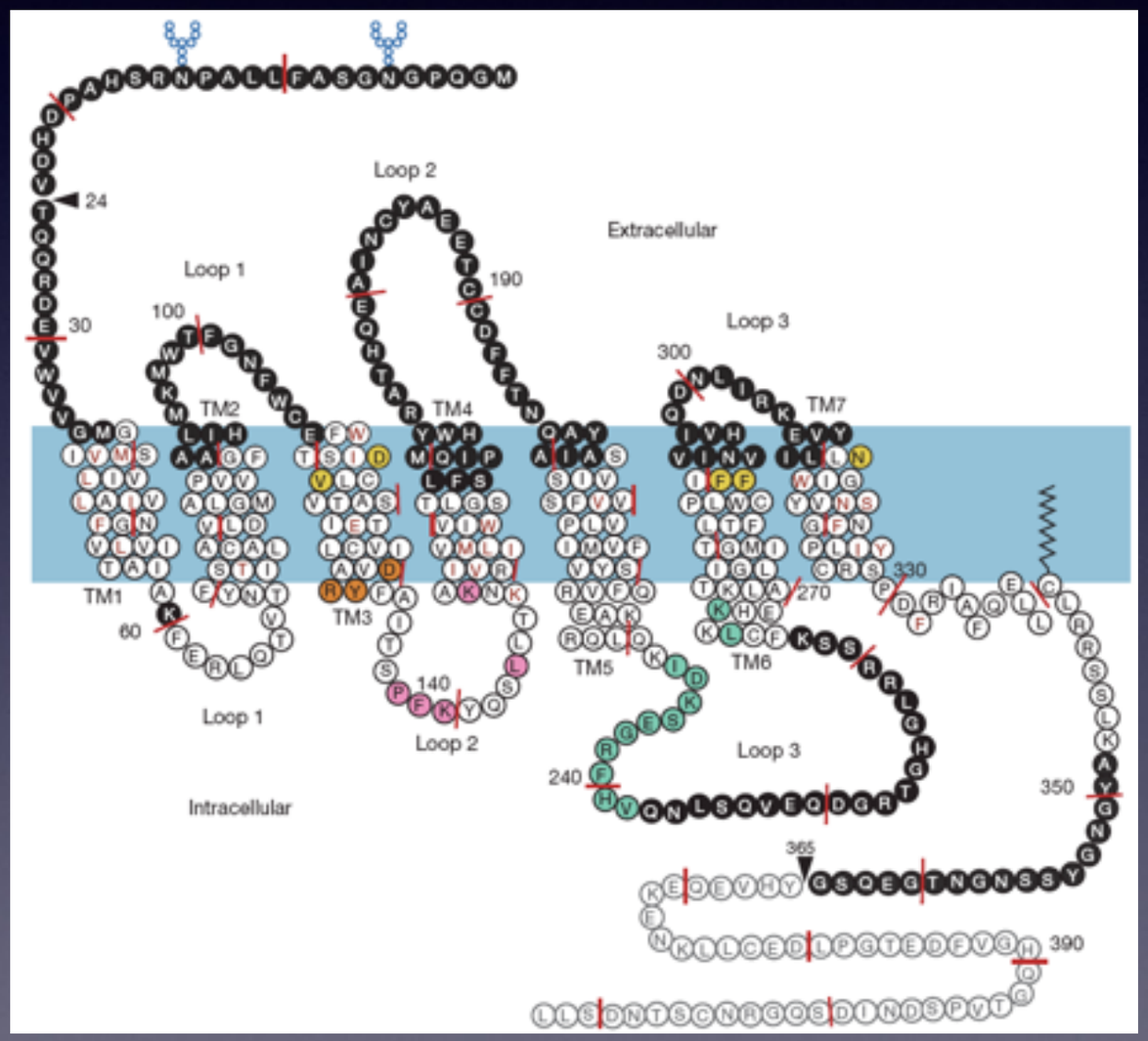


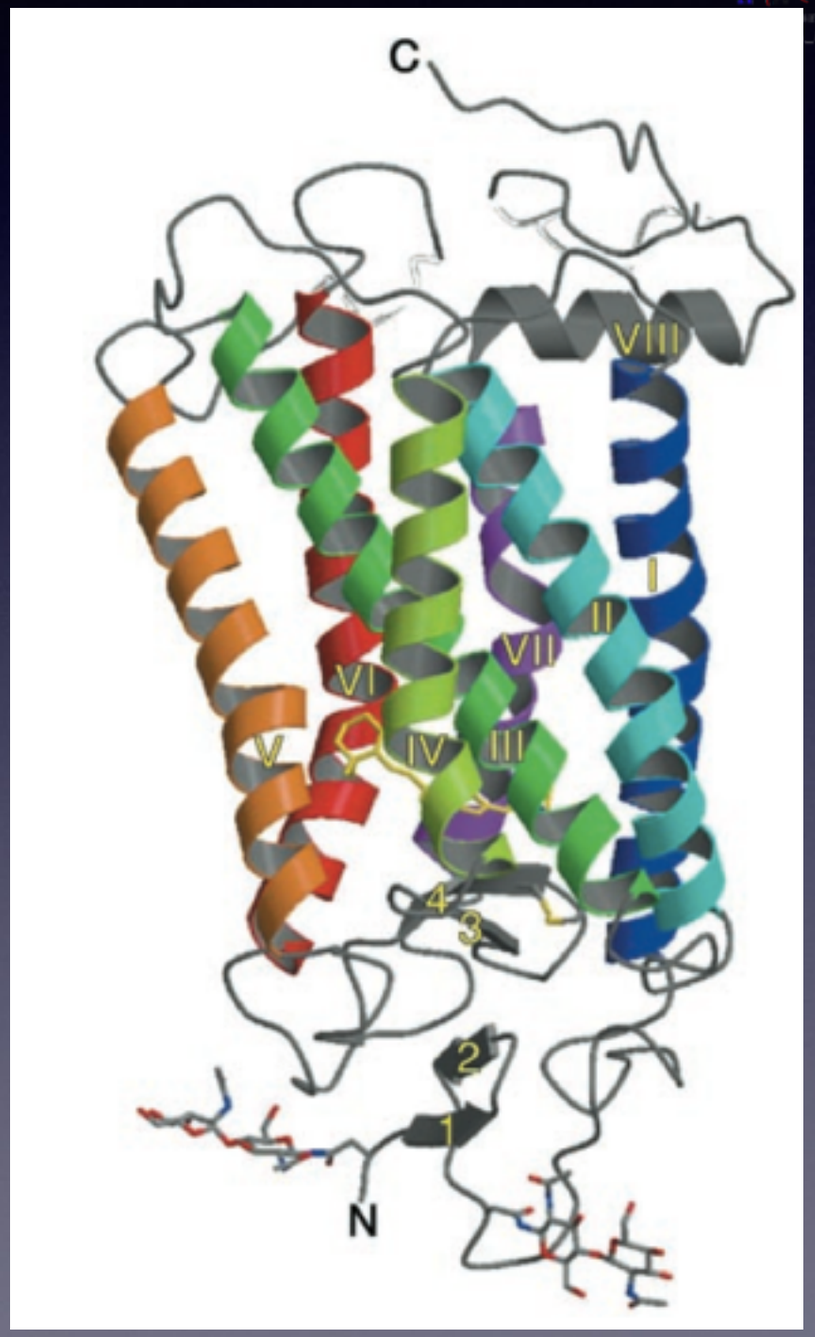




- ~900 genes in human
- Known unique structures: 25

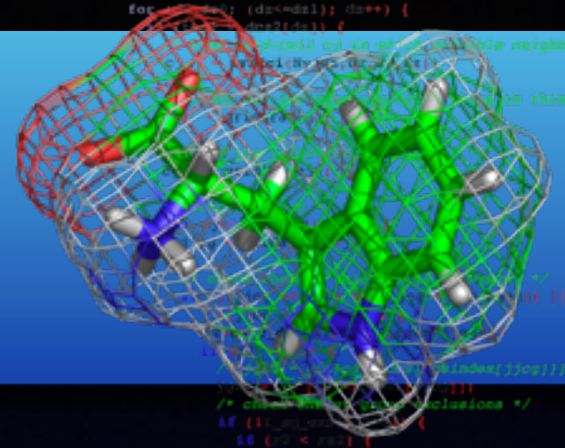
# Structure



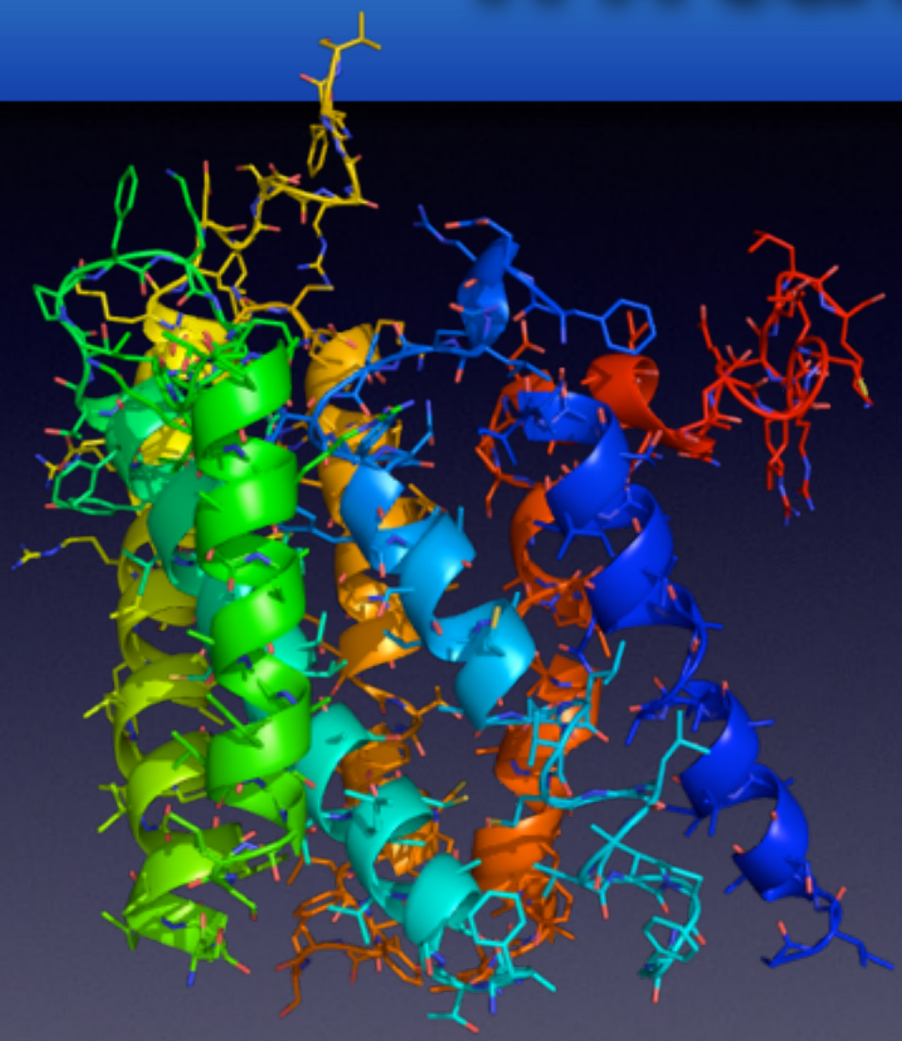

Loop 2
Extracellular
Loop 1
Loop 3
TM2
TM4
TM7
TM1
TM3
TM5
TM6
Loop 1
Loop 2
Loop 3
Intracellular
24
30
60
100
140
190
240
270
300
330
350
365
390


C
VIII
I
II
VII
VI
V
IV
III
4
3
2
1
N

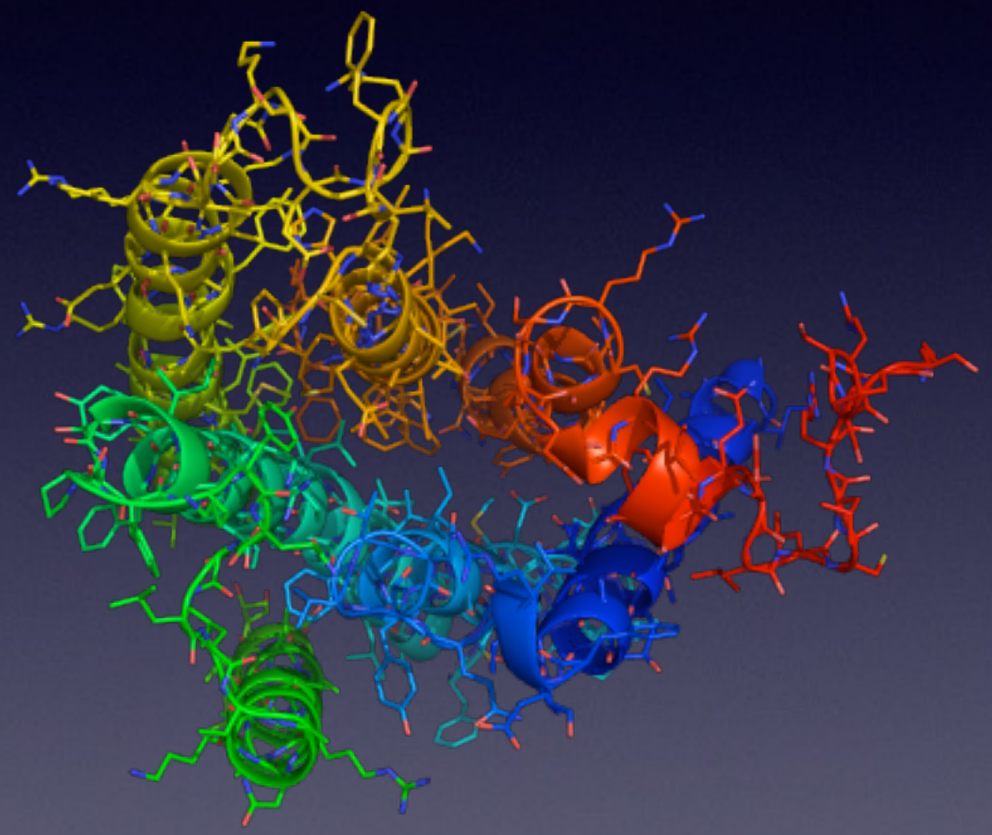
# Inital success!



## Human $\beta_2$ adrenergic GPCR

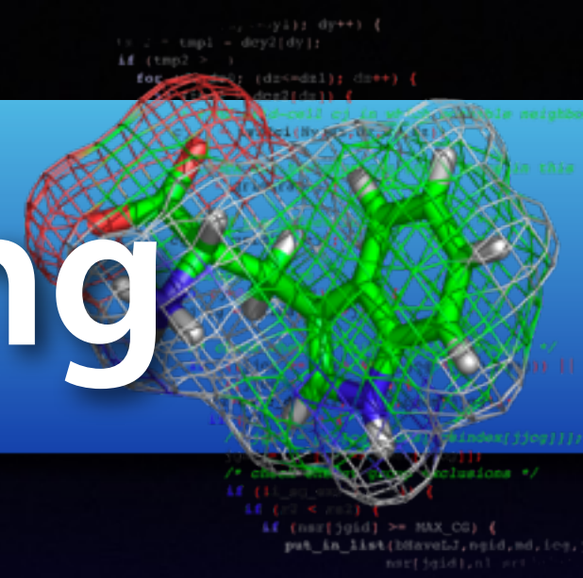


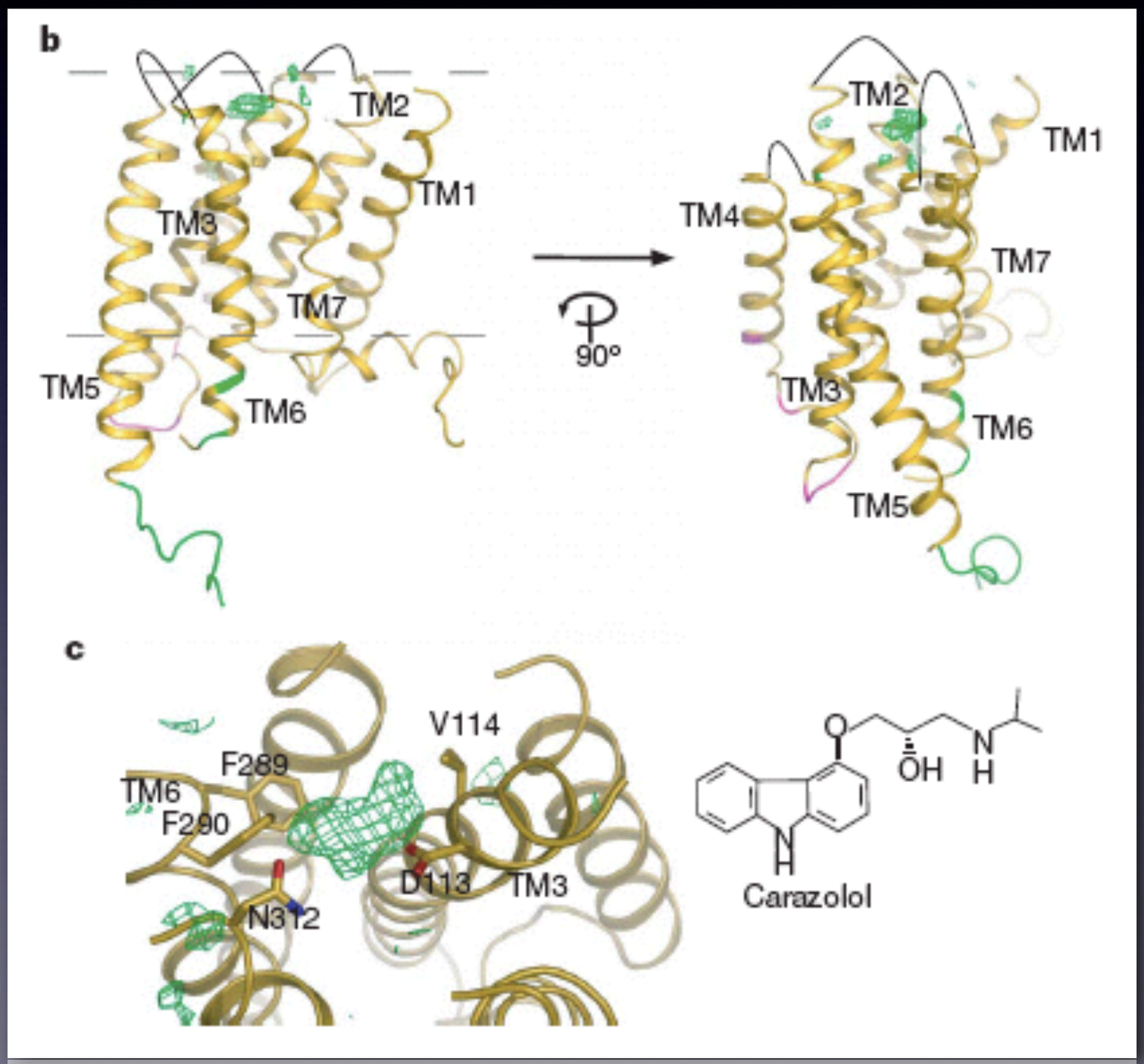
Nature, 15 Nov 2007:
Brian Kobilka
@ Stanford University



Science, 23 Nov 2007:
Ray Stevens
@ Scripps

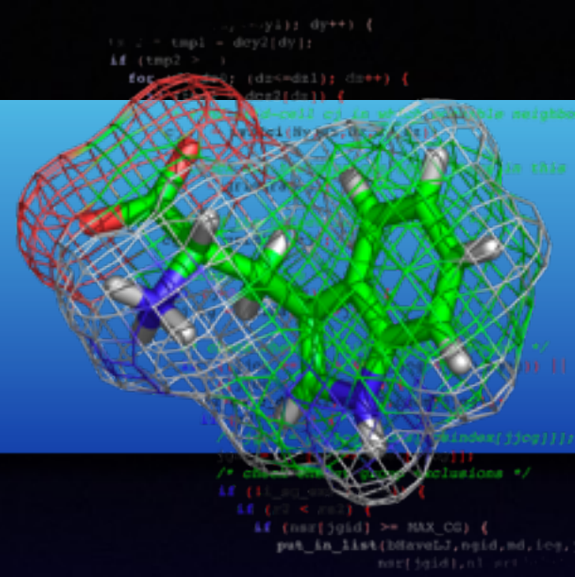
# GPCR drug binding

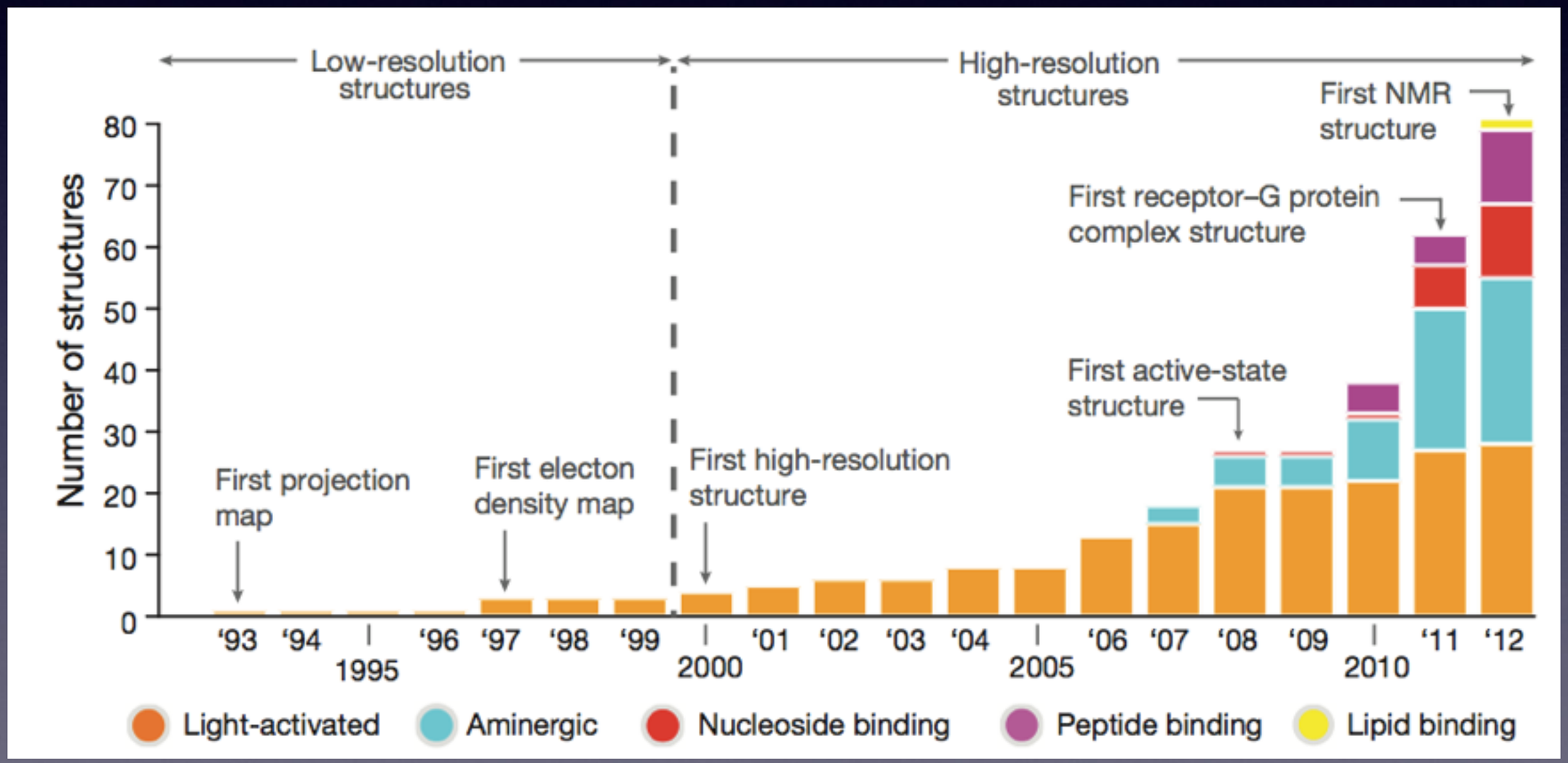




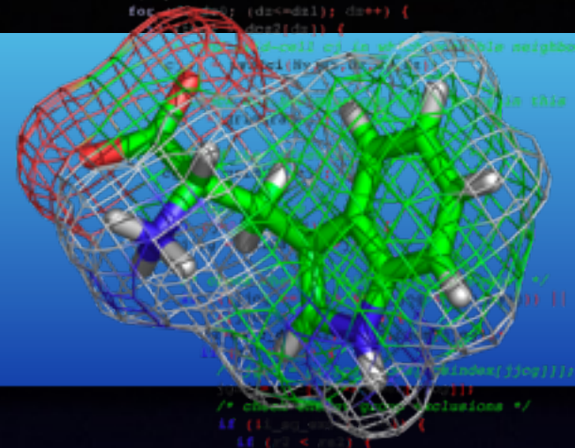

## Carazolol binding to Human $\beta_2$ adrenergic GPCR

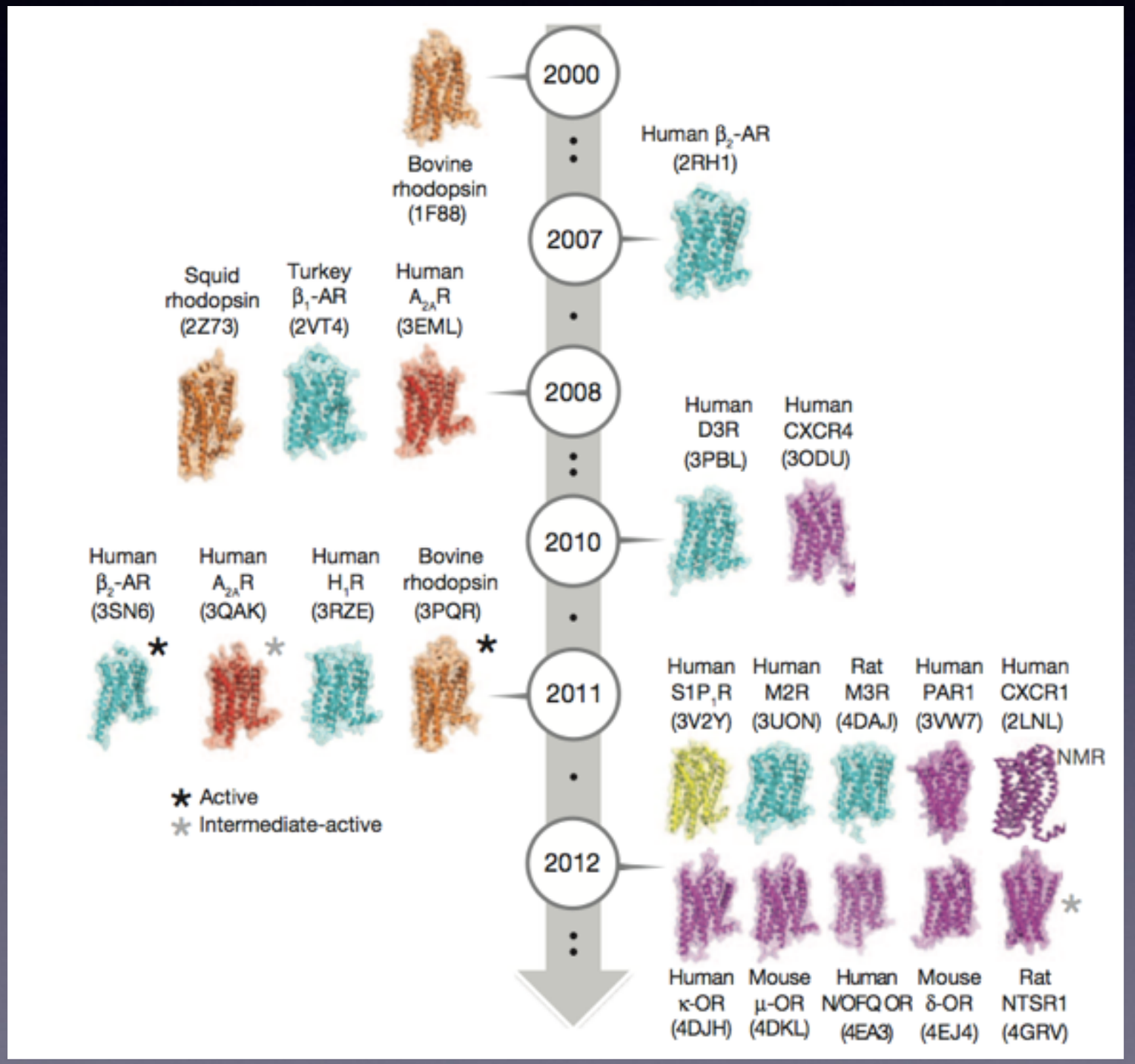
# Explosion!



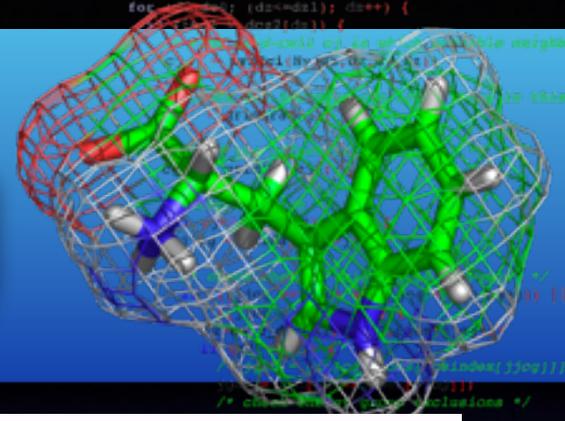

Low-resolution structures
High-resolution structures
First NMR structure
First receptor–G protein complex structure
First active-state structure
First projection map
First electon density map
First high-resolution structure
Number of structures
0
10
20
30
40
50
60
70
80
'93
'94
1995
'96
'97
'98
'99
2000
'01
'02
'03
'04
2005
'06
'07
'08
'09
2010
'11
'12
Light-activated
Aminergic
Nucleoside binding
Peptide binding
Lipid binding

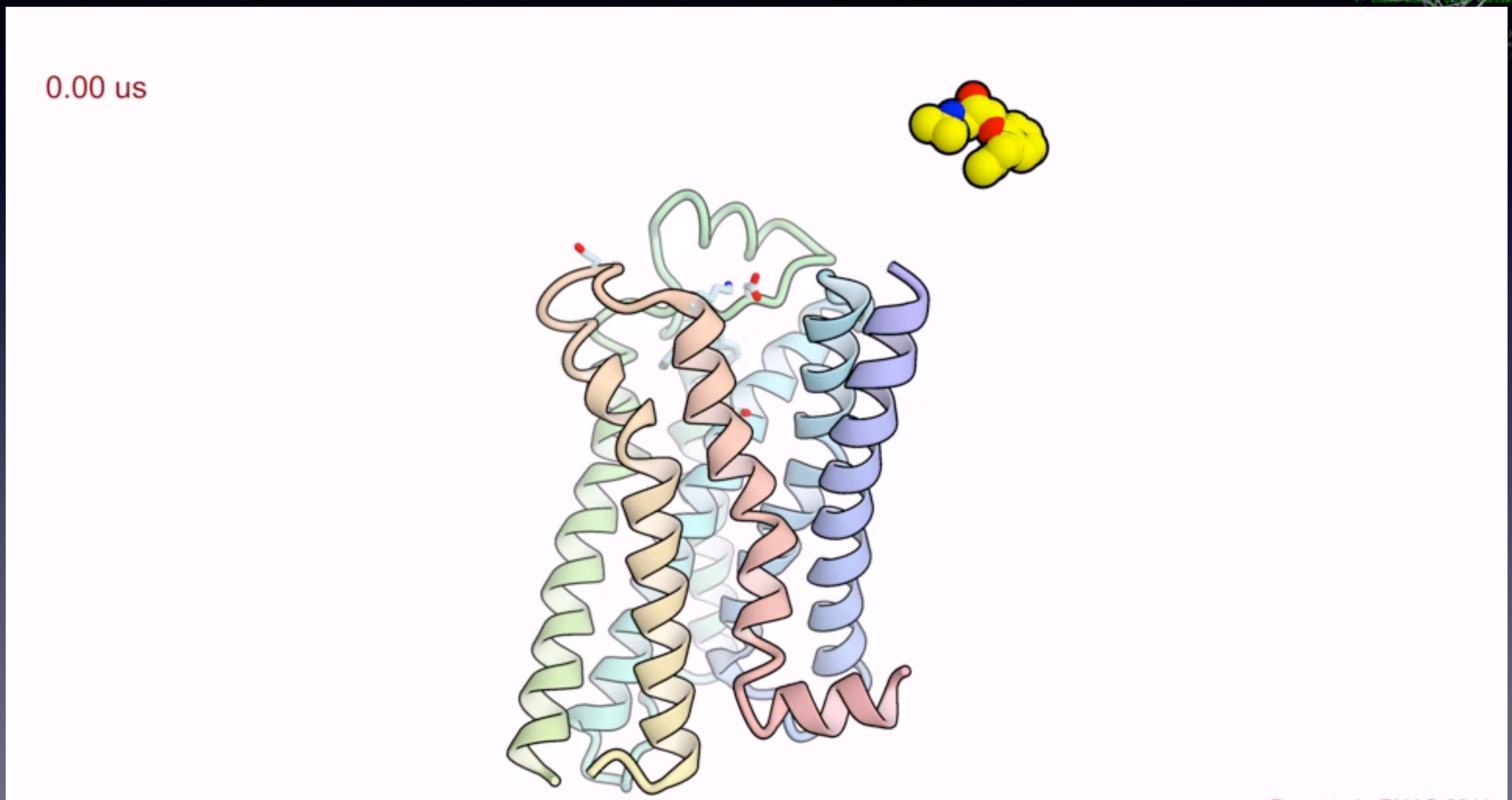
# Explosion




2000
Bovine rhodopsin (1F88)
Human $\beta_2$-AR (2RH1)
2007
Squid rhodopsin (2Z73)
Turkey $\beta_1$-AR (2VT4)
Human $A_{2A}$R (3EML)
2008
Human D3R (3PBL)
Human CXCR4 (3ODU)
2010
Human $\beta_2$-AR (3SN6)
Human $A_{2A}$R (3QAK)
Human $H_1$R (3RZE)
Bovine rhodopsin (3PQR)
2011
Human $S1P_1$R (3V2Y)
Human M2R (3UON)
Rat M3R (4DAJ)
Human PAR1 (3VW7)
Human CXCR1 (2LNL)
NMR
2012
Human κ-OR (4DJH)
Mouse μ-OR (4DKL)
Human N/OFQ OR (4EA3)
Mouse δ-OR (4EJ4)
Rat NTSR1 (4GRV)
Active
Intermediate-active

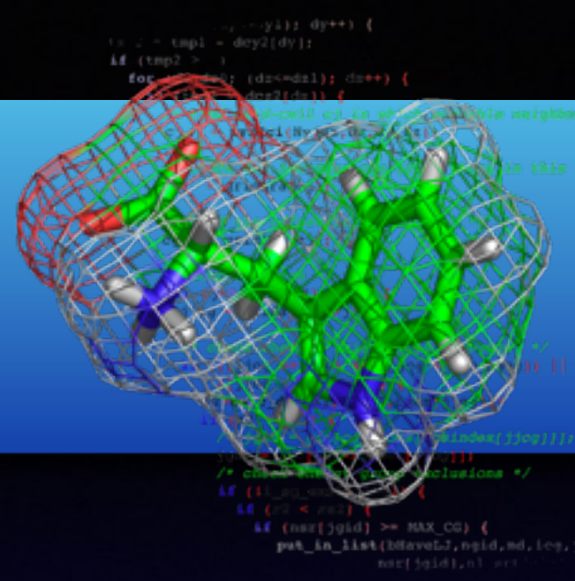
# Ligand docking

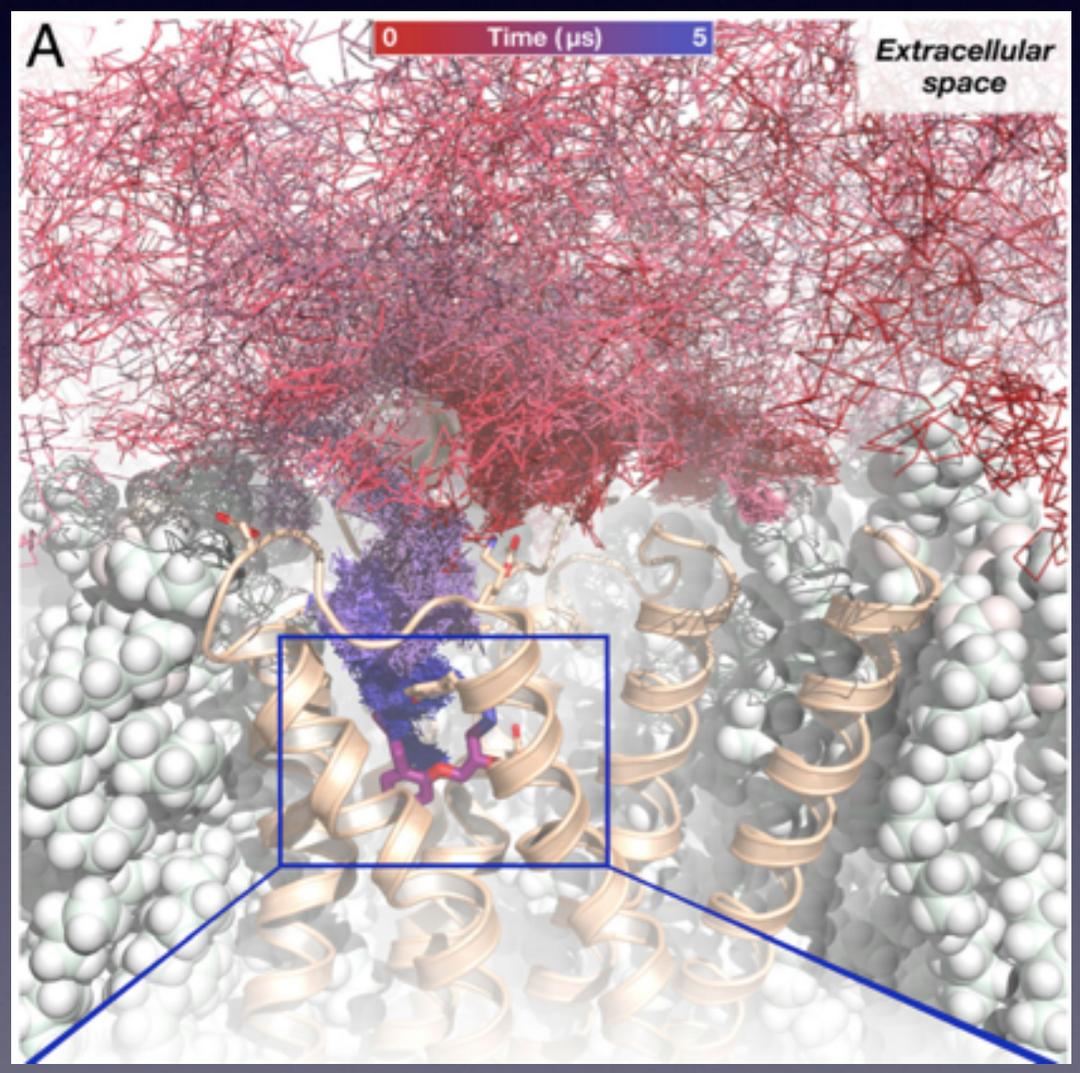


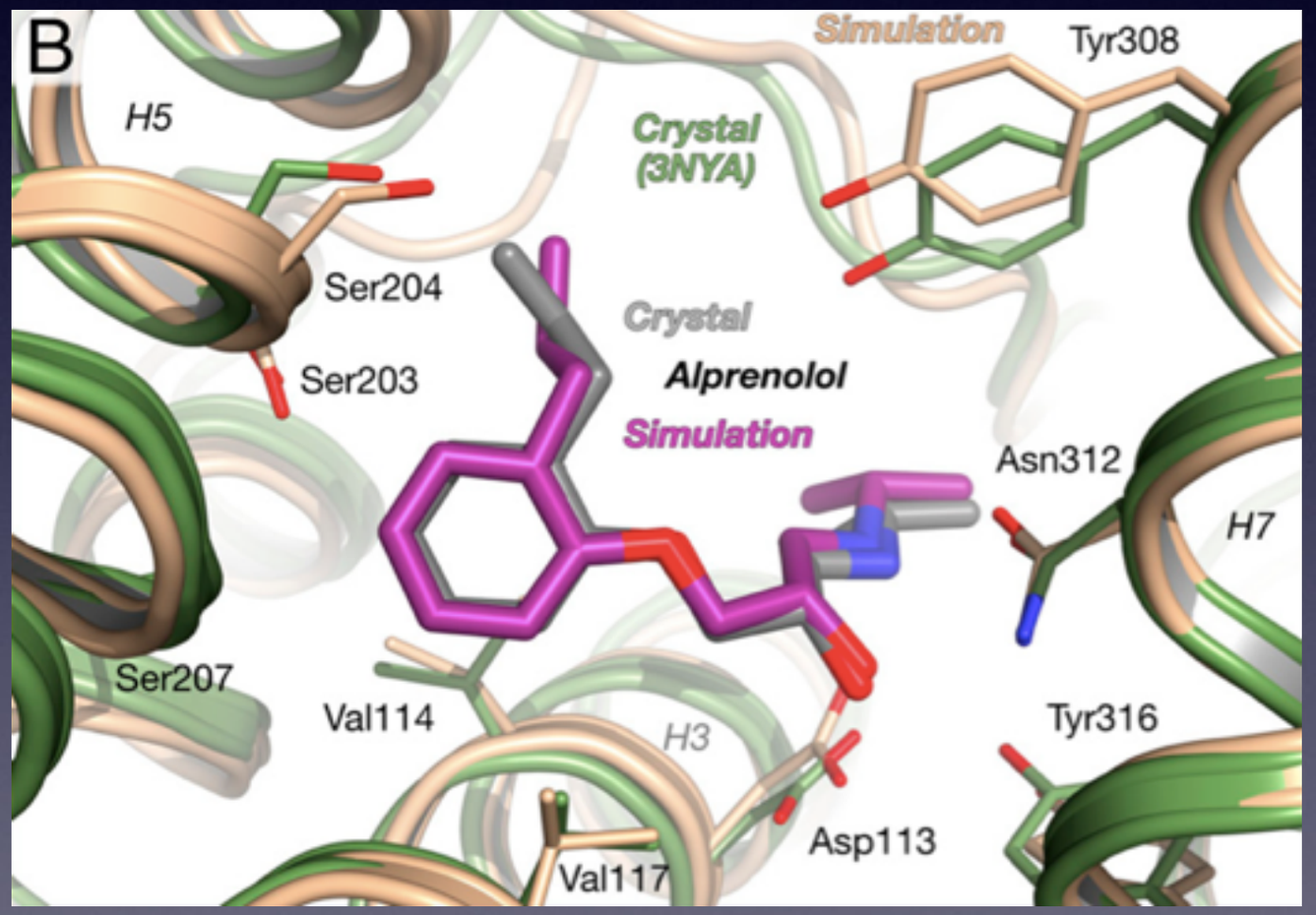



Dror et al., PNAS 2011

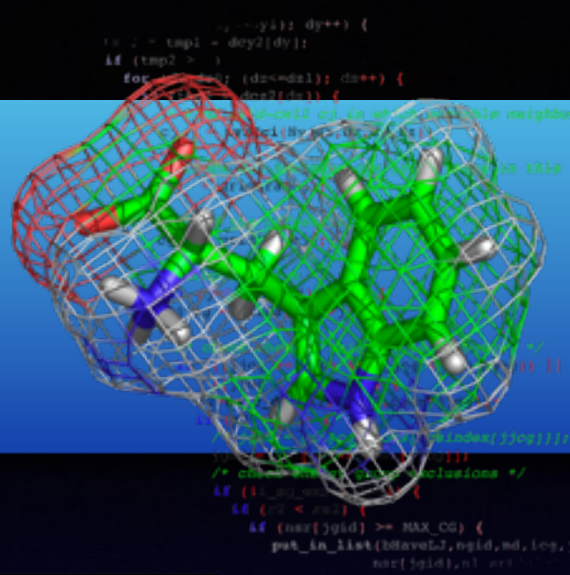
# Binding site

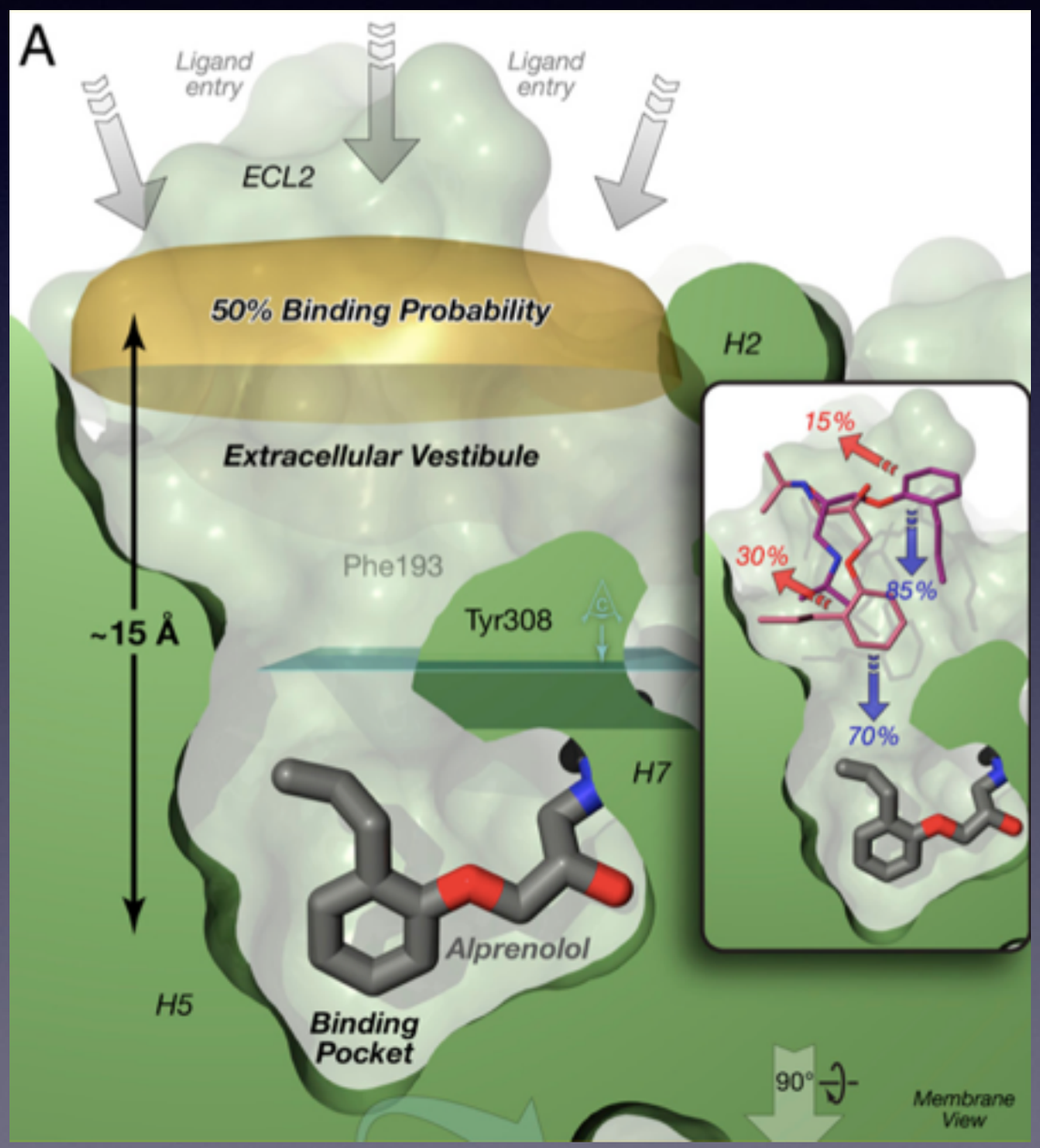


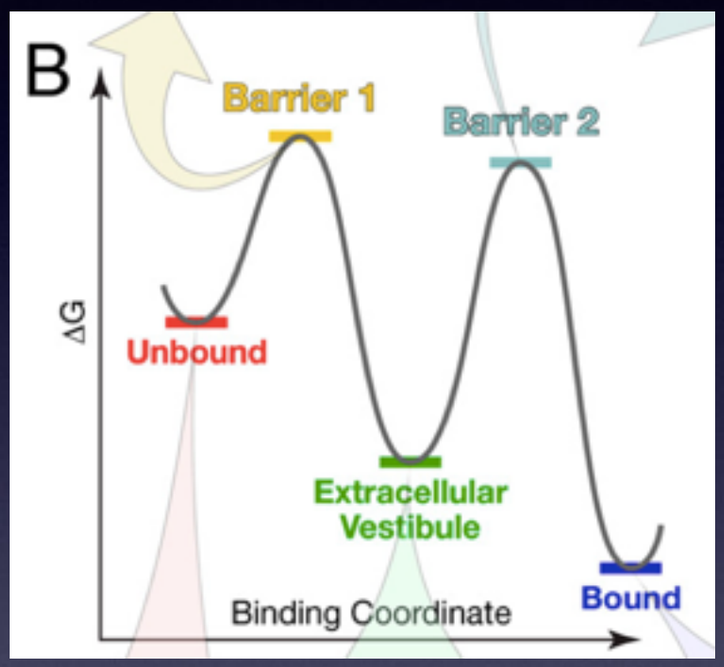


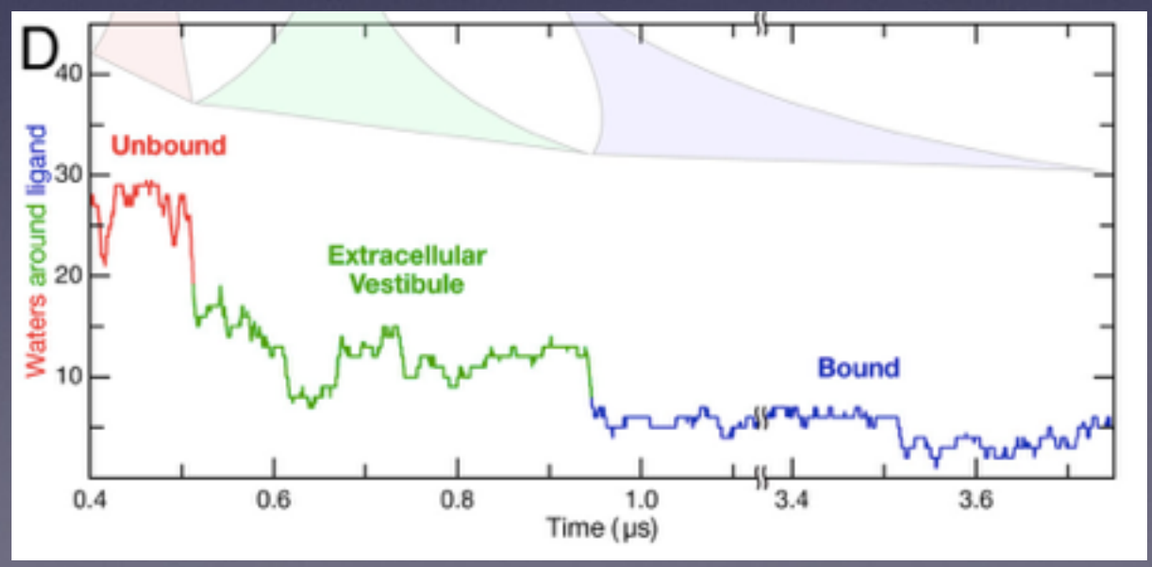



# Barrier 1

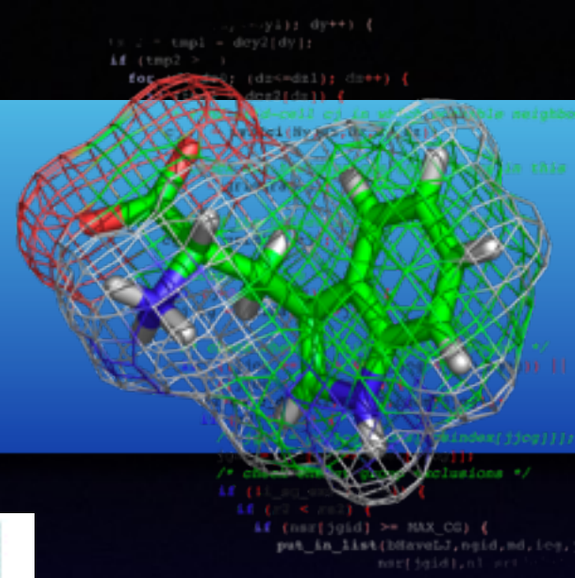


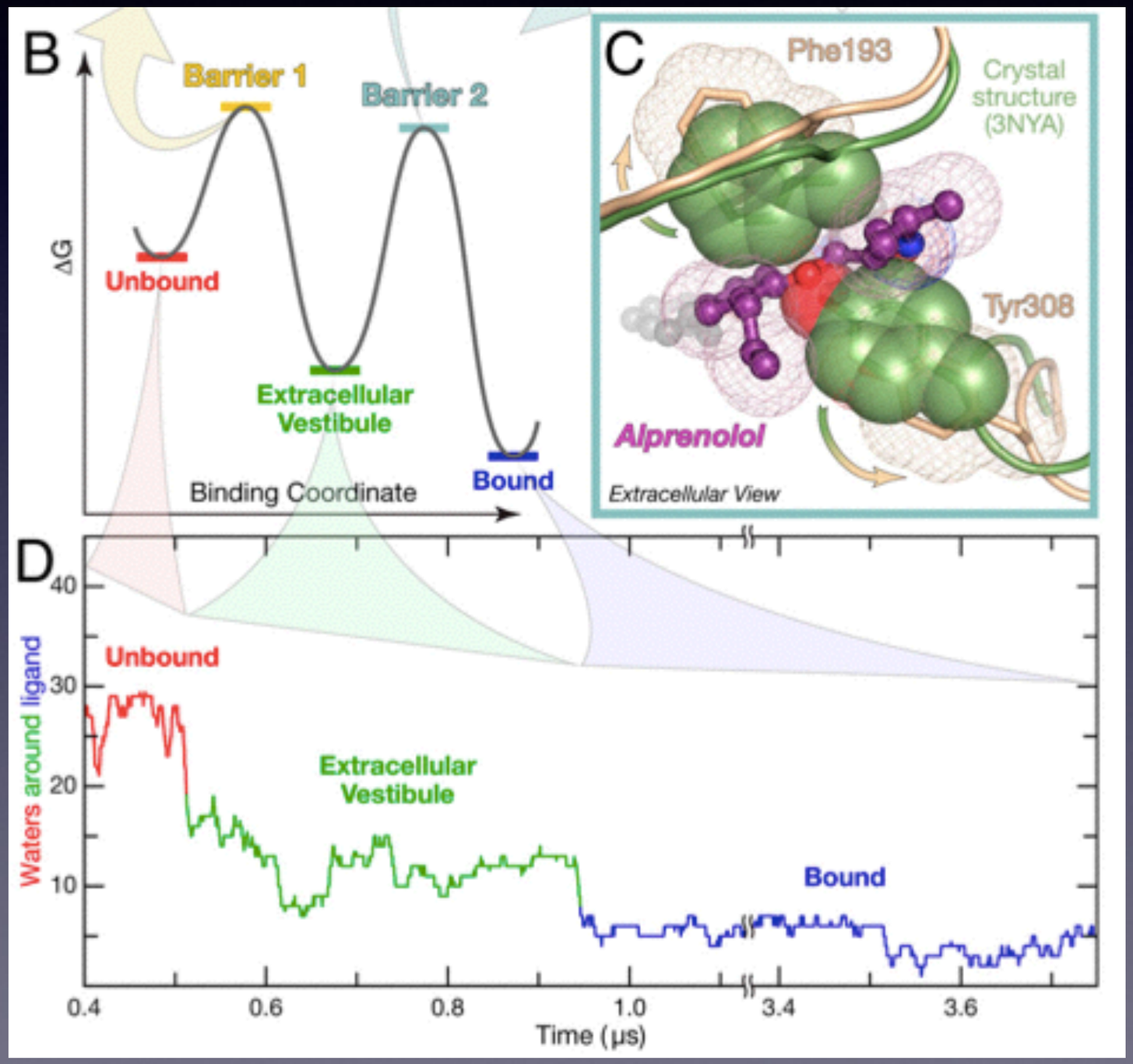

A
Ligand entry
Ligand entry
ECL2
50% Binding Probability
H2
Extracellular Vestibule
15%
30%
85%
70%
Phe193
Tyr308
~15 Å
H7
Alprenolol
H5
Binding Pocket
90°
Membrane View


B
Barrier 1
Barrier 2
ΔG
Unbound
Extracellular Vestibule
Bound
Binding Coordinate

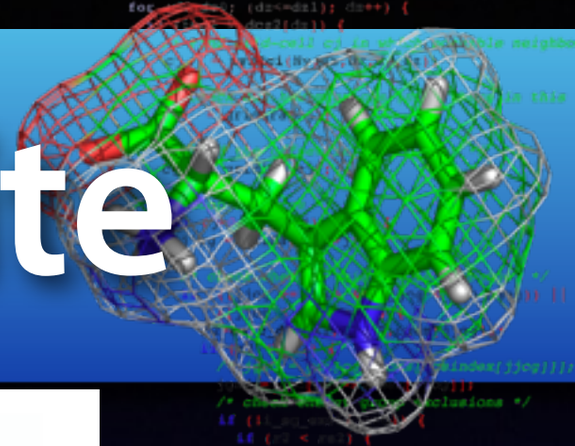

D
Waters around ligand
40
30
20
10
Unbound
Extracellular Vestibule
Bound
0.4
0.6
0.8
1.0
3.4
3.6
Time (μs)

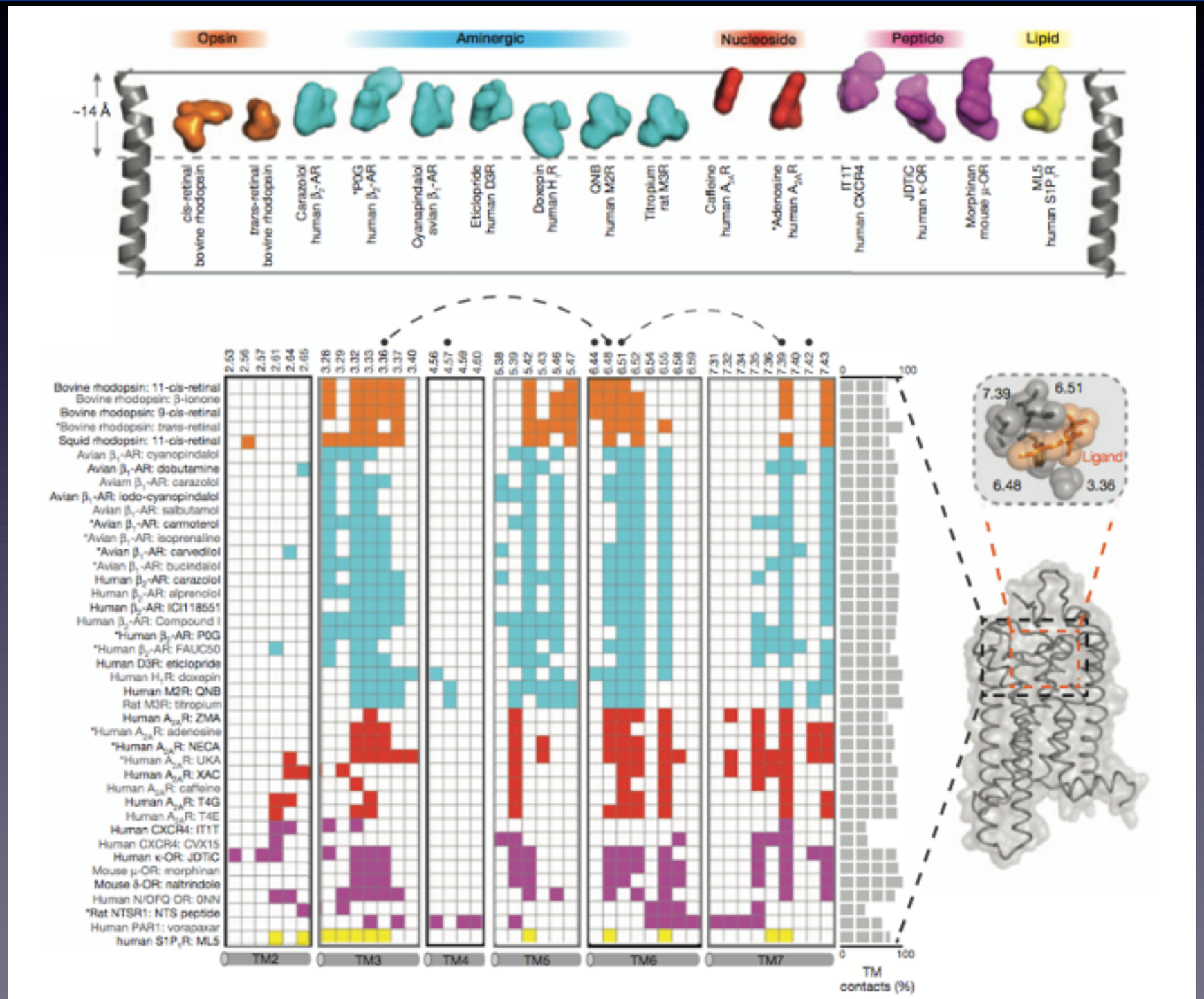
# Barrier 2



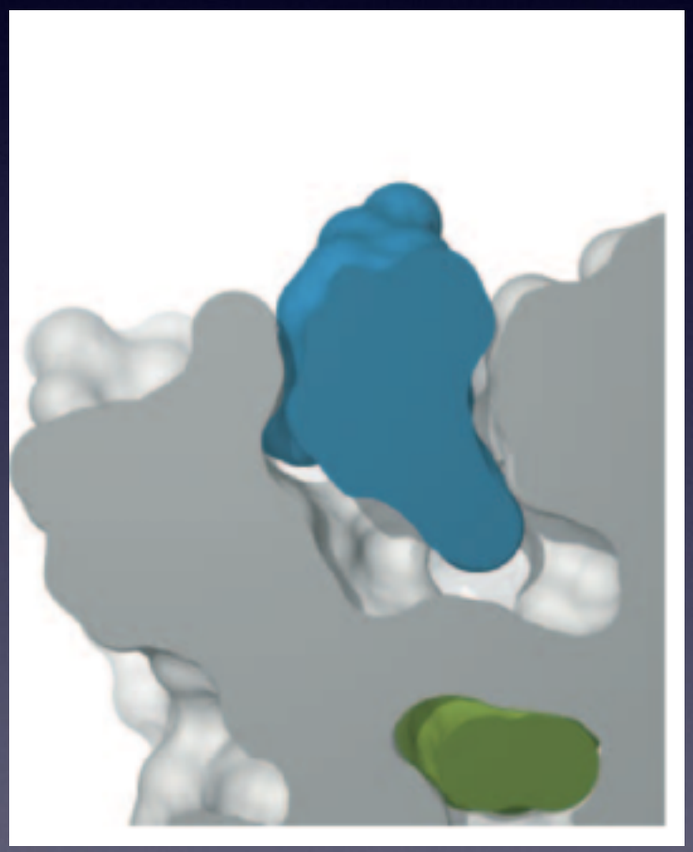

B
ΔG
Barrier 1
Barrier 2
Unbound
Extracellular Vestibule
Bound
Binding Coordinate
C
Phe193
Crystal structure (3NYA)
Tyr308
Alprenolol
Extracellular View
D
Waters around ligand
40
30
20
10
Unbound
Extracellular Vestibule
Bound
0.4
0.6
0.8
1.0
3.4
3.6
Time (μs)

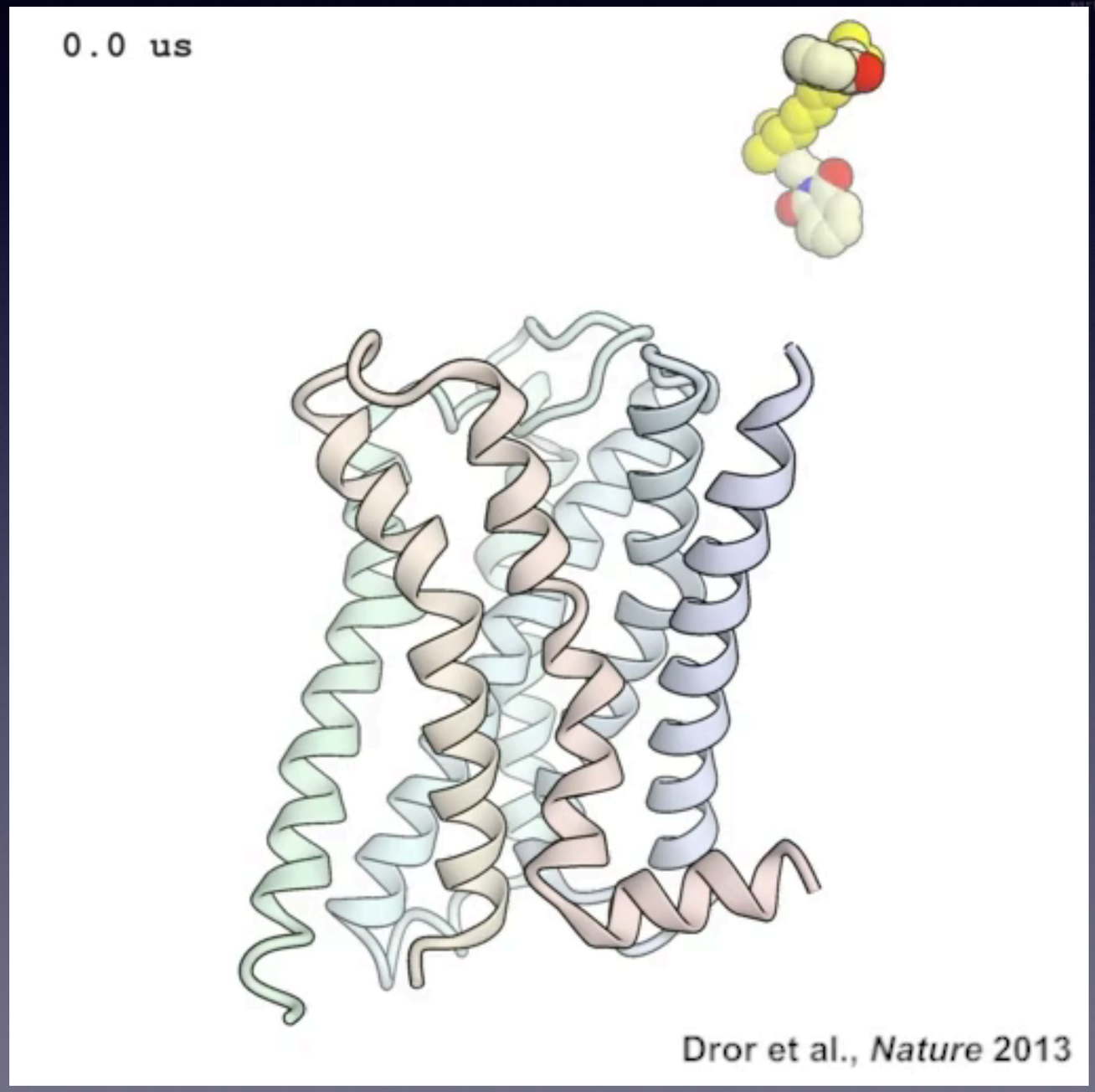
# Ligand-binding site



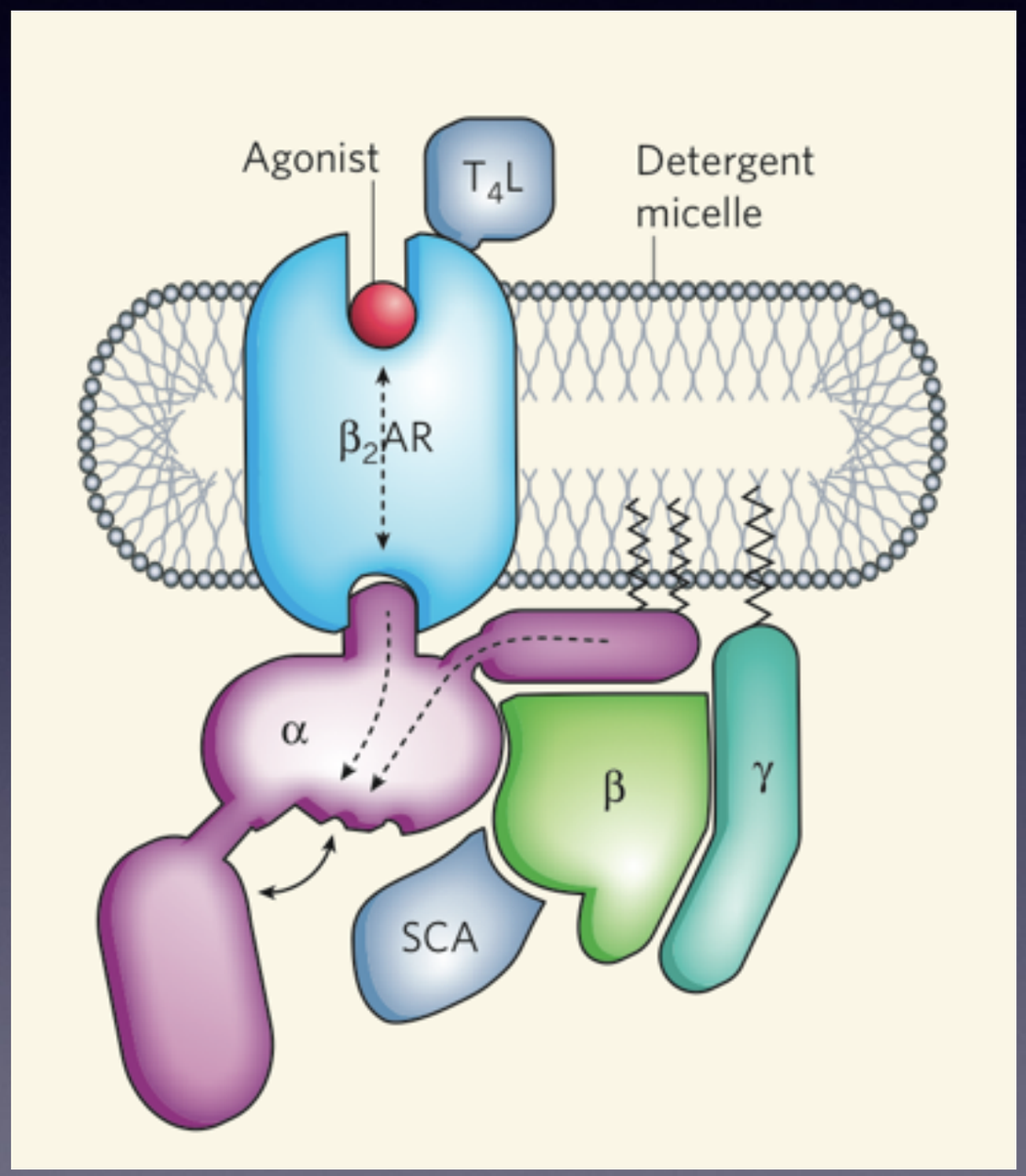


# Ortho/Allo-steric sites



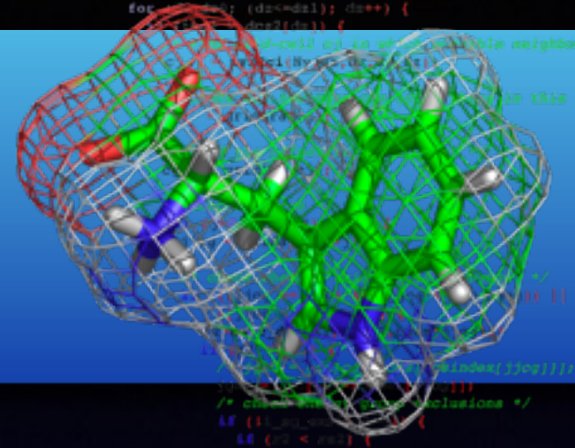


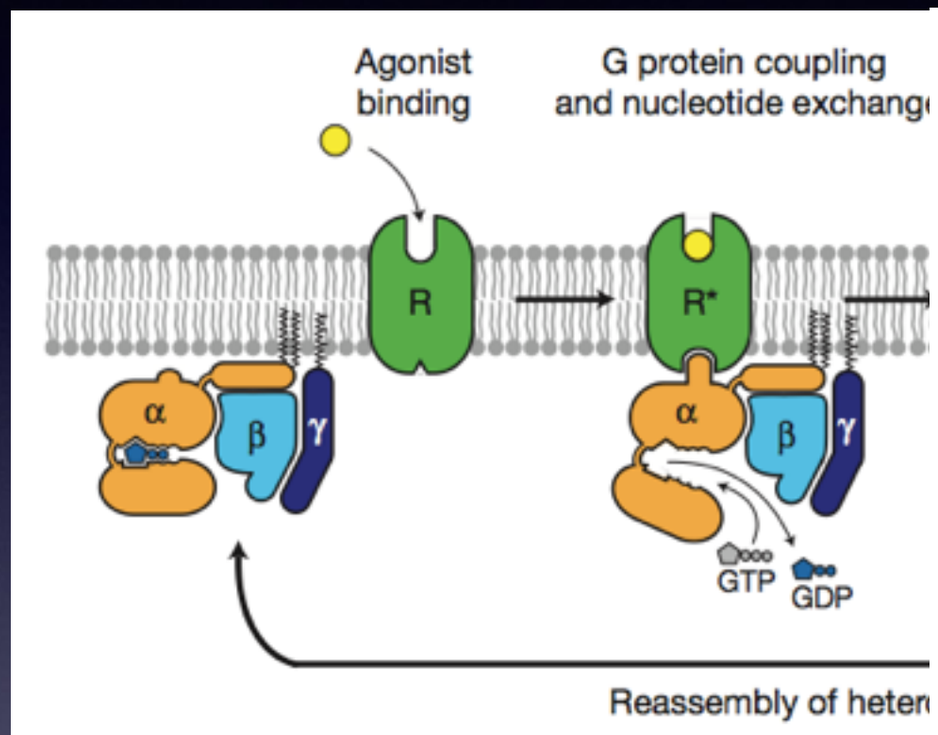
# Activation & Signalling

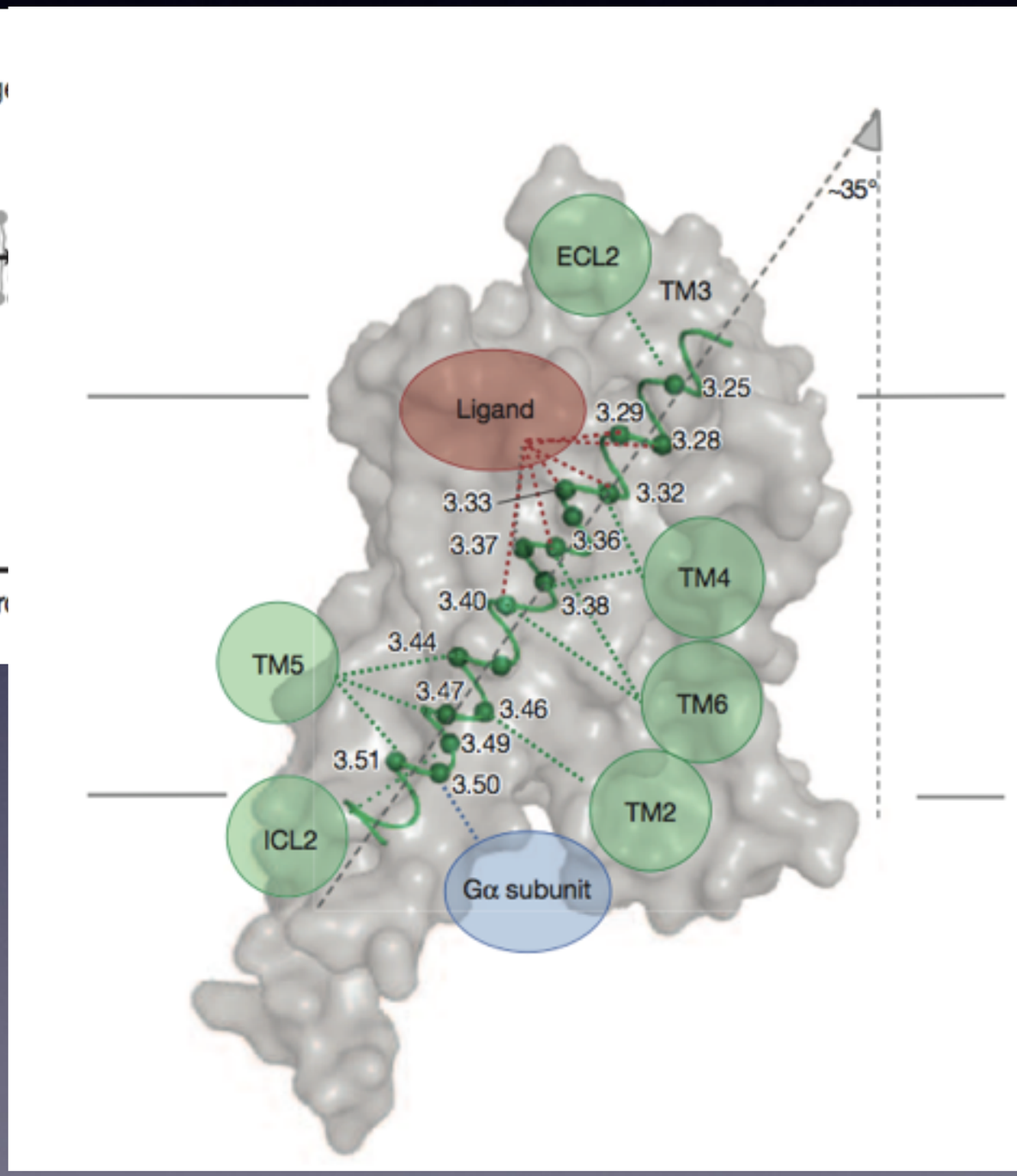



Kobilka 2011

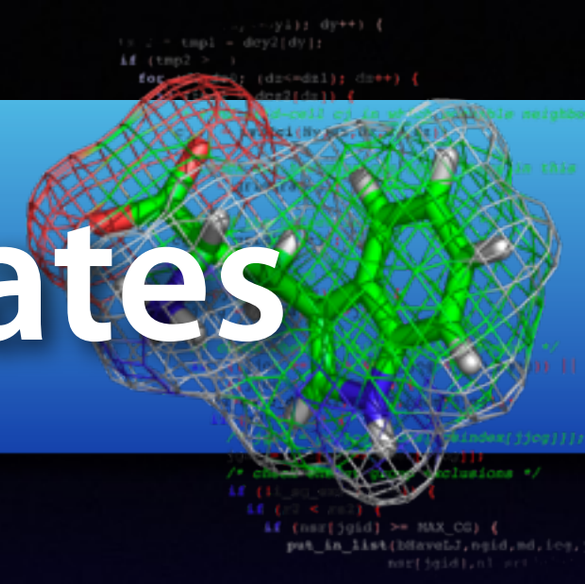
# Signalling

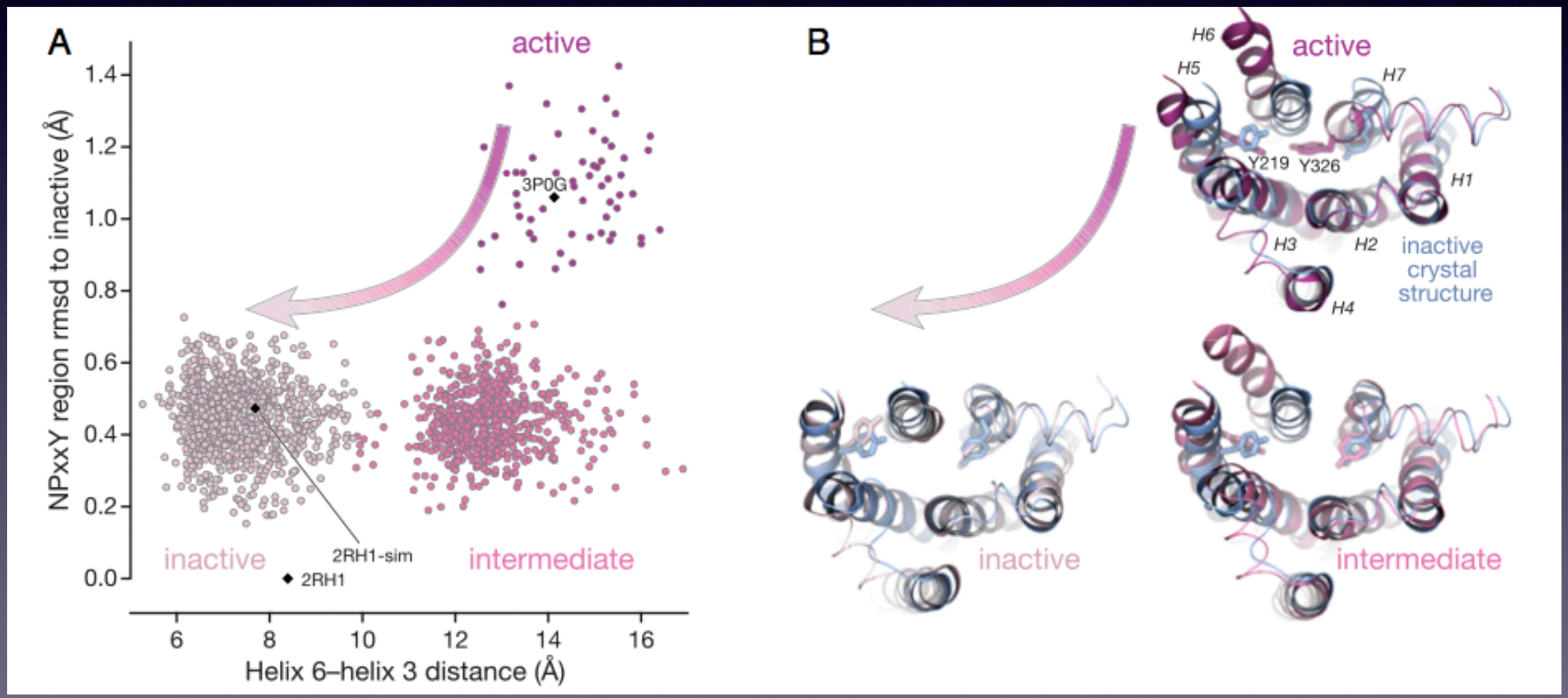



Agonist binding
G protein coupling and nucleotide exchang
R
R*
α
β
γ
α
β
γ
GTP
GDP
Reassembly of heter


~35°
ECL2
TM3
Ligand
3.25
3.29
3.28
3.33
3.32
3.37
3.36
TM4
3.40
3.38
3.44
TM5
3.47
3.46
TM6
3.49
3.51
3.50
TM2
ICL2
Gα subunit

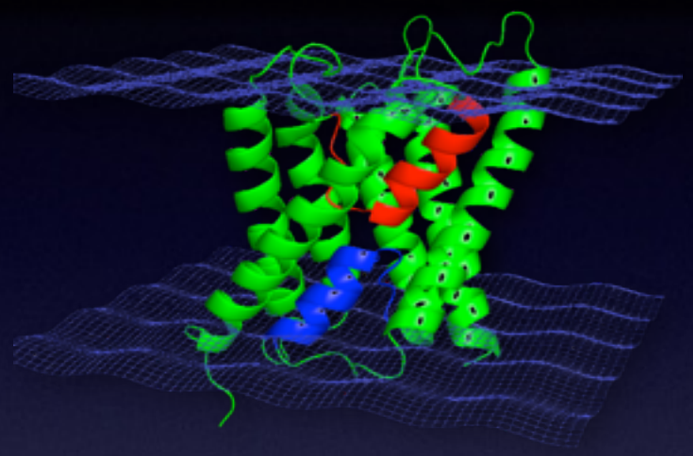
# Activation intermediates



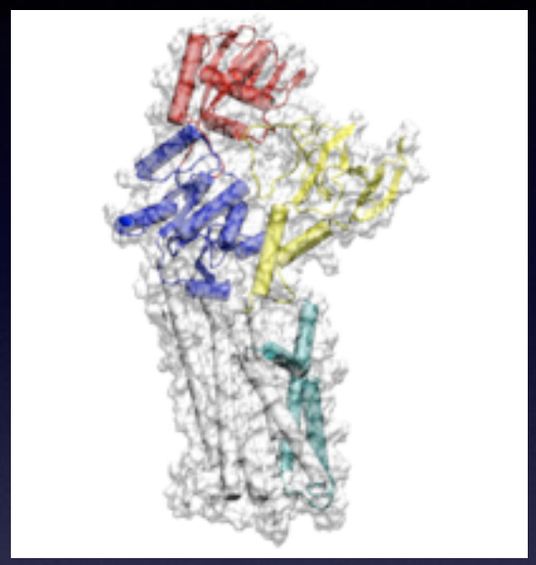



Dror et al. PNAS 2011

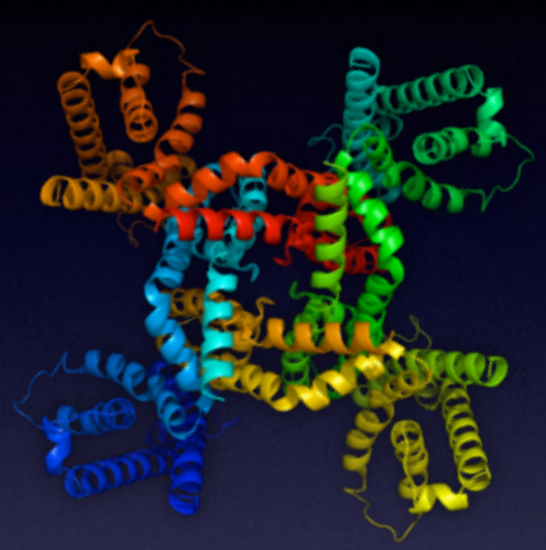
# Master thesis



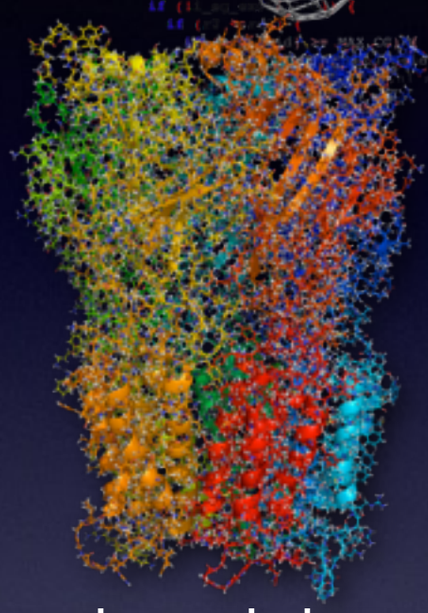
Aquaporins



P-type ATPases



Voltage-gated channels



Ligand-gated channels