# Using Supercomputers for Computer Modeling of Biomolecules & Drug Design

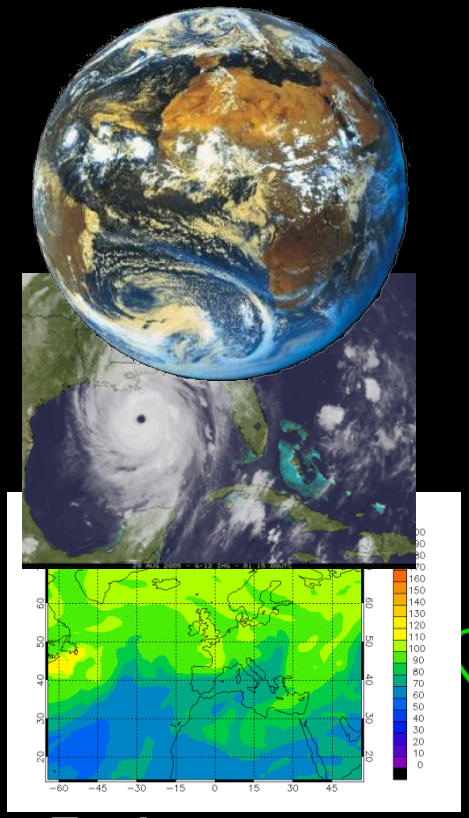
**Zoe Cournia**

*Biomedical Research Foundation, Academy of Athens*

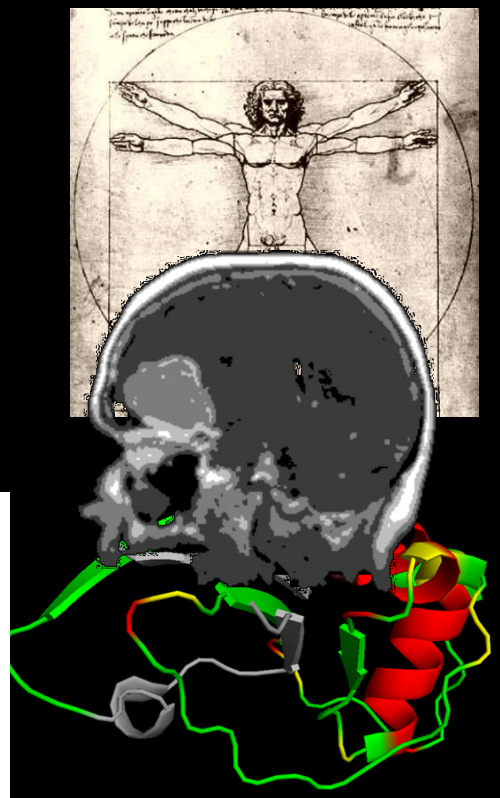
**VI-SEEM Life Sciences Training Event, 19 October 2016**

*zcournia@bioacademy.gr* *twitter: @zoecournia*

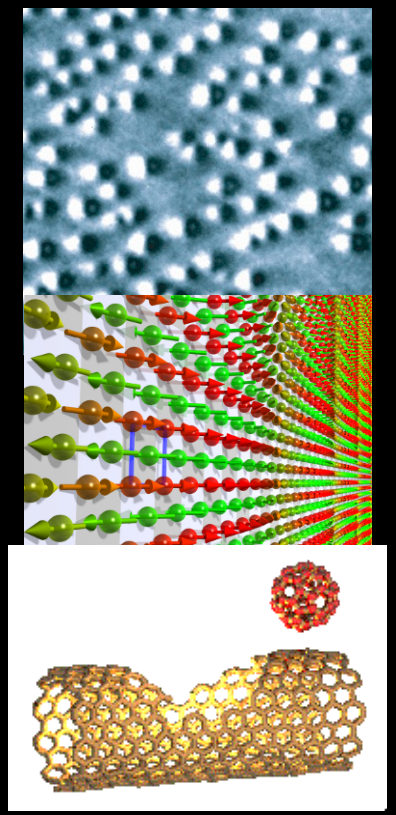
# Supercomputing Drives Science through Simulation



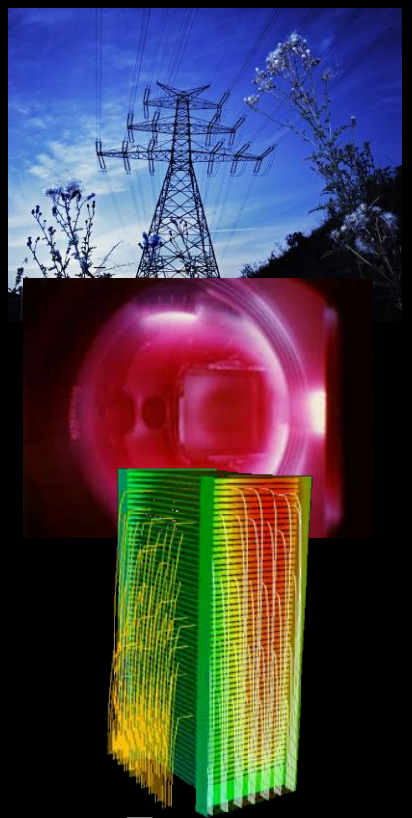
**Environment**
Weather/ Climatology
Pollution / Ozone Hole



**Finding Cures**
Medicine
Biology

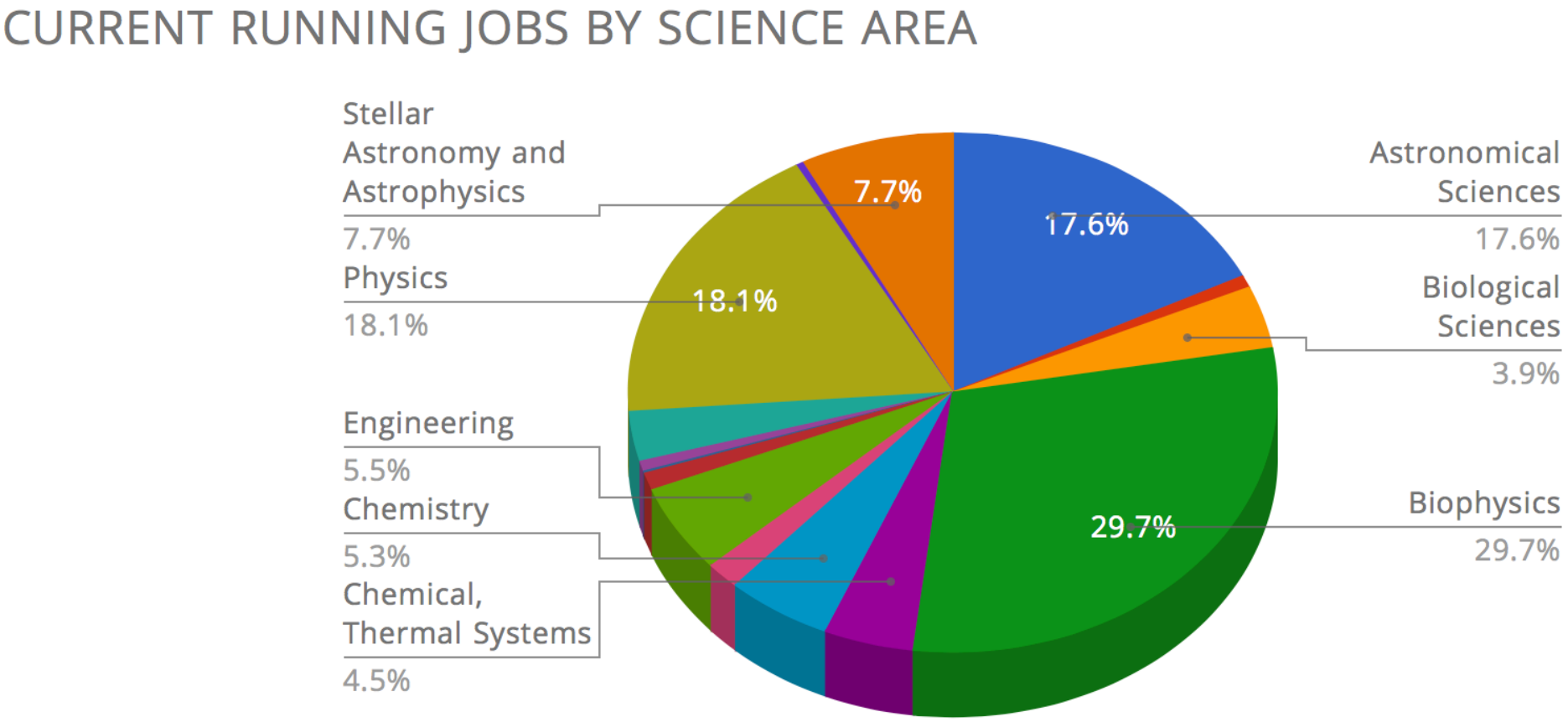


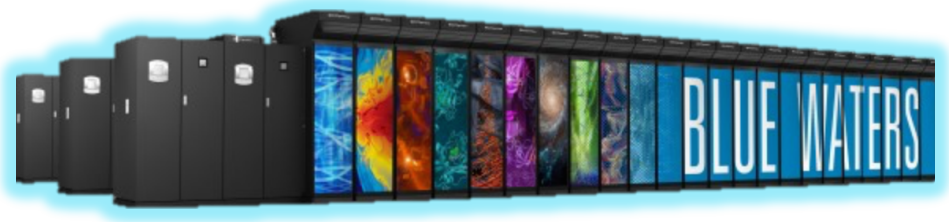
**Materials/ Inf. Tech**
Spintronics
Nano-science



**Energy**
Plasma Physics
Fuel Cells

# Distribution of HPC based on Science Area




Source: https://bluewaters.ncsa.illinois.edu

# Computing is transforming biomedical research




Exascale

Compute Power

time

**HP 735**
**12 CPUs**

1993

protein
10k atoms
100s ps

**SGI Origin**
**128 CPUs**

1997

ion channel
100k atoms
1 ns

**LeMieux**
**3k CPUs**

2002

ATPase
500k atoms
10s ns

**Ranger**
**60k CPUs**

2007

ribosome
2 mil atoms
100s ns

**LCF**
**Mira 786k cores**
**Titan 280k cores**
**+ GPUs**

2015

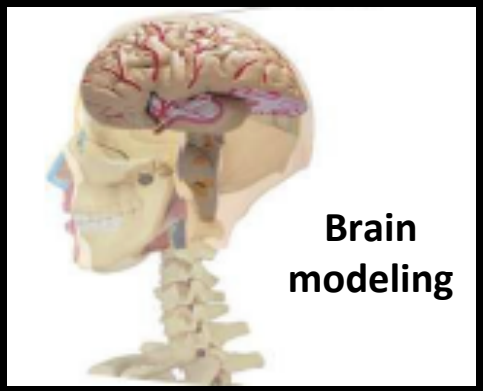
Enveloped virus
200 mil+ atoms
1-100 μs

**Performance (in FLOPS):**

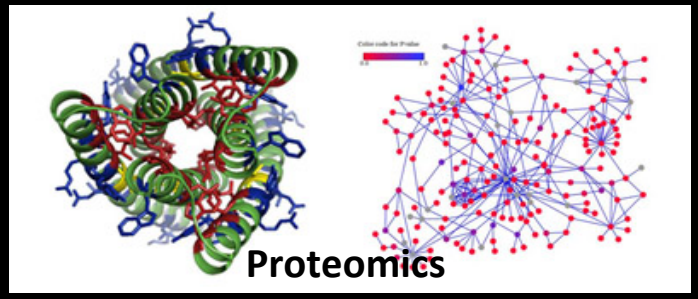
| | |
|---|---|
| Megaflop | $10^6$ |
| Gigaflop | $10^9$ |
| Teraflop | $10^{12}$ |
| Petaflop | $10^{15}$ |

Source: adapted from Prof. Rommie Amaro

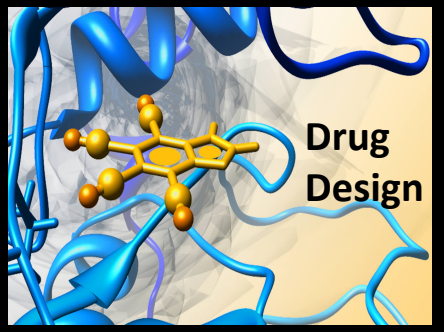
# Key areas of biomedical research where HPC is key

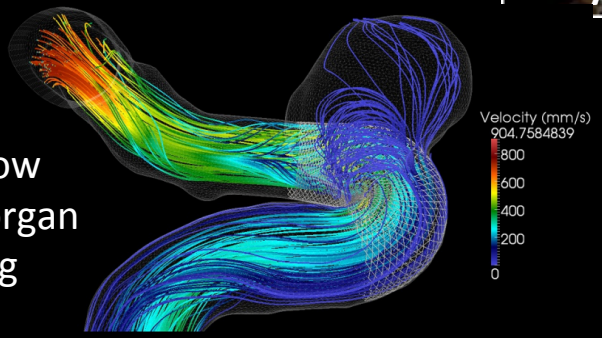

Protein Biophysics

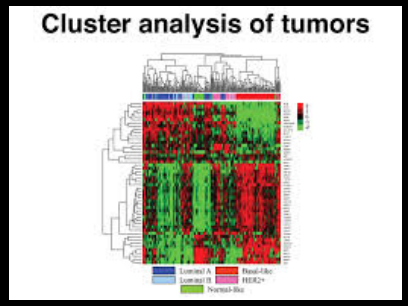




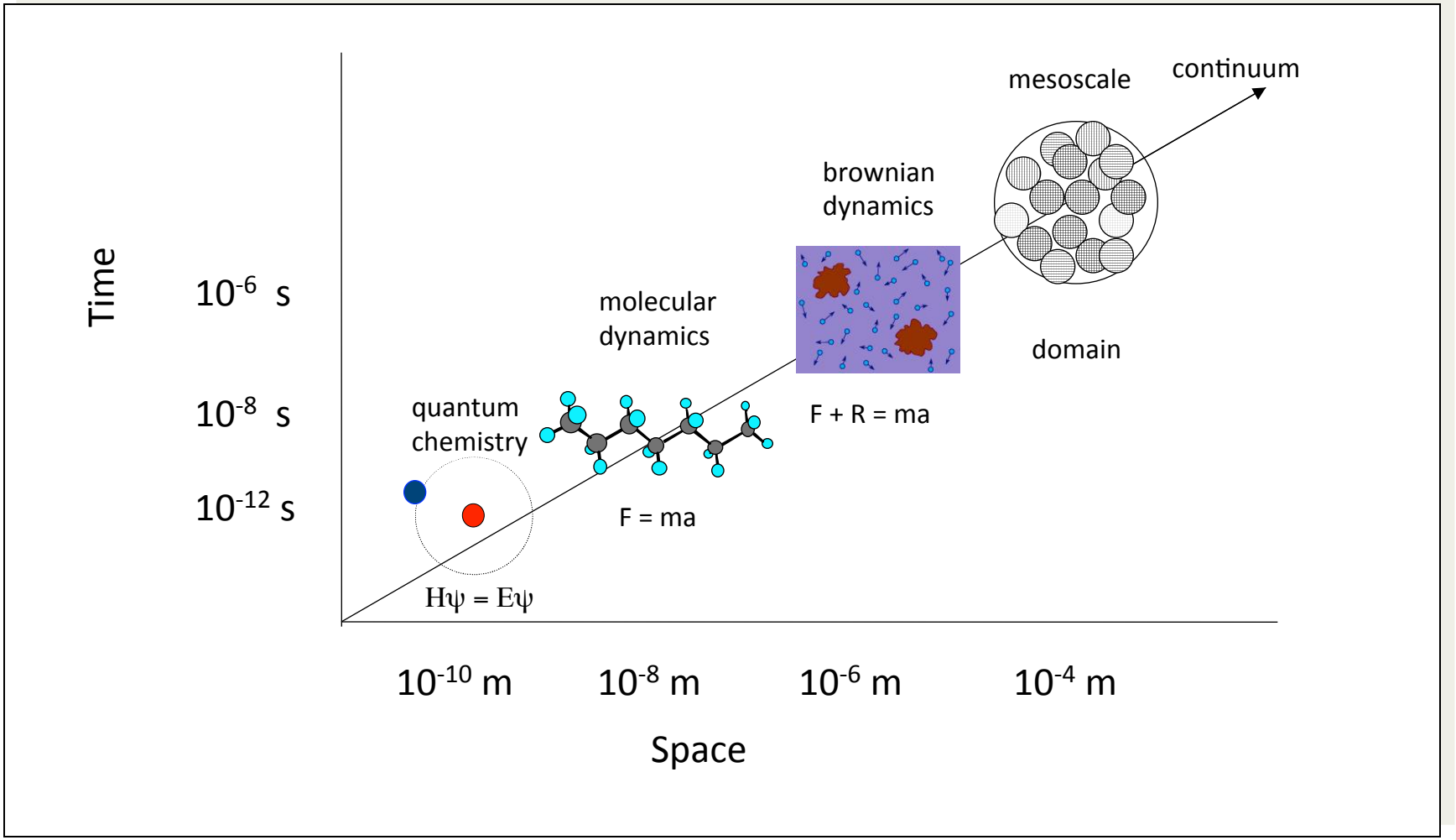



Blood flow
Whole organ
modeling



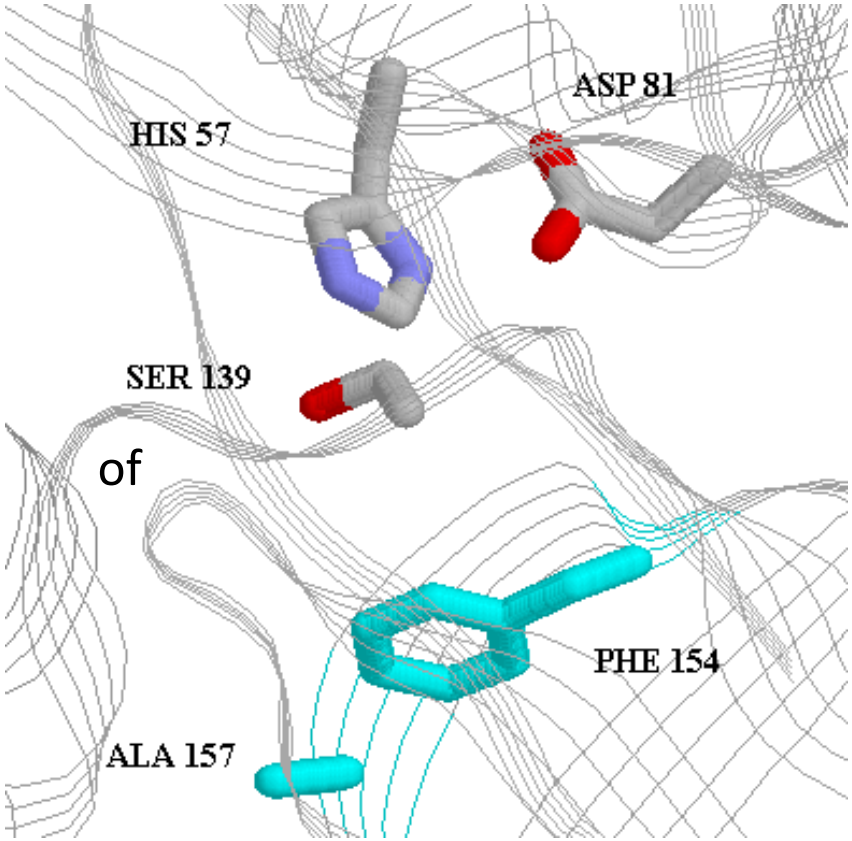
# Molecular Simulations across scales



# Molecular Modeling

*dynamics*

Structure --------> function

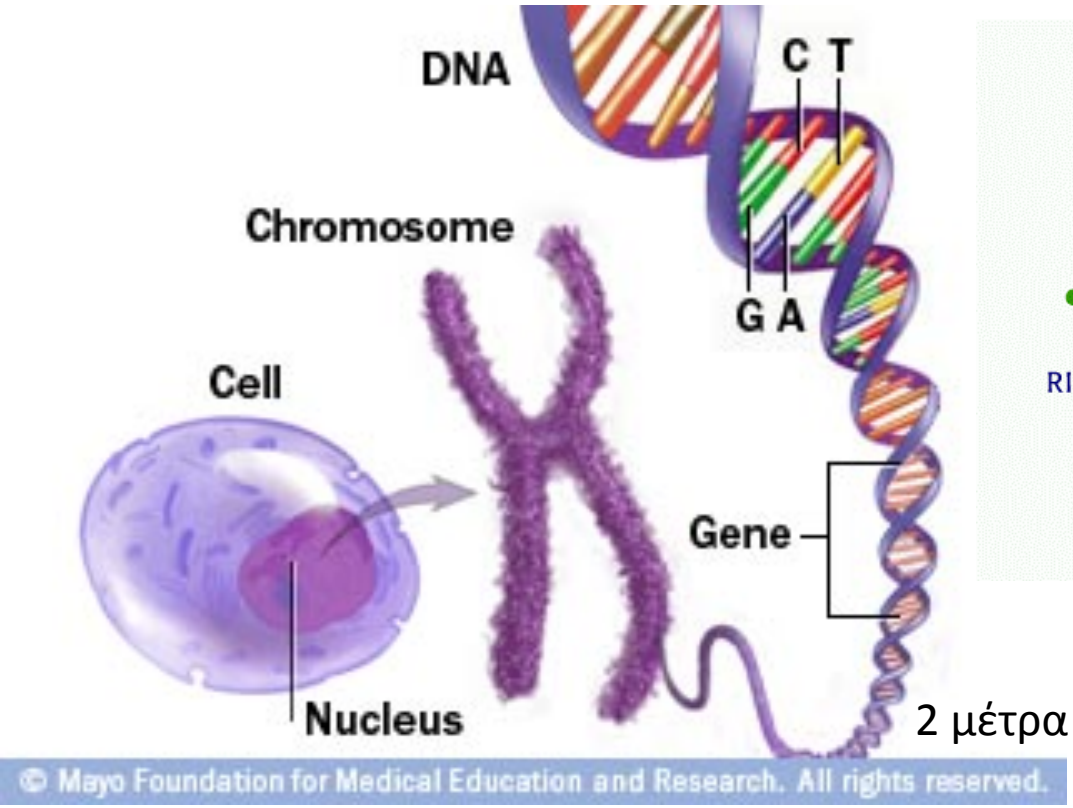




## Molecular Dynamics

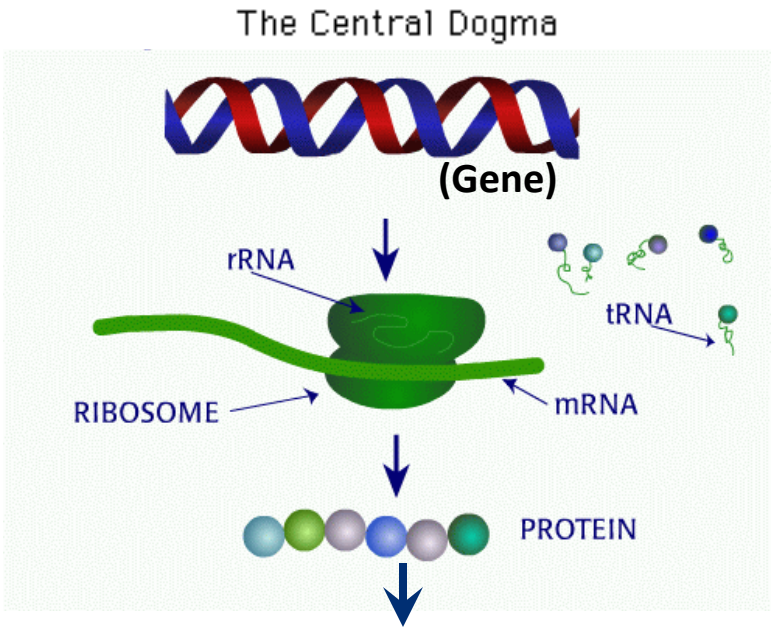
of

- molecular/atomic level picture structure and dynamics
- property prediction
- ion transport
- solvent effects
- protein stability / conform. changes, …

# From DNA, to genes and proteins




2 μέτρα!

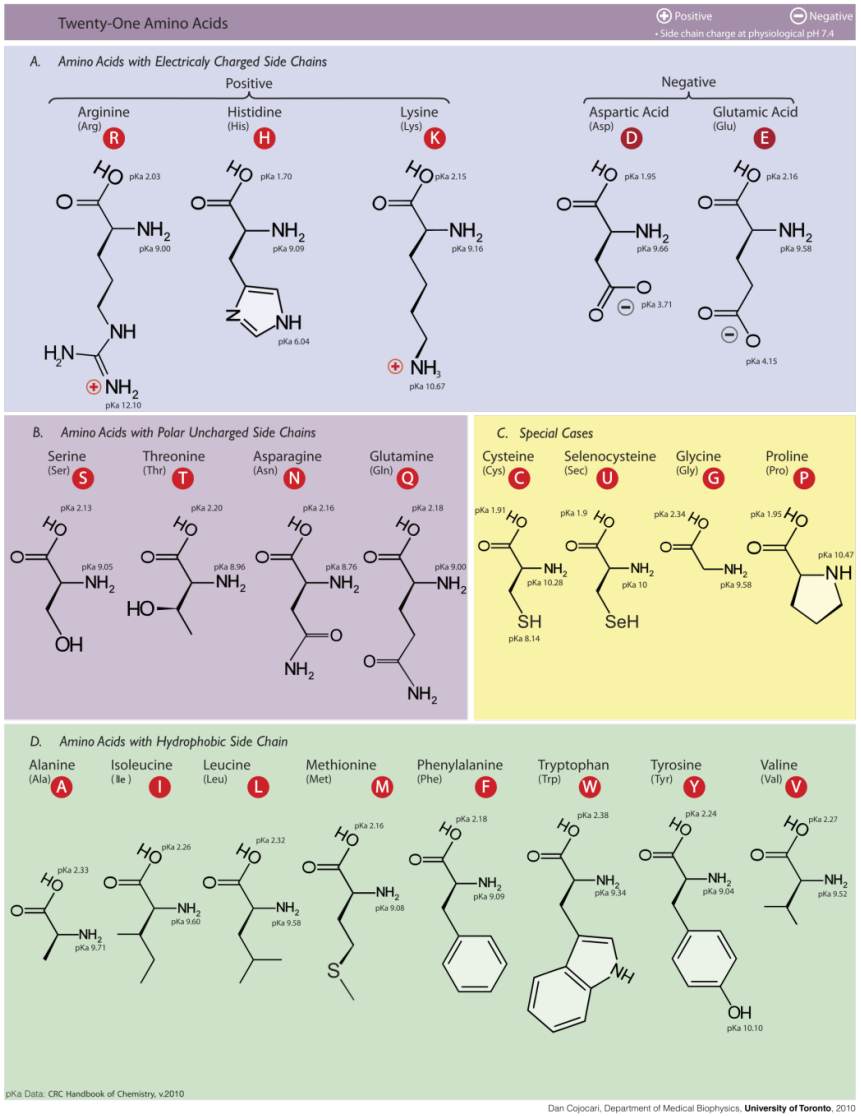


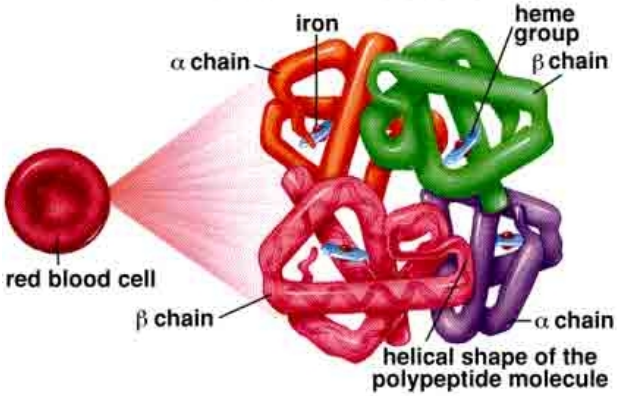

*X - Rays = Protein Pictures*

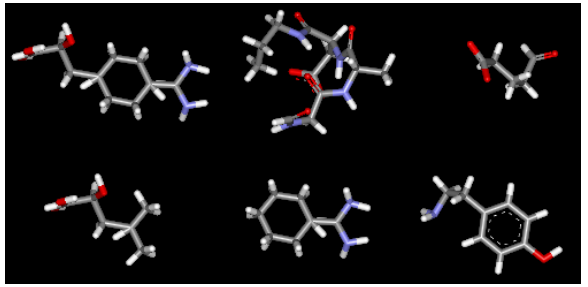
**20.000 genes in the nuclei of our cells**
**→ PROTEINS**

- **Proteins are the means of expression of genes to functional molecules**
- **Proteins perform essential functions in the cell**

# Protein Modeling, Protein-Drug Modeling



**Drugs associate with proteins through Intermolecular Interactions!**

***Hydrogen Bonds***

***Electrostatic Interactions***

***van der Waals Forces***

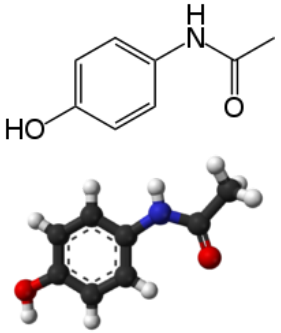
***π – π Interactions***

# Drugs

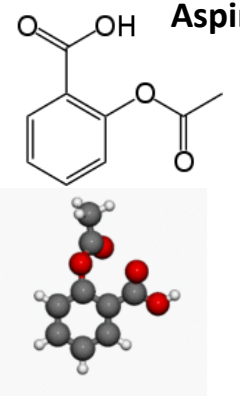
**Normally they are small organic molecules**

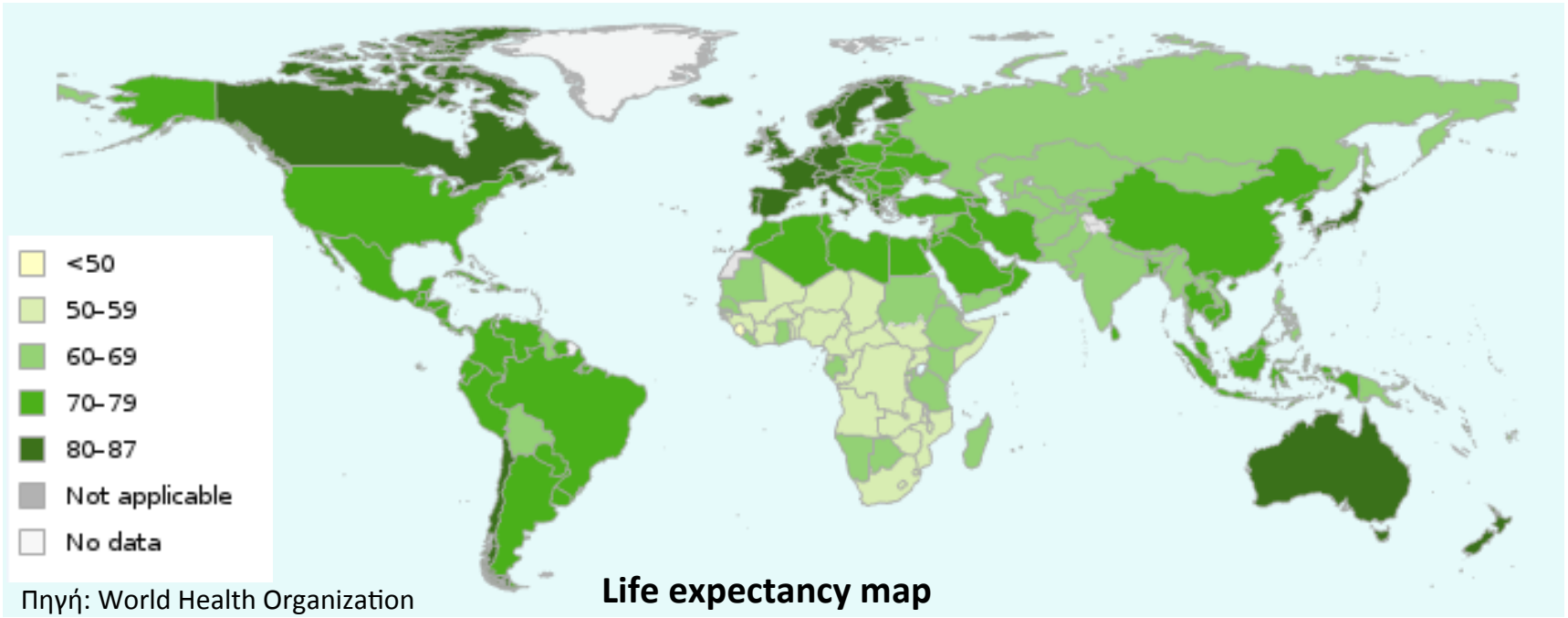
- **Therapy**
- **Relief**
- **Prevention**
- **Quality of life improvement**
- **Life expectancy prolongation**

**Paracetamol (Depon)**




**Aspirin**

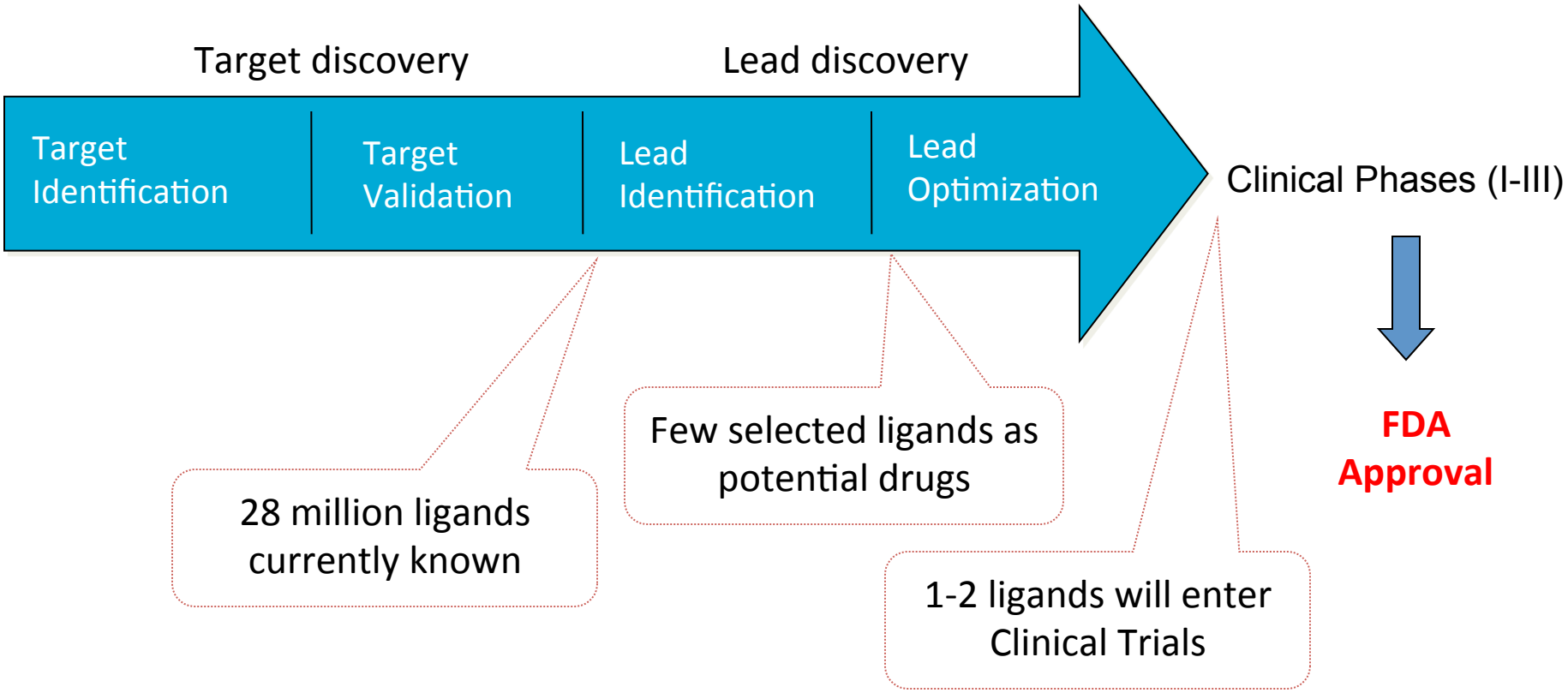



Πηγή: World Health Organization

**Life expectancy map**

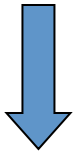
# Phases of Pharmaceutical Development




Duration: 12 – 15 years, Cost: ~ 1 billion US $

# Traditional Drug Discovery

- Random screening of hundreds of thousands of molecules with High Throughput Screening (HTS) for combating the pathogen
- Random discoveries (i.e. penicillin, viagra)
- Trying out existing drugs and modifications
- Estimated number of small molecules that can act as drugs **$10^{66}$**
- Estimated number of atoms in the world **$10^{50}$**

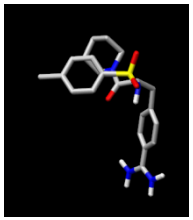


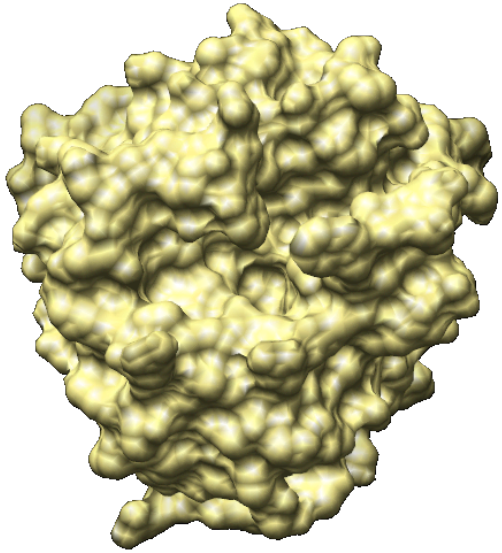
**Structure-based approaches + Targeted Therapy**

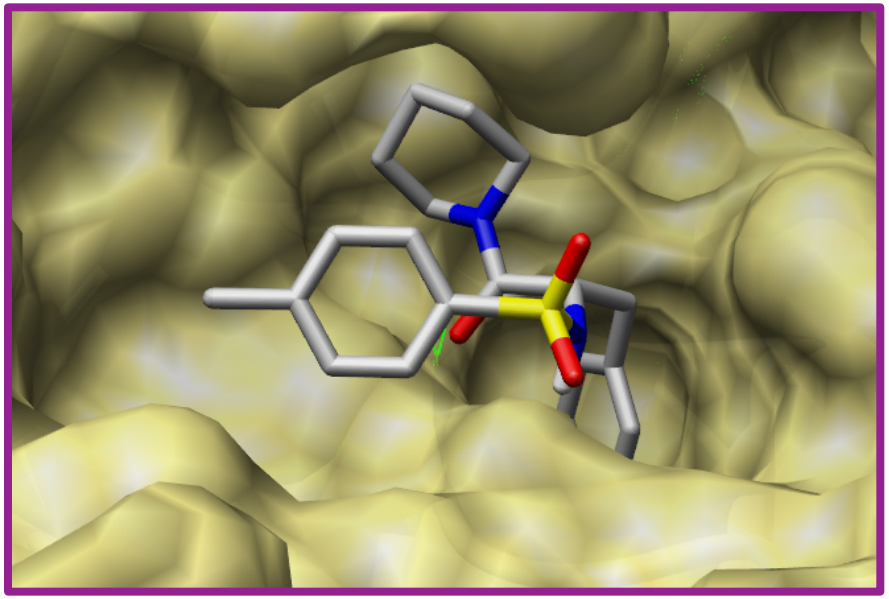
# *Rational* Drug Discovery

- **Identify important genes for a diseases**
- **Targeting/inactivating genes (proteins) of the pathogen with small molecules = drugs**

**TARGETED THERAPY!**

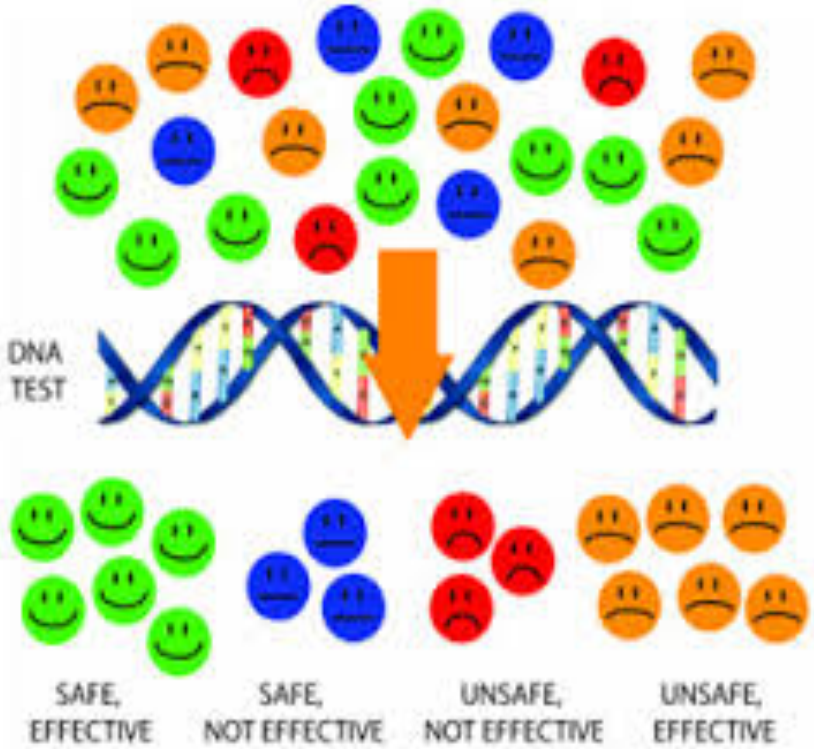







Curr Opin Drug Discov Devel. 2002 May; 5(3): 355–360

# The era of Personalized Medicine




**Lung Cancer**

↓

genotyping

↓

4% of patients with
non-small cell lung carcinoma
Rearrangement in **ALK protein**

↓

carcinogenesis

↓

Drug design for this specific
subset of patients

↓

**Crizotinib for ALK+**
**lung cancer patients**

# Personalized Medicine: new drug generation

Table 1 | **Selected oncology agents in Phase III biomarker-driven clinical trials**

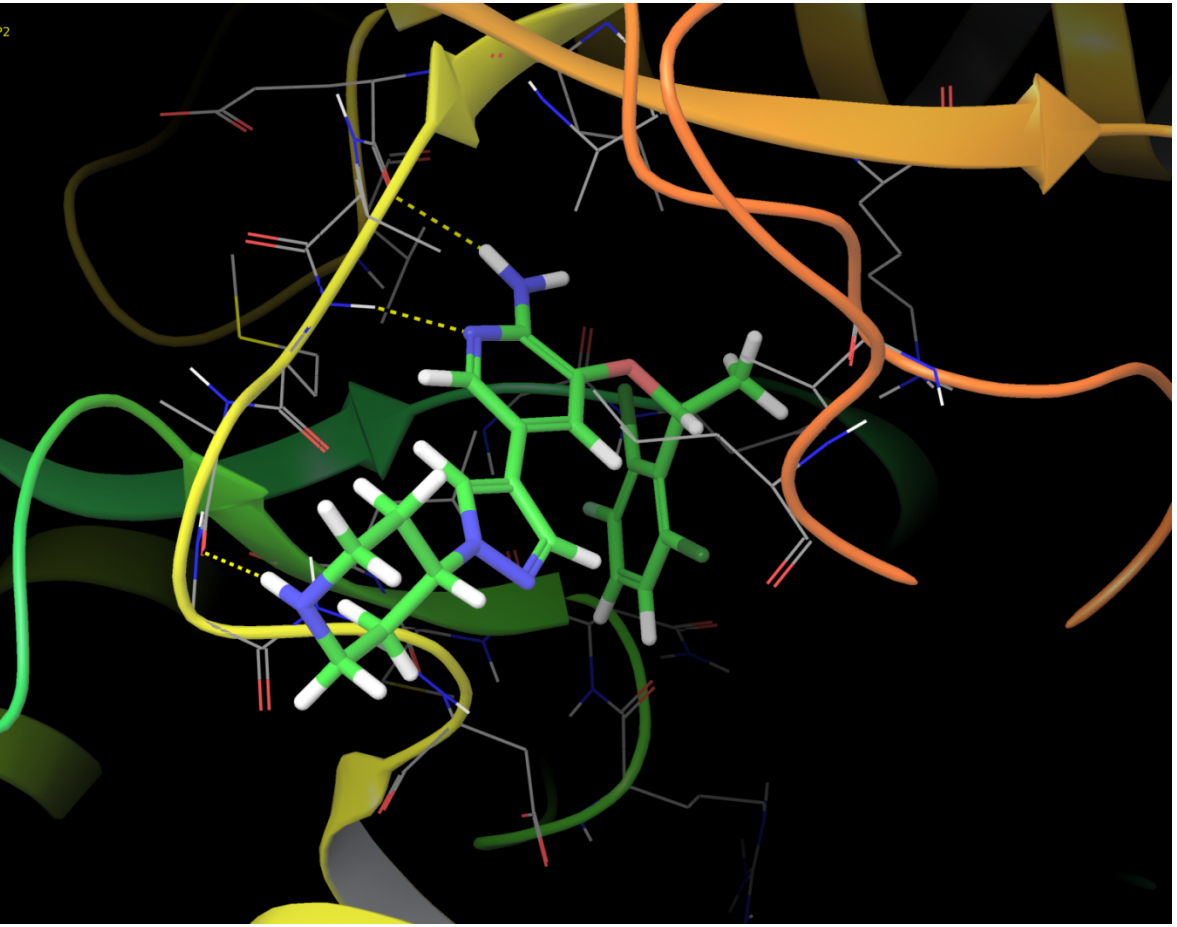
| Drug | Company | Indication | Biomarker |
|---|---|---|---|
| *Precedented biomarkers* | | | |
| Iniparib | Sanofi/BiPar Sciences | Breast cancer | Triple-negative |
| Pertuzumab | Roche/Chugai | Breast cancer | HER2 |
| Neratinib | Pfizer | Breast cancer | HER2 |
| Bosutinib | Pfizer | CML | Philadelphia |
| Nimotuzumab | YM BioSciences | Breast cancer | HER2 |
| Afatinib | Boehringer Ingelheim | NSCLC | EGFR |
| Dacomitinib | Pfizer | NSCLC | EGFR and KRAS |
| *Novel biomarkers* | | | |
| Midostaurin | Novartis | AML | FLT3 |
| Cilengitide | Merck Serono | Glioblastoma | Methylated *MGMT* |
| Trabedersen | Antisense Pharma | Glioma | TGFβ2 |
| GSK2118436 | GlaxoSmithKline | Melanoma | BRAF |
| GSK1120212 | GlaxoSmithKline | Melanoma | BRAF |

AML, acute myeloid leukaemia; CML, chronic myeloid leukaemia; EGFR, epidermal growth factor receptor; FLT3, FMS-like tyrosine kinase 3; MGMT, 6-*O*-methylguanine-DNA methyltransferase; NSCLC, non-small-cell lung cancer; TGFβ2, transforming growth factor-β2.

Chiang and Million, Nature 2011

# Crizotinib (Xalkori, Pfizer)

**Structure of the anaplastic lymphoma kinase (ALK)**
**Complexed with the drug crizotinib – (PDB ID: 2XP2)**



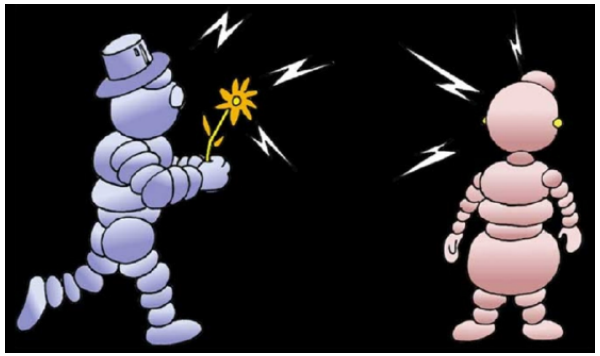
**Protein-Ligand interactions:**

**Intermolecular Interactions (Enthalpy)**

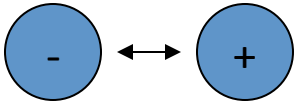
***Hydrogen Bonds***
***Electrostatic Interactions***
***van der Waals Forces***
***π – π Interactions***

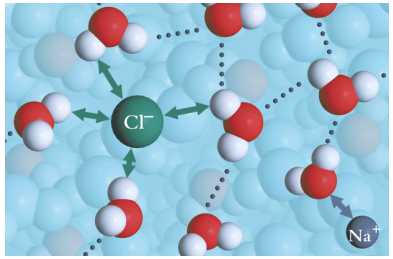
**Entropy**

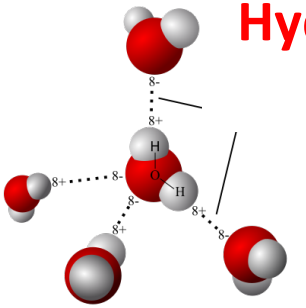
# Intermolecular Interactions



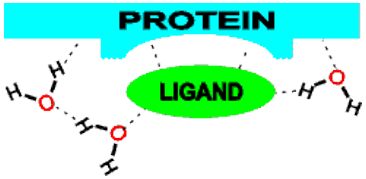
## Electrostatic Interactions

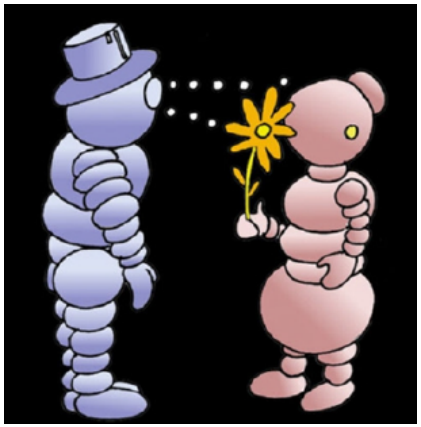


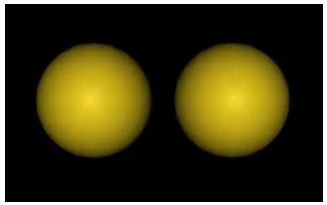



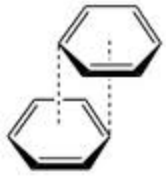
## Hydrogen Bonds

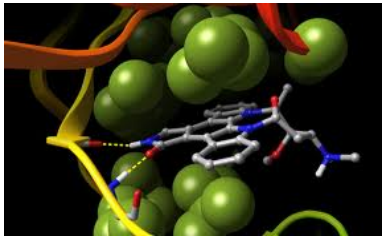


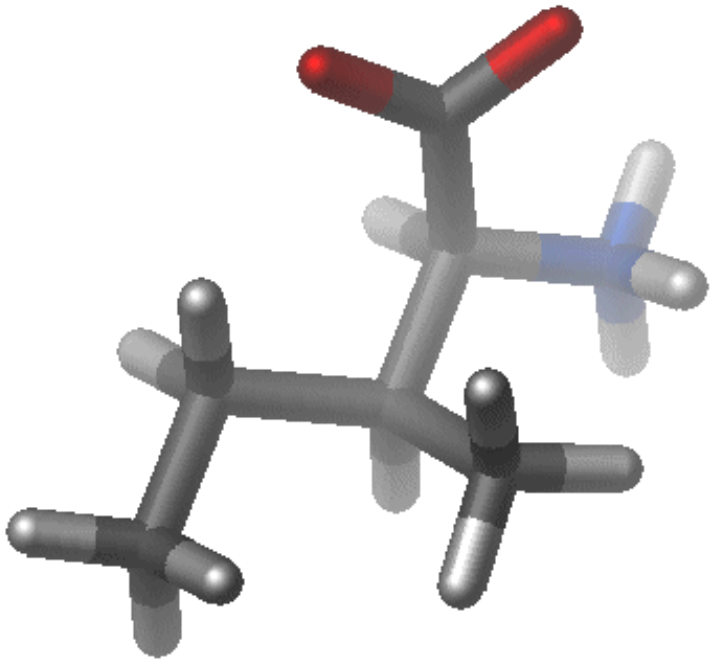
## van der Waals Forces
## π – π, cation - π interactions







# Molecular Simulations?



**Nobel Prize in Chemistry 2013**

Nobel Prizes and Laureates
Chemistry Prizes ▼ 2013

▼ About the Nobel Prize in Chemistry 2013
Summary
Prize Announcement
Press Release
Advanced Information
Popular Information
Greetings

▶ Martin Karplus
▶ Michael Levitt
▶ Arieh Warshel

All Nobel Prizes in Chemistry
All Nobel Prizes in 2013

The Nobel Prize in Chemistry 2013
Martin Karplus, Michael Levitt, Arieh Warshel

# The Nobel Prize in Chemistry 2013


© Nobel Media AB
**Martin Karplus**


Photo: Keilana via Wikimedia Commons
**Michael Levitt**


Photo: Wikimedia Commons
**Arieh Warshel**

**The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.**

# MD Simulations study structure + dynamics

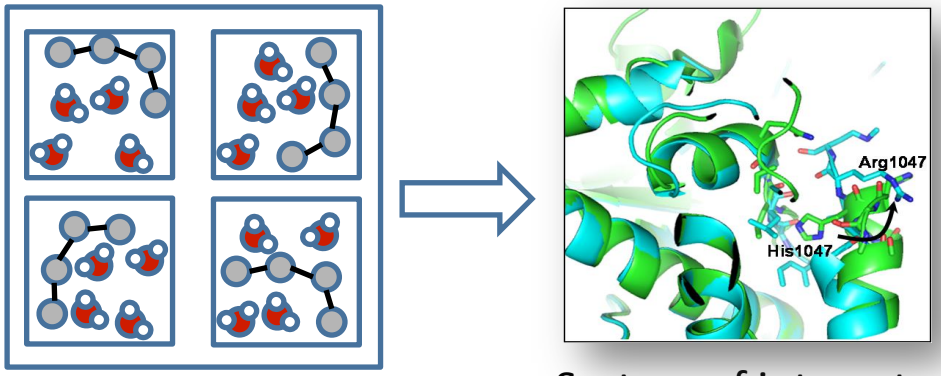
Is there a fast and efficient way to study the structure and dynamics of biomolecules in atomic-level detail?

## Molecular Dynamics simulations

**Step 1.** Model the potential energy and use coordinates from experimental structures and assign initial velocities (Etotal = Epotential + Ekinetic)

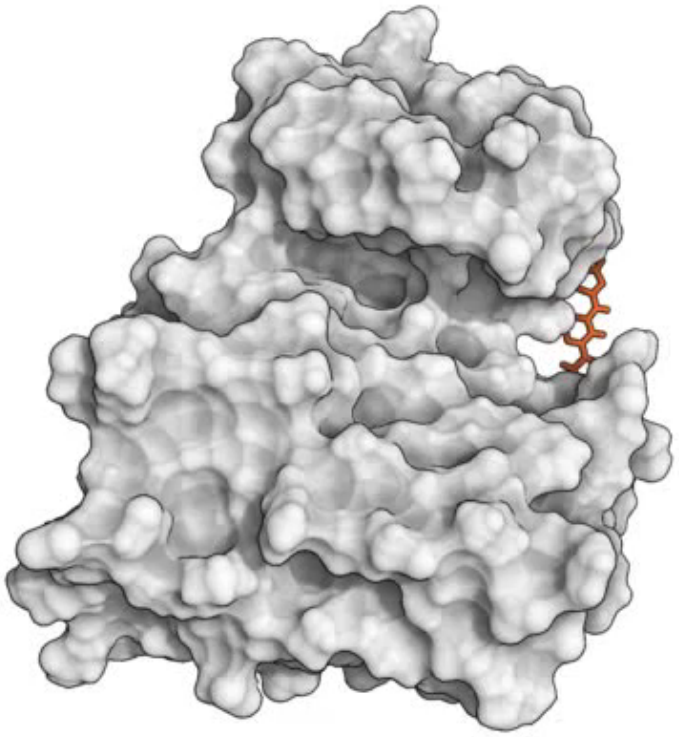
**Step 2.** Integrate Newton's second law and get the new velocities (**v**) of the system and the new coordinates (**r**) of the atoms

**Step 3.** Macroscopic properties can be expressed through **v** and **r** *via statistical mechanics*

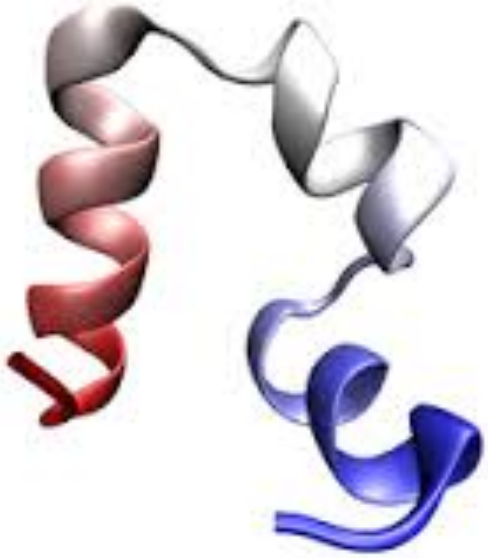



System of interest

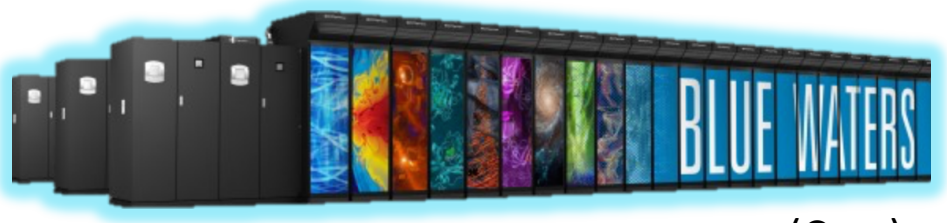
# Examples of MD simulations of proteins



Shan et al (2011)
**Cancer drug dasatinib binding on Src kinase**



Schulten et al (2012)
**Folding of the Villin Headpiece protein**

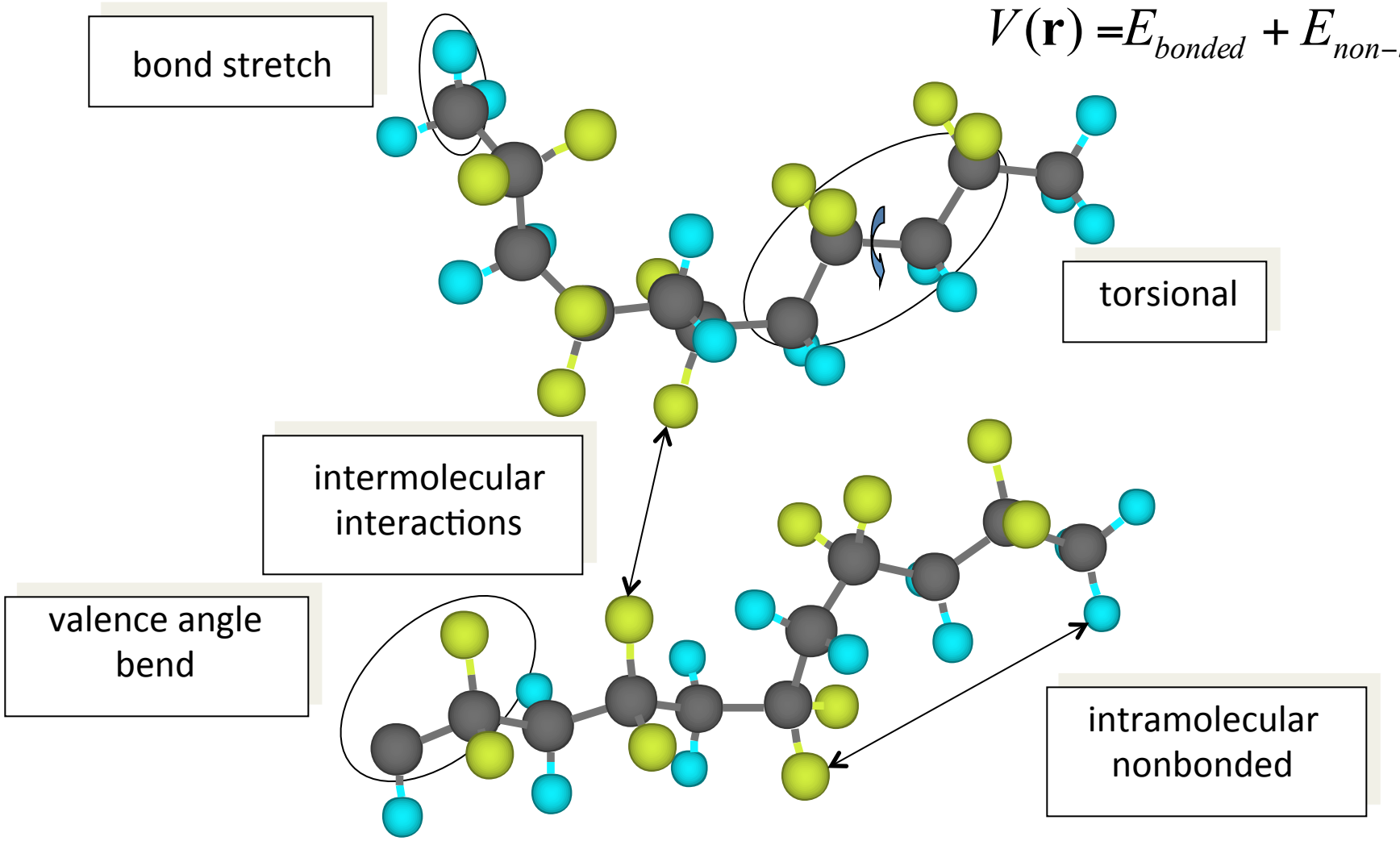

(Cray)

# The Potential Energy Function (Force Field)

The energy of the system is represented by the Hamiltonian:

$$H = K + V = \frac{1}{2} m\mathbf{v}^2 + V(\mathbf{r})$$

$$V(\mathbf{r}) = E_{bonded} + E_{non-bonded}$$

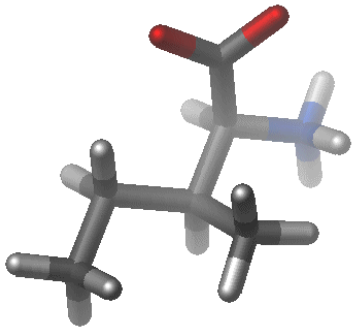


# Modeling the Potential E: Bond stretch potential

- Molecules undergo vibrational motion, which is modeled as a harmonic potential according to HOOKE's law

$$F = -kx = -\nabla V(x)$$

$$V(x) = E_{bond-stretch} = \sum_{1,2\,pairs} k_b (b - b_0)^2$$



- $K_b$ represents the force constant and $b_0$ represents the equilibrium value around which the bond oscillates

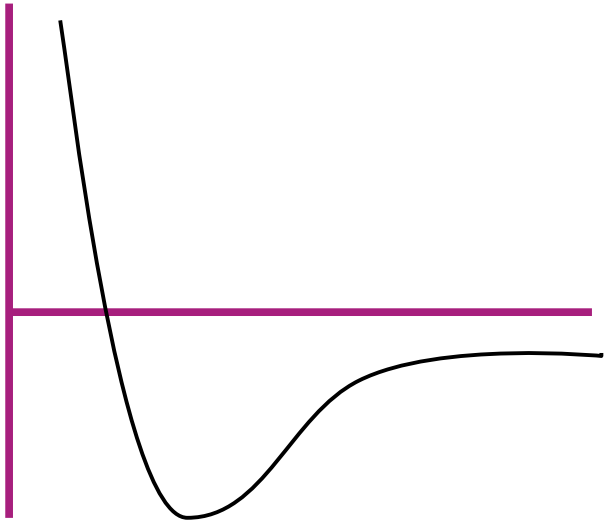
- This harmonic potential is valid only for deviations of 0.1 Å or less

# Harmonic vs Morse potential

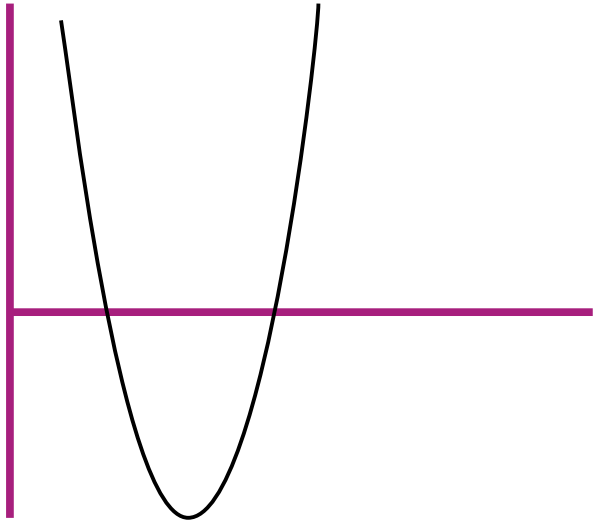
- The Morse term is more accurate, however it is generally not used in MD simulations since it requires 3 parameters to be specified for each bond

$$v(l) = D_e \left\{ 1 - \exp\left[ -a(l - l_0) \right] \right\}^2$$

- The Morse potential would allow a bond to stretch to an unrealistic length and break



Morse potential for a C-H bond



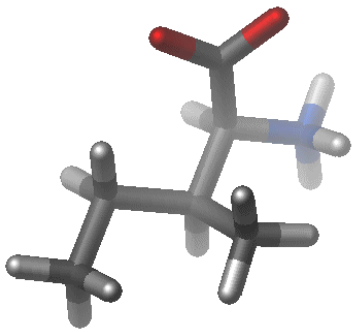
Harmonic potential for a C-H bond

# Bond angle potentials

- Describe the deviation from an ideal bond angle geometry

$$E_{bond-bend} = \sum_{angles} K_{\theta}(\theta - \theta_0)^2$$

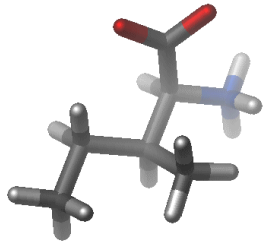
- $K_\theta$ represents the angle bending constant, $\theta_0$ represents the deviation from the ideal bond angle



# Torsion angle potentials

- This terms models the steric barrier between atoms separated by 3 covalent bonds

$$E_{rotate-along-bond} = \sum_{1,4\ pairs} K_{\phi}\left(1-\cos\left(n\phi\right)\right)$$



- The motion associated is rotation, described by a dihedral angle around the middle bond

- The potential is assumed to be periodic and expressed as a cosine function

- $K_{\varphi}$ represents rotation constant, n represent the periodicity of the rotational barrier and $\varphi$ the dihedral angle

# Electrostatic interactions: The Coulomb potential

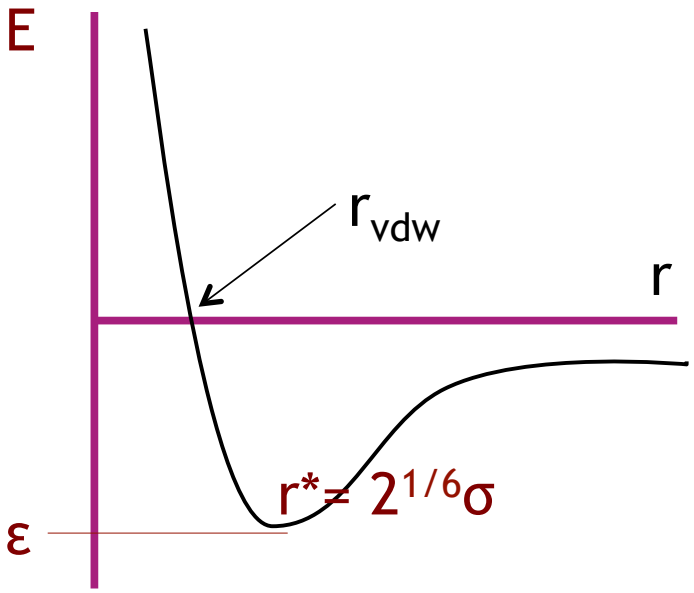
- Electrostatic interaction decays slowly with distance, considered long range interactions. Can be modeled by Coulomb's law.

$$E_{electrostatic} = \sum_{i,j} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}$$

- $r_{ij}$ represents the distance between two atoms having charges $q_i$ and $q_j$

- $\varepsilon_0$ represents the vacuum permittivity, a number relating the ability of a material to carry current

# The van der Waals potential: Lennard-Jones

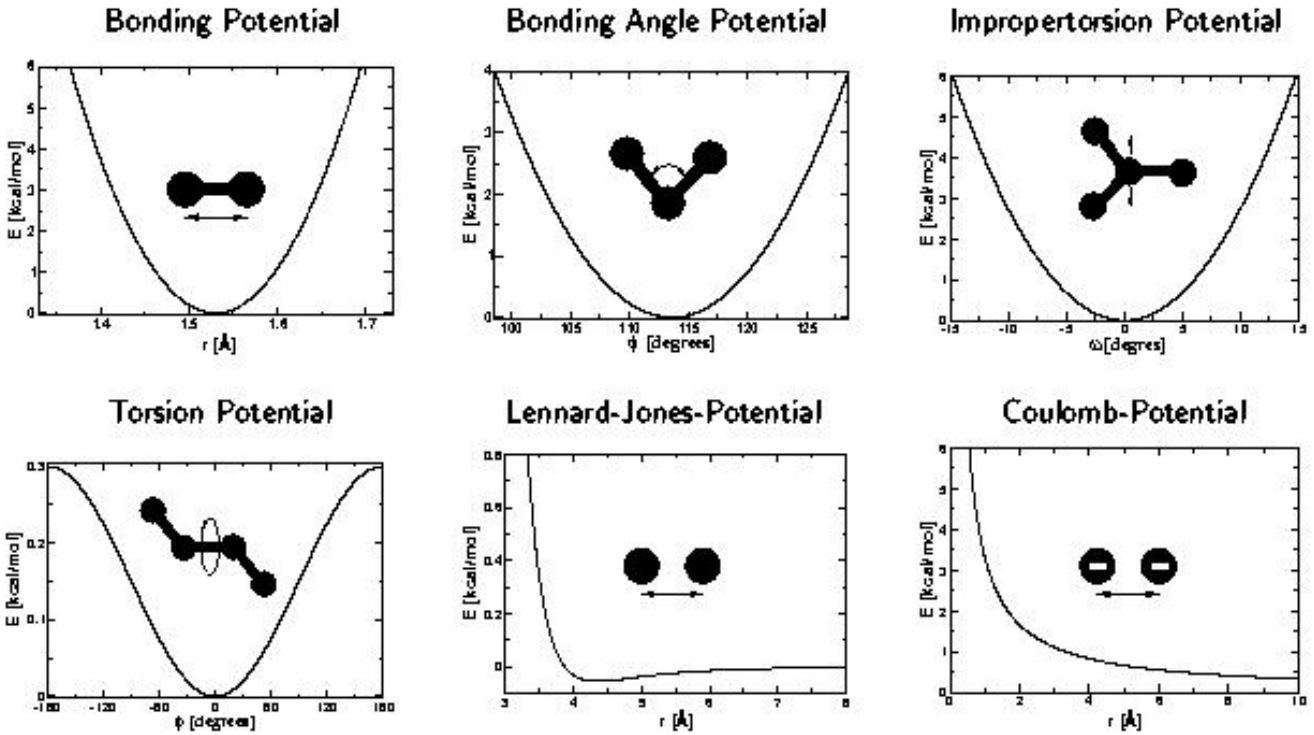



VdW energy best described by a Lennard-Jones potential

$$E_{vdw} = \sum_{i,j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

- Expresses the interaction energy between two atoms
- Contains an attractive part and a repulsive part
- Attractive forces due to London forces (dipole –dipole interaction)
- Repulsive part due to Pauli-exclusion principle and inter-nuclear repulsion
- ε is the depth of the potential well, σ is the finite distance at which the inter-particle potential is zero

# The Potential Energy Function (Force Field)



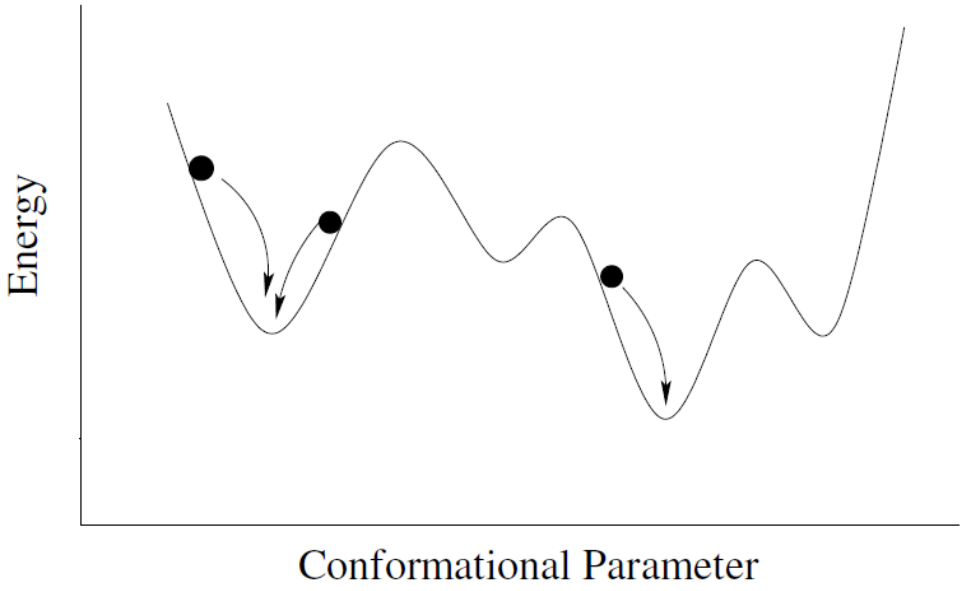

$$E = \frac{1}{2} m\mathbf{v}^2 + V(\mathbf{r}) \qquad \mathbf{F}_i = -\nabla V(\mathbf{r}) \qquad V(\mathbf{r}) = E_{bonded} + E_{non-bonded}$$

$$E_{bonded} = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \delta]) + \sum_{impropers} k_\omega (\omega - \omega_0)^2$$

$$E_{non-bonded} = \sum_{i,j} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i,j} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}$$

# Energy minimization of the potential E function

- Prior to starting a simulation it is advisable to do energy minimization
- Useful in correcting flaws such as steric clashes between atoms and distorted bond angles/lengths
- Goal of energy minimization is to find the local energy minimum to start an MD simulation from a realistic structure *or to perform normal mode analysis to analyze the system vibrations at the energy minimum*




$$\frac{\partial f}{\partial x_i} = 0; \frac{\partial^2 f}{\partial x_i^2} > 0$$

Most common minimization algorithms use derivatives of the energy with respect to the coordinates to predict the location of the closest minimum.

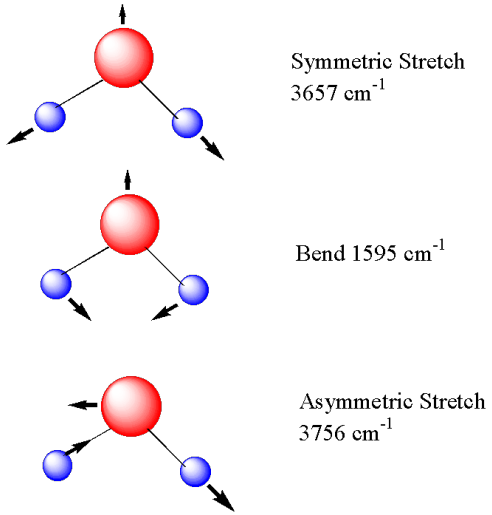
# Application of E min: Normal Mode Analysis

**Normal modes are characteristic vibrations of a system's structure around a local energy minimum. In molecular mechanics, bonds are represented as springs.**

- Characteristic Vibrations
- Harmonic Potential
- At the potential minimum

- Orthogonal normal modes
- Conformational fluctuation = a superposition of normal modes

Water normal modes



# Molecular Dynamics Simulations

- Born – Oppernheimer Approximation

- The molecular dynamics simulation is based on Newton's law of motion: **F=ma**

- By knowing the force that is acting on each atom it is possible to determine the acceleration of each atom.

- Integration of the equation yields a trajectory that describes the positions, velocities and acceleration of each atom as they are varied with time

- Once the positions and velocities of each atom are known, the state of the system can be predicted at any time

- Energy of the system:

$$H = T + V = \frac{1}{2} m\mathbf{v}^2 + V(\mathbf{r})$$

- Initial coordinates are taken from experimental structures and velocities from a distribution, e.g. Maxwell-Boltzmann

# MD Formalism

- Initial coordinates are taken from experimental structures and velocities from a distribution, e.g. Maxwell-Boltzmann
- Newton's equation of motion

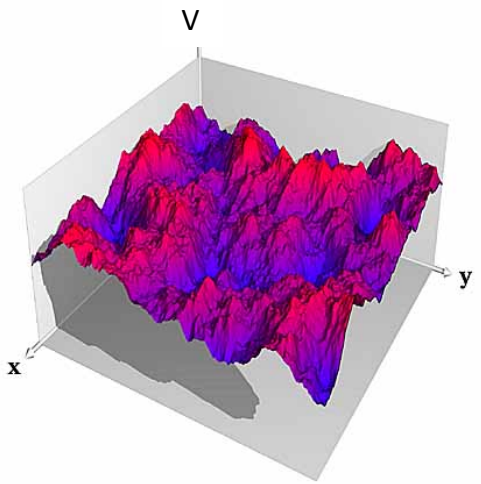
$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2}$$

- The force can be written as the gradient of the potential energy
- Combine the two equations to get

$$\mathbf{F}_i = -\nabla_i V(\mathbf{r})$$

- A trajectory is obtained by solving this differential eq

$$\frac{dV(\mathbf{r})}{d\mathbf{r}} = -m_i \frac{d^2 \mathbf{r}_i}{dt^2}$$



# How to integrate Newton's equation of motion?

- The potential energy is a function of the atomic positions of all the atoms in the system.
- Due to this complexity there is no analytical solution
- Use algorithms to obtain the positions, velocities, accelerations at a later time $t + \delta t$ to a sufficient degree of accuracy
- $\delta t$ is limited by the fastest vibration of the system, ie. the C-H bond ($\delta t$ = 1 fs = $10^{-15}$ s)
- An estimate of the positions, velocities, etc may be obtained with **Taylor's expansion**

$$\underset{\text{new position}}{\mathbf{r}(t+\delta t)} = \underset{\text{old position}}{\mathbf{r}(t)} + \underset{\text{old velocity}}{\delta t \mathbf{v}(t)} + \frac{1}{2}\underset{\text{acceleration}}{\delta t^2 \mathbf{a}(t)} + \ldots$$

$$\underset{\text{new velocity}}{\mathbf{v}(t+\delta t)} = \underset{\text{old velocity}}{\mathbf{v}(t)} + \underset{\text{acceleration}}{\delta t \mathbf{a}(t)} + \ldots$$

# Examples of numerical algorithms: Verlet

- Common use is the **VERLET** algorithm.
- For a differential equation of second order of the type $\frac{d^2\mathbf{r}(t)}{dt^2} = V(\mathbf{r}(t))$

with initial conditions $\mathbf{r}(t_0) = \mathbf{r}_0$ and $\frac{d\mathbf{r}(t_0)}{dt} = \mathbf{v}_0$ , an approximate numerical solution $\mathbf{r}_n \approx \mathbf{r}(t_n)$ at the times $t_n = t_0 + n\delta t$ may be obtained by the method:

- set $\mathbf{r}_1 = \mathbf{r}_0 + \mathbf{v}_0\delta t + ½\, V(\mathbf{r}_0)\delta t^2$
- for $n$ = 1,2 iterate:

$$\mathbf{r}_{n+1} = 2\mathbf{r}_n - \mathbf{r}_{n-1} + \mathbf{v}(\mathrm{r}_n)\delta t^2$$

- In MD, each position is determined from the current position and position at time $t - \delta t$

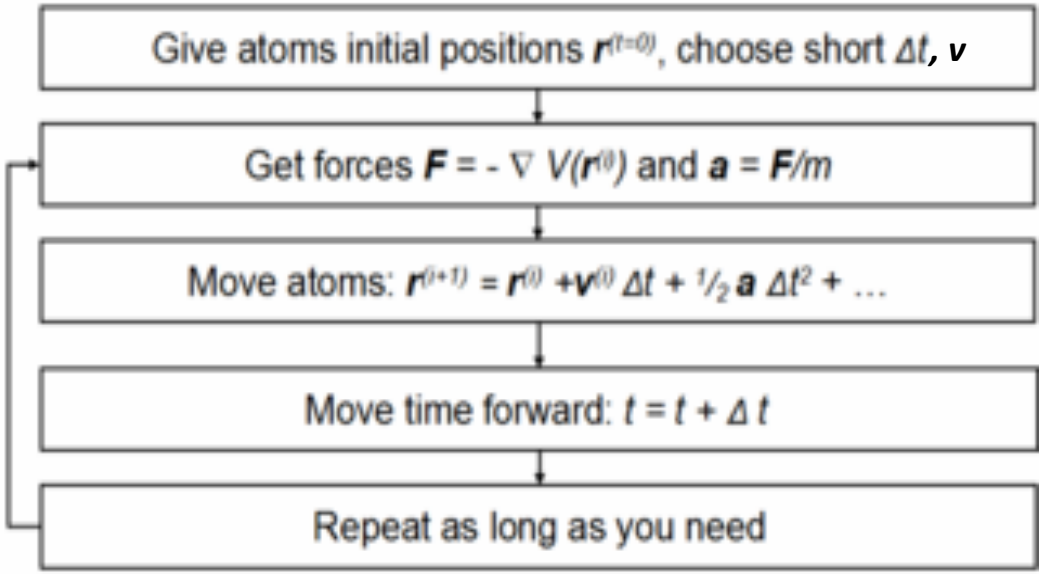
$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \mathbf{a}(t)\delta t^2 + \ldots$$

- Vecocities calculated from

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t}$$

# Molecular Dynamics Simulations

- Integration broken down to many small stages: **δt**
- The total force on each particle in the configuration at a time t is the vector sum of its interactions with other particles.
- From the force determine the acceleration of the particles and combine it with positions and velocities at time t to calculate at time t +δt
- The force is constant during the time step

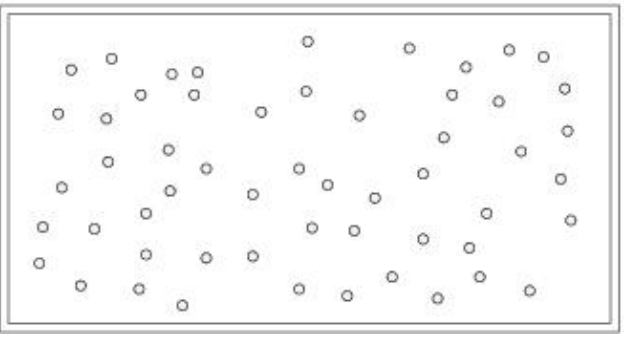


# Statistical Ensembles & Phase Space

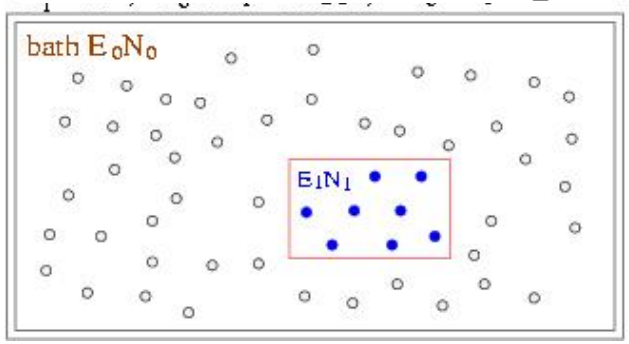
*A statistical ensemble is a set of microscopic states corresponding to a given macroscopic state.*

A thermodynamic ensemble is in statistical equilibrium and is used to derive the properties of thermodynamic systems from the laws of classical or quantum mechanics

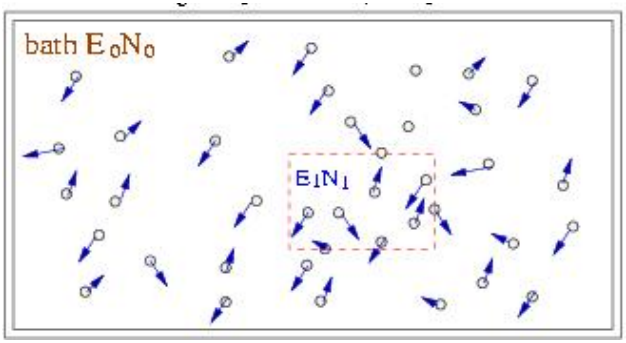
Microcanonical, NVE

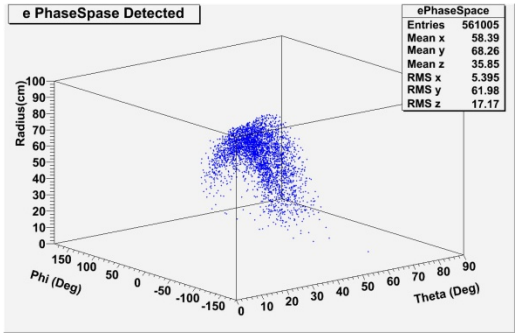


Canonical, NVT




Grand Canonical, μVT

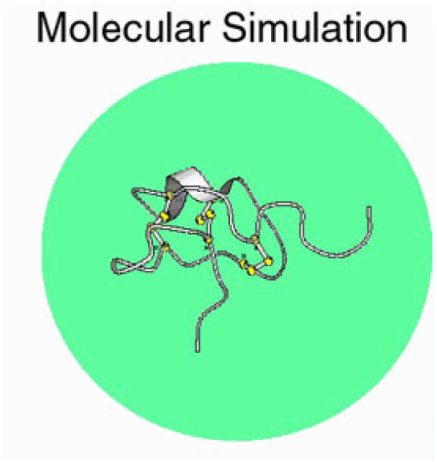



A statistical ensemble is a set of representative representative points in the 6N-dimensional phase space.

The microstates move through phase space, describing a ***dynamic trajectory*** and are described by coordinates and momenta evolving in time.

# Statistical Mechanics





Microscopic Description

Macroscopic Description

*Quantum Mechanics:*
Eigenvalues $E_i$ and eigenfunctions $\Psi(r_1,r_2,\ldots,r_N)$ of Schrodinger's equation

*Molecular Mechanics:*
Kinetic and Potential energy E(**r,v**)

*Thermodynamics:*

Relations for the system at equilibrium and non-equilibrium states

- Use statistical mechanics to derive macroscopic properties from the microscopic picture

# The partition function

Boltzmann law: $\frac{n_i}{n_j} = e^{-(\varepsilon_i - \varepsilon_j)/kT}$ $n_i$, $n_j$ populations of energy states

For the lowest energy state: $n_i = n_0 e^{-\beta\varepsilon_i}, \beta = 1/kT$ $n_0 = \frac{N}{\sum e^{-\beta\varepsilon_i}}$

The molecular partition function determines how particles distribute themselves over accessible states

$$n_i = \frac{N e^{-\beta\varepsilon_i}}{\sum e^{-\beta\varepsilon_i}} \qquad q = \sum e^{-\beta\varepsilon_i}$$

**Partition function per particle**

The above treatment applies to quantum statistical mechanics (discrete eigenstates)
In classical statistical mechanics, the *position* and *momentum* variables of a particle can vary continuously, so the set of microstates is actually uncountable.
Describe the **partition function** using an integral rather than a sum.

$$\mathbf{Q} = \frac{1}{N!h^{3N}} \int \exp[-\beta H(p_1 \cdots p_N, x_1 \cdots x_N)]\, d^3p_1 \cdots d^3p_N\, d^3x_1 \cdots d^3x_N$$

# Themodynamic propertis are a function of Q

Canonical partition function, N,V,E constant

$$Q(N,V,E) = \frac{1}{h^{3N}N!}\sum(N,V,E) = \frac{1}{h^{3N}N!}\int d^{3N}\mathbf{q}\,d^{3N}\mathbf{p}$$

$$E = \frac{N}{q}\sum_i \varepsilon^i e^{-\beta\varepsilon^i} \qquad U - U_0 = -\left(\frac{\partial \ln Q}{\partial \beta}\right)_V \qquad S = \frac{U - U_0}{T} + k\ln Q$$

Connection to macroscopic thermodynamics: $S(N,V,E) = k_B \ln Q(N,V,E)$

Thermodynamic properties can be derived from the above, e.g.

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_{N,V} \Rightarrow \frac{1}{T} = k_B\left(\frac{\partial \ln Q}{\partial E}\right)_{N,V}$$

# Statistical Ensembles

***NVE***

H = K + V = total energy = constant

$$K = \sum_i \frac{1}{2} m_i v_i^2 \;,\; V = \sum_{i,j} U(r_{ij})$$

$$\frac{dr_j}{dt} = \frac{\partial H}{\partial p_j} = v_j \;,\; \frac{dp_j}{dt} = -\frac{\partial H}{\partial r_j} = -\frac{\partial V}{\partial r_j} = F_j$$

***NVT***

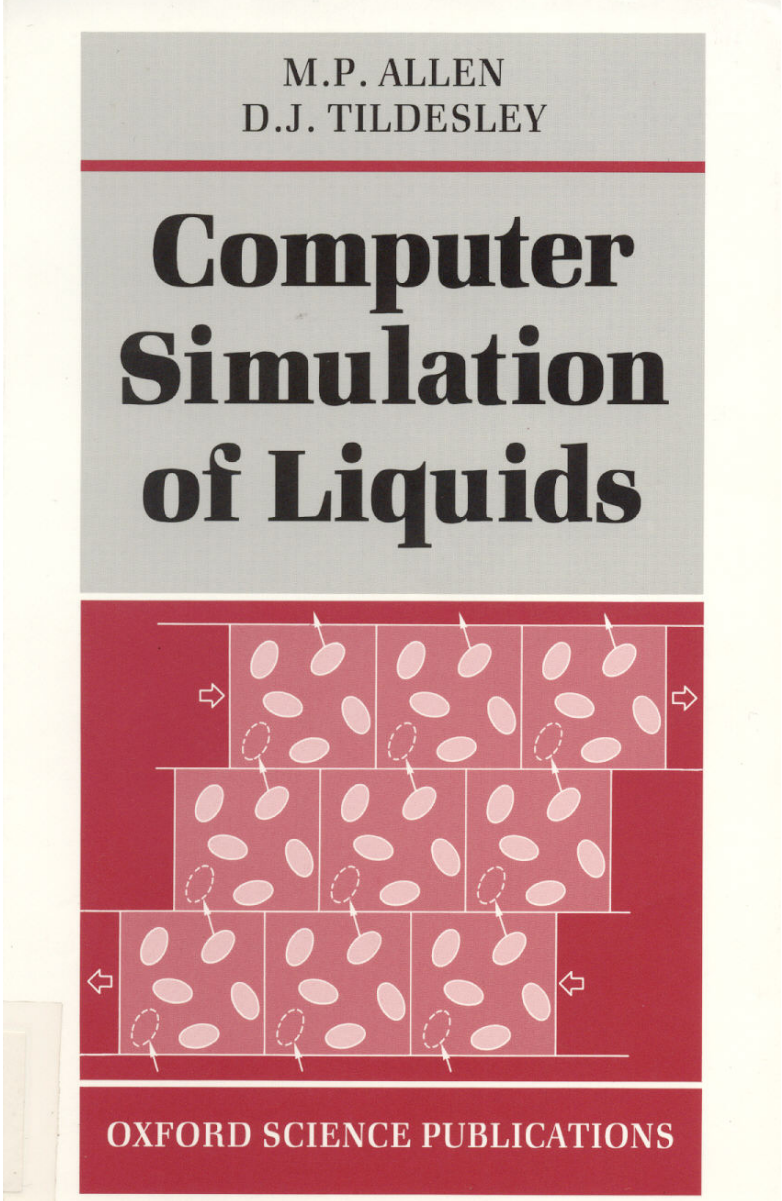
$H = K + V + K_s + V_s = \text{constant}$

$K_s = \frac{1}{2} Q p_s^2$

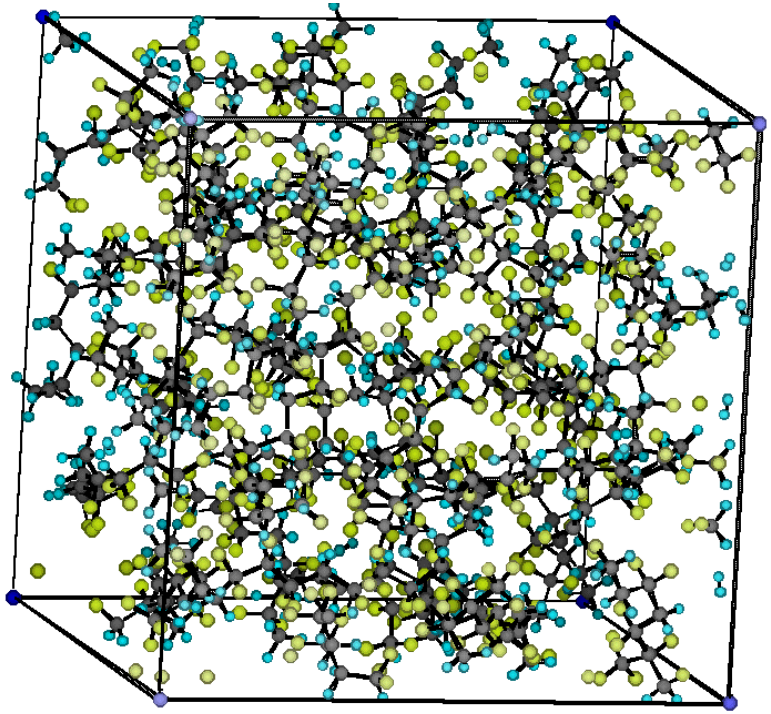
$V_s = (f+1) k_B T \ln(s)$

$p_s$ = conjugate momentum of thermostat
Q = conjugate mass
s = generalized coordinate of the thermostat
$f$ = number of degrees of freedom

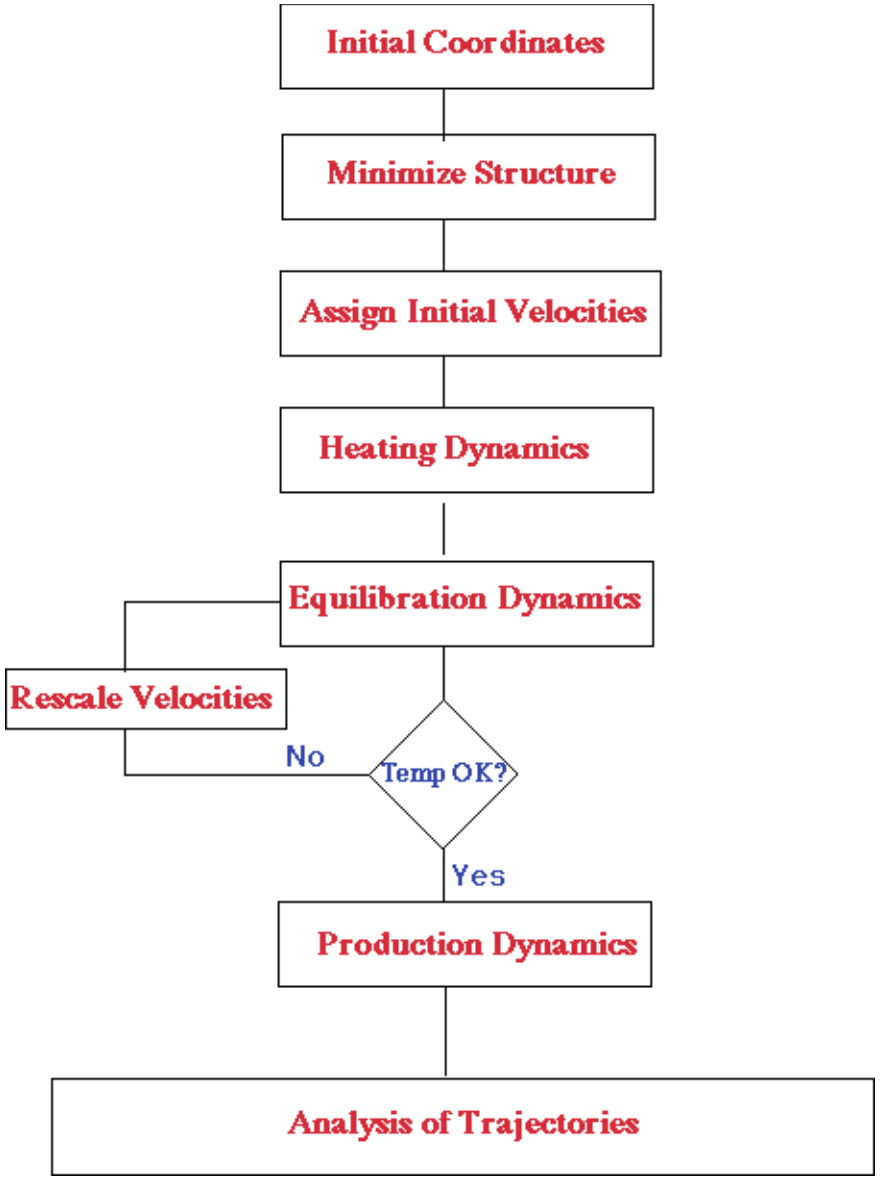
# Periodic Boundary Conditions




- *Goal:* To simulate 'infinite' system
- Particles experience forces as if they were in a bulk fluid
- If one molecule leaves the box then it is replaced by an image particle that enters from the opposite side



# Running an MD simulation



Academy of Athens

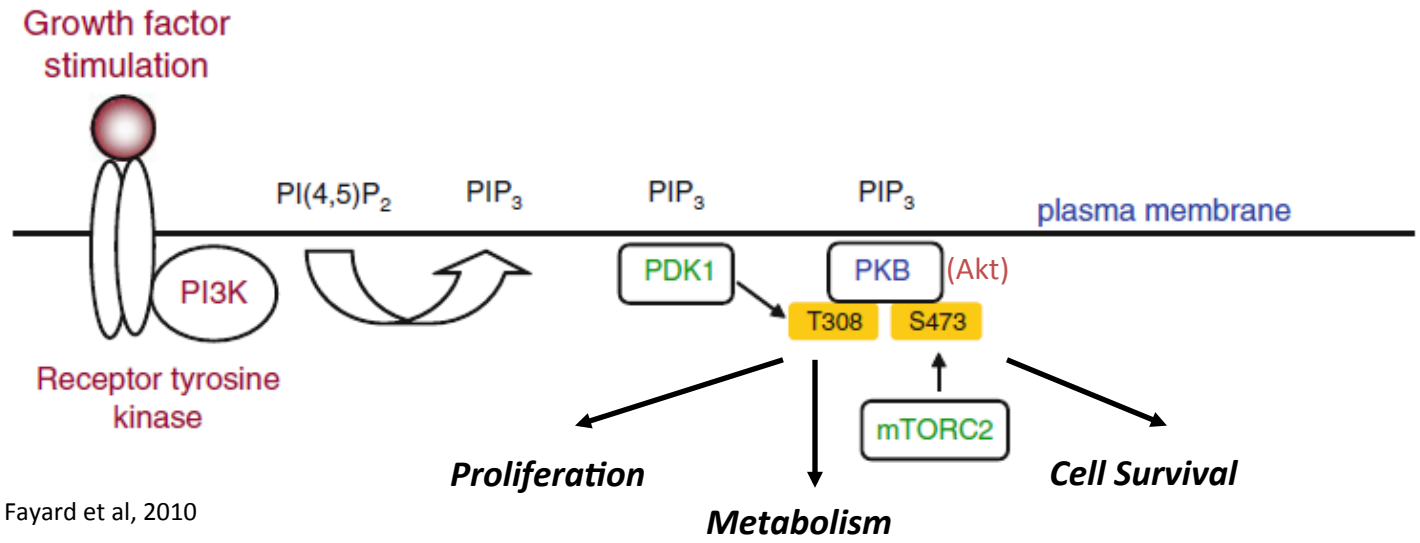
Biomedical Research Foundation

# Case Study:
# *Targeting the mutated PI3Kα*

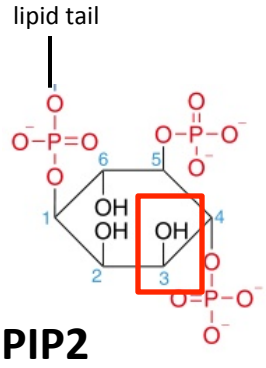
Zoe Cournia
zcournia@bioacademy.gr

**12 December 2014**

# PI3Kα is a lipid kinase that promotes cell survival




Fayard et al, 2010



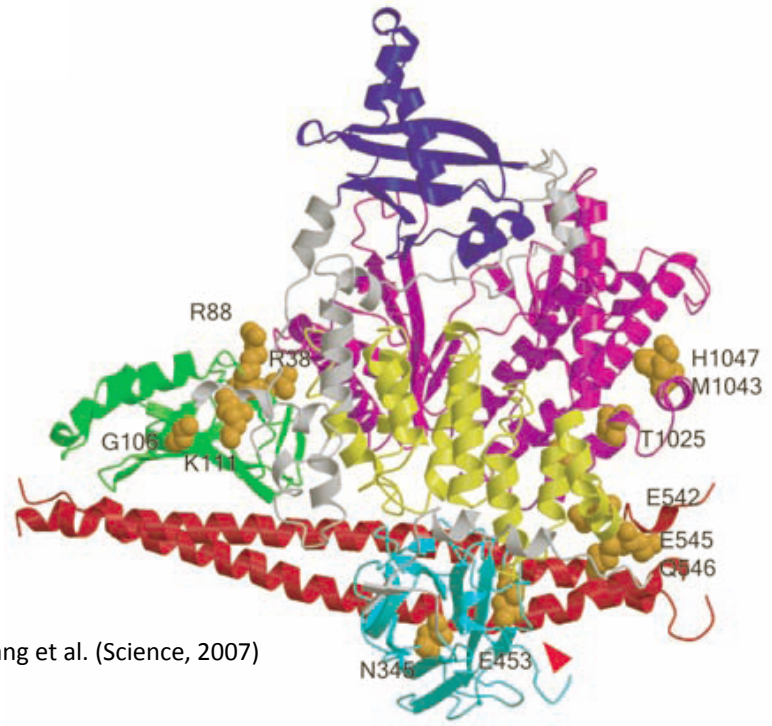

- Active PI3Ka phosphorylates PIP2 to PIP3 at the plasma membrane.

- PIP3 recruits Akt close to PDK1.

- Co-localization of these proteins leads to phosphorylation of residues, which in turn leads to proliferation, growth, survival.

# PI3Kα: most commonly mutated kinase in cancer

- PI3Ka is a membrane-associated lipid kinase
- Involved in cell growth, proliferation, differentiation
- Most commonly mutated kinase in the human genome ⇒ cancer



Huang et al. (Science, 2007)

80% of all mutations:

Glu545Lys | His1047Arg

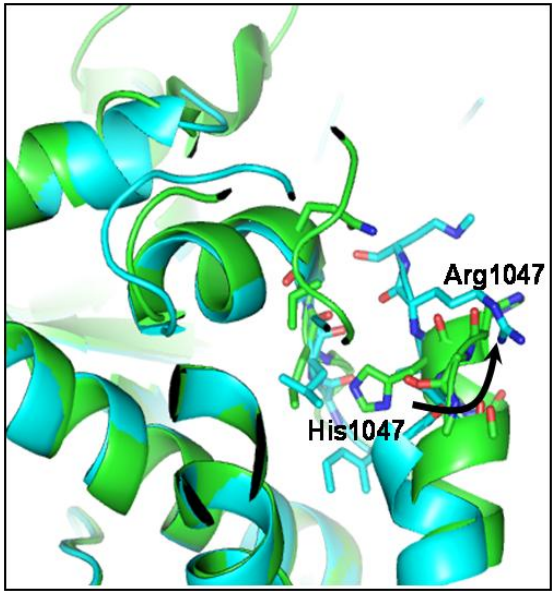
30% of breast cancer patients

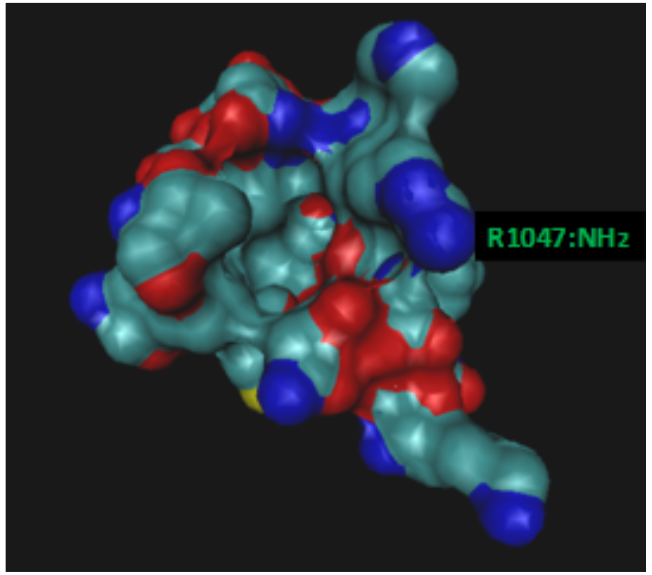
**Mechanism of overactivation?**

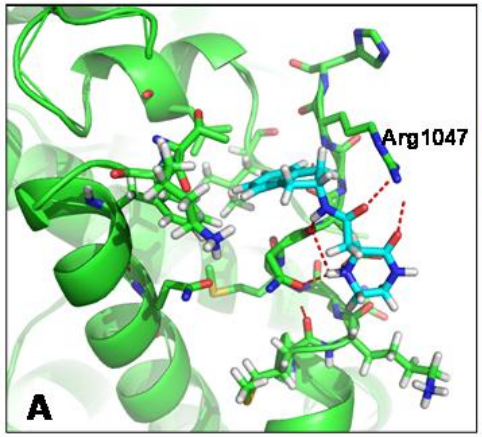
**Mutant and isoform specific therapies?**

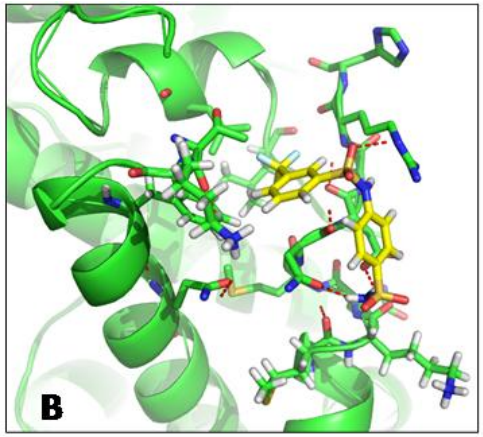
**MD Simulations**
**Virtual screening**
**Property prediction**
***In vitro* & *In vivo* assays**
**Lead Optimization**

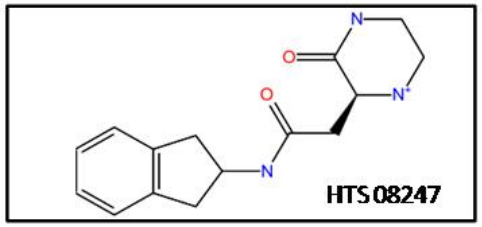
# The H1047R mutant of PI3Kα opens up a crevice

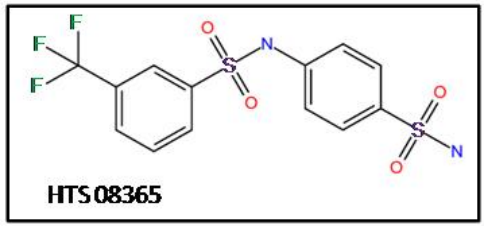


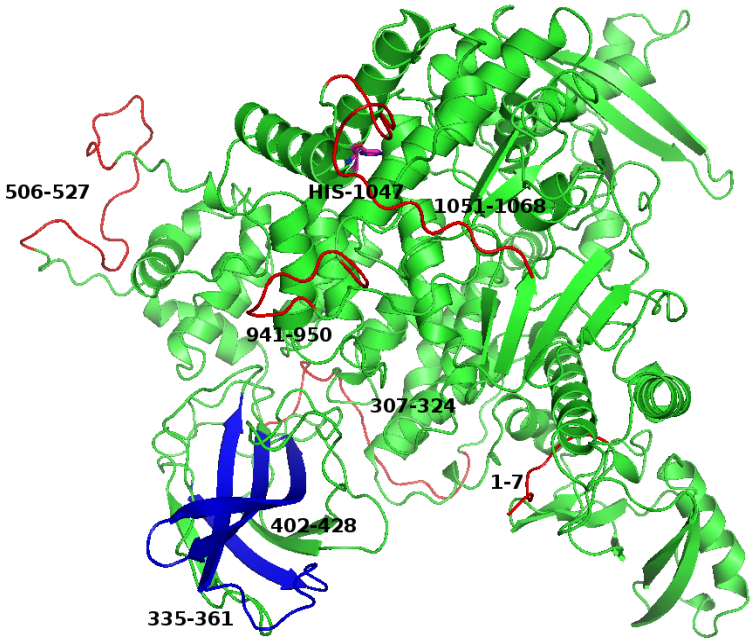



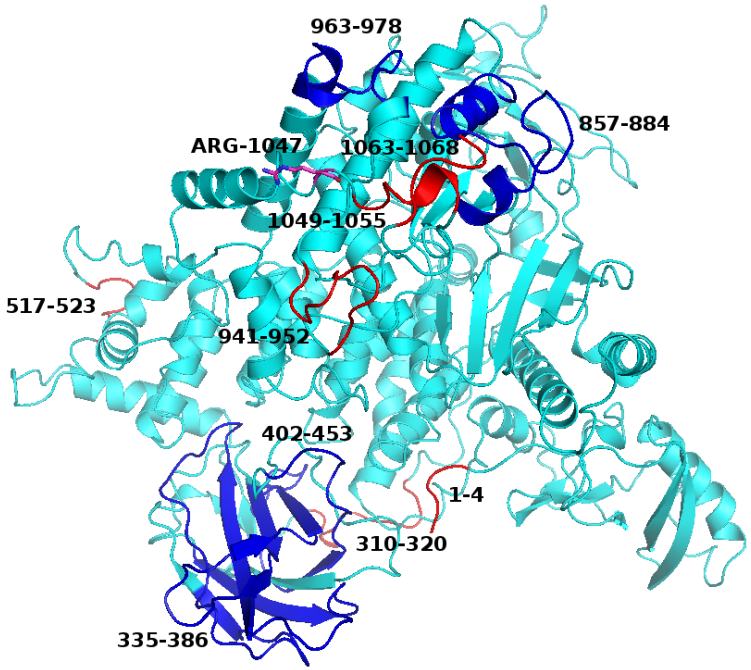



# MD Simulations of WT and H1047R PI3Kα




**Model of the WT p110a subunit based on 2RD0 X-ray structure**

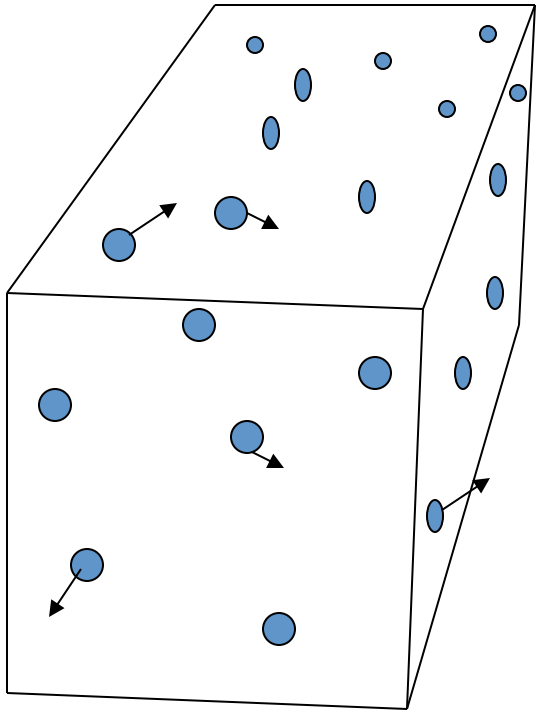



**Model of the H1047R p110a subunit based on 3HIZ X-ray structure**

- WT and H1047R PI3Kα (modeling of p110α), 300K atoms
- 100-150 ns equilibration, 100 ns production run, NPT, NAMD+CHARMM
- **FIVE** independent MD simulations of each protein
- Total simulation time (~1μs)

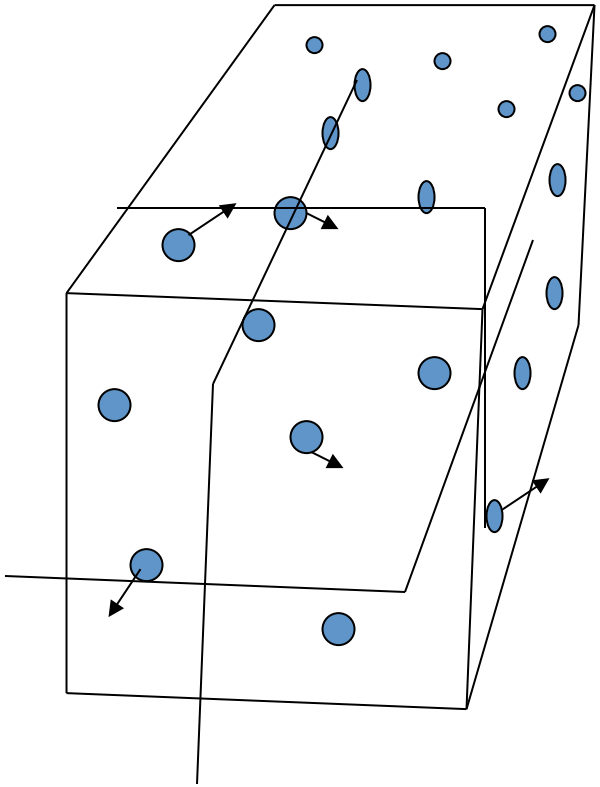
# MD Simulator requirements

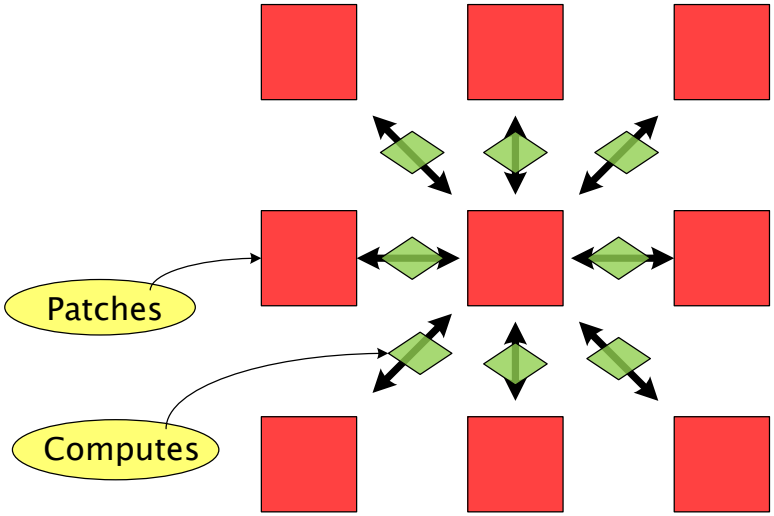
- Parallelization
- (getting an idea of the level of computation needed)
- For every time step, every atom must communicate within its *cutt-off radius* with every other atom.
- A lot of inter-processor communication that can be scaled well is needed.



# MD Simulator requirements

- Parallelization
- (getting an idea of the level of computation needed)
- Whole System is broken down into boxes (processing nodes)
- Each node handles the bonded interactions within a cutoff





# Access to HPC-How many cores should be used?

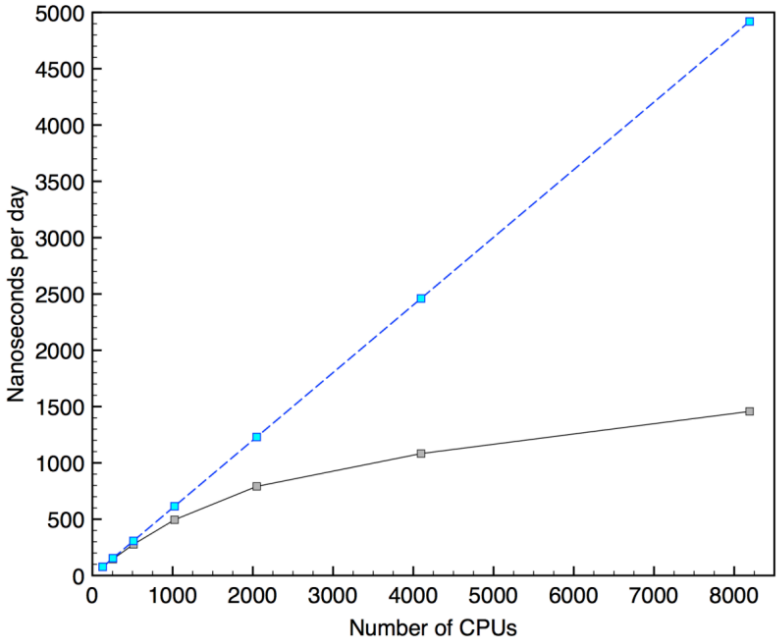
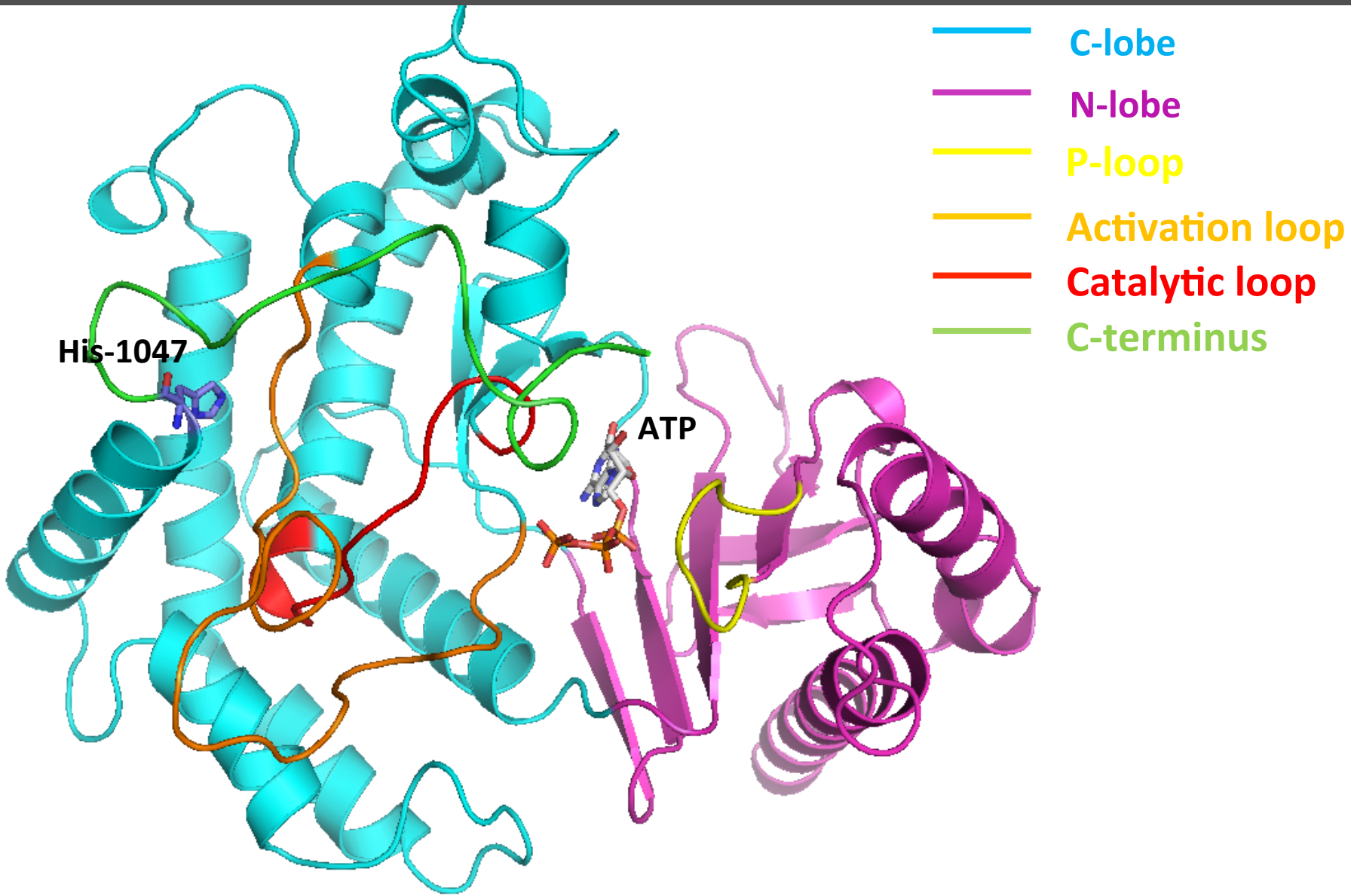



**Figure 2**: GROMACS performance in CURIE Thin Nodes for a system of 2M particles.

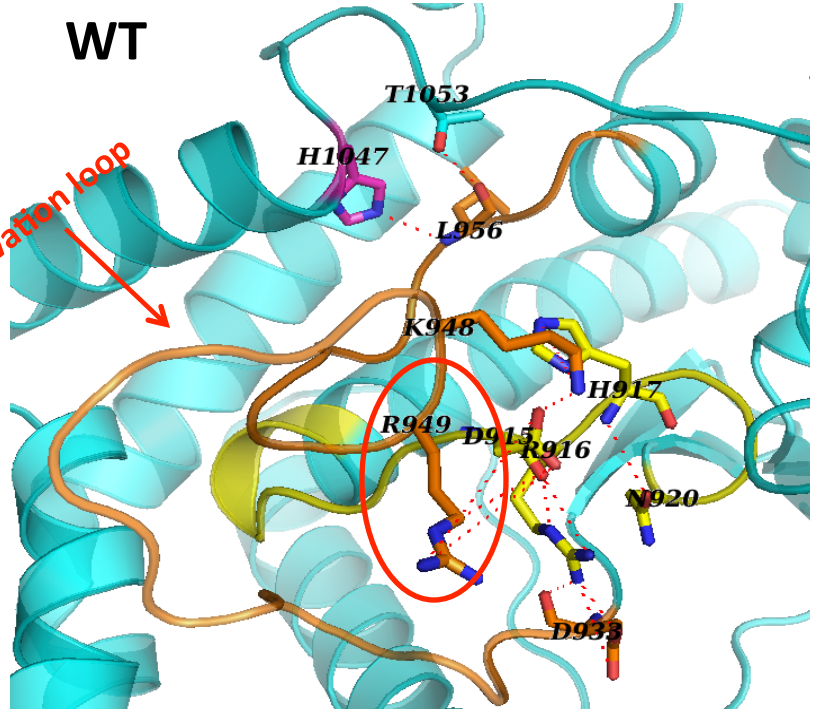
**Table 1.** Benchmark of a 2M-particle system on Curie Thin Nodes.

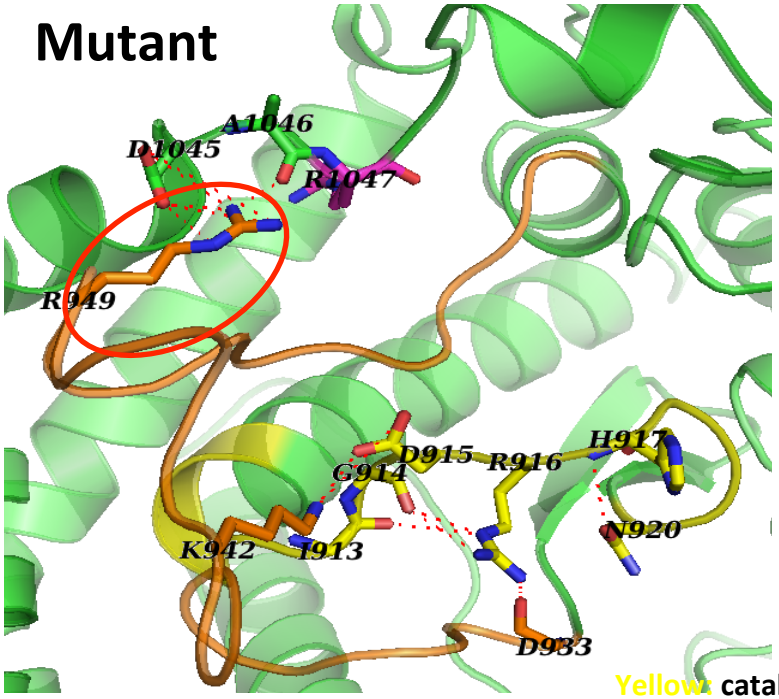
| # cores | absolute timing (s) | speedup |
|---|---|---|
| 128 | 1124.022 | 1.0 |
| 256 | 590.897 | 1.9 |
| 512 | 312.562 | 3.6 |
| 1024 | 174.532 | 6.5 |
| 2048 | 109.207 | 10.3 |
| 4096 | 79.805 | 14.1 |
| 8192 | 59.291 | 19.0 |

# Kinase Domain Organization



# Hydrogen Bond Analysis



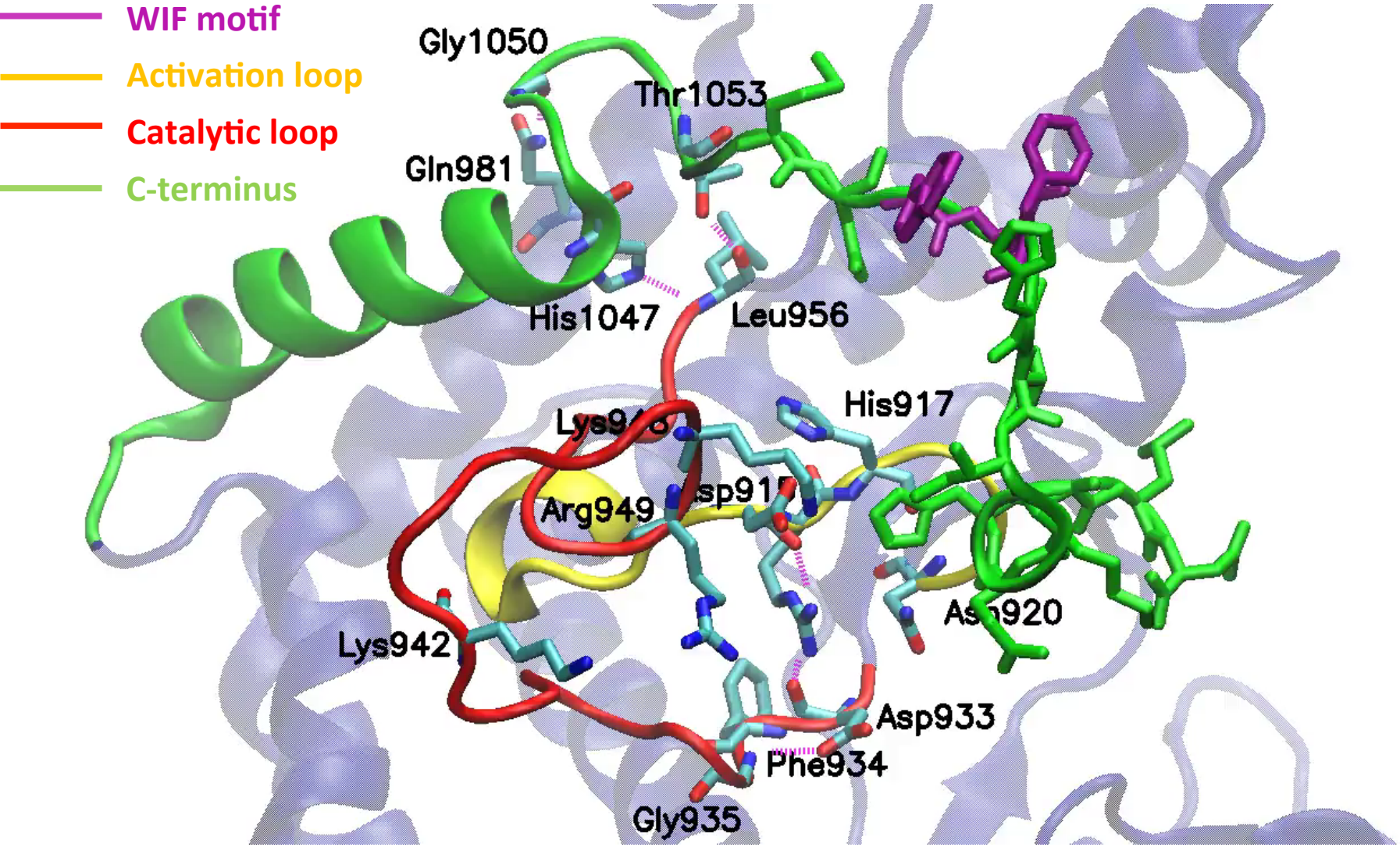




Yellow: catalytic loop
Orange: activation loop
Magenta: residue 1047

- The Hbond between activation loop Leu956 and His1047 breaks
- The α-helix of H1047 partially unfolds in the presence of 1047R
- Displacement of Arg949 creates a different Hbond network in the mutant, which changes the activation and catalytic loop positions

**H917, RESPONSIBLE FOR ATP HYDROLYSIS, IS ORIENTED TOWARD THE CATALYTIC POCKET IN THE MUTANT AND AWAY FROM THE POCKET IN THE WT**

# Simulation of the normal protein




**His-917 points away from the active site, while the C-terminus prevents the catalytic loop from reaching the ATP-binding site.**

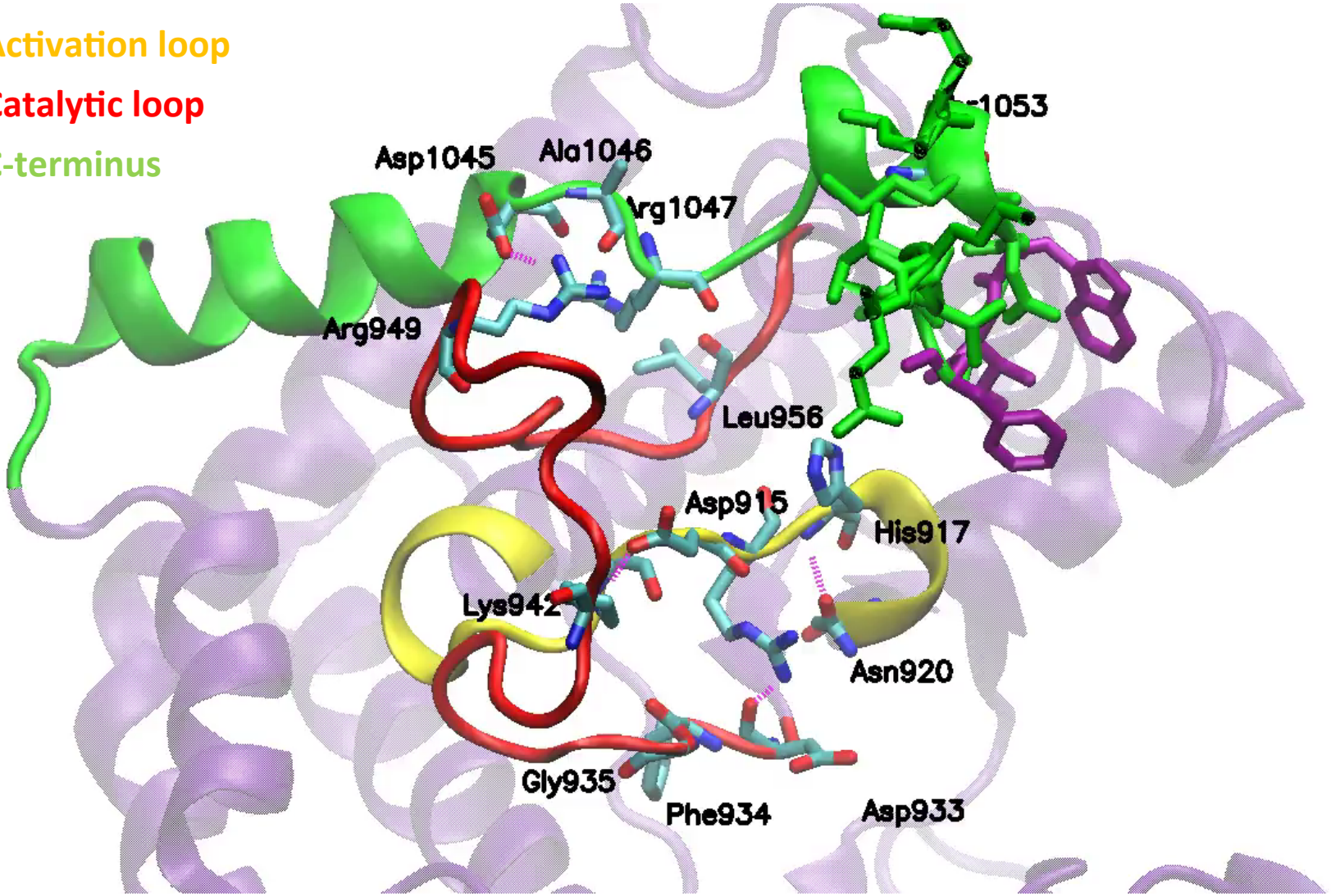
# Simulation of the mutated protein

WIF motif

Activation loop

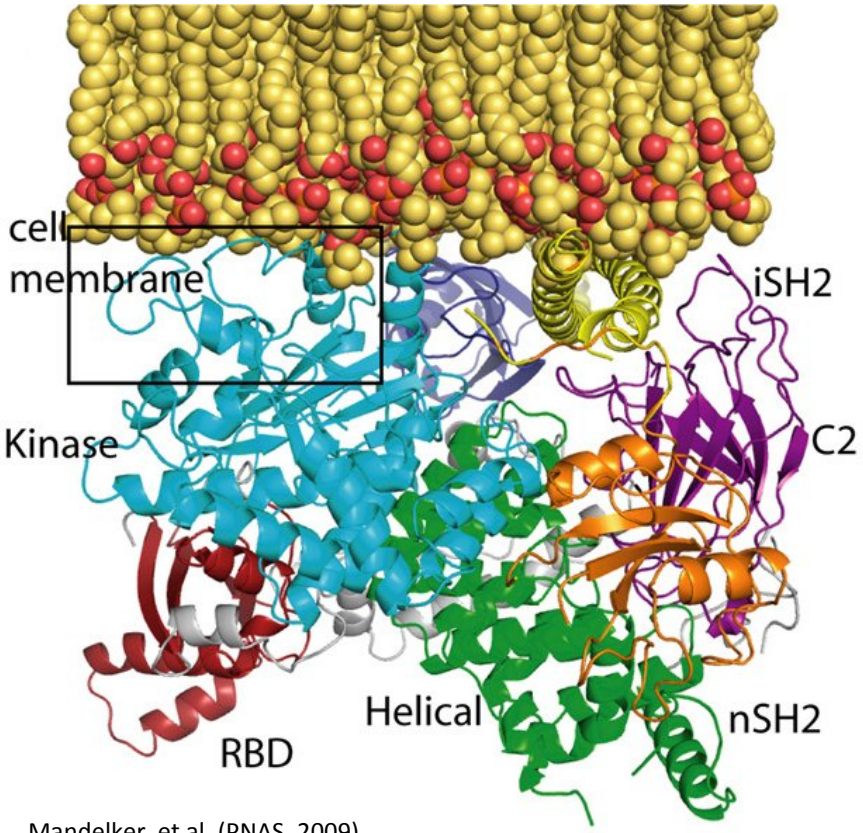
Catalytic loop

C-terminus




**His-917 points towards the active site, while the C-terminus does not interfere with the access of the catalytic loop to the ATP-binding site.**

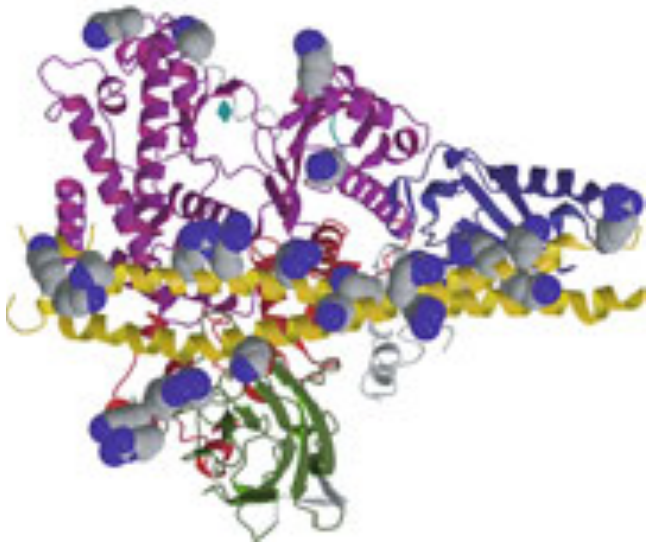
# PI3Kα – membrane interactions




Mandelker et al. (PNAS, 2009)

- PI3Ka binds to the membrane to convert PIP2 to PIP3

- *The activation of the His1047Arg mutant takes place through a change in the way p110a interacts with the cell membrane (Mandelker et al, 2009).*

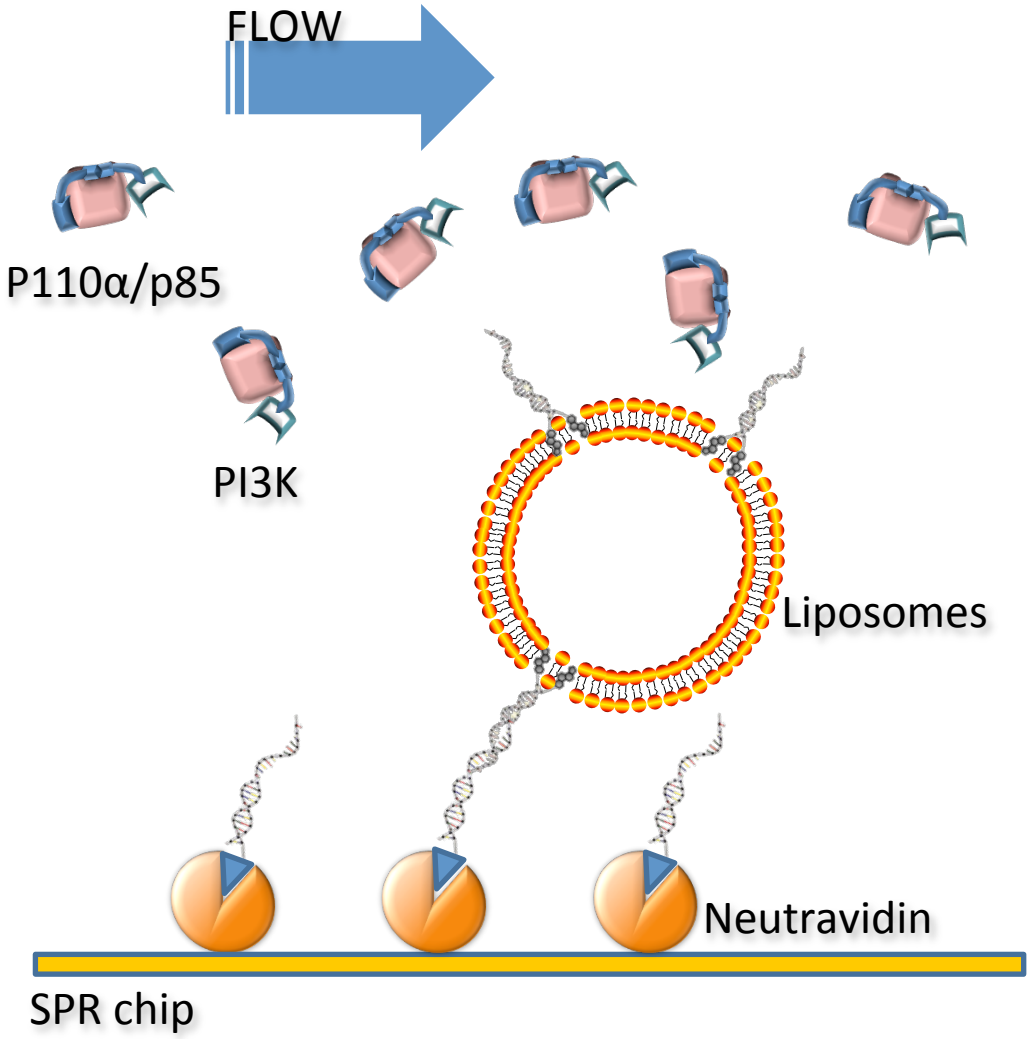


Gabelli et al. (2009)

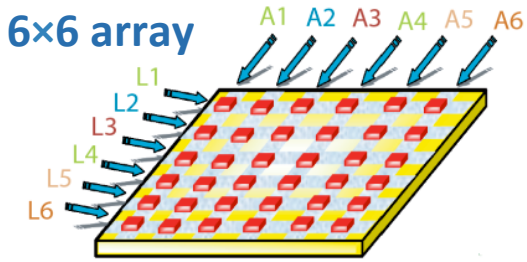
- His1047Arg mutant has a 2-fold increase in lipid kinase activity.

- Mechanism of action of mutant?

# SPR Experiments for membrane-protein binding

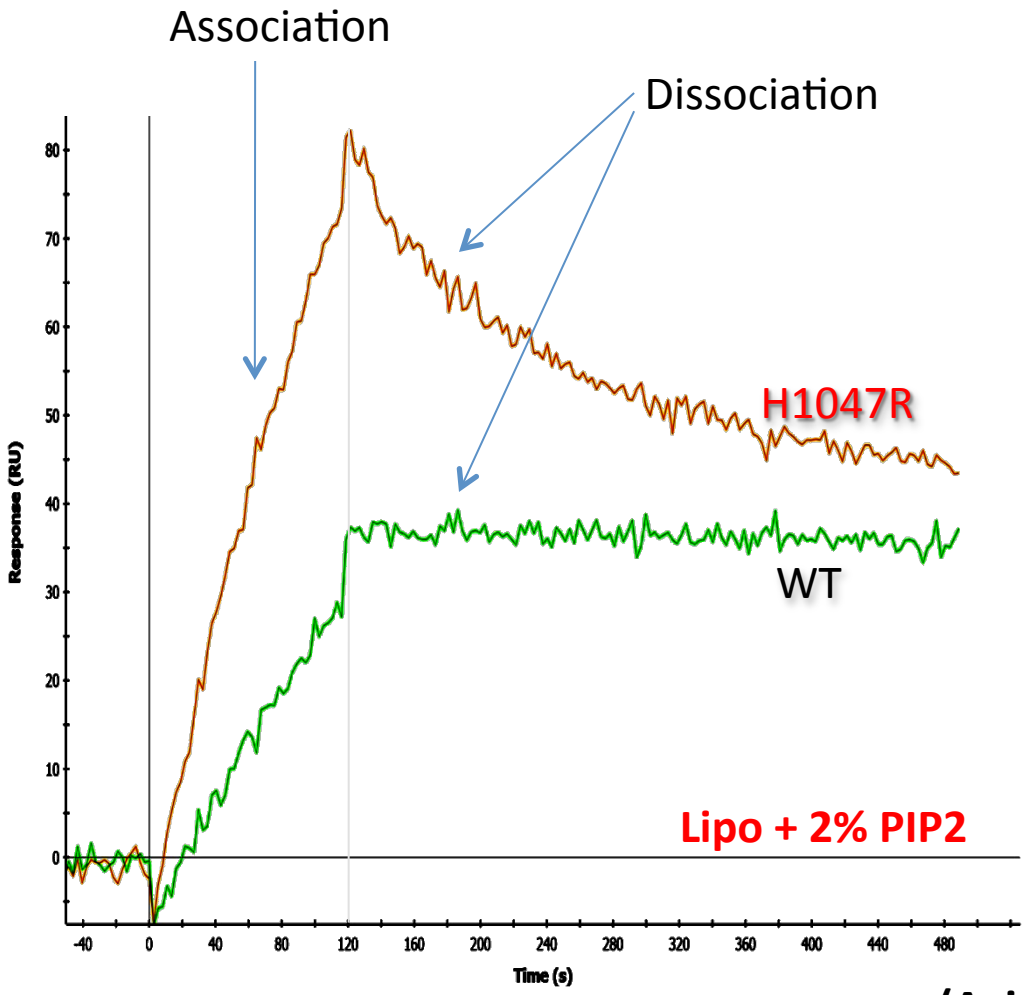



**Experiments & Simulations show that the mutated protein binds more to cell membranes than the normal one**




**(Agianian lab, University of Thrace, Greece)**

# WT vs H1047R membrane binding kinetics



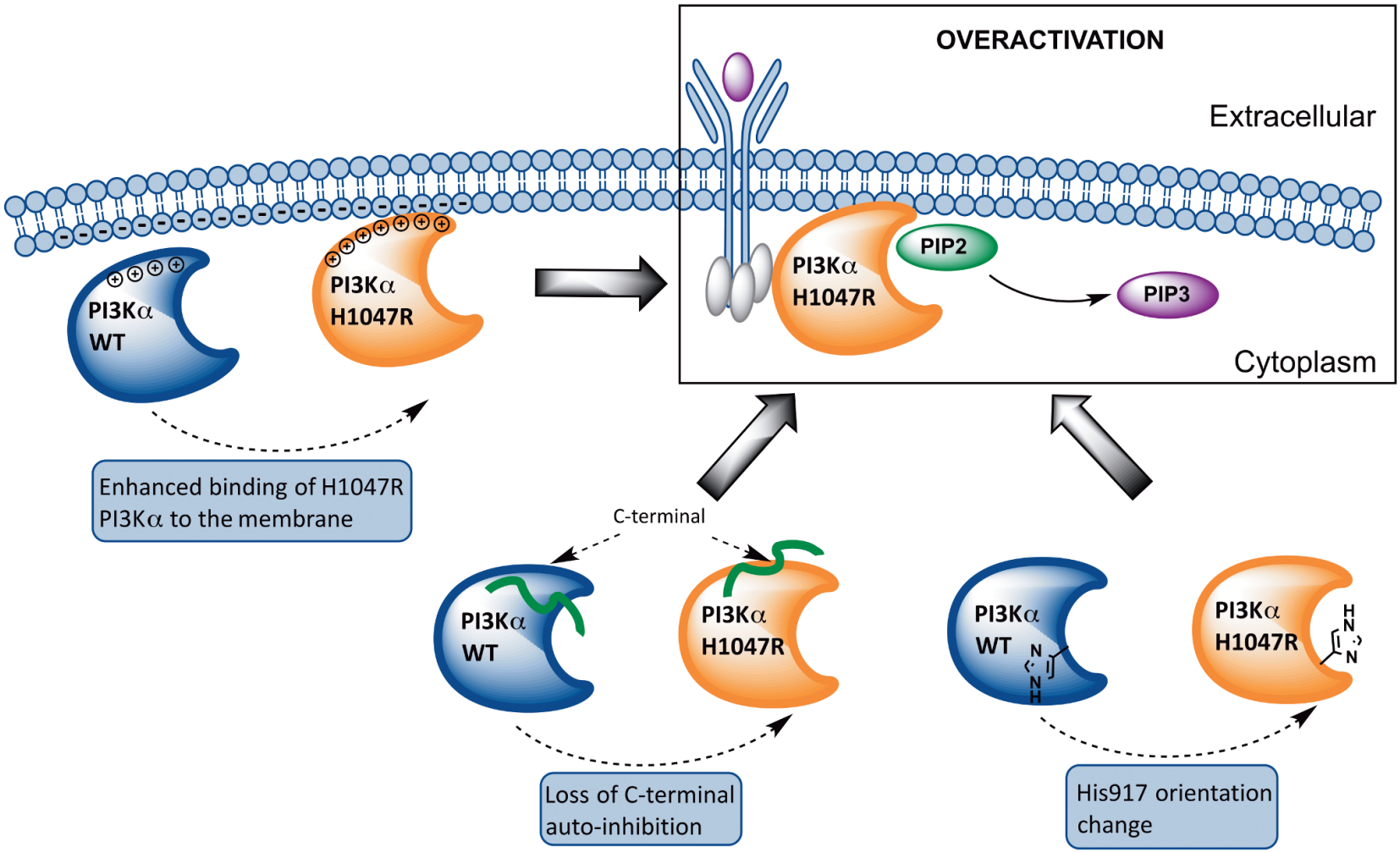


**(Agianian lab, University of Thrace)**

Association: level of binding during association

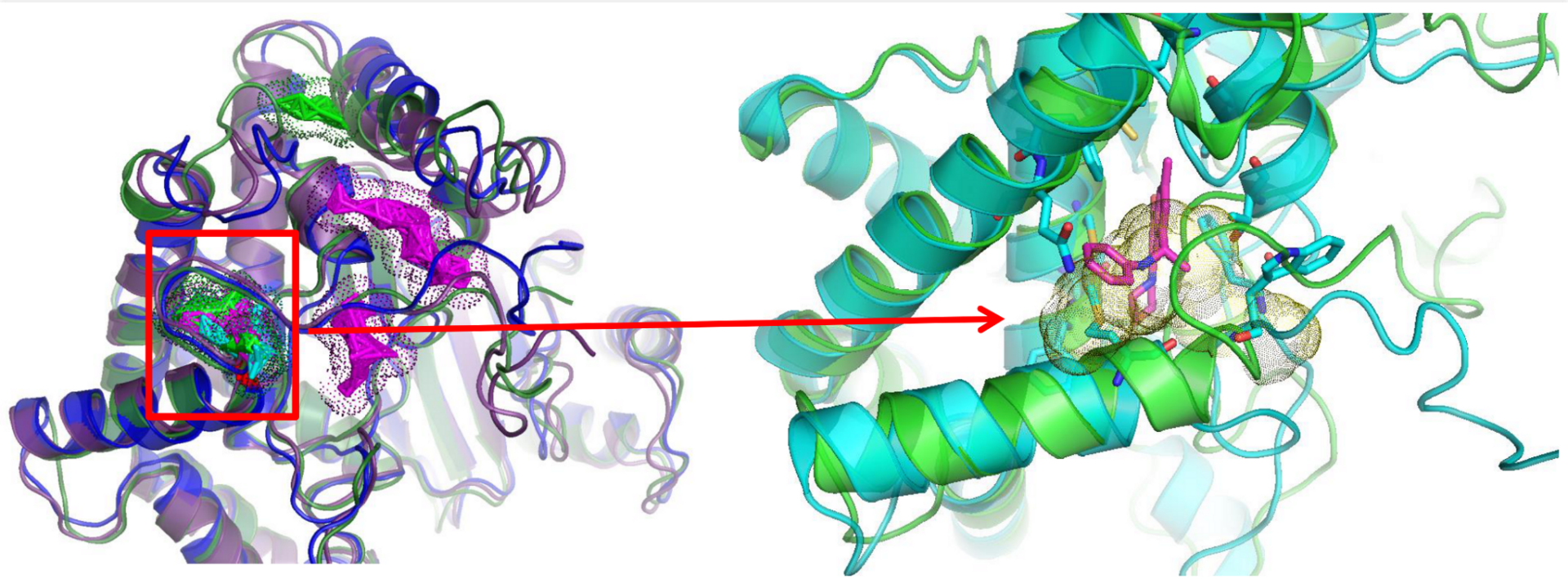
Residual Bound: residual protein binding to the membrane after buffer injection

# Proposed mechanism of H1047R overactivation




Gkeka et al, PLOS Comput Biol (2014)

# Binding site identification on PI3Kα conformers



**Binding site prediction on PI3Kα representative structures**

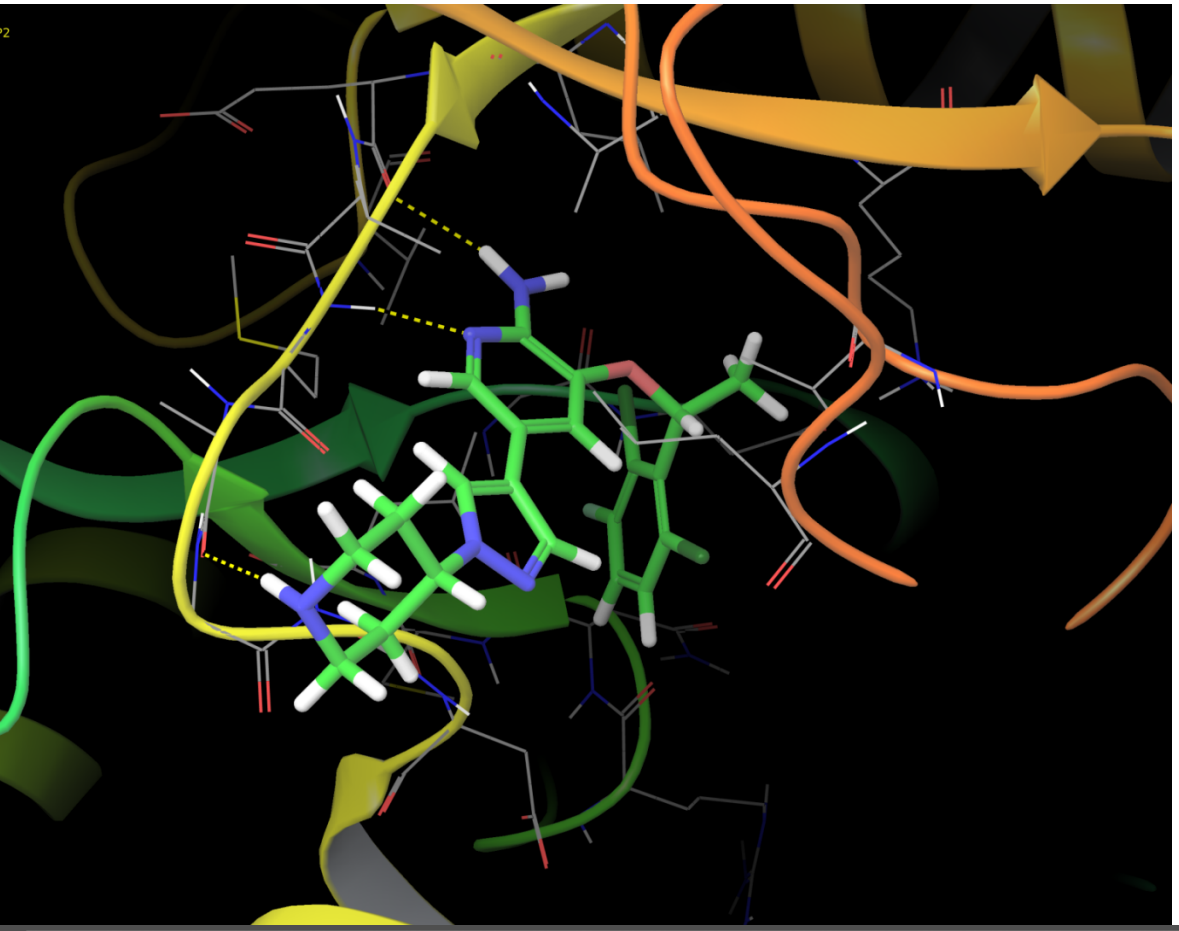
**Blue: WT Crystal Structure by Hon et al (2011)**

**Green: Cluster conformation from MD**
**Dots: Predicted binding site**

**Does this binding site also exist in the mutant form and can it be exploited for selective drug design?**

# Drugs bind on protein pockets through chemical interactions

**Structure of the anaplastic lymphoma kinase (ALK)**
**Complexed with the drug crizotinib – (PDB ID: 2XP2)**



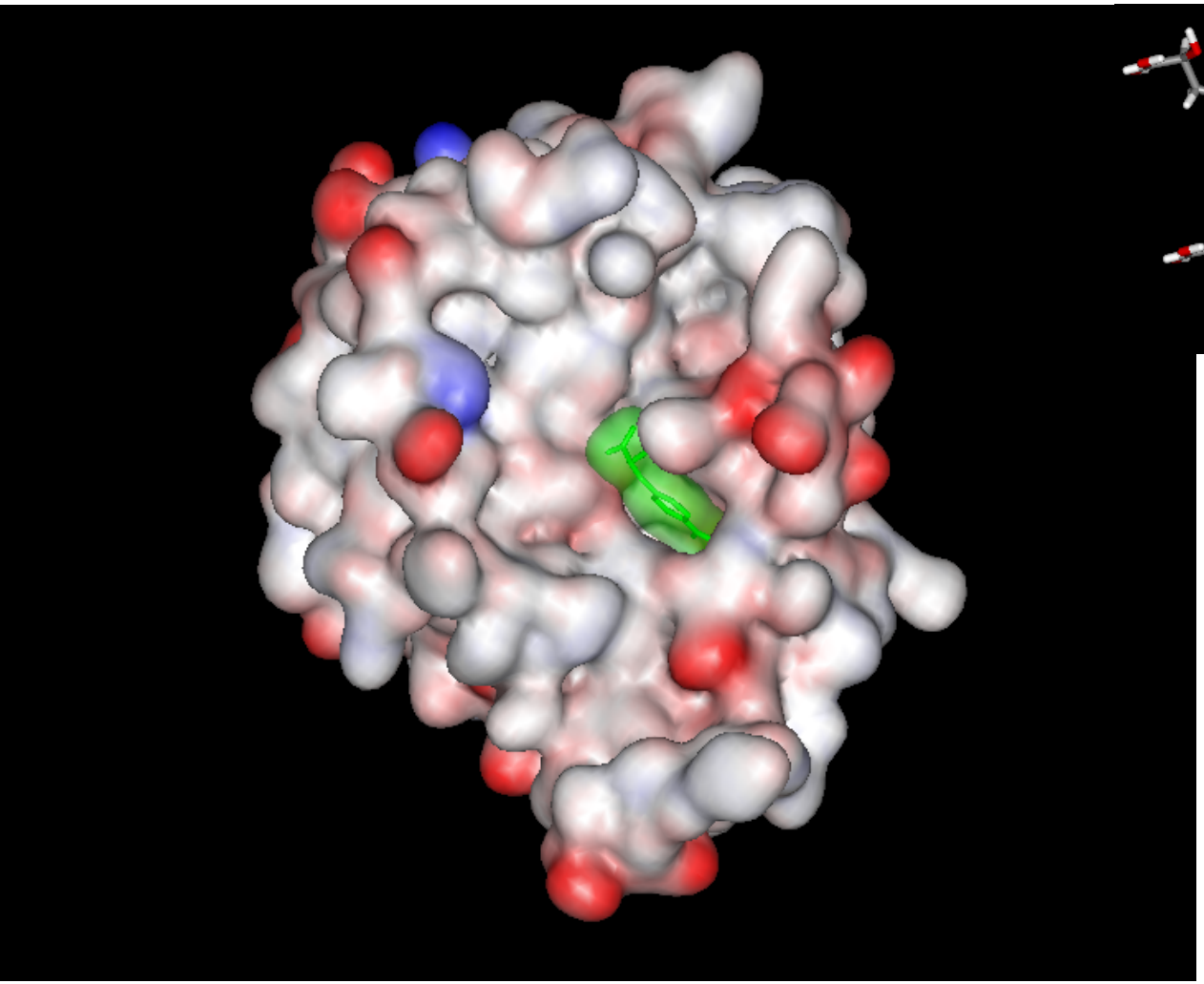
**Protein-Ligand interactions:**

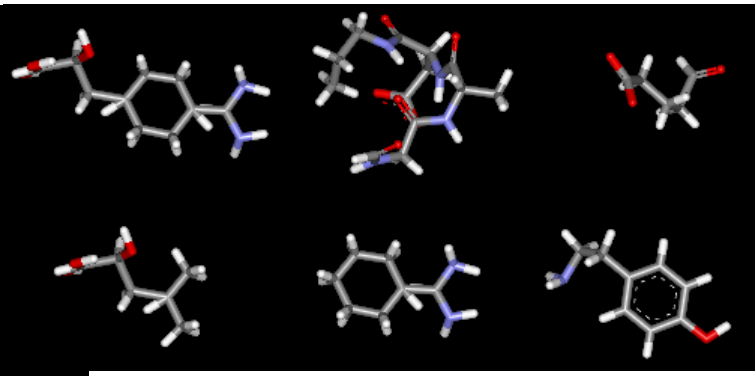
**Intermolecular Interactions (Enthalpy)**

***Hydrogen Bonds***
***Electrostatic Interactions***
***van der Waals Forces***
***π – π Interactions***

**Entropy**

# Computing protein-drug structure

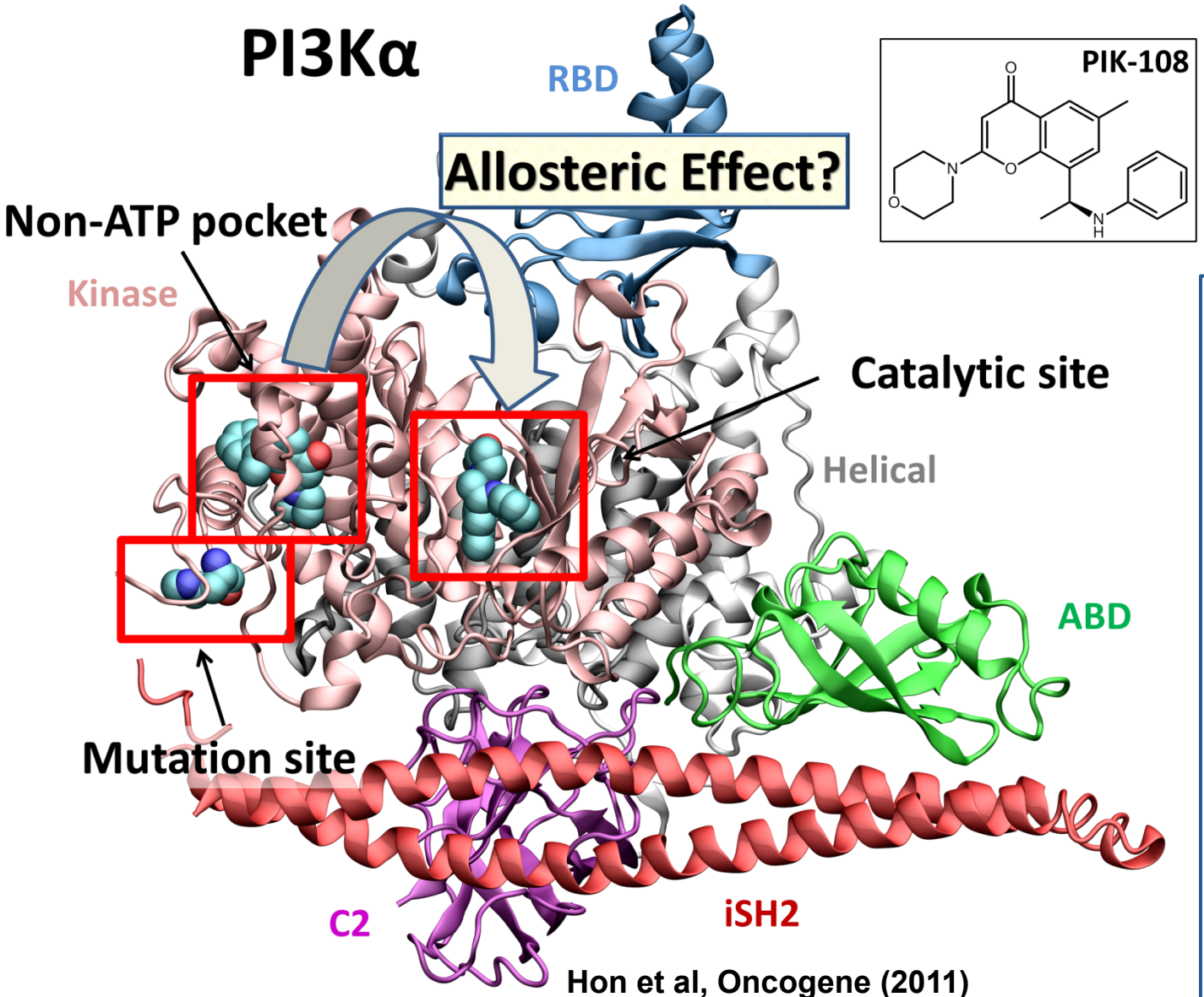




*Virtual Screening*

https://www.youtube.com/watch?v=u49k72rUdyc

# Mutant-specific drugs in “allosteric” pockets

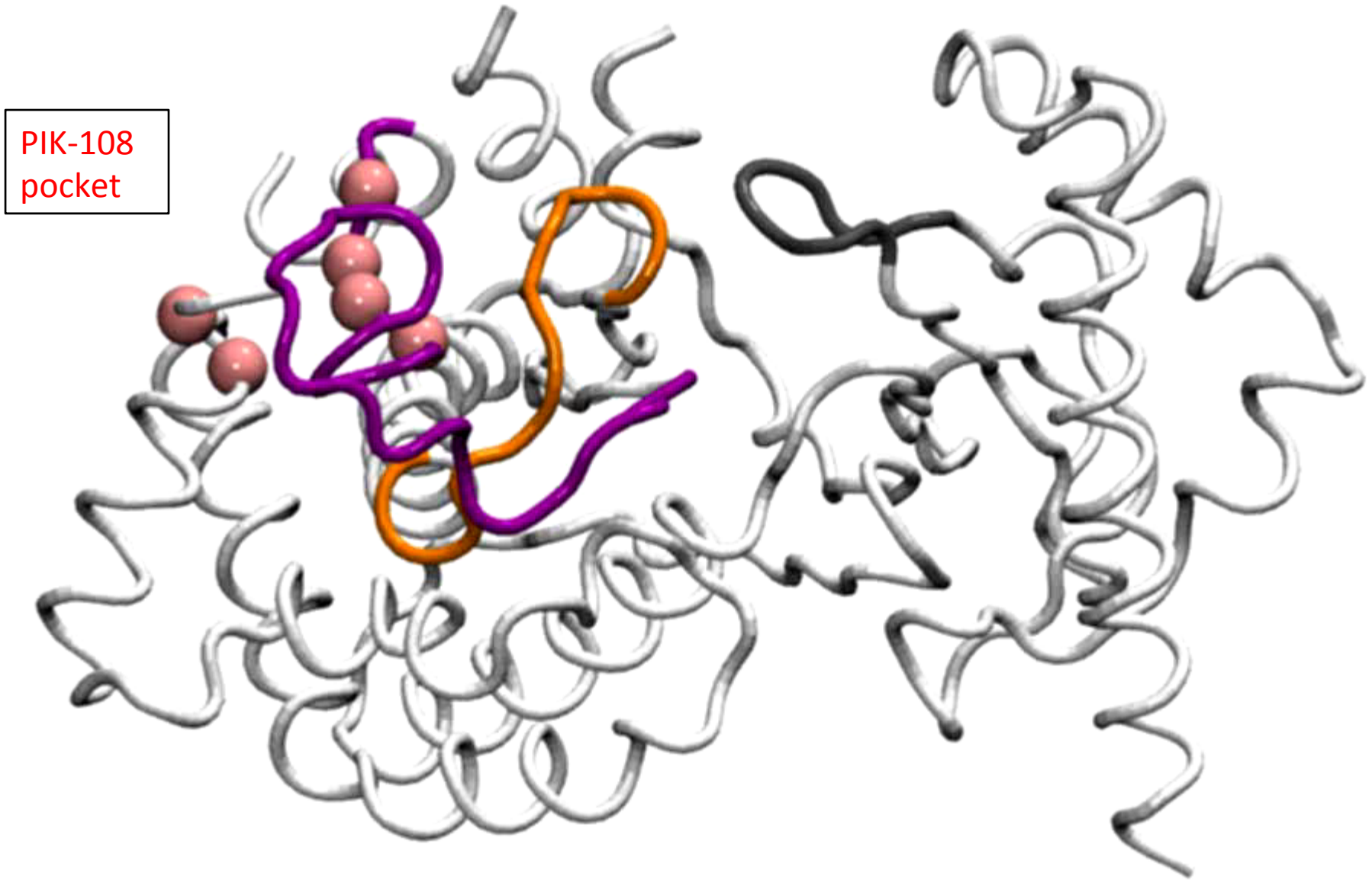






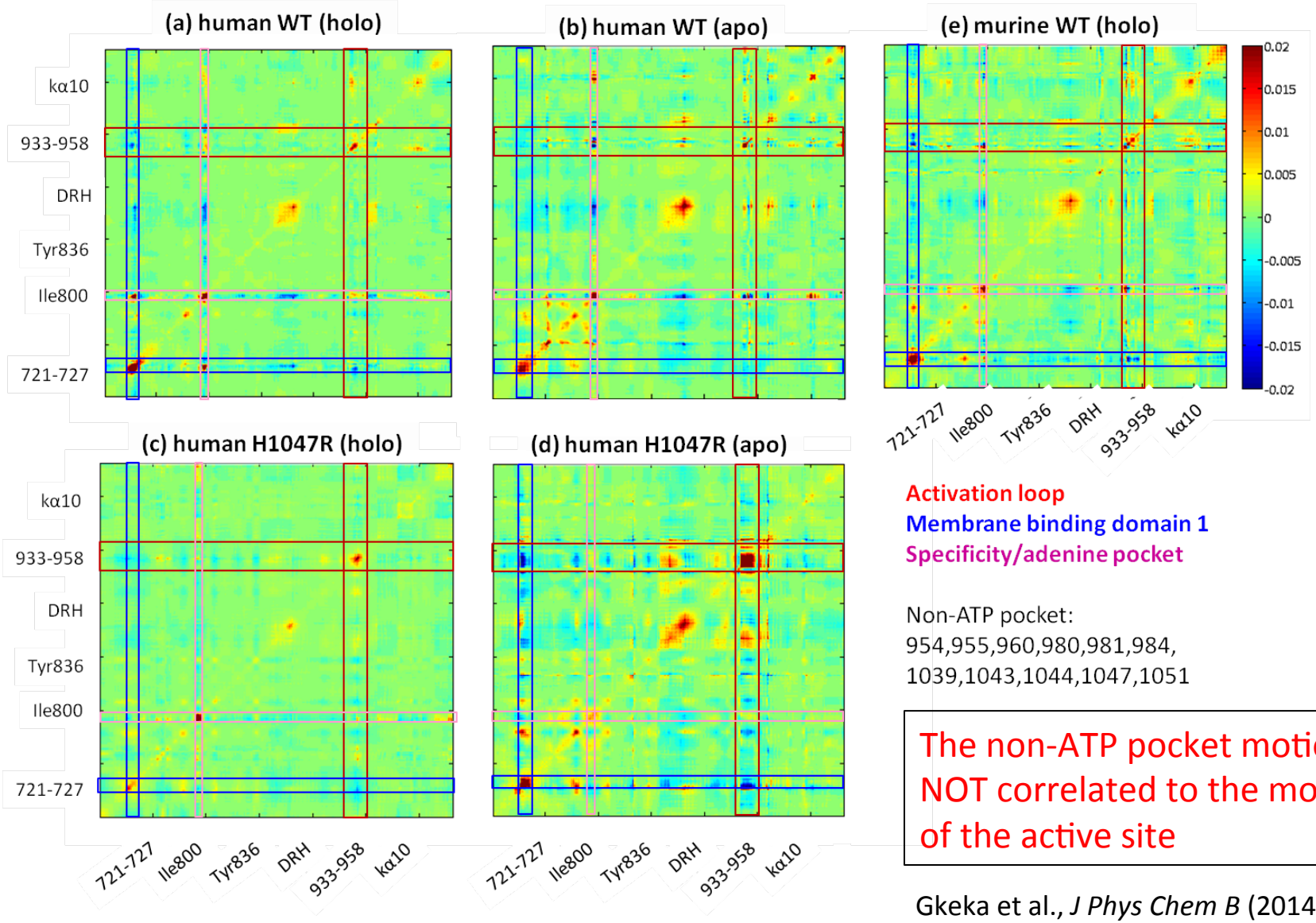
- Active site and non-ATP pocket occupied by **PIK-108**
- MD simulations of WT, H1047R apo and holo forms (100ns production run)
- Is the non-ATP pocket allosteric?
- **Can we discover allosteric pockets with simulations?**

# Assessment of allosteric pockets with PCA



# Correlation of non-ATP pocket motion to active site?

**Positional covariance matrix (through PCA)**




Activation loop
Membrane binding domain 1
Specificity/adenine pocket

Non-ATP pocket:
954,955,960,980,981,984,
1039,1043,1044,1047,1051

The non-ATP pocket motion is NOT correlated to the motion of the active site

Gkeka et al., *J Phys Chem B* (2014)

# Is the non-ATP pocket an allosteric site?

Allosteric inhibitors bind to a different site than the active site and influence the active site conformation.

**Simulations show that the non-ATP pocket is not allosteric**
***How can we measure this experimentally?***

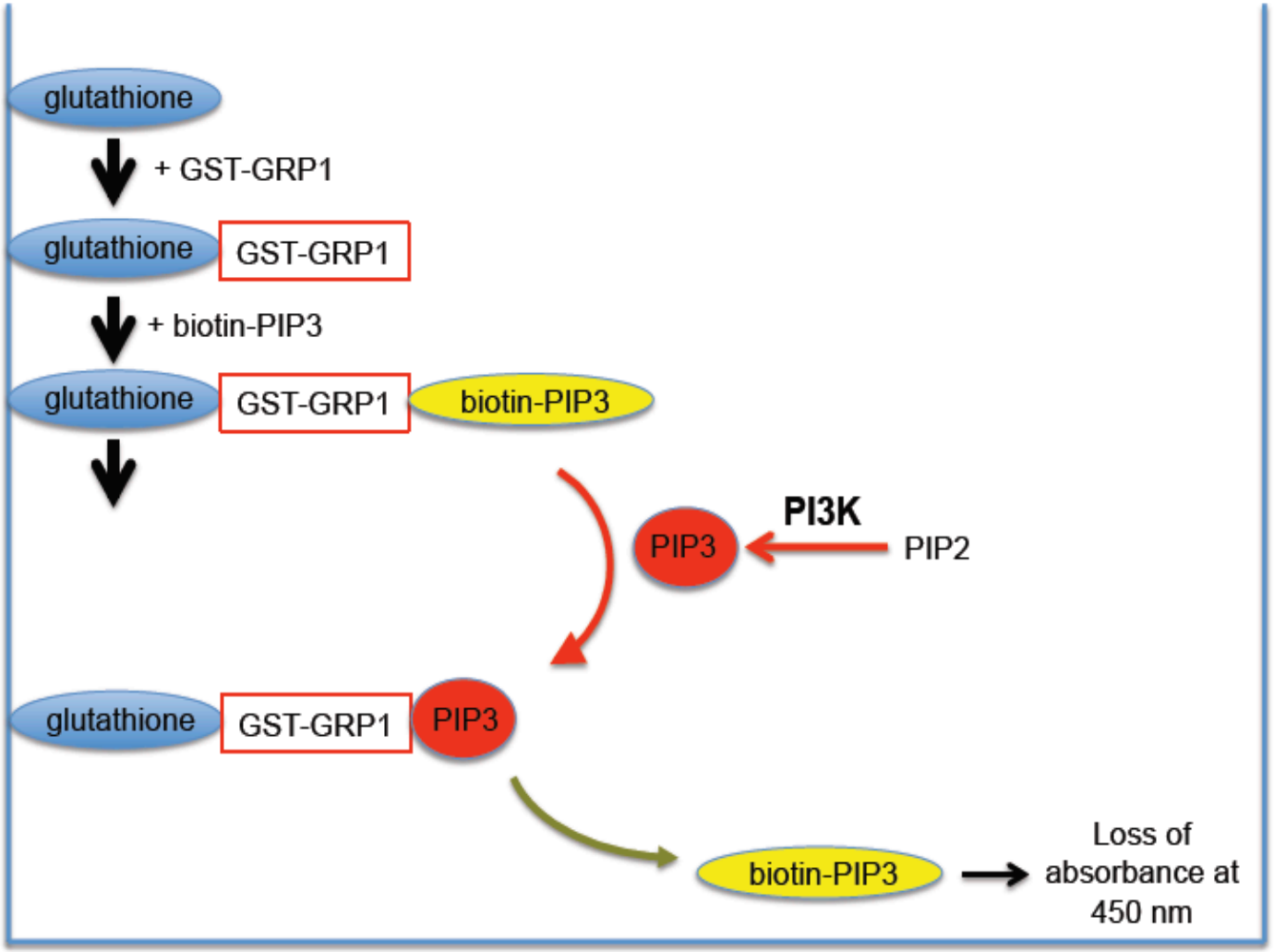
- *Competitive inhibitors* of the active site will be influenced by high ATP concentration (they will lose activity)

- *Non-competitve (allosteric)* inhibitors will not be affected by a high ATP concentration (they will not lose activity)

=> Perform in vitro activity measurement of PI3Kα:

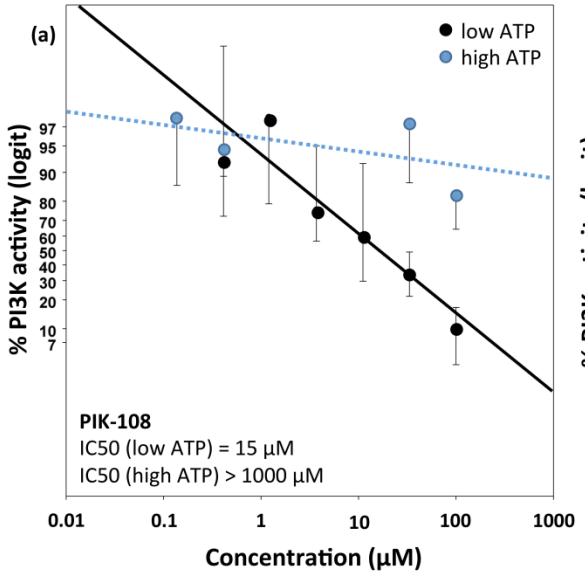
- **Low ATP (100uM)**
- **High ATP (2mM)**

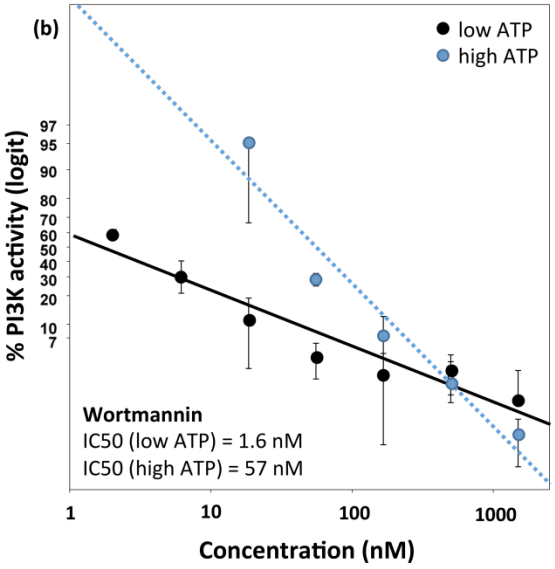
# In vitro cell-free assay with cancer liposomes

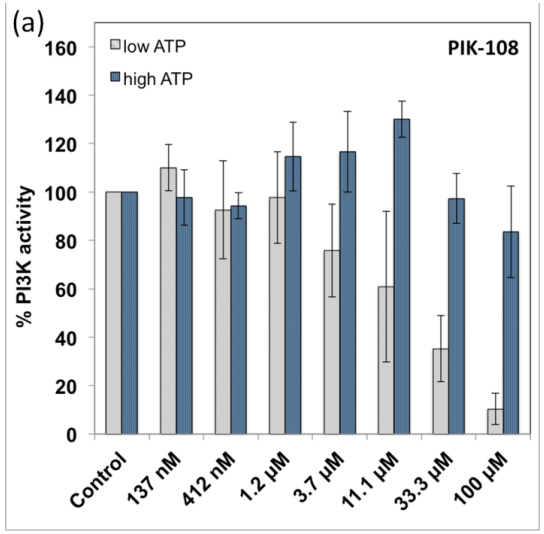


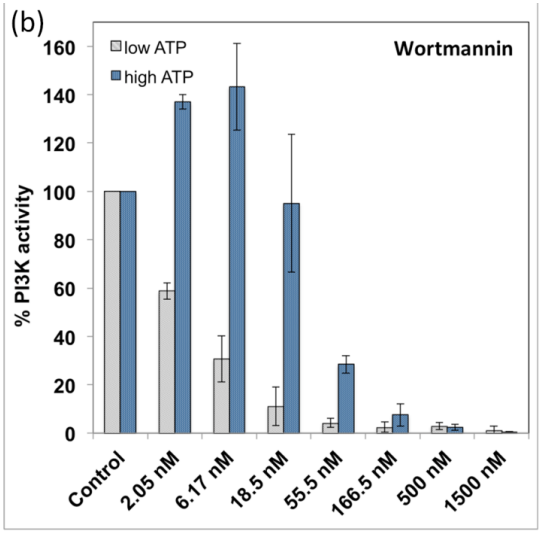

Christoforidis lab, University of Ioannina, in vitro assays
Couladouros lab, University of Athens, synthesis of PIK-108

# The IC50 of PIK-108 depends on ATP concentration




(Christoforidis lab, U Ioannina)

Gkeka et al., *J Phys Chem B* (2014)

Perform in vitro activity measurement of PI3Ka:

- Low ATP (100uM)
- High ATP (2mM)

- *Competitive inhibitors* of the active site will be influenced by high ATP concentration

- PIK-108 and wortmannin dramatically change IC50 with a change of ATP

- Wortmannin is well-known competitive inhibitor

- ***PIK-108 is a competitive inhibitor (not allosteric)***

# Binding site Prediction

Crystal structure

↓

Molecular Dynamics

↓

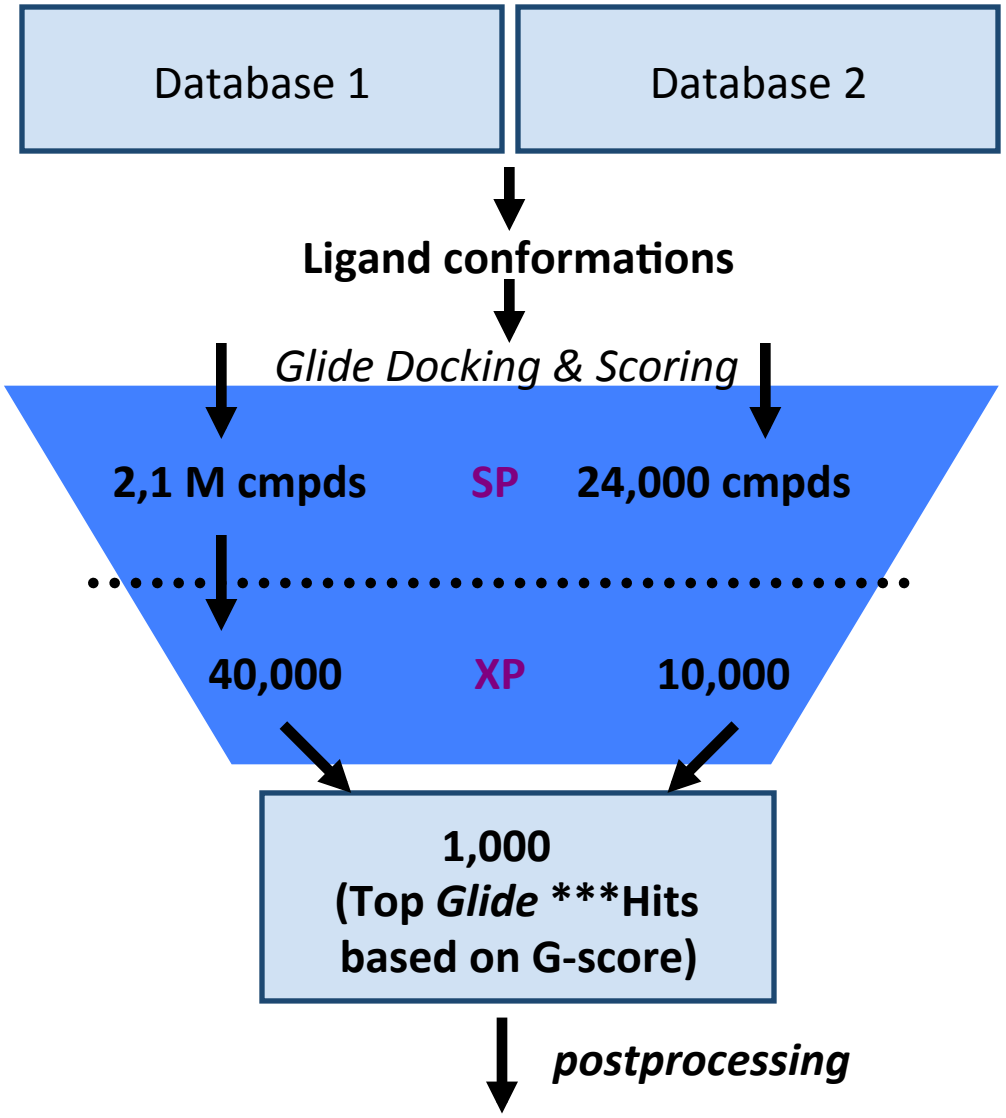
Cluster conformations from trajectory (*RMSD*)

↓

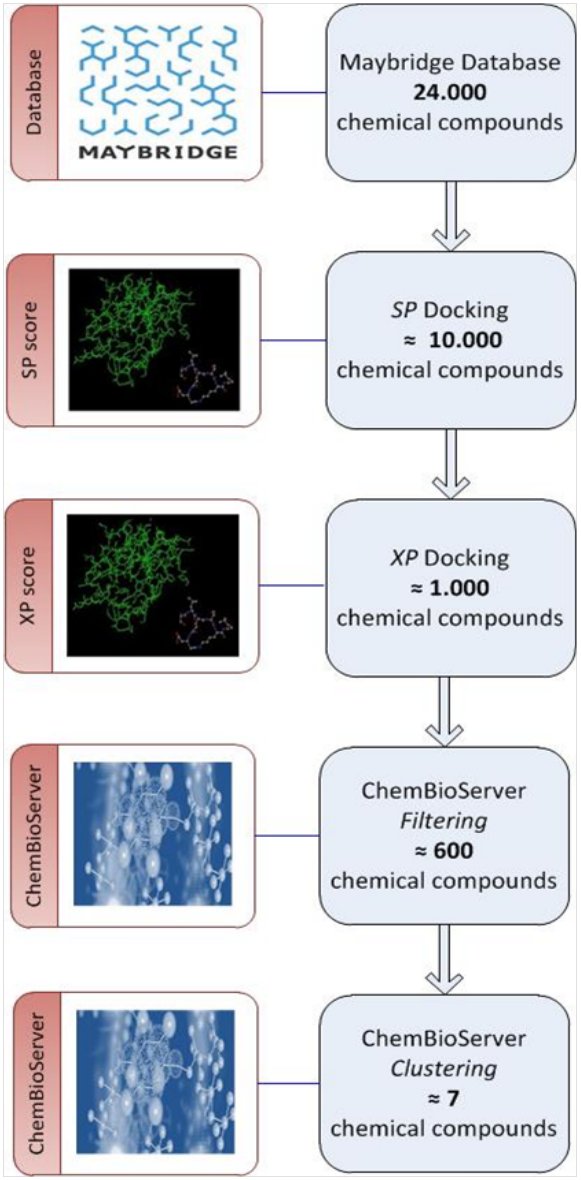
Binding site prediction and *confirmation for allosteric pockets*
**(Q-Site Finder, PCA)**

Lionta et al, *Curr Top Med Chem* (2014)

# Virtual Screening



# How are compounds selected for assaying?



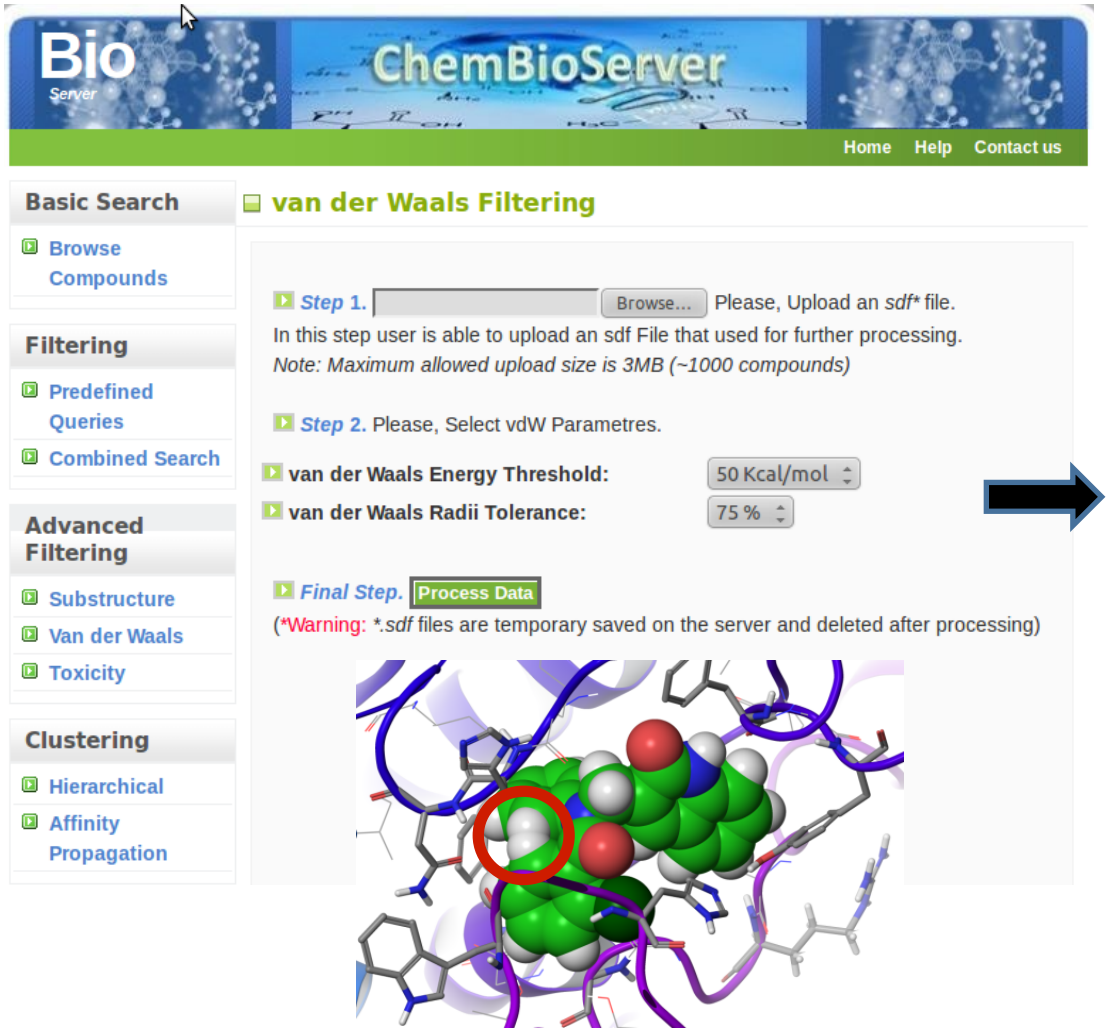

- **Library docking using Glide SP, XP**
- **1000 Top-scored XP compounds**
- **Postprocessing with ChemBioServer**
- **Calculate ADME/tox properties**
- **Check for bad vdW contacts**
- **Hierarchical Clustering**
- **Affinity Propagation (exemplars)**
- **Visualization: check for compound conformations**

**http://bioserver-3.bioacademy.gr/Bioserver/ChemBioServer/**

Athanasiadis, Cournia, Spyrou, *Bioinformatics* (2012)

# Pre/Postprocessing with ChemBioServer

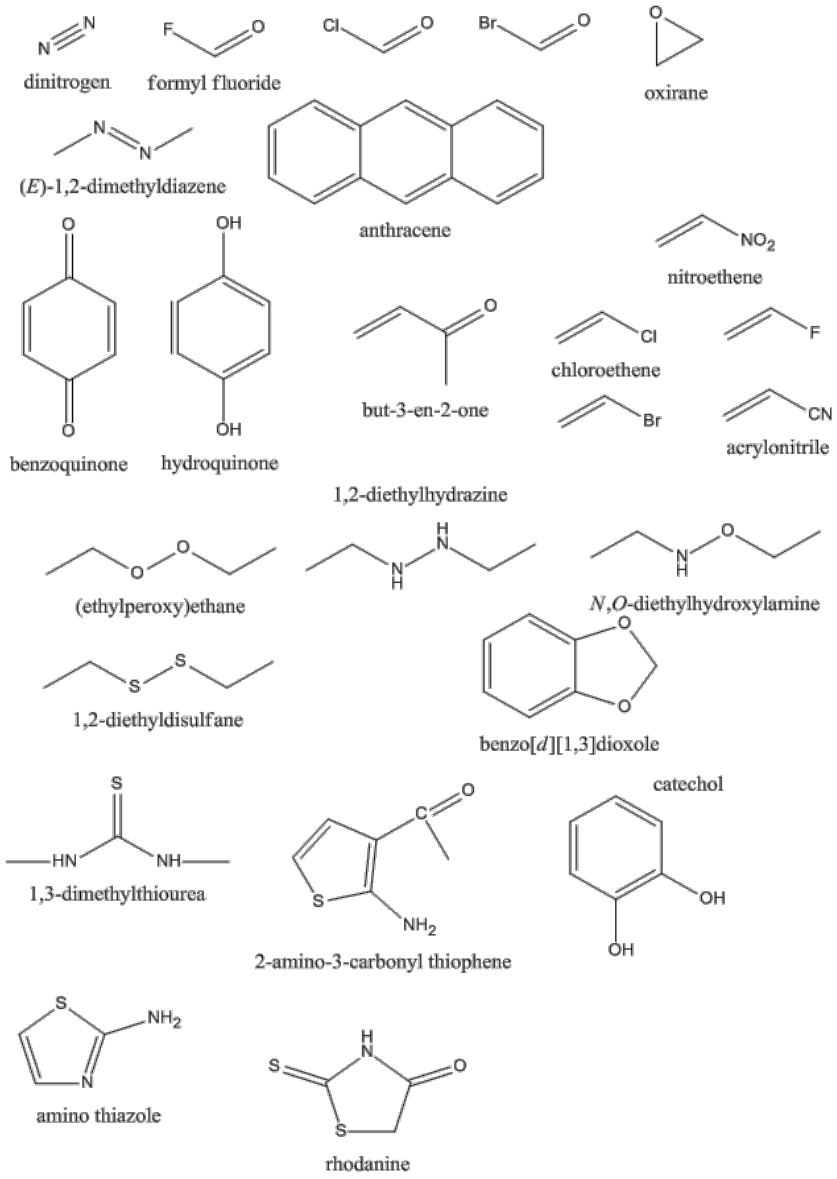
ChemBioServer post-processes virtual screening results



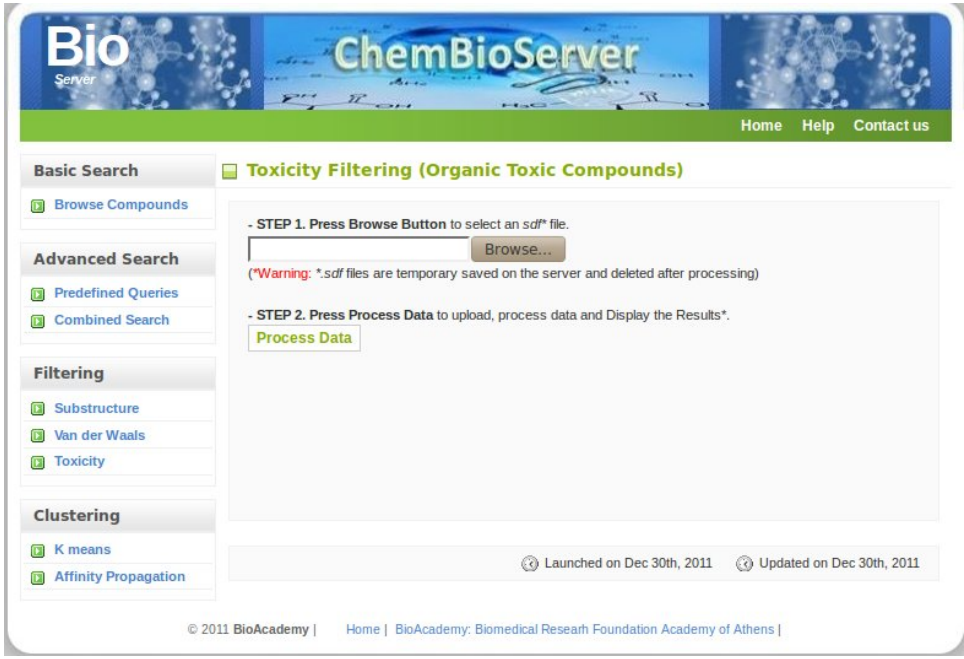

| Compound ID | VDW Energy Test | VDW Distance Test |
|---|---|---|
| Compound: 1 AW 00785 | - PASS AW 00785 - Browse List For Details... | - FAIL AW 00785 - Browse List For Details... |
| Compound: 2 AW 00788 | - PASS AW 00788 - Browse List For Details... | - FAIL AW 00788 - Browse List For Details... |
| Compound: 3 AW 00785 | - PASS AW 00785 - Browse List For Details... | - FAIL AW 00785 - Browse List For Details... |
| Compound: 4 AW 00939 | - PASS AW 00939 - Browse List For Details... | - FAIL AW 00939 - Browse List For Details... |
| Compound: 5 AW 00694 | - PASS AW 00694 - Browse List For Details... | - FAIL AW 00694 - Browse List For Details... |
| Compound: 6 CD 10205 | - PASS CD 10205 - Browse List For Details... | - PASS CD 10205 - Browse List For Details... |
| Compound: 7 GK 02096 | - PASS GK 02096 - Browse List For Details... | - FAIL GK 02096 - Browse List For Details... |
| Compound: 8 HTS 01561 | - PASS HTS 01561 - Browse List For Details... | - FAIL HTS 01561 - Browse List For Details... |
| Compound: 9 MWP 00404 | - PASS MWP 00404 - Browse List For Details... | - FAIL MWP 00404 - Browse List For Details... |
| Compound: 10 NRB 02577 | - PASS NRB 02577 - Browse List For Details... | - FAIL NRB 02577 - Browse List For Details... |

*Athanasiadis, Cournia, Spyrou, Bioinformatics (2012)*

# Pre/Postprocessing with ChemBioServer

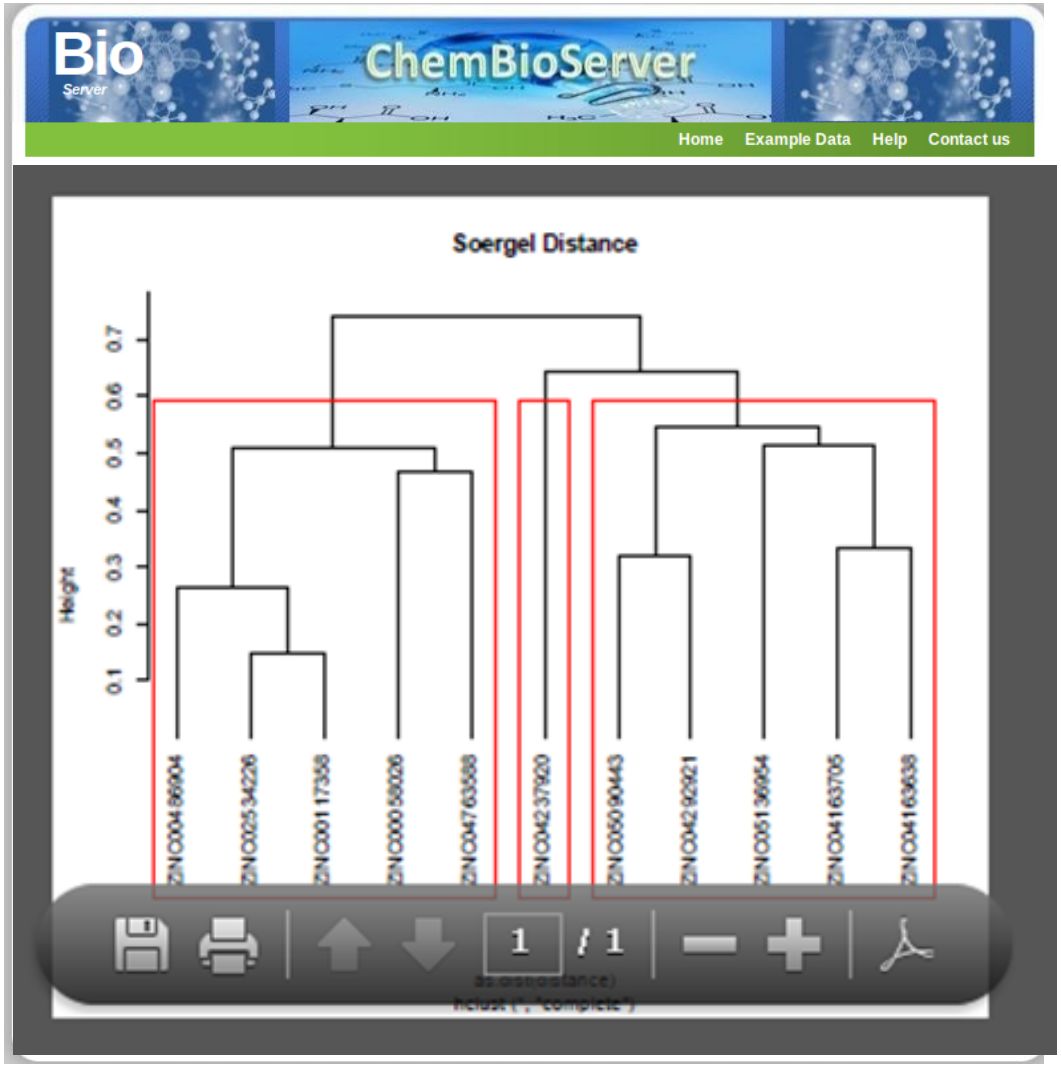



**.gr/Bioserver/ChemBioServer/**



# Clustering and molecular similarity

**Similar structures and properties ⇒ similar activity**




**500- 1,000 compounds**

↓

**Approximately 15 representative clusters**

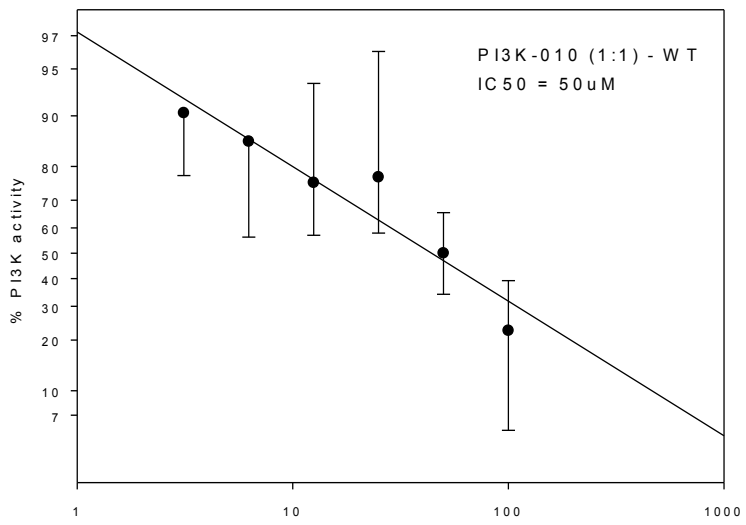
↓

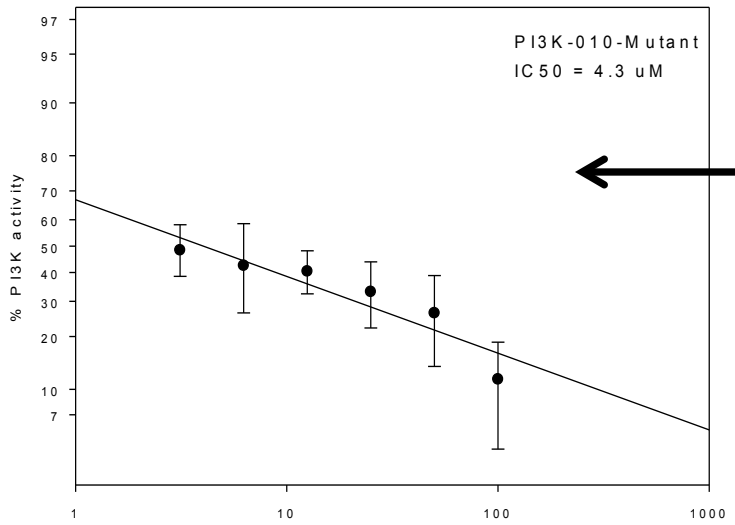
**200 exemplars**

↓

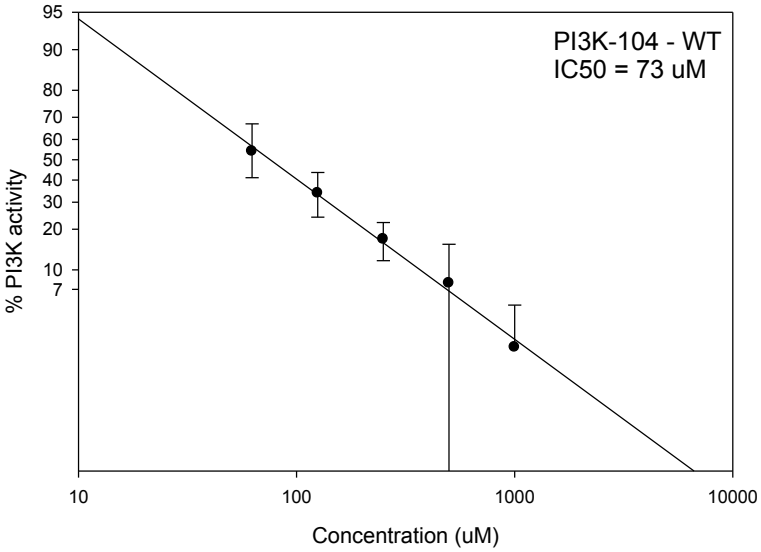
**Visualization, purchase ~10 compounds**

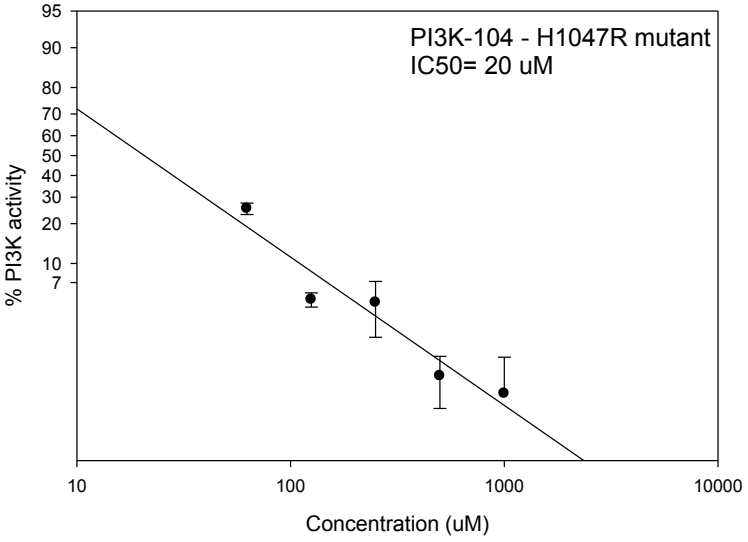
# In vitro cell-free assay with cancer liposomes



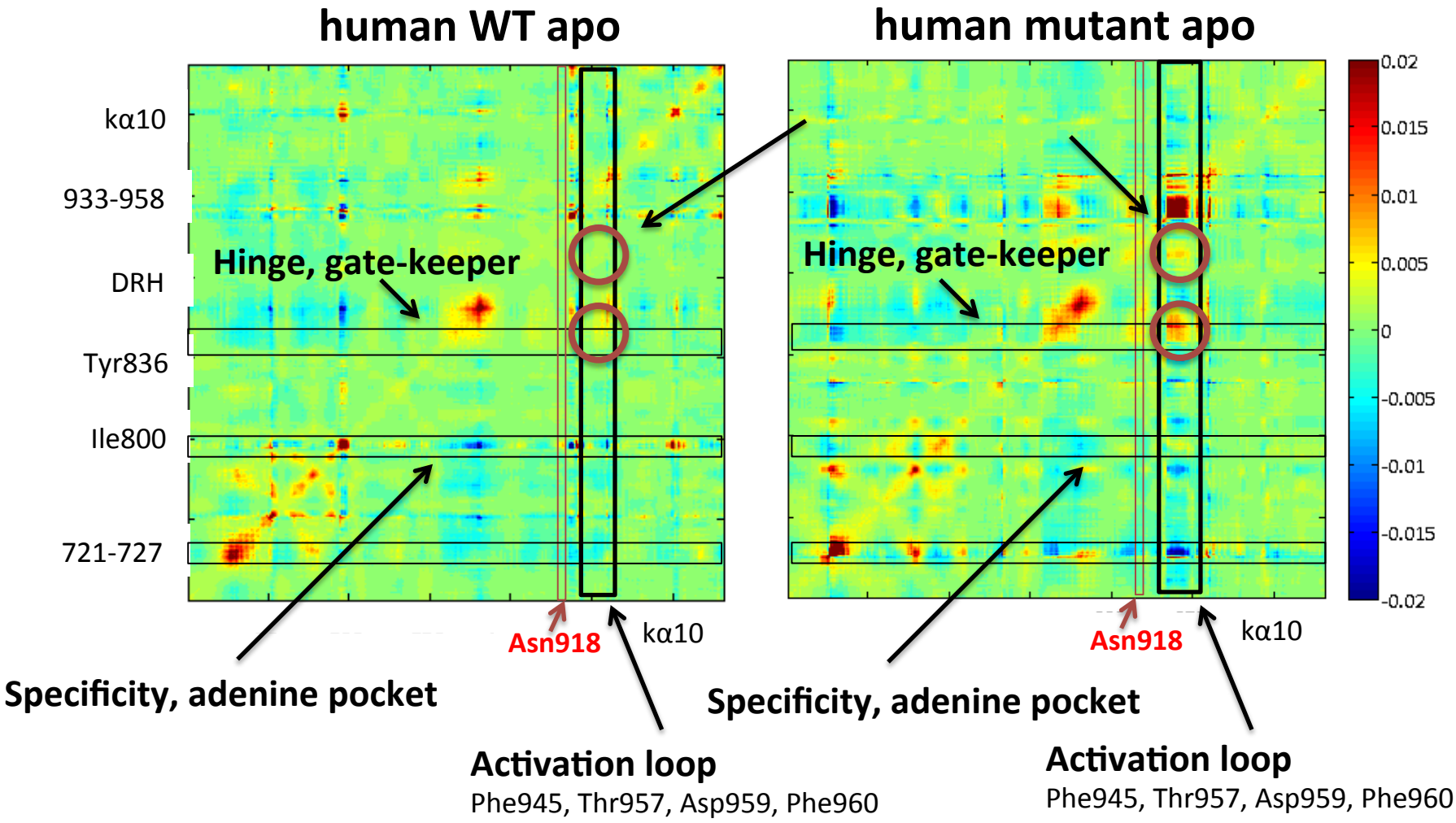

**11-fold selectivity of the mutant vs the WT**



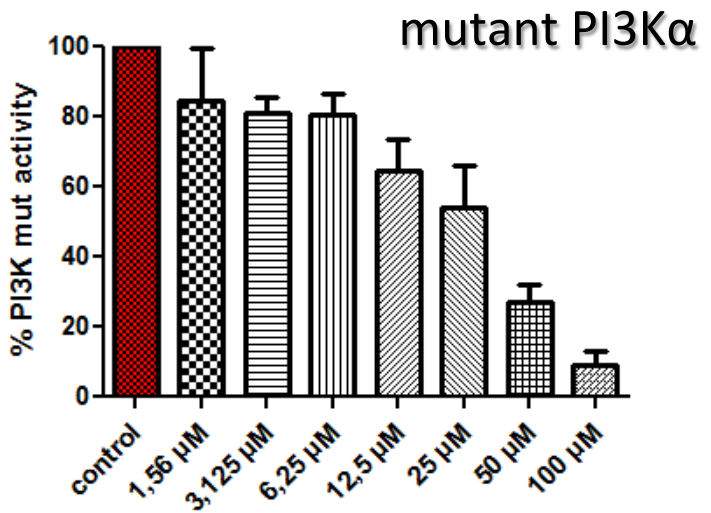

IC50 = the concentration of the compound required to inhibit the protein by 50%

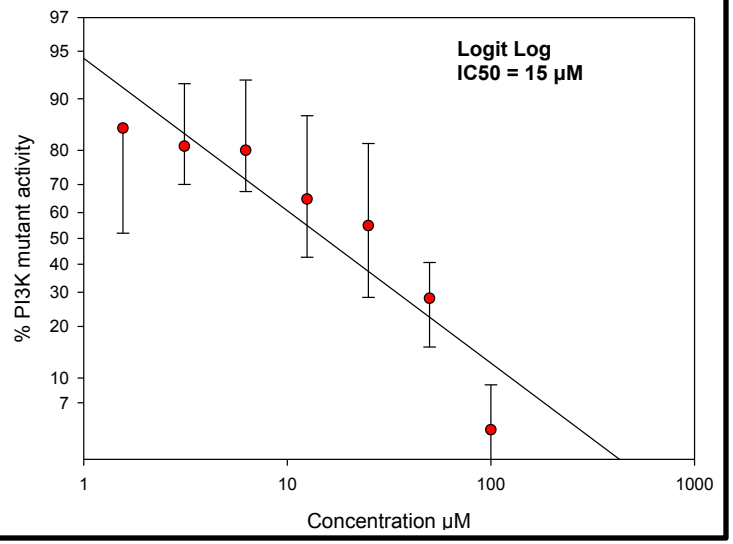
# Is PI3K-010 an allosteric (non-competitive) inhibitor?




**The motion of the pocket where PI3K-010 resides IS correlated to the motion of the active site.**

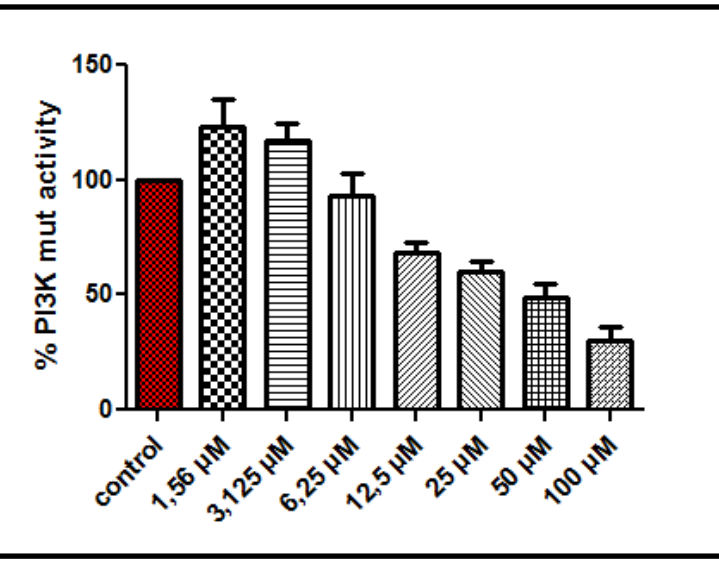
# Is PI3K-010 an allosteric (non-competitive) inhibitor?

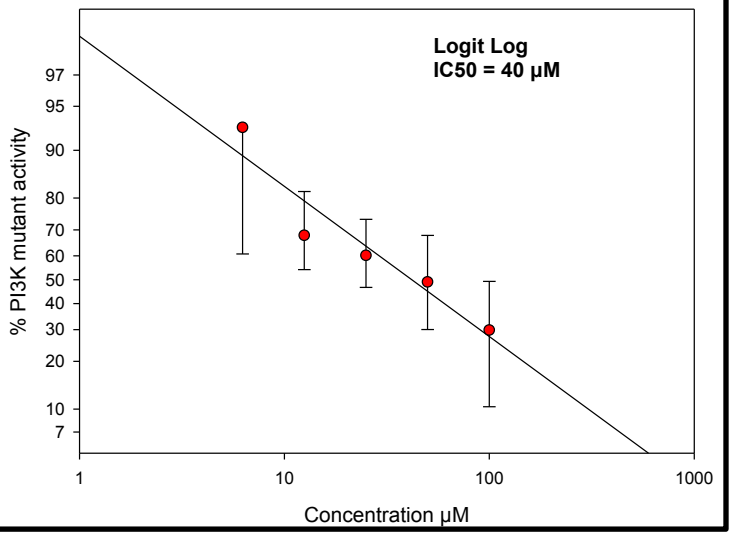



**Low ATP (100µM):**
**IC50 = 15 µM**

**High ATP (2mM):**
**IC50 = 40 µM**



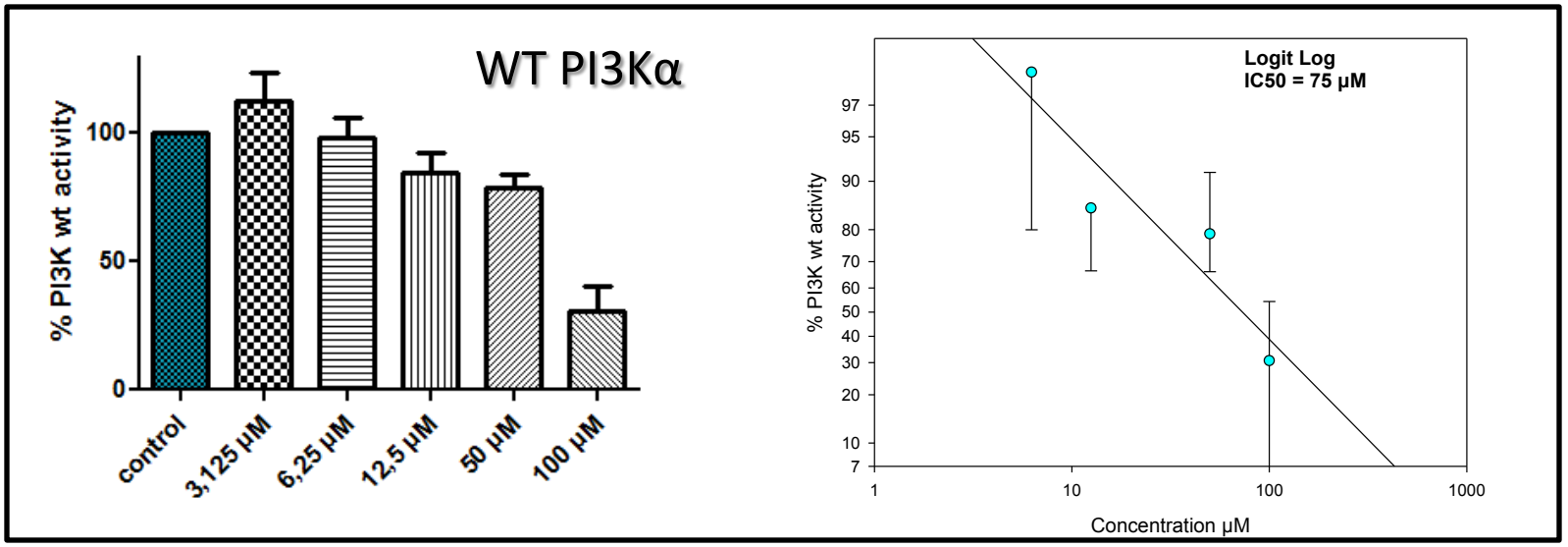

**PI3K-010 IC50 is not influenced by ATP concentration**

**Could be considered allosteric**

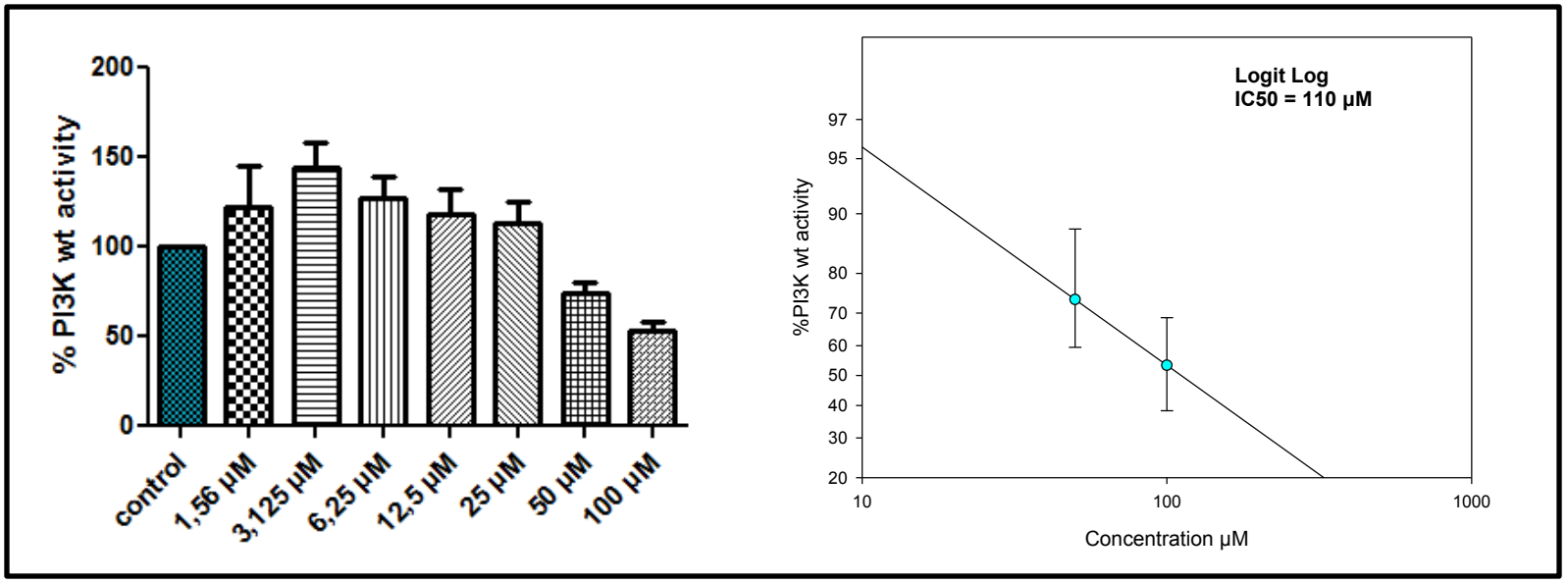
**2 experiments low ATP, 4 experiments high ATP**

# Is PI3K-010 an allosteric (non-competitive) inhibitor?

**Low ATP (100 µM)**

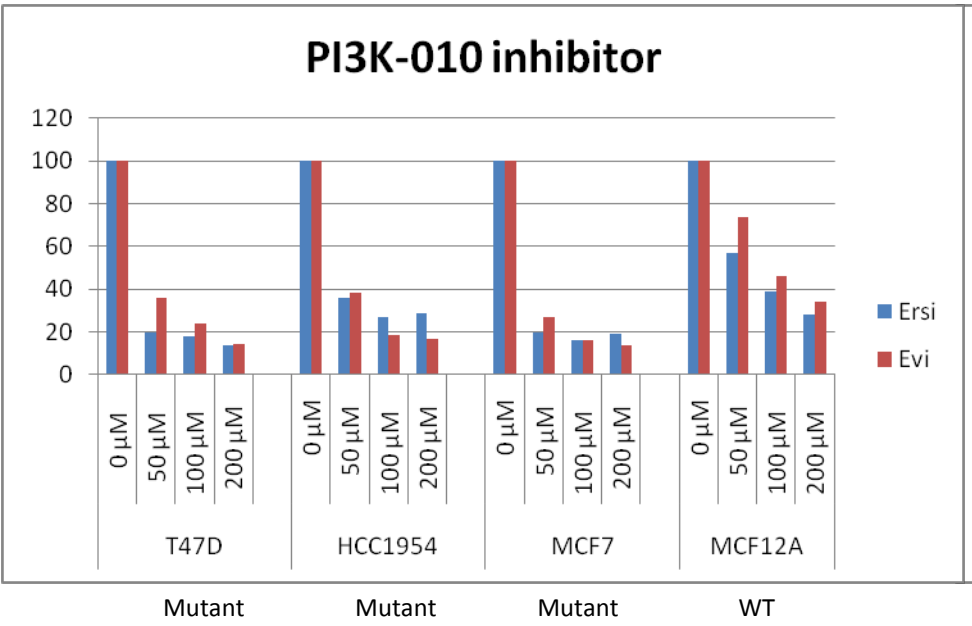


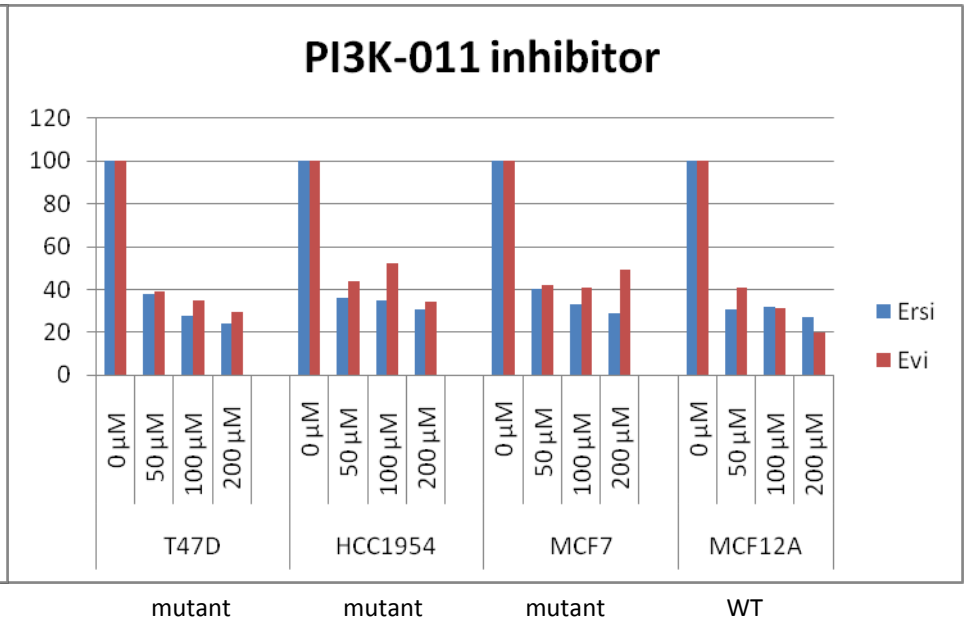

**High ATP (2mM)**




**2 experiments low ATP, 3 experiments high ATP**

# MTT assay on mutant and WT PI3Kα

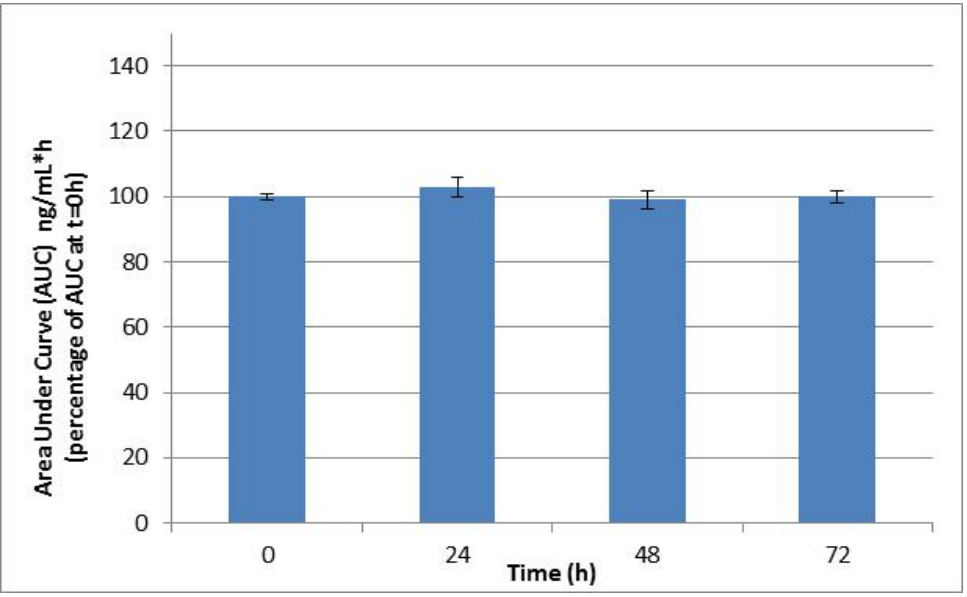



- Mutant-specific inhibition is possible
- IC50 WT = 7uM
- IC50 H1047R = 1uM

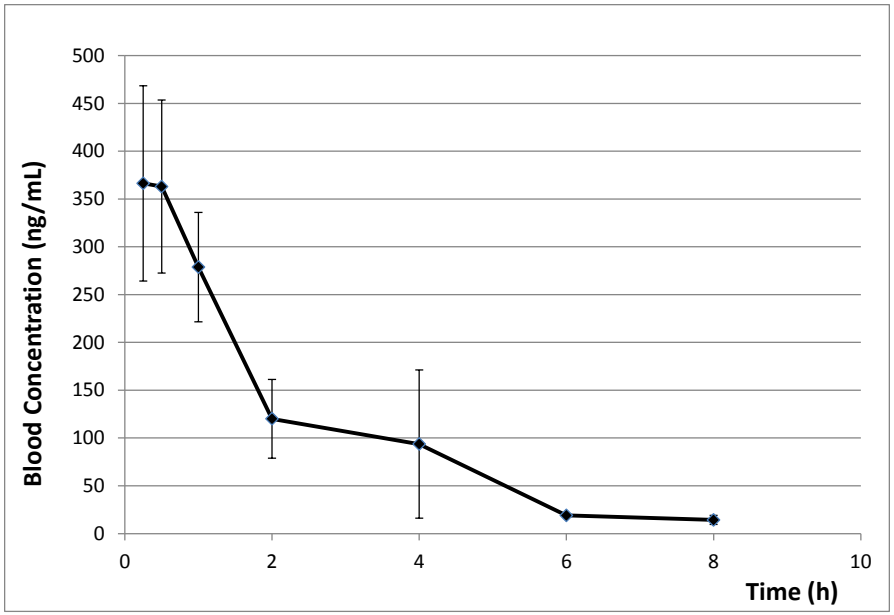
# Pharmacokinetic experiments on PI3K-010

*Stability of compound PI3K010 in cell conditioned- medium*




(Tamvakopoulos lab, BRFAA)

*Mean blood concentrations of PI3K010 in corn oil following oral dosing in mice (10 mg/Kg).*




Cmax of 396 ng/mL (~ 1 µM)
4 h post-dose - average concentrations of 100 ng/mL (~ 0.3 µM).

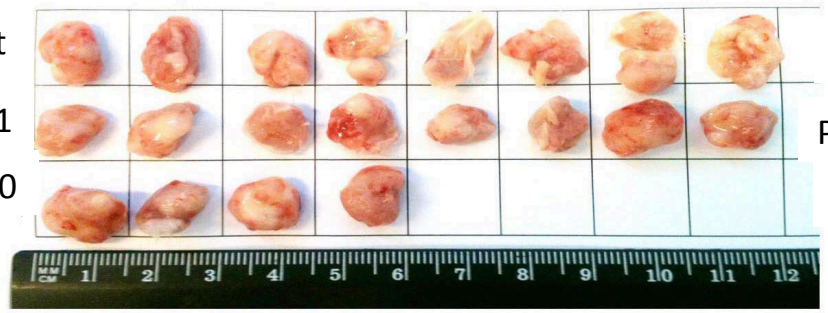
# Preclinical study of PI3K-010 (xenografts)

## MDA-231-MB (PI3Kα WT)

Solvent

PI3K-021

PI3K-010

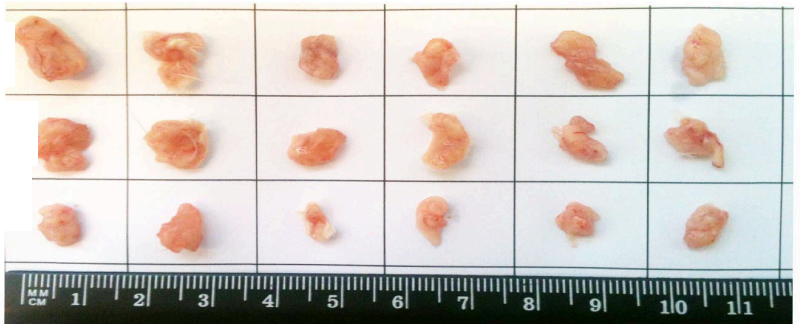


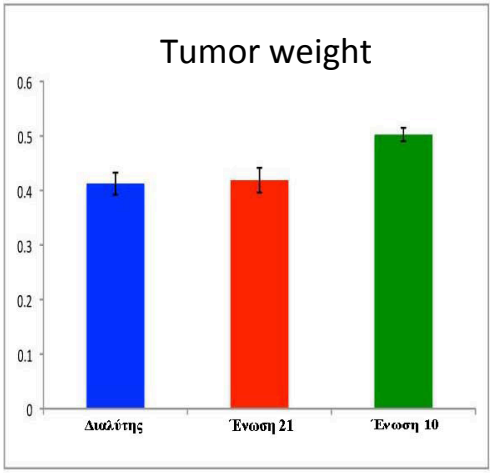
## HCC1954 (H1047R PI3Kα mutant)

Solvent

PI3K-021

PI3K-010

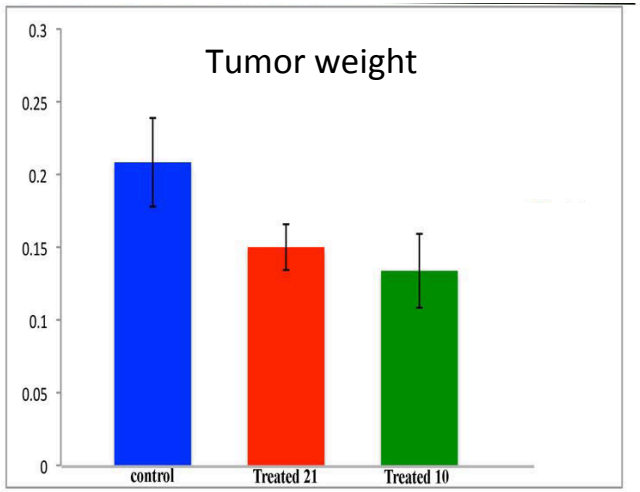





Solvent

PI3K-021

PI3K-010




(D. Stellas, Klinakis & Efstratiadis labs)

***PI3K010 in corn oil following oral dosing in mice (100 mg/Kg).***

# Lead optimization of PI3K-010

**Synthesis of analogs**

**Compound PI3K-021**

**In vitro cell-free assay**

**IC50 WT: > 1000 μM**

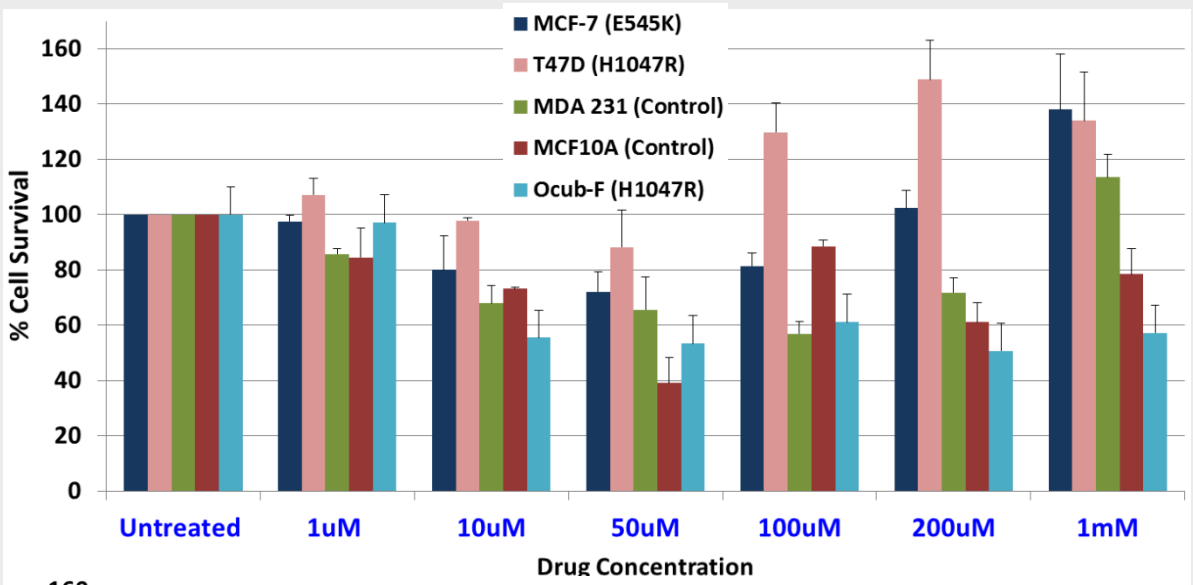
**IC50 Mutant: 13.5 μM**

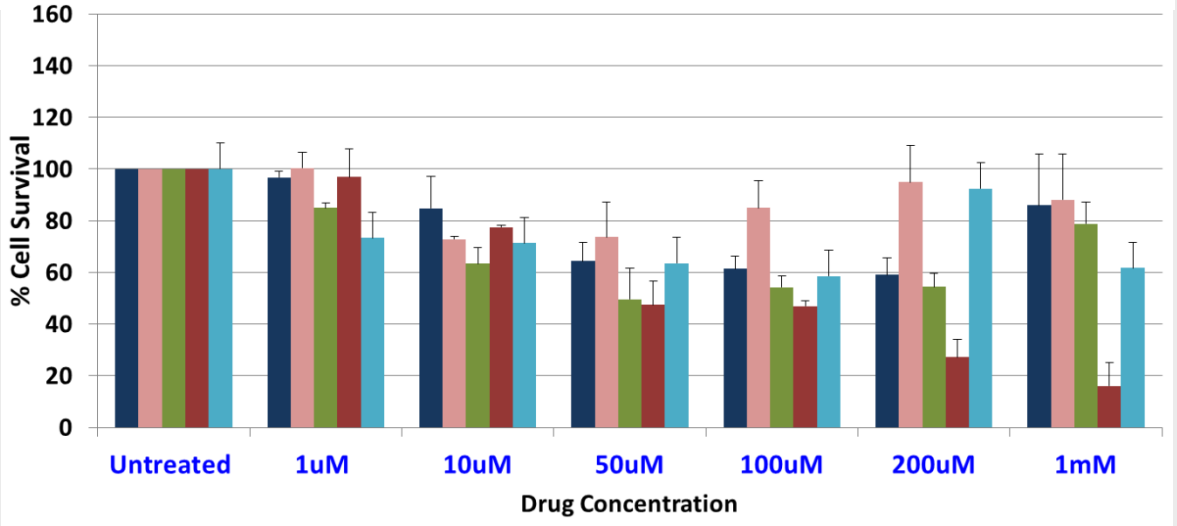
**Selectivity > 100 fold**

Solubility issues with PI3K-021



Optimization of pchem properties

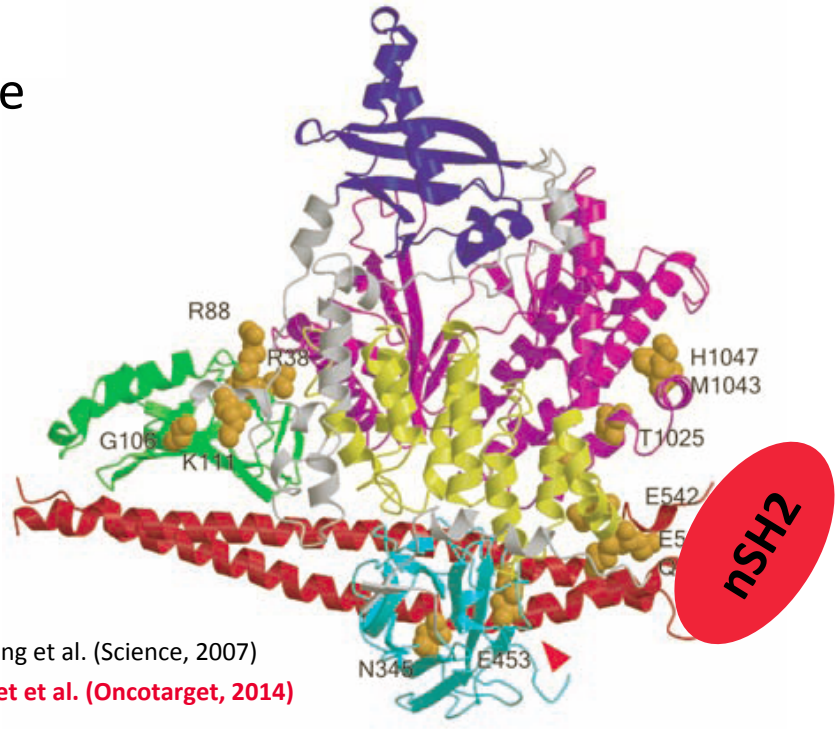



Couladouros lab, University of Athens, synthesis

# PI3Kα: most commonly mutated kinase in cancer

- PI3Ka is a membrane-associated lipid kinase
- Involved in cell growth, proliferation, differentiation
- Most commonly mutated kinase in the human genome ⇒ cancer




Huang et al. (Science, 2007)

Millet et al. (Oncotarget, 2014)

80% of all mutations:

Glu545Lys His1047Arg

30% of breast cancer patients

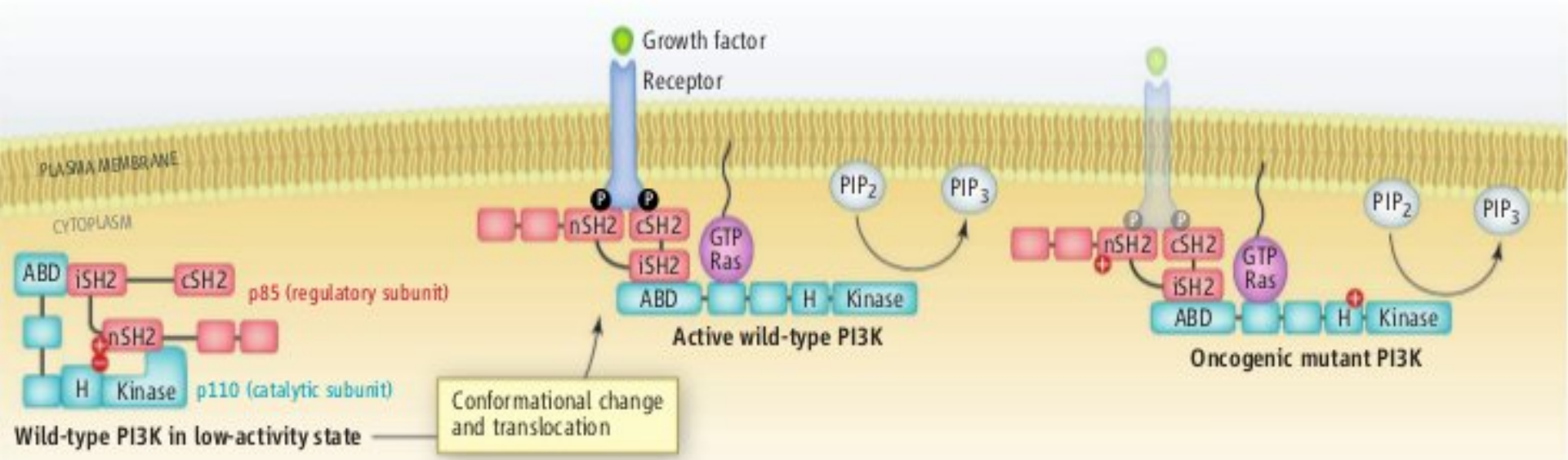
**Mechanism of overactivation?**

**Mutant and isoform specific therapies?**

**MD Simulations**
**Virtual screening**
**Property prediction**
***In vitro* assays**
**Lead Optimization**

# PI3Kα E545K proposed mechanism of activation

Phosphopeptides bind to the nSH2 domain and release the inhibitory contact thus activating the protein




**Phosphopeptide binding**
**nsh2 detachment**
**=> activation**

Glu545 is located precisely where the phosphopeptide binds to nSH2. Glu545Lys changes the charge-charge interactions of the nSH2-helical domains leading to enzyme overactivation.

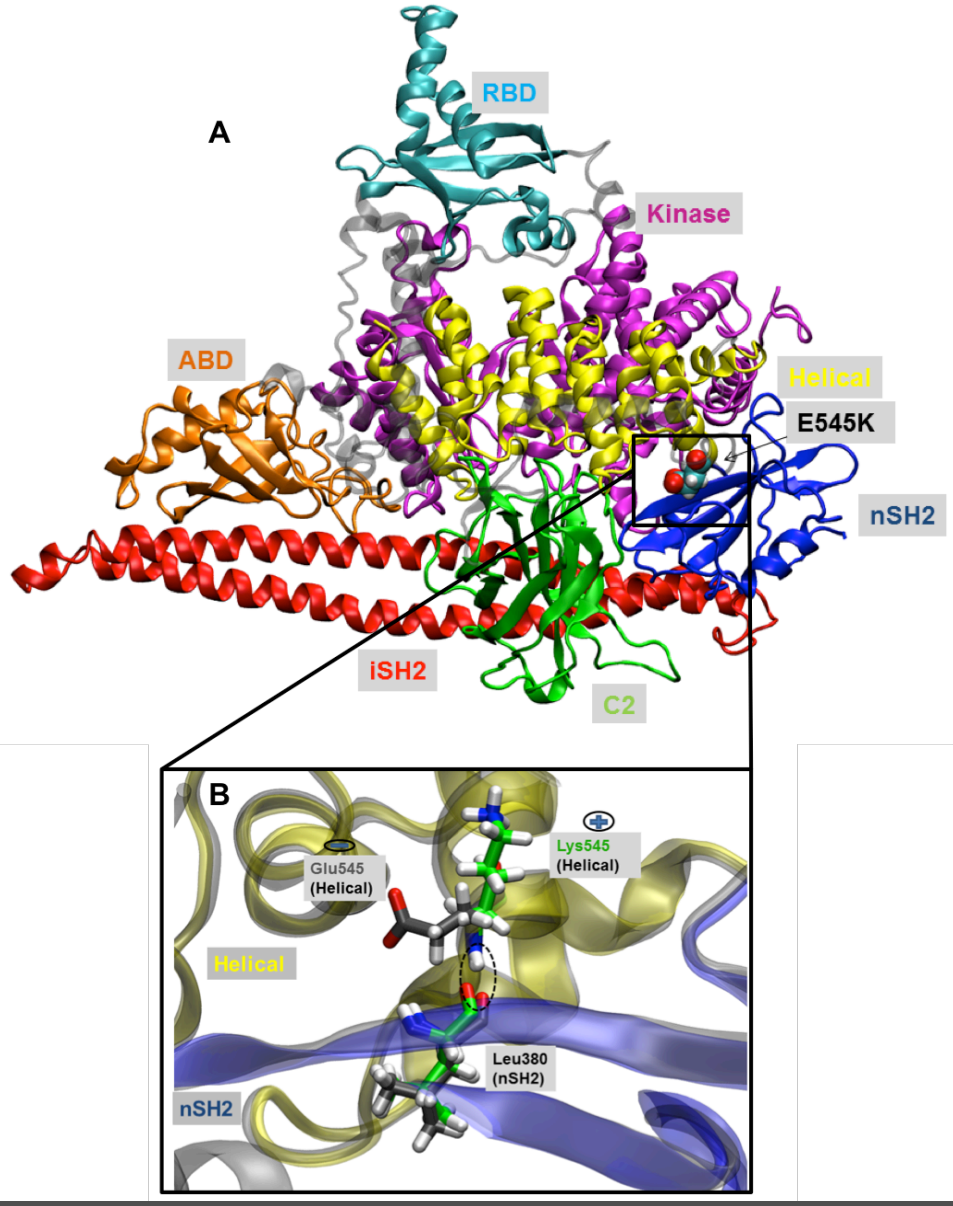
*Burke, et. al., PNAS, 2012*
*Lee, et. Al. Science, 2007*

# PI3Kα WT and E545K mutant MD simulations

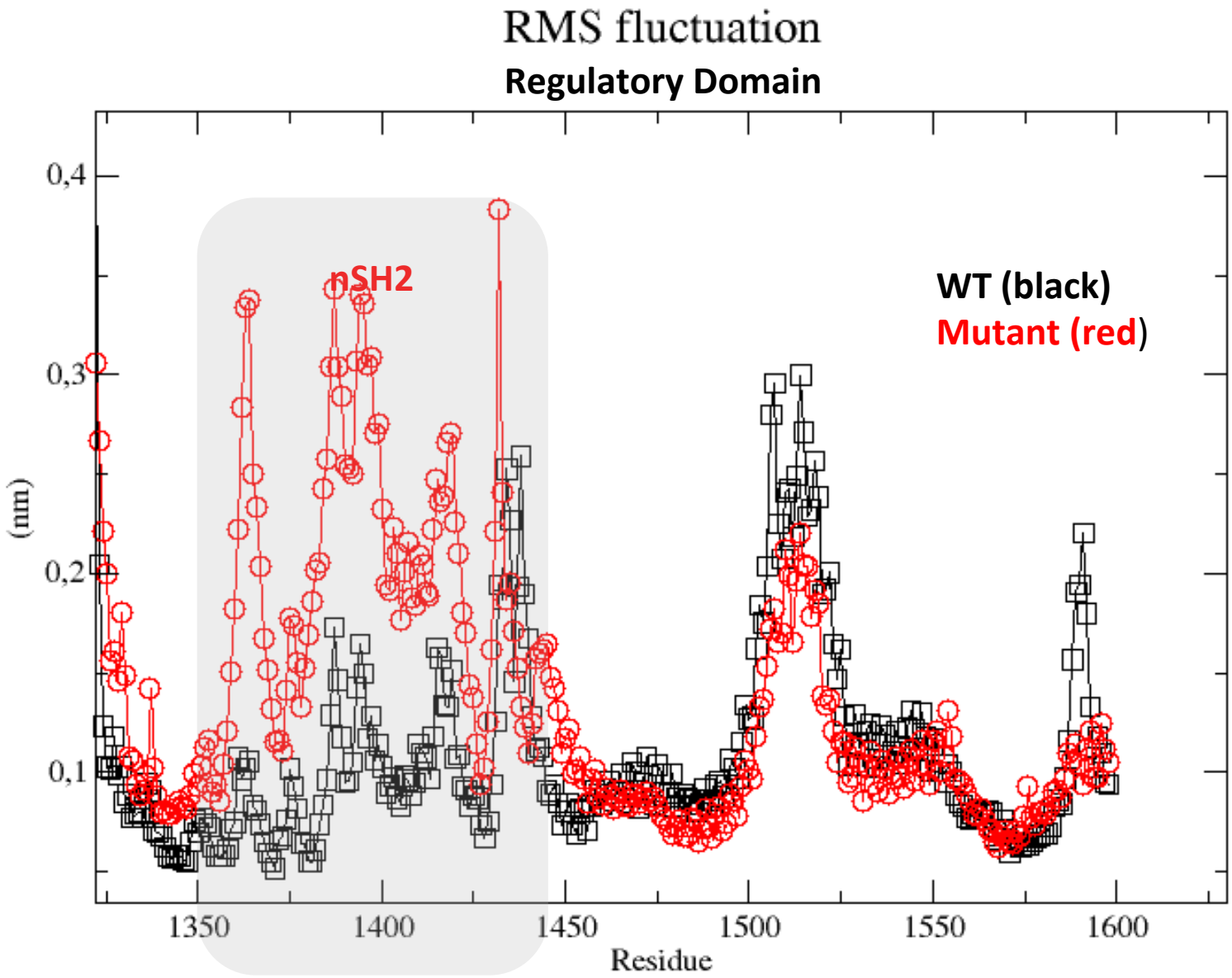
Crystal structure of E545K mutant not available
=> Mutation of WT (PDB ID: 4OVU) residue 545 to Lysine

System size:
355.000 atoms
(1344 a.a + water+ions)

| Simulation | Time (ns) |
|---|---|
| WT replica 1 | **800** |
| WT run 1 | **900** |
| Mutant run 1 | **900** |
| Mutant replicate 1 | **900** |
| Mutant replicate 2 | **900** |
| Mutant replicate 3 | **900** |



# RMSF of the E545K regulatory domain



# Map of Distance Fluctuations (MDF)

$$A_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle$$

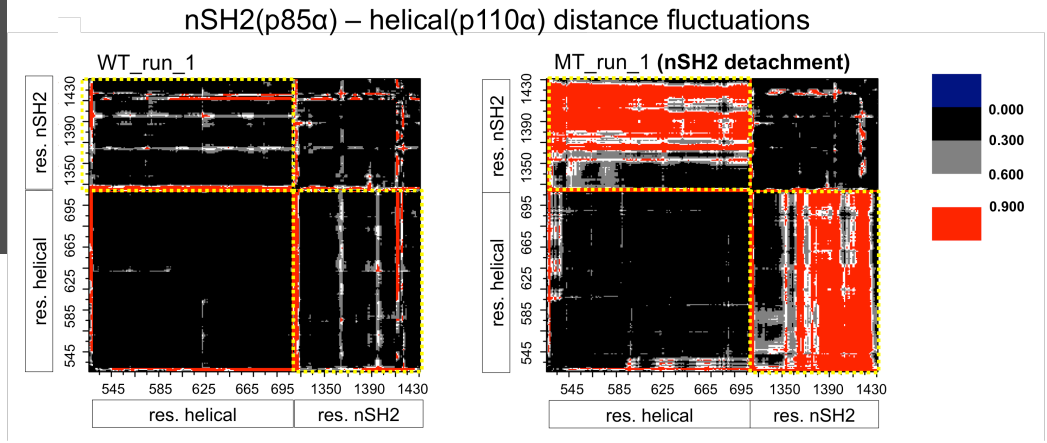
dij is the time dependent distance between Calpha of residue i-j

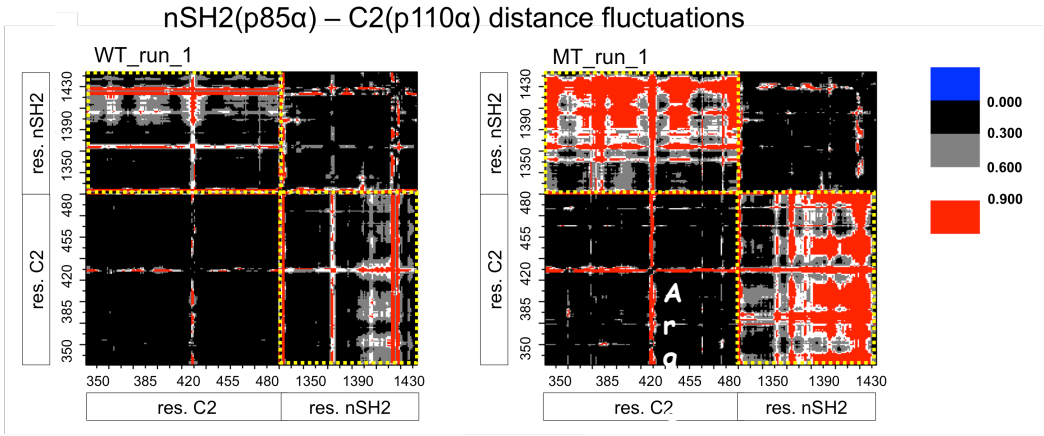
Good communication between residues → if the dij fluctuates less (black)

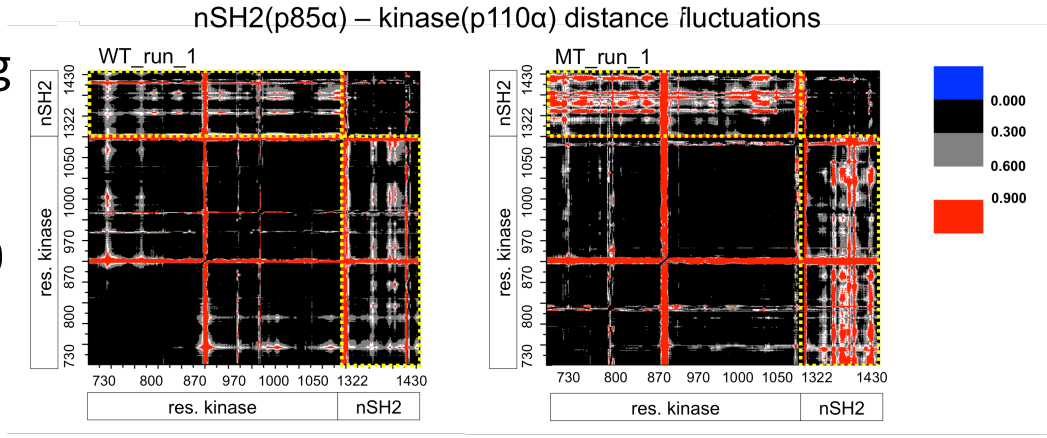
The A matrix is used to characterize the elasticity of a protein undergoing structural fluctuations

Trajectories projected on the first 10 PCs (Ca)

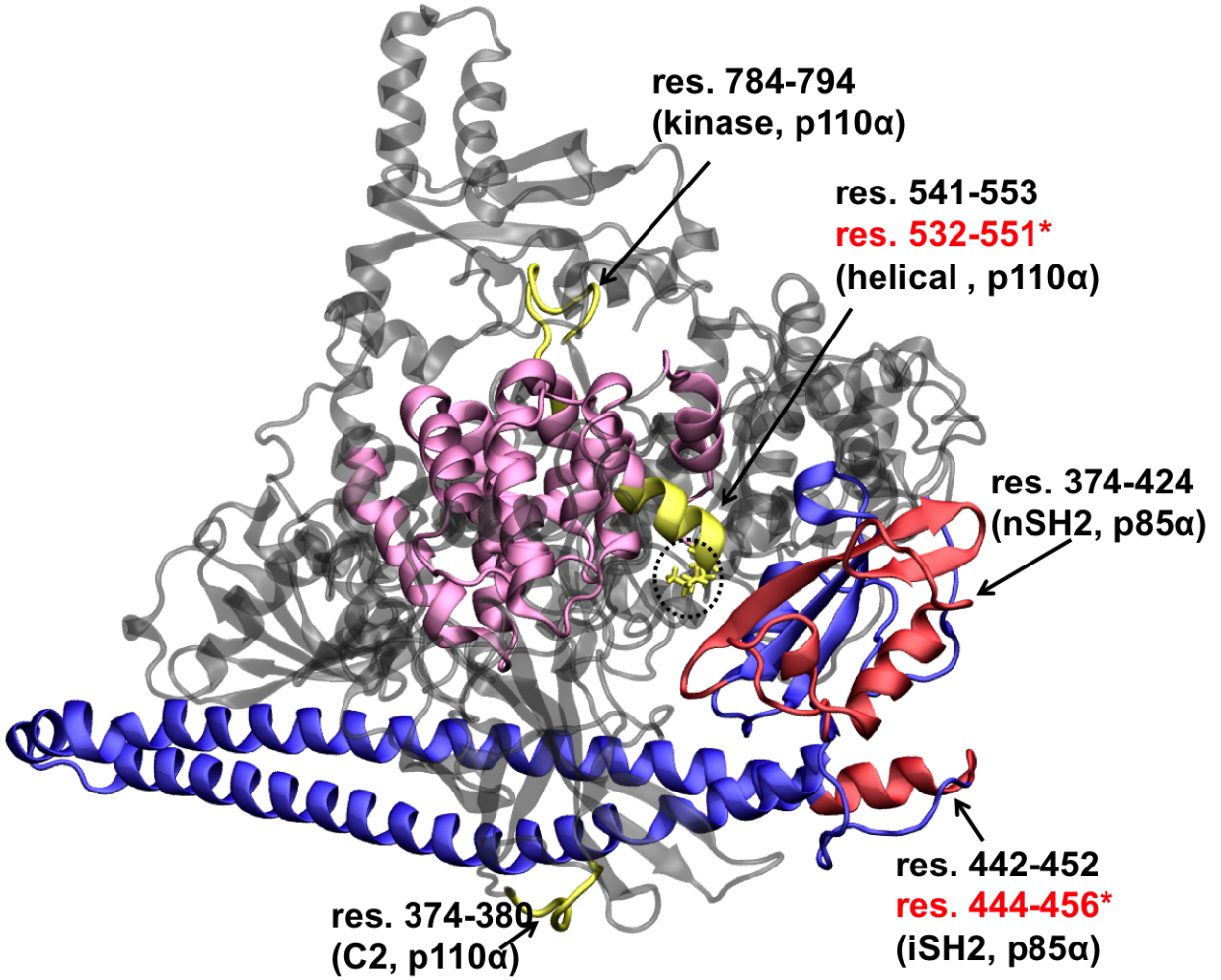
*Morra et al, PLOS Comp Biol, Colombo lab (2012)*



# Enhanced Flexibility of nSH2, kinase, helical regions



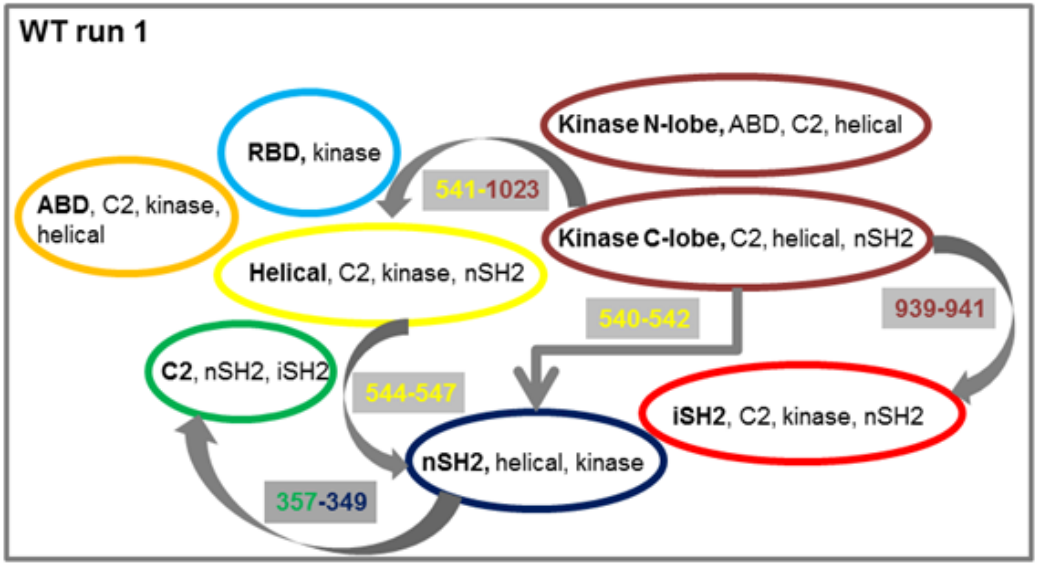

RMSF calculations showed enhanced flexibility of

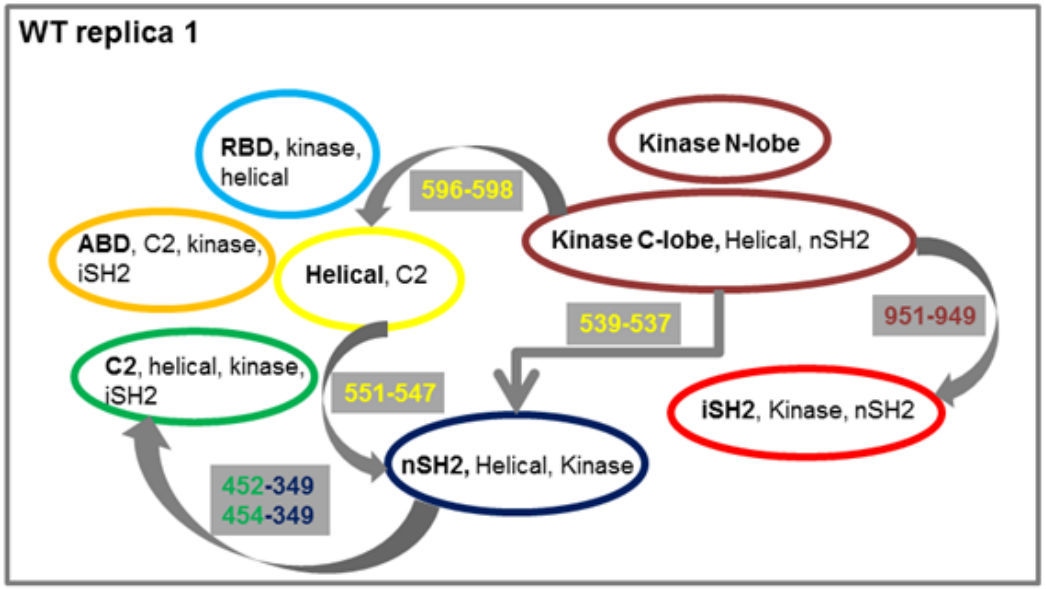
nSH2
Kinase
Helical regions

in accordance to HDX experiments

(Burke et al 2015)

# Dynamical Network Analysis



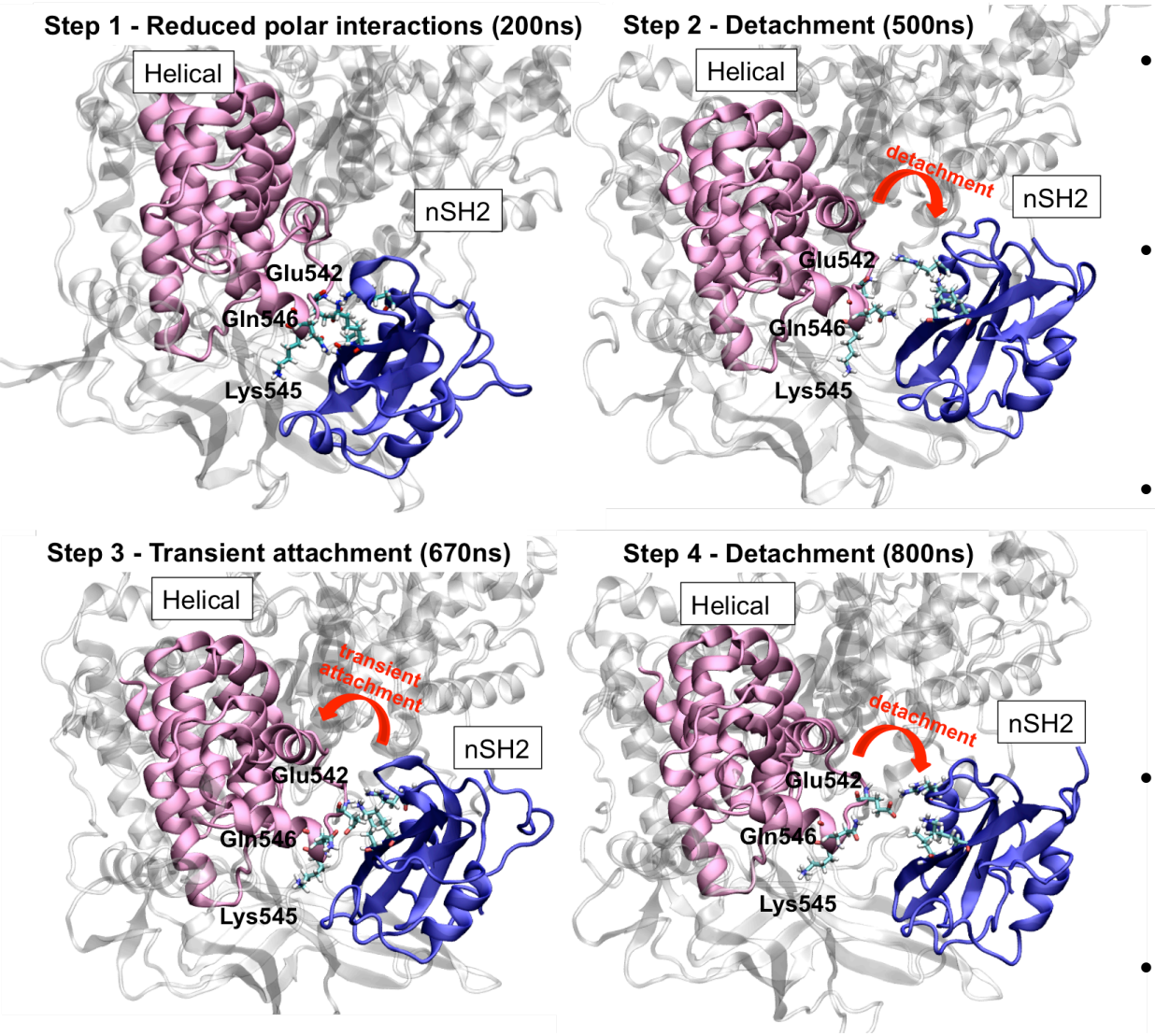




Dynamical network analysis showed that critical nodes of communication between helical, kinase and nSH2 are clustered around the mutation:

537, 539-542, 544, and 547

*(Seth et al, 2012)*

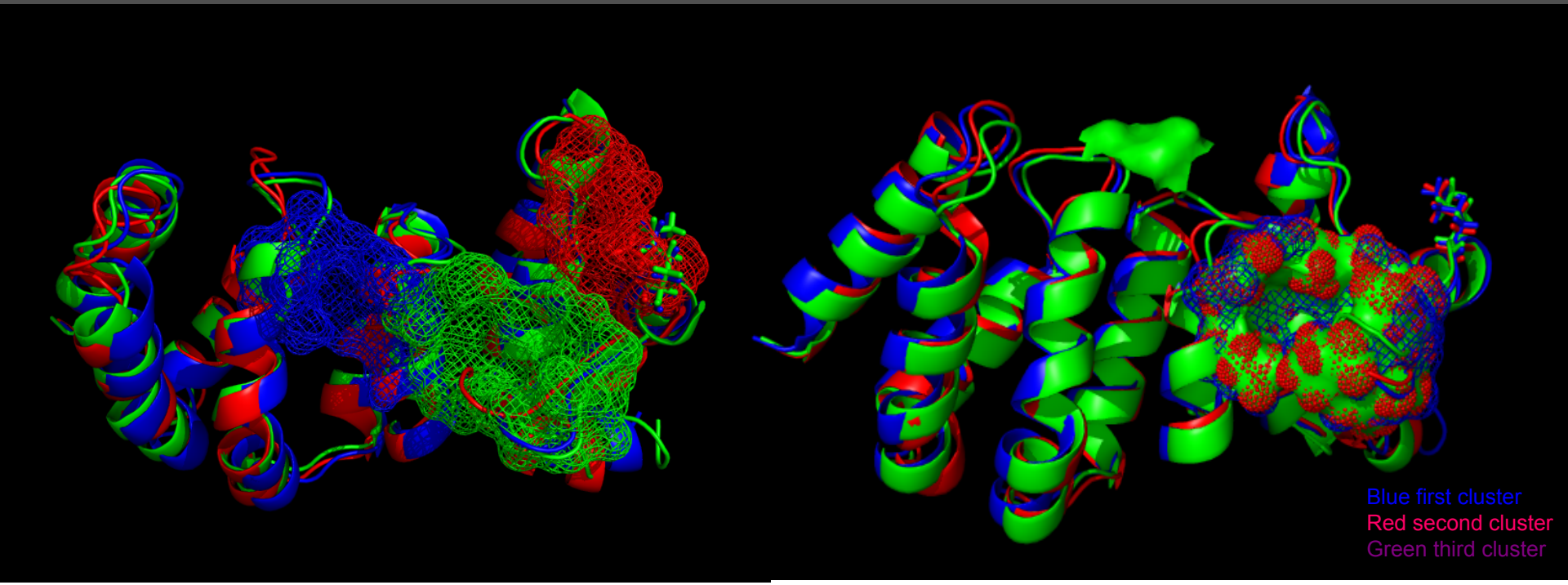
# Proposed mechanism of activation




*Leontiadou & Cournia, submitted*

- Breaking of crucial hydrogen bonds between nSH2 and helical domains
- Enhanced flexibility of certain areas of the nSH2, helical and iSH2 domains, in agreement with HDX experiments
- Map od distance fluctuations indicates significant loss of communication between the regulatory and catalytic subunits of the kinase
- Dynamical netowrk analysis: critical nodes : 537, 539-542, 544, and 547
- Catalytic loop freezes in an active conformation after the huge conformational change

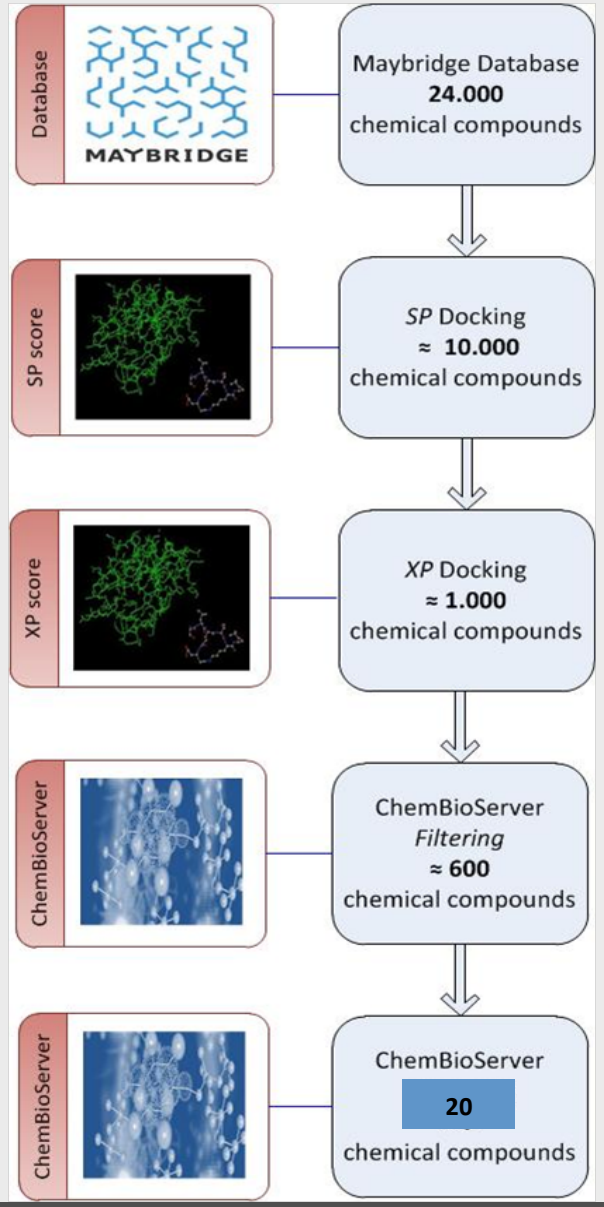
# Binding Site Prediction on WT and E545K




WT Mutant

- Cluster analysis was performed
- Binding site prediction on cluster representatives.
- Binding cavities discovered in a region close to the mutation site
- Screening (Glide docking, Maybridge) was performed in the discovered cavity for the WT and Mutant

# Virtual Screening of the Maybridge Database

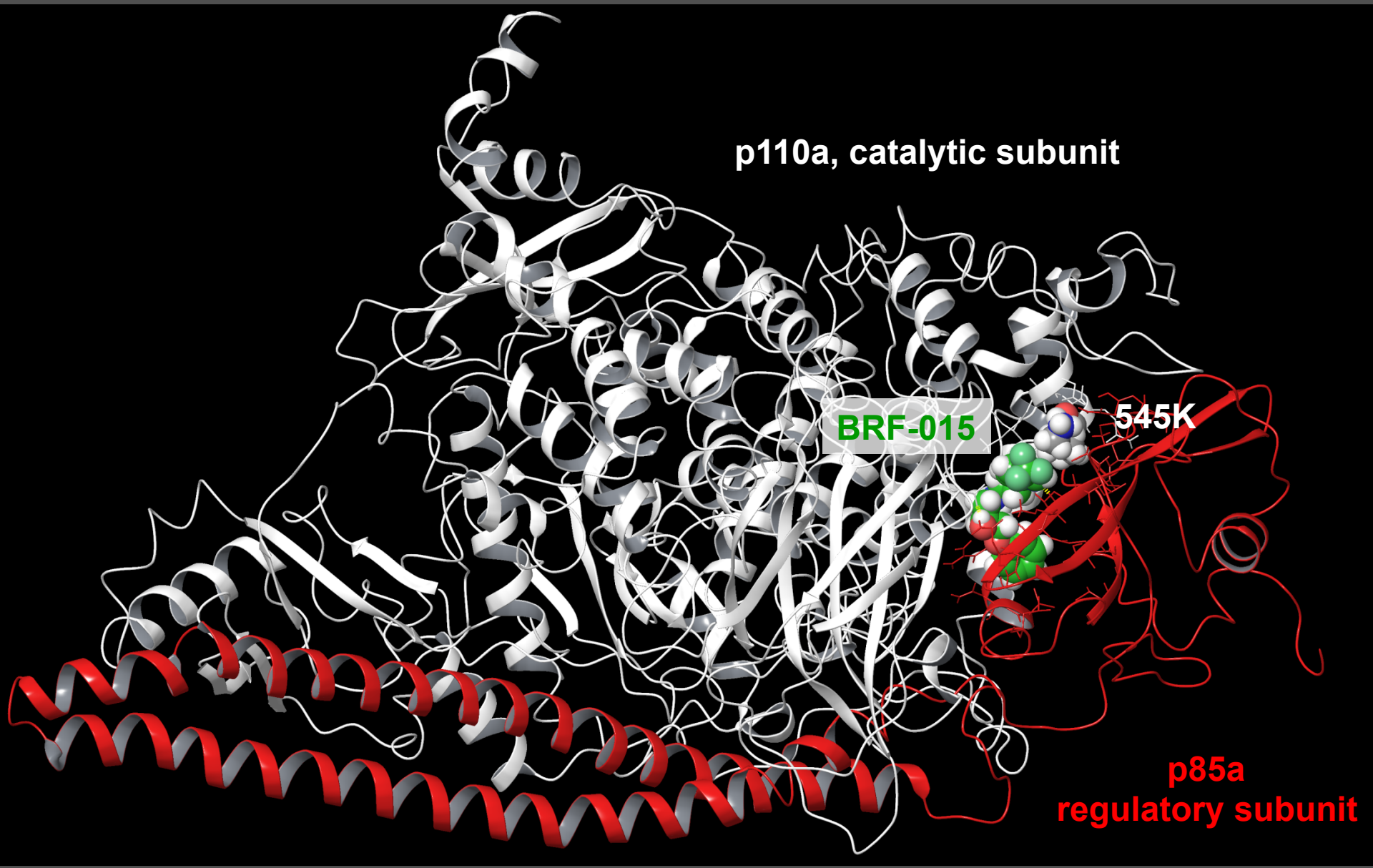



- **Docking of compounds using Glide**
  - **SP (faster – first filtering)**
  - **XP (more accurate)**

**Hydrogen bonds, vdW, electrostatics, strain in protein and ligand, hydrophobic effect**

- **Calculate ADME/tox/pchem properties**
- **Metabolic liabilities**
- **Check for compound conformations**
- **Clustering based on similarity**

# Targeting the p110a + p85a interface

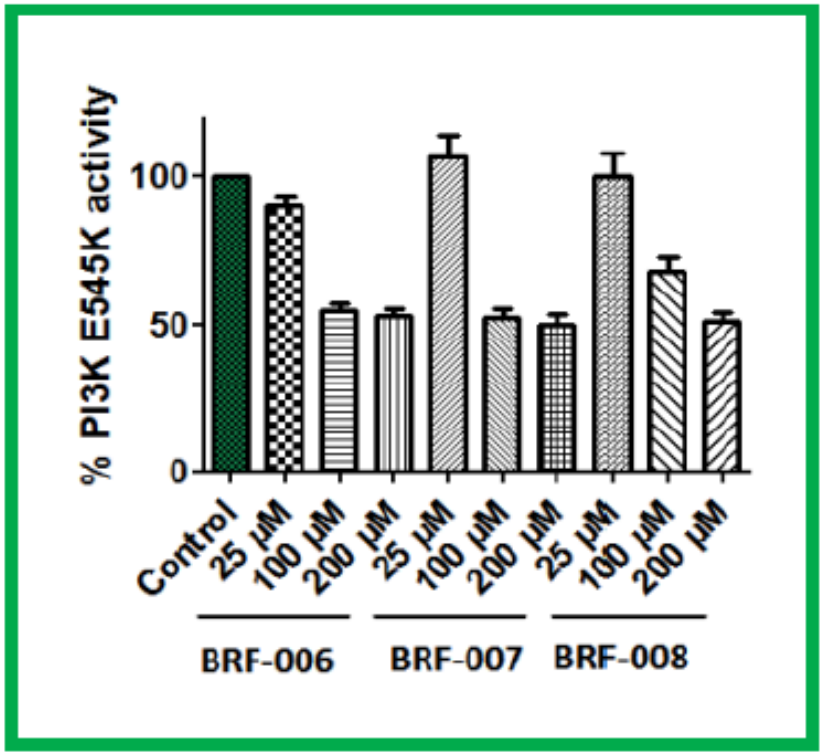


# In vitro cell-free assay with cancer liposomes

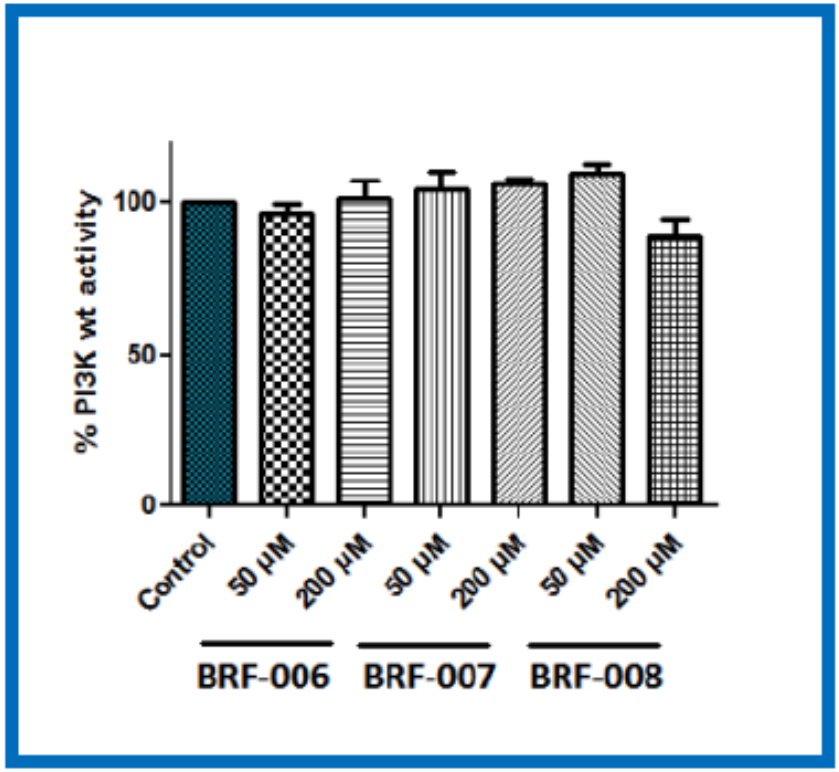
## Compounds BRF 006-007-008

*Christoforidis lab, U Ioannina*

### E545K PI3K




### wt PI3K

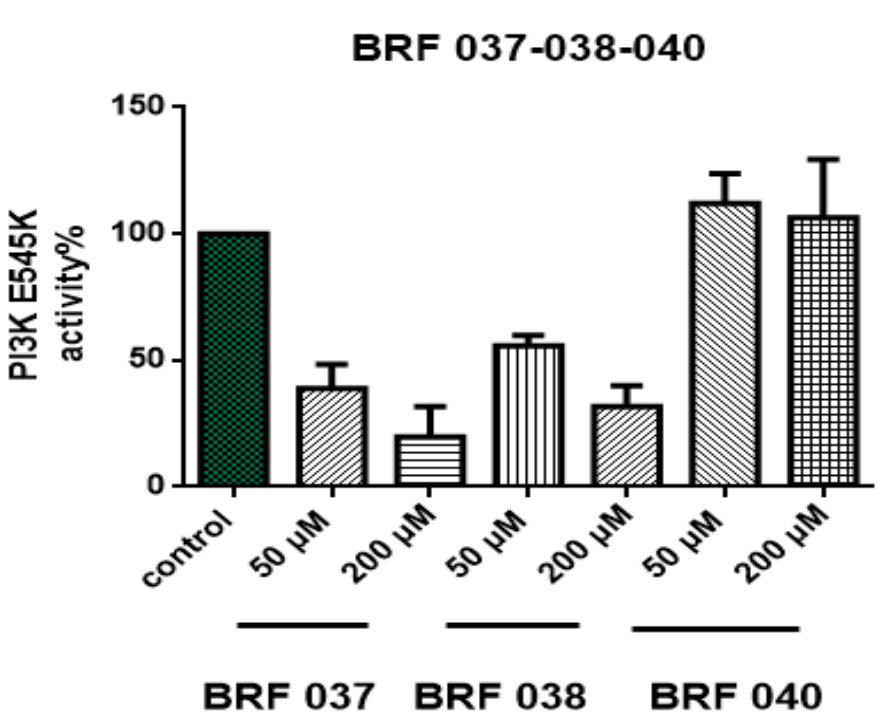



***Achieved selective inhibition of E545K mutant***

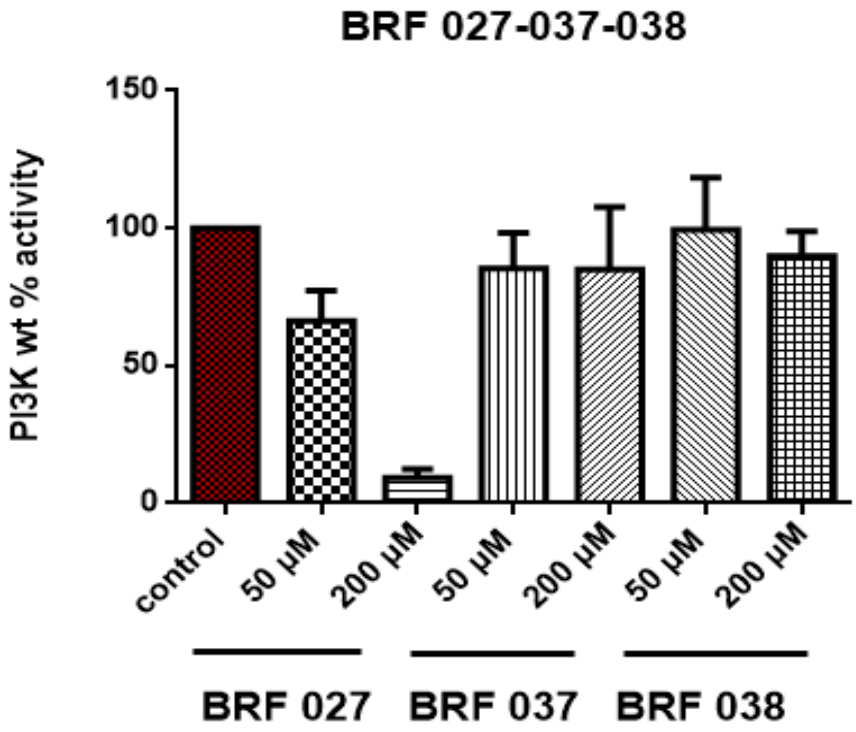
***Activity is in mid-μM range => need for lead optimization***

# In vitro cell-free assay with cancer liposomes

**E545K PI3Ka**




**WT PI3Ka**




***Christoforidis lab, U Ioannina***

# Project Team

**BRFAA**
**Cournia lab (MD, drug design, cells)**
**Dr. Evi Gkeka**
**Dr. Hari Leontiadou**
**Thomas Evangelidis**

**Efstratiadis & Klinakis labs (cells+mice)**
**Dr. Ersi Tsellou**
**Dr. Dimitris Stellas**

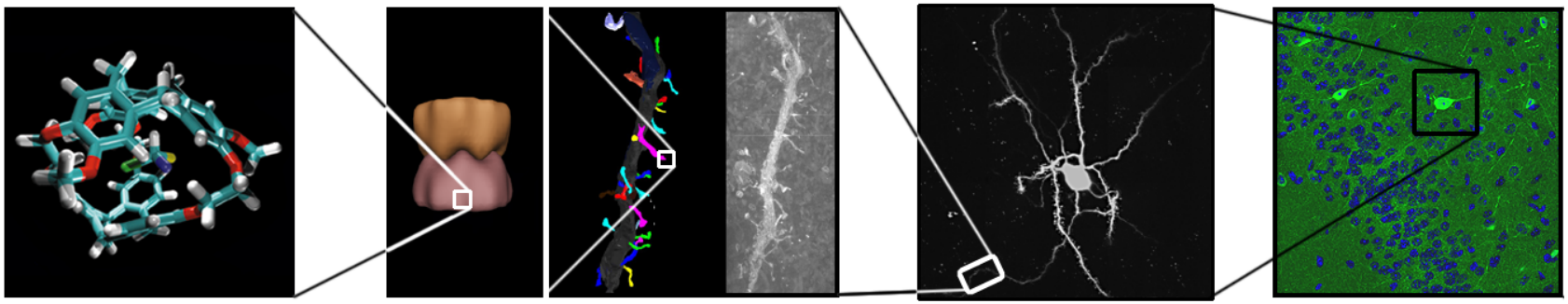
**NCSR Demokritos**
**Couladouros lab**
**Anna Kapela**
**Maria Ouzouni**

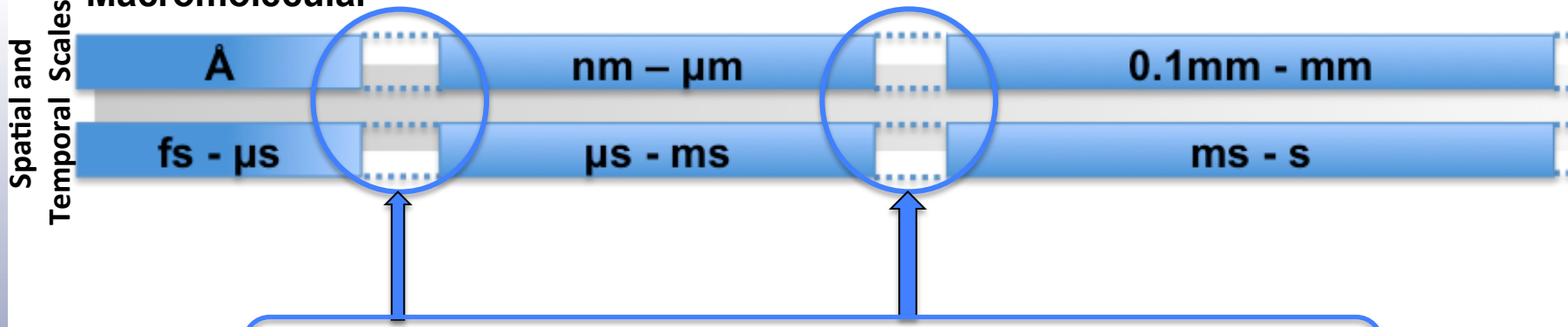
**University of Thrace**
**Agianian lab**
**Dr. Maria Pavlaki**

**University of Ioannina**
**Christoforidis lab (cell-free assays)**
**Alexandra Papafotika**
**Dr. Vasiliki Lazani**

American Association for Cancer Research

# Mind the Gaps




**Challenge areas define the future of computational chemistry & biophysics**

***e.g., Can we understand the drug target in its real environment? Can we understand the molecular and chemical mechanisms underlying disease?***

Source: adapted from Prof. Rommie Amaro

# Coarse-graining membranes: MARTINI FF

C - apolar

P - polar

N - nonpolar

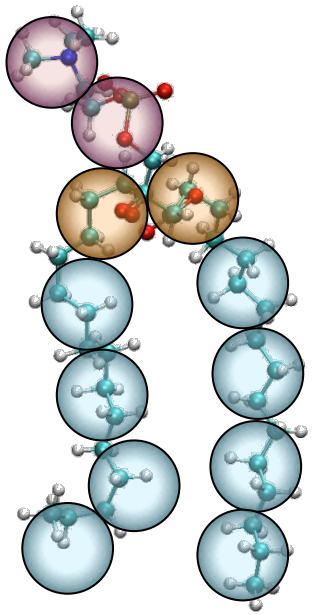
Q - charged

18 subtypes

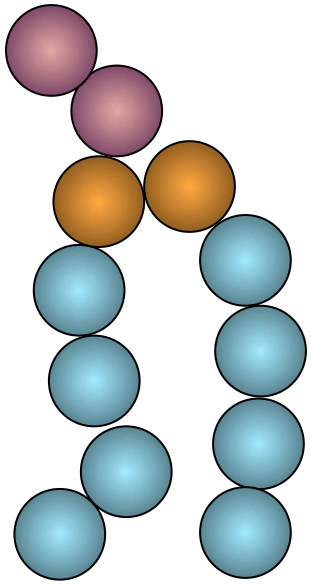
- Hydrogen bond capabilities: d, a, da, 0
- Degree of polarity: 1, low…. 5, high

$$V = V_{bond} + V_{angle} + V_{id} + U_{LJ} + U_{el}$$

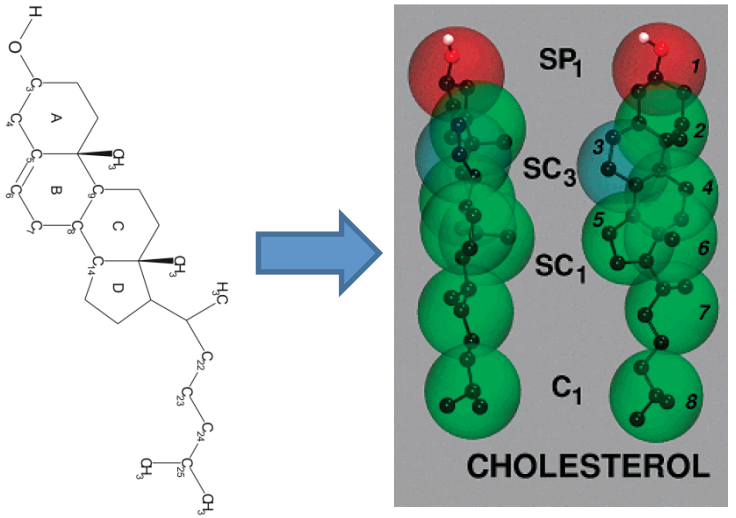
DPPC lipid



130 atoms

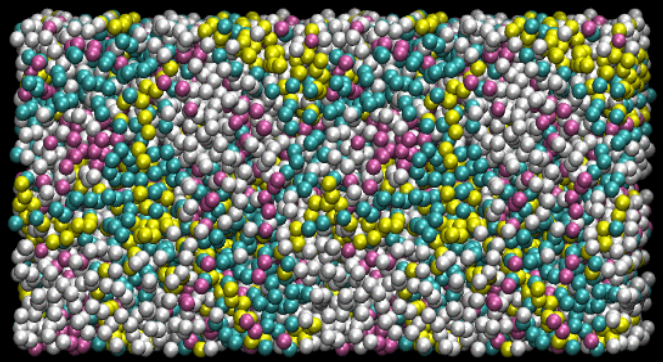


12 particles



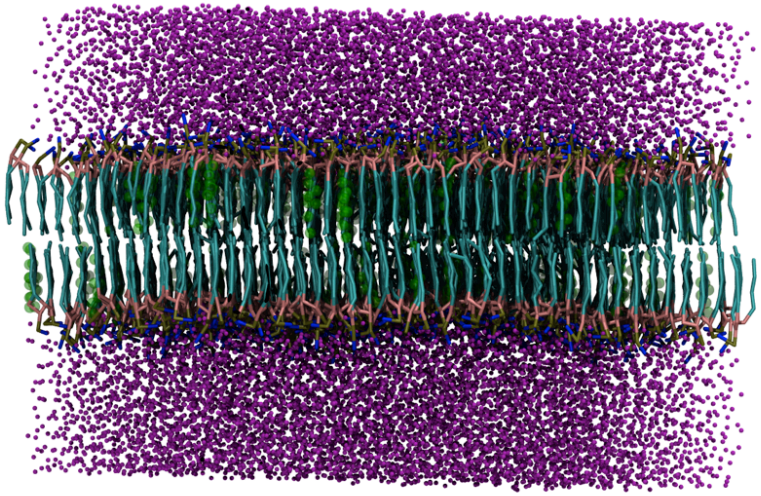

S. J. Marrink *et al.*, *J. Phys. Chem. B,* vol. 111, pp. 7812-7824, 2007.

# Lipid bilayer formation

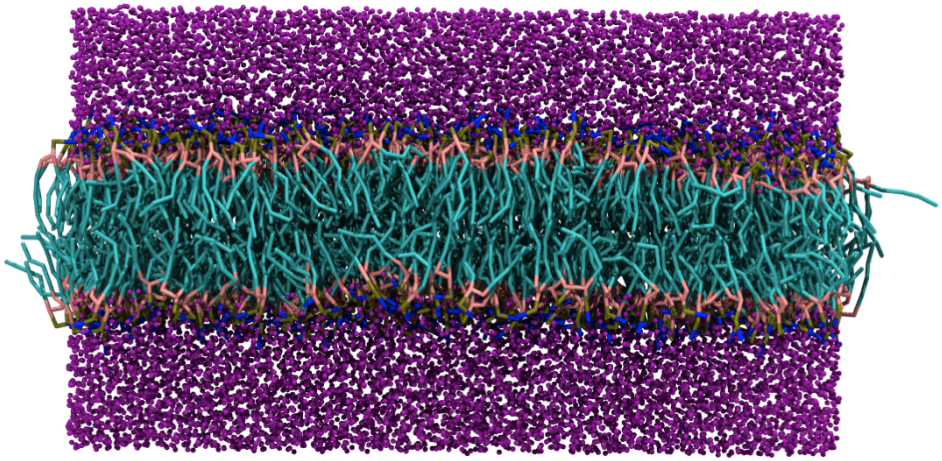


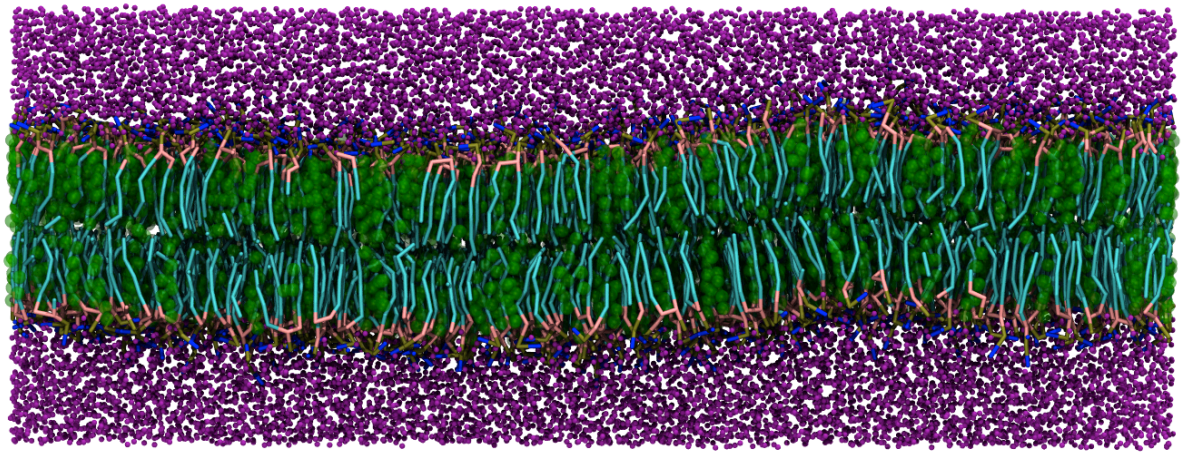
# Three lipid bilayer phases observed with CG-MD

**Gel phase (T=290K, 10% mol. chol.)**



**Liquid phase (T=323K, 0% mol. chol.)**



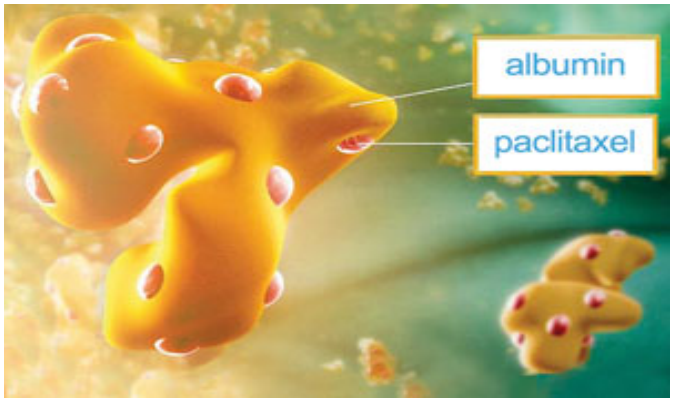


**Liquid-ordered**
**T=323K**
**50% mol. chol.**

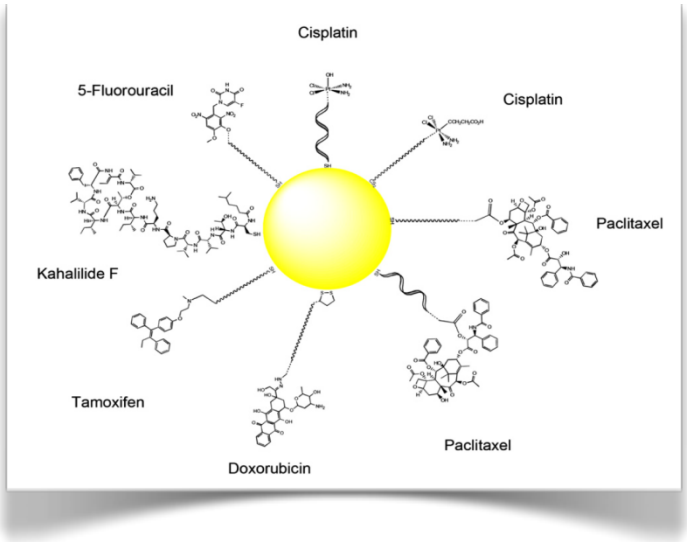
# Nanoparticle applications in medicine

## Nanoparticle albumin-bound paclitaxel (Abraxane®)

http://www.abraxane.com/



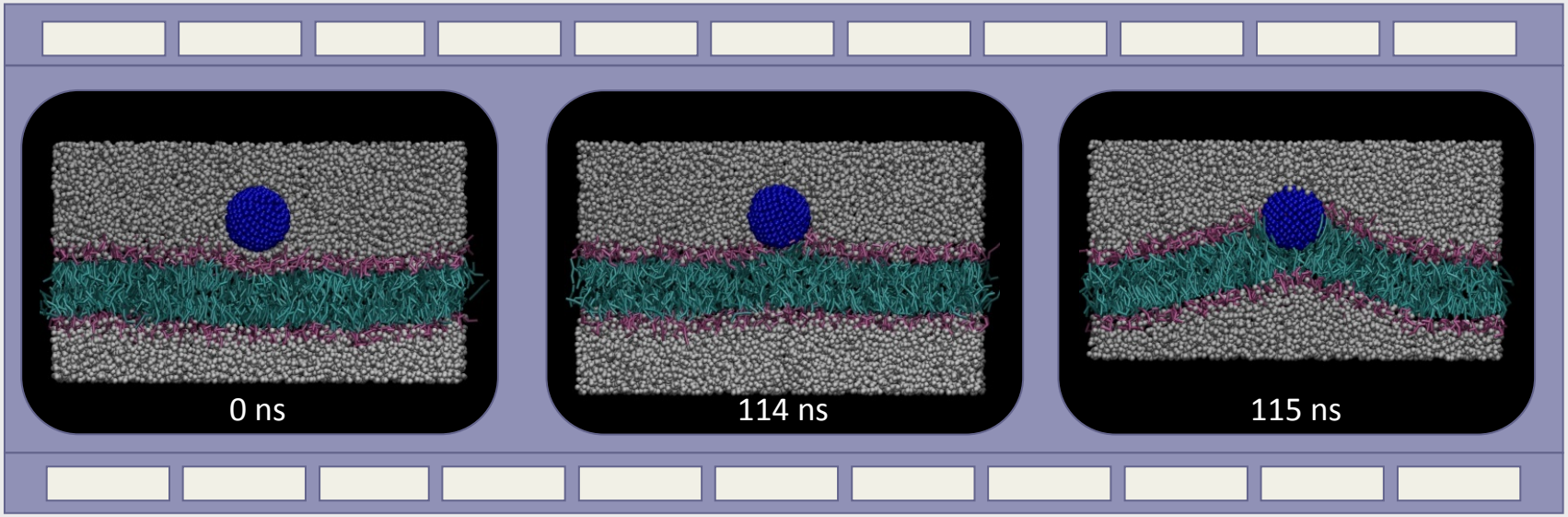

*Targeted Drug Delivery*

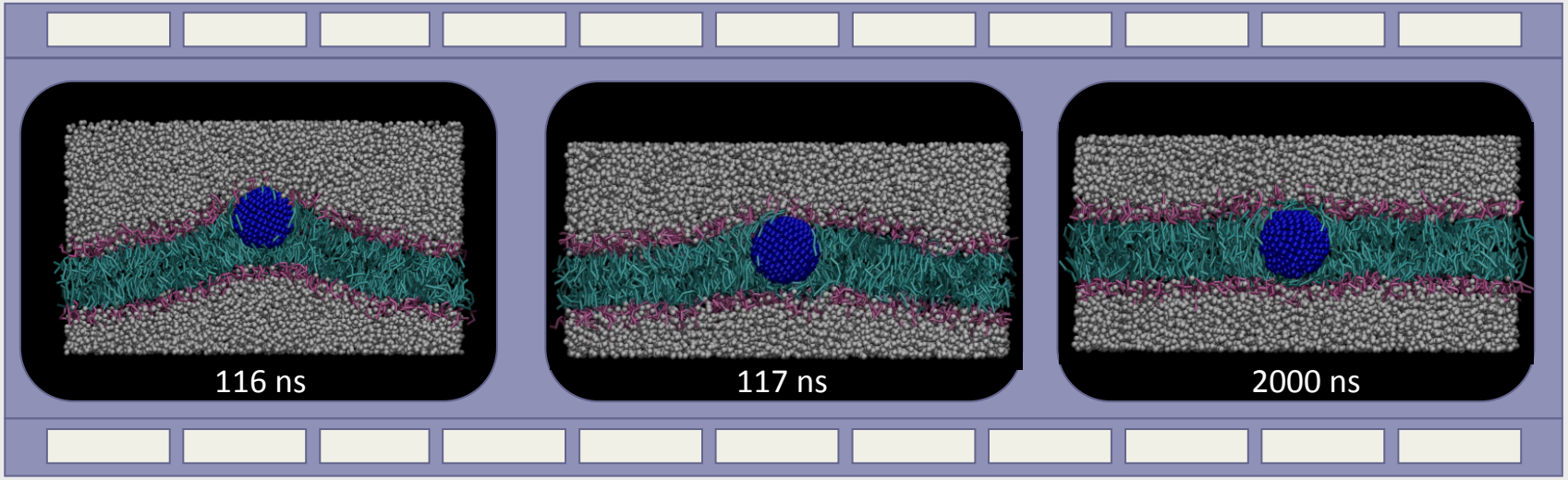



## Anticancer drugs covalently conjugated to gold nanoparticles

*L. Vigderman, E.R. Zubarev / Advanced Drug Delivery Reviews 65 (2013) 663–676*

# The fully hydrophobic nanoparticle

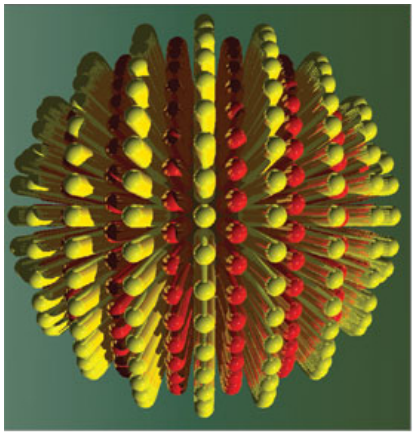


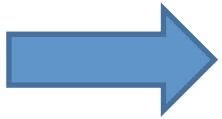
# The “hairy” nanoparticle

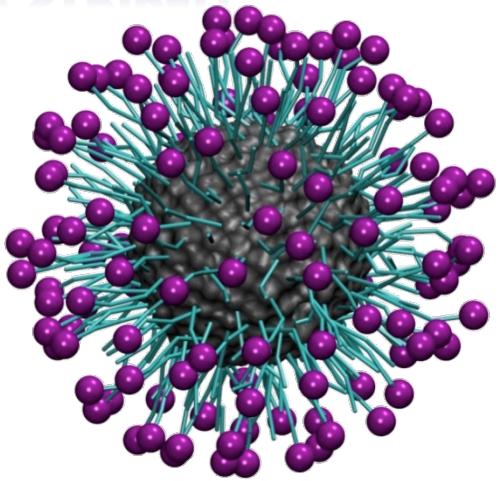
## STRIPED NP IS NOT REALLY STRIPED

Evi Gkeka



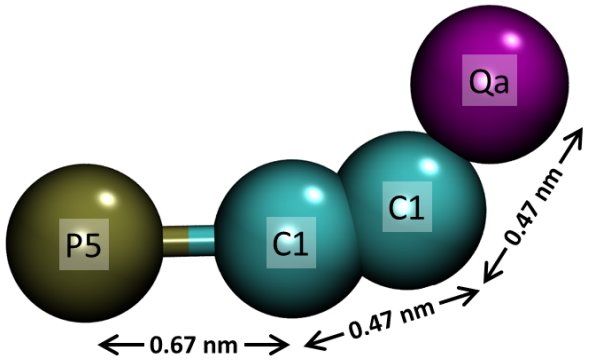
Verma *et al.* Nature Materials 2008



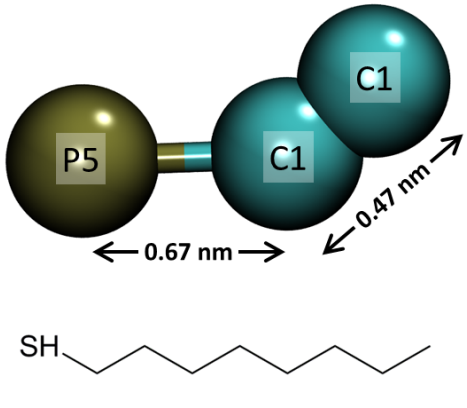


Gkeka et al. PLOS Comput Biol 2014

**11-mercapto-1-undecanesulphonate (MUS)**




**1-octanethiol (OT)**




**MARTINI modeling of NP surface ligands**

Harmonic bond potential

$k^{P5-C1}$=12,500 kJ mol$^{-1}$

$k^{C1-C1}$=1,250 kJ mol$^{-1}$

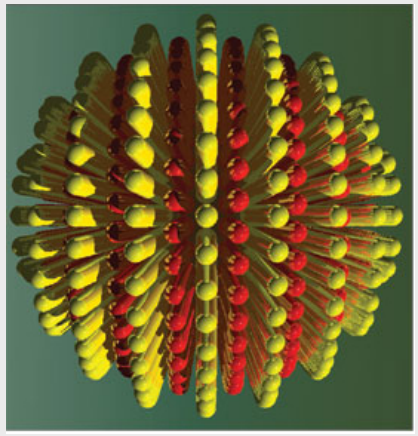
$k^{C1-Qa}$=1,550 kJ mol$^{-1}$

Cosine based angle potential

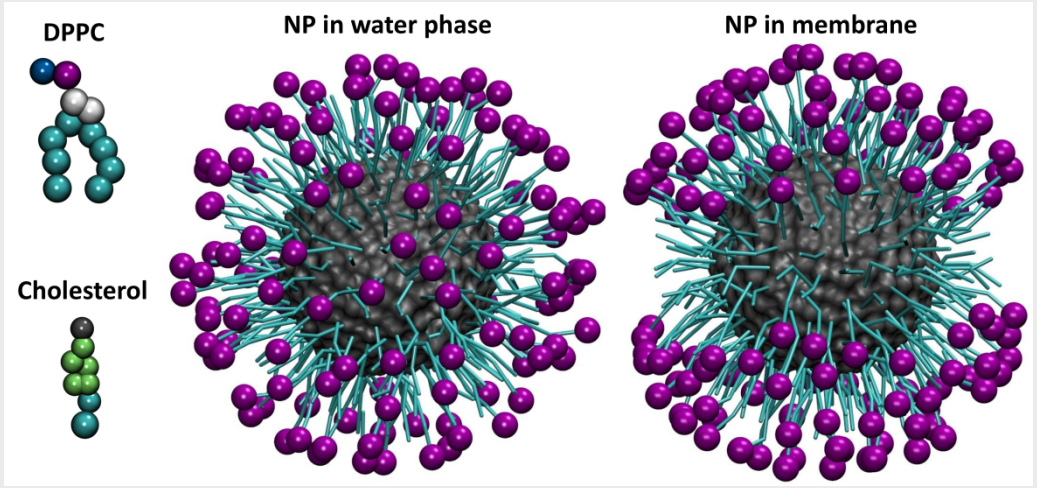
$\theta_0$ =180°

k=25 kJ mol$^{-1}$

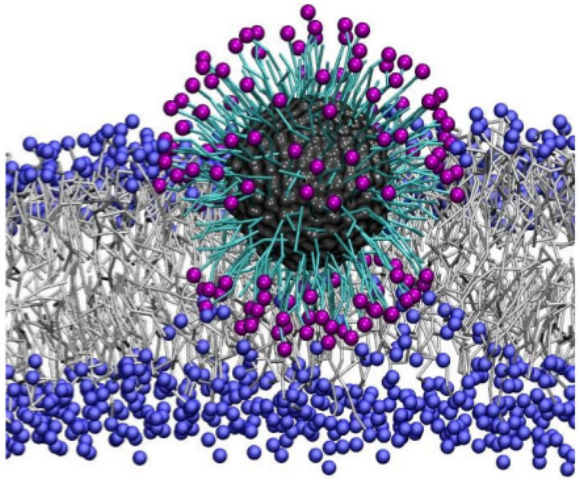
# The effect of cholesterol on NP insertion



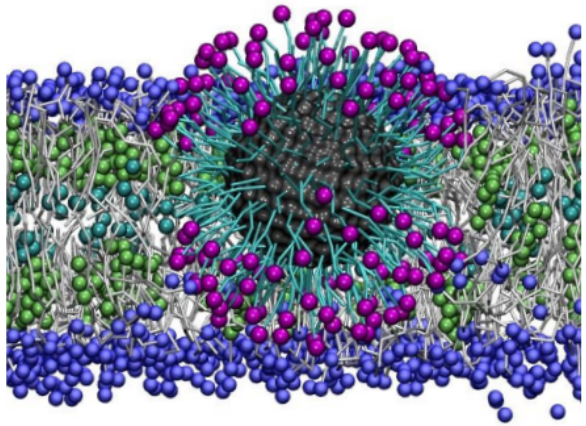
Verma *et al.* Nature Materials 2008



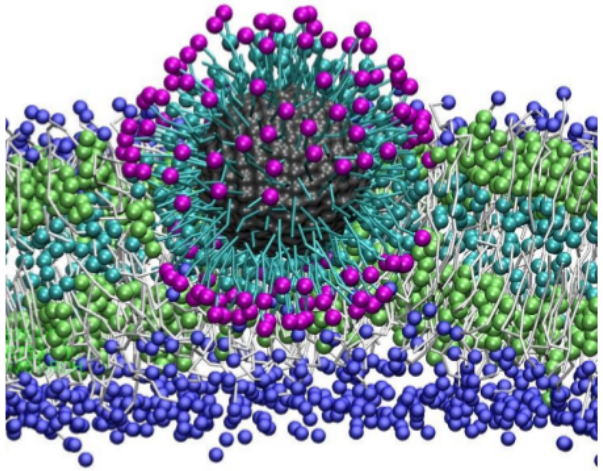

**6 different membrane cholesterol concentrations: 0%, 10%, 20%, 30% , 40%, and**



0% cholesterol

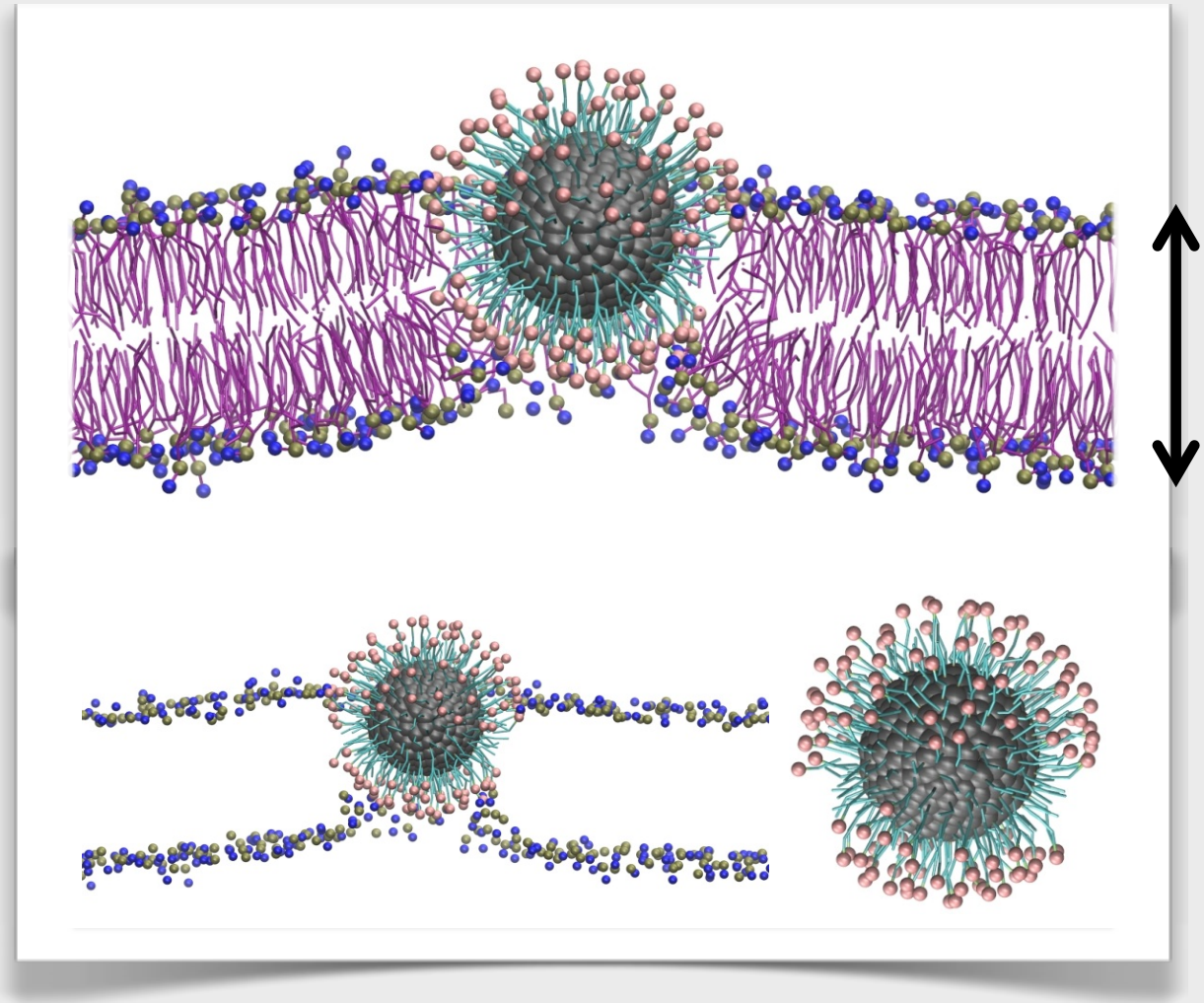


30% cholesterol



50% cholesterol

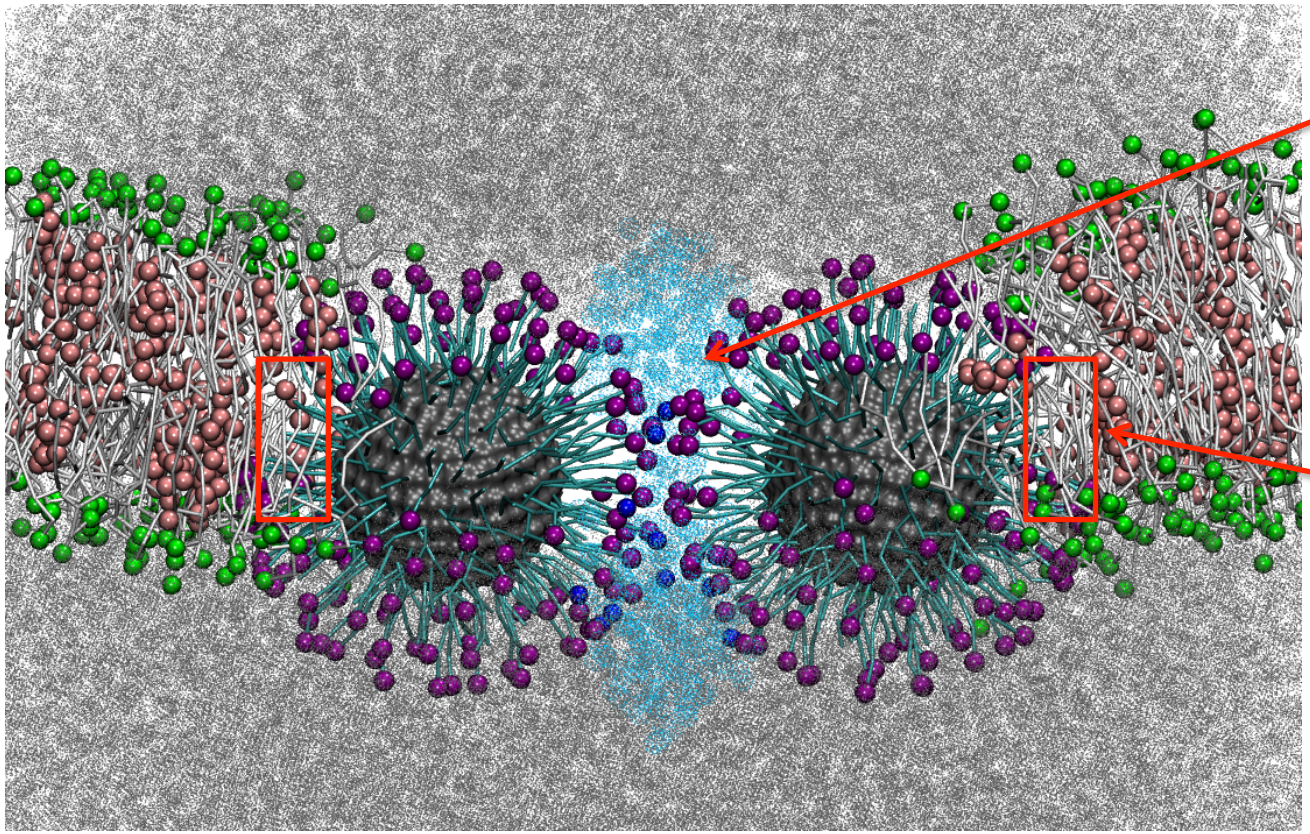
# NP is thinning the membrane




Measure the bilayer thickness at the area of NP penetration and in the bulk lipid bilayer

# NP-NP interface in the cell membrane

**A water pore is formed at the NP-NP interface**



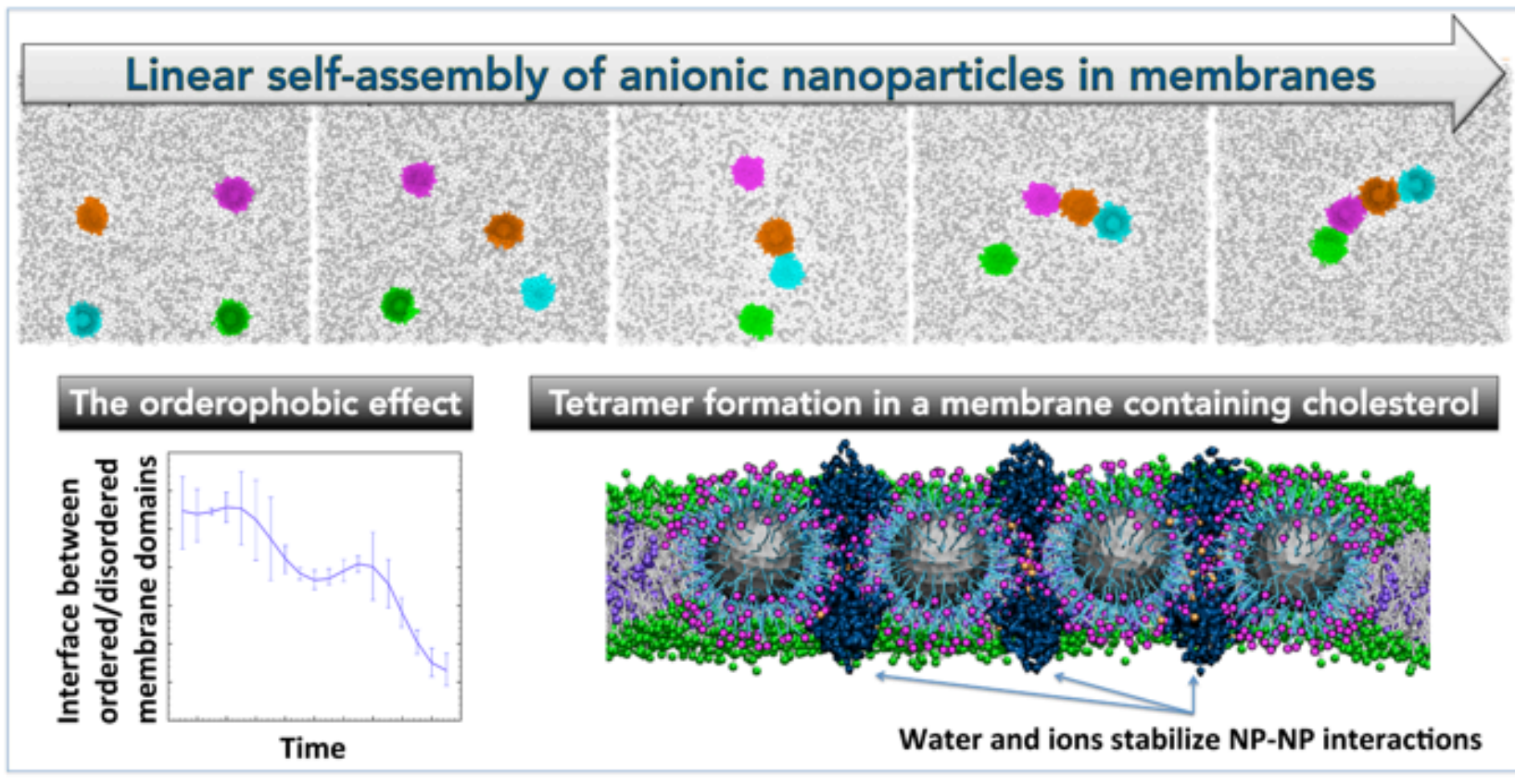
Water and ions lie at the interface between the two NPs

The snorkeling effect is still evident at the side of the NP that is interacting with the cell membrane

**Evi Gkeka**

NC3 polar ligand ends hydrophobic tails cholesterol water ions

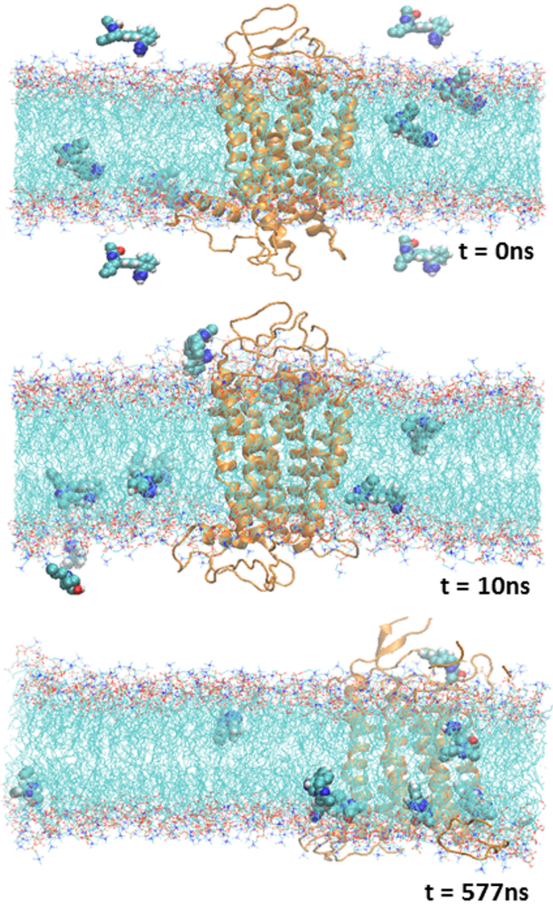
# Nanoparticles as drug delivery systems




Angelikopoulos *et al.* submitted

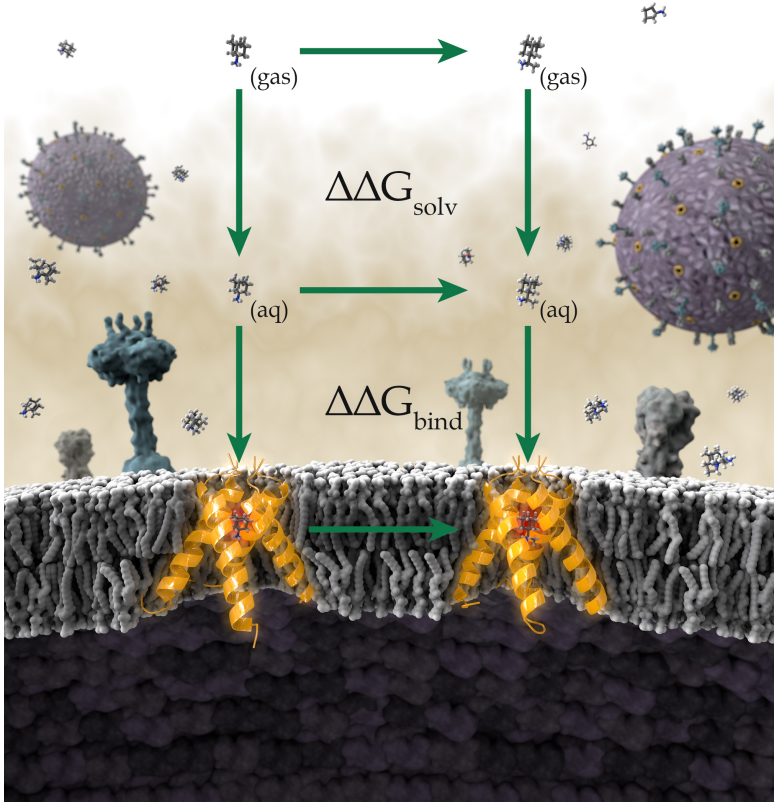
# Targeting membranes/membrane interfaces for computer-aided drug design and drug delivery

AT1 receptor



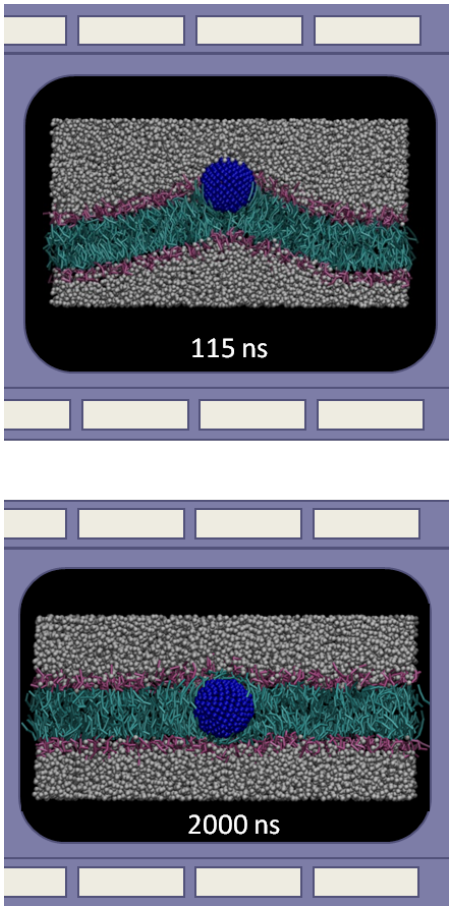

BBA - Biomembranes (2014)

M2TM Influenza A



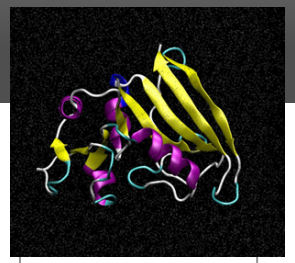

J Chem Theor Comput (2013)

NPs as drug delivery systems

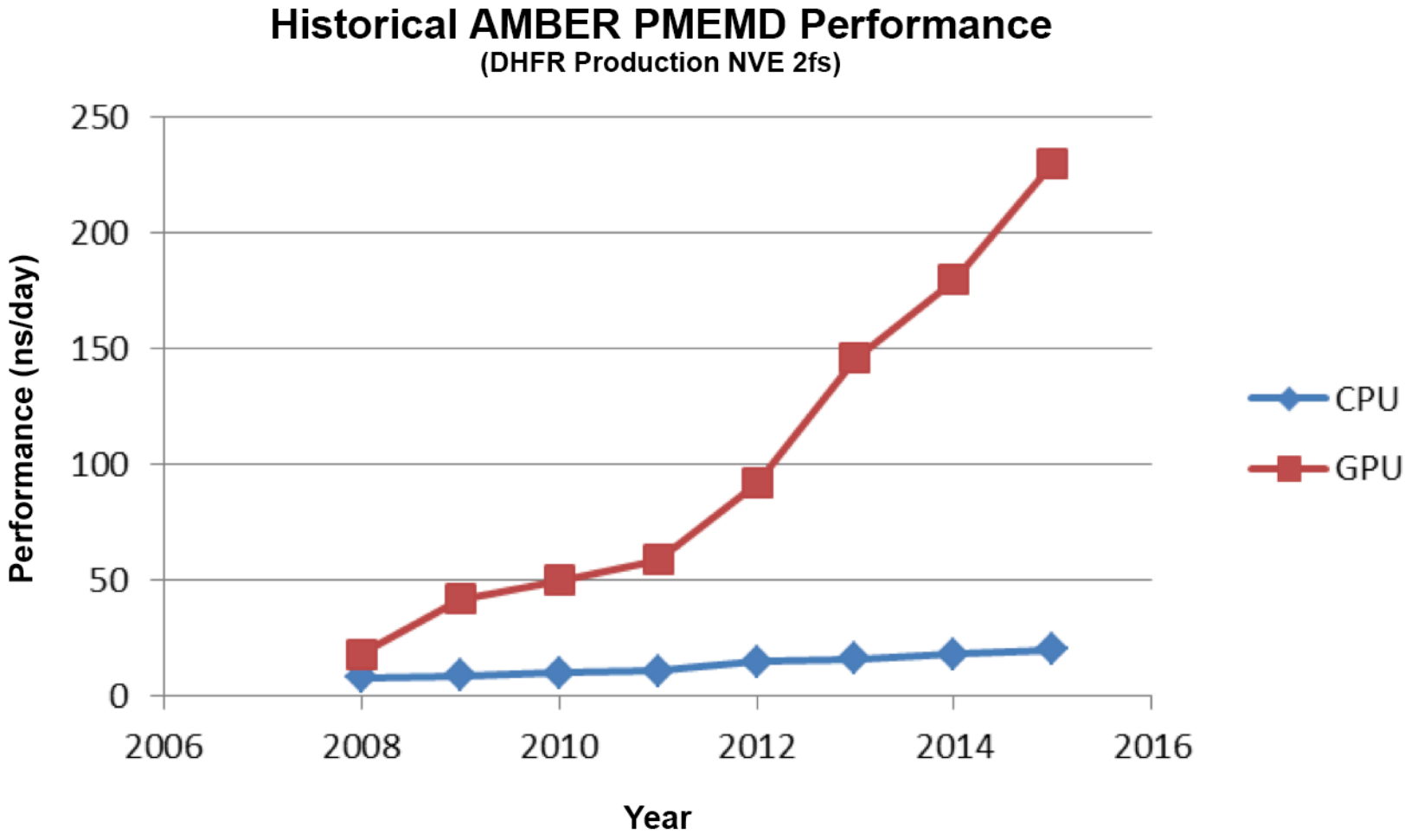



Plos Comput Biol (2014)

# The new era: GPU acceleration



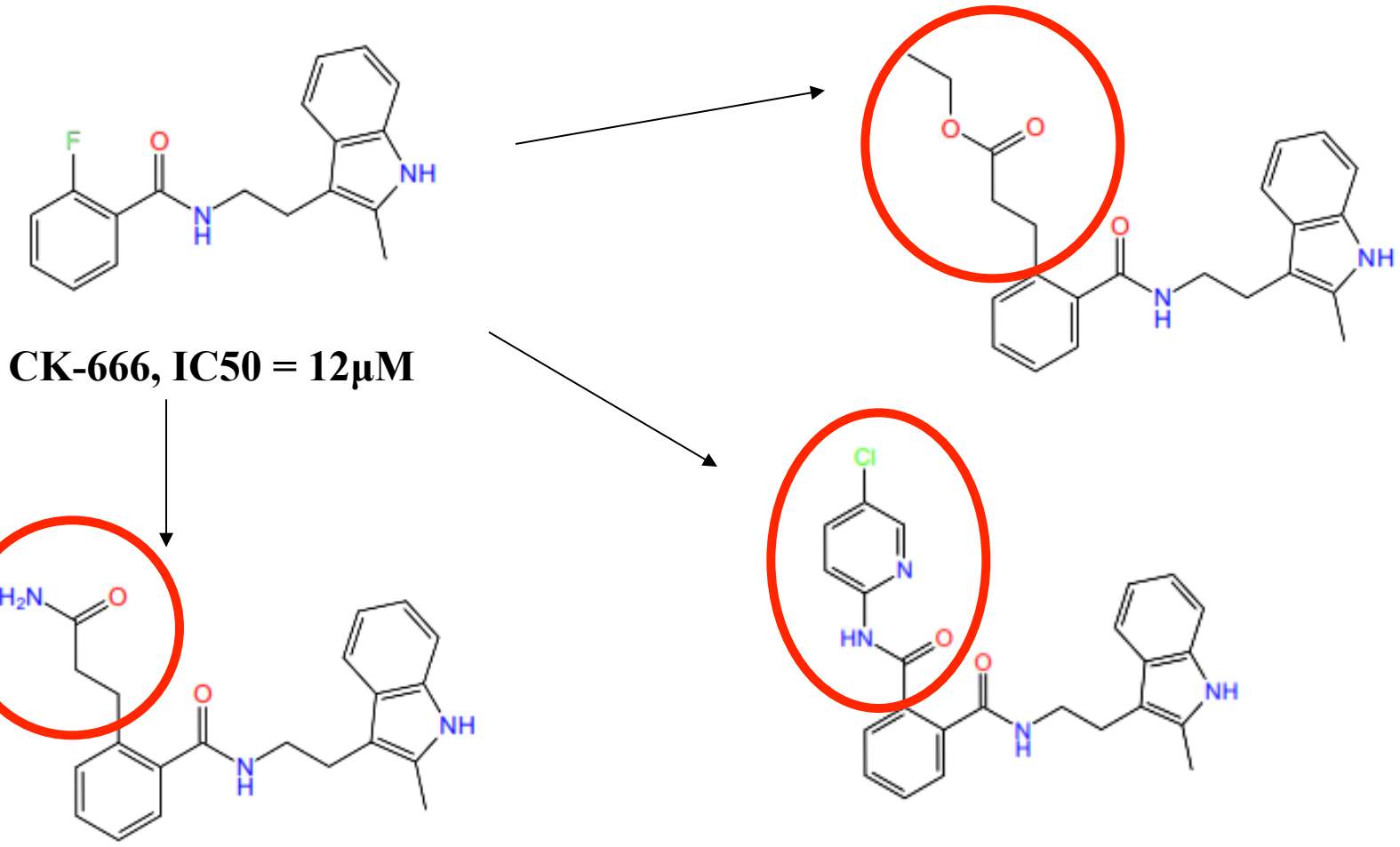
# Historical Single Node / Single GPU Performance




**Credit: Professor Ross Walker, UCSD Supercomputing Center, AMBER developer**

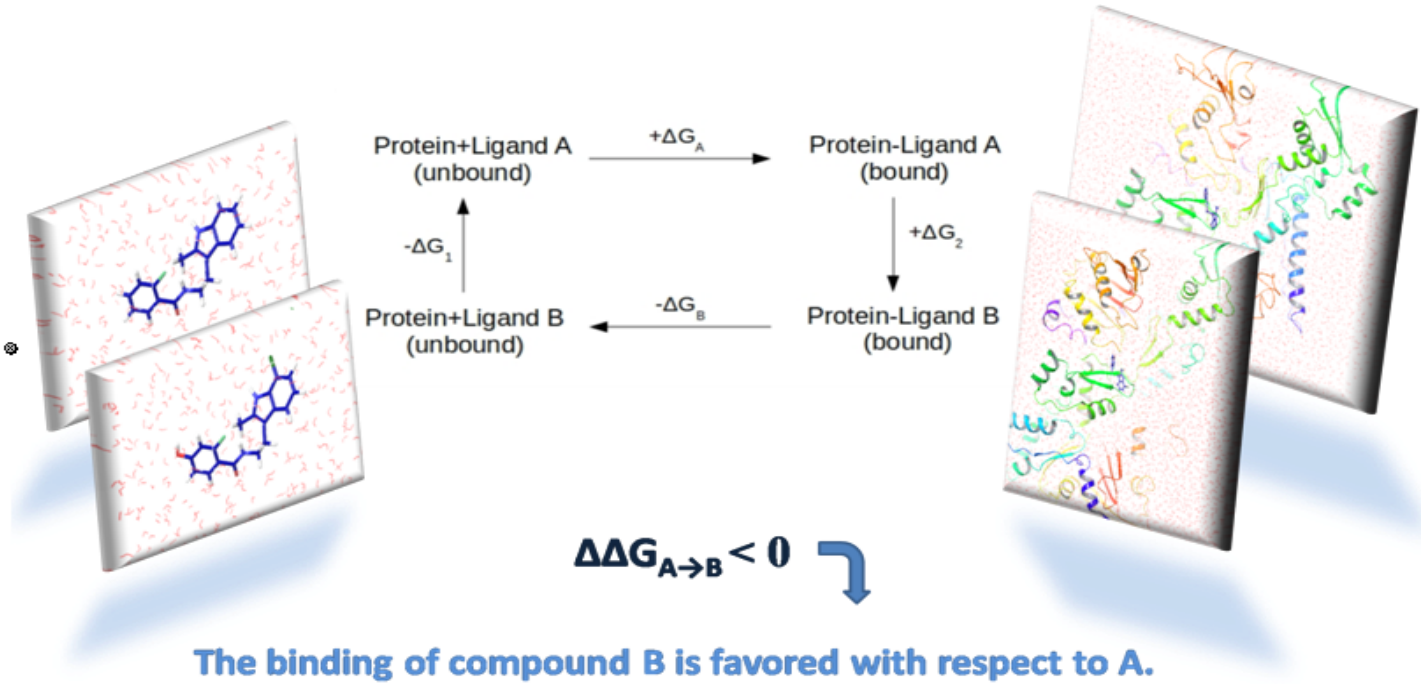
# GPU Acceleration: Example on Drug Design

**Christina Athanasiou**



# Free Energy Perturbation Calcualtions

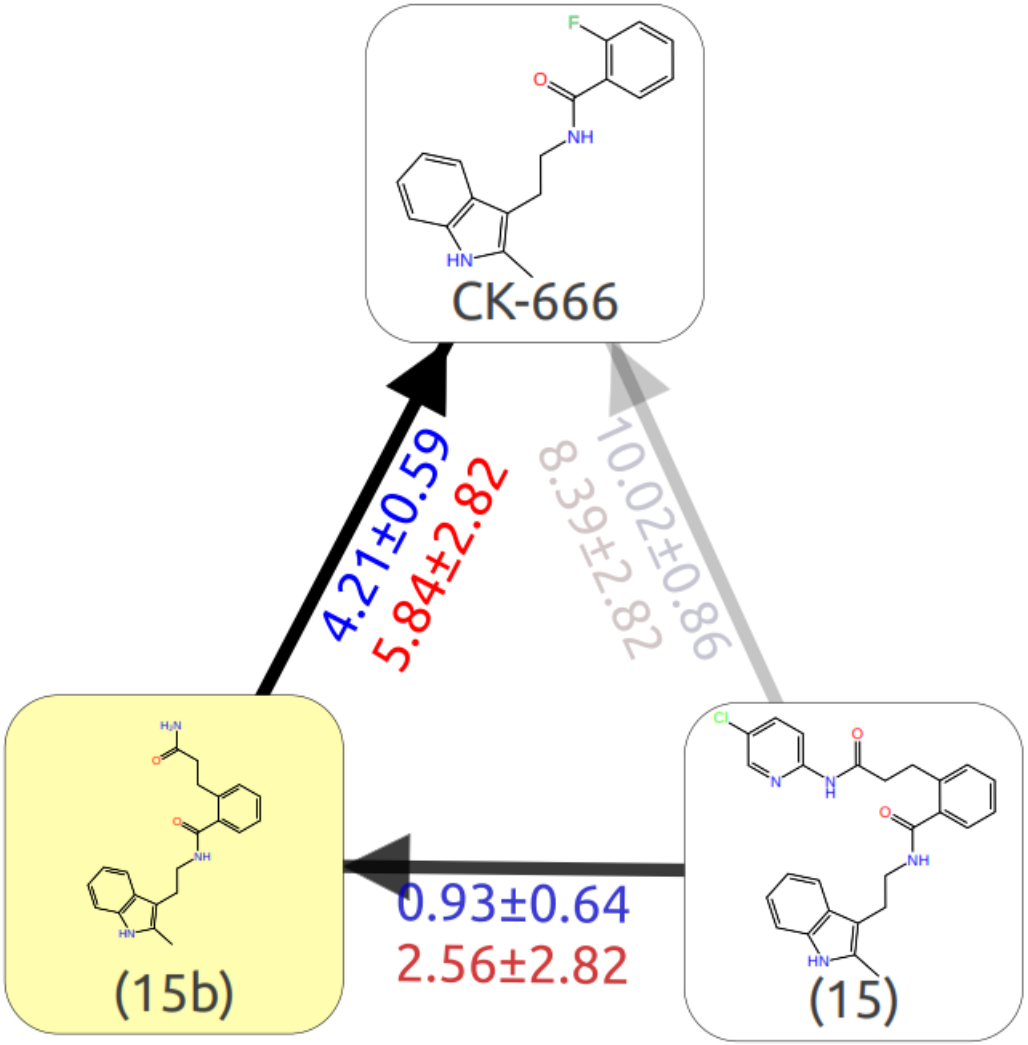
$$\textit{Zwanzig's formula:}\quad \Delta G(A \to B) = G_B - G_A = -kT \ln \left\langle \exp\left( -\frac{V_B - V_A}{kT} \right) \right\rangle_A$$




$$\Delta\Delta G_{binding} = \Delta G_2 - \Delta G_1 = \Delta G_A - \Delta G_B$$

$\Delta G_A$ and $\Delta G_B$ are the free energies of **transfer** of A and B from the unbound to the bound state.
$\Delta G_1$ and $\Delta G_2$ are the free energy differences of the **mutation of A into B** in solvent and bound to protein

# FEP: GPU Acceleration




**FEP chemical transformation in protein and in water**

***CPU: 24 hours on 768 cores per transformation***

***GPU: 7 hours on 1 GPU per transformation!!***

Simulations performed in ARIS - GRNET

# Project Team

**BRFAA**
**Cournia lab (MD, drug design, cells)**
**Dr. Evi Gkeka**
**Dr. Hari Leontiadou**
**Thomas Evangelidis**

**Efstratiadis & Klinakis labs (cells+mice)**
**Dr. Ersi Tsellou**
**Dr. Dimitris Stellas**

**NCSR Demokritos**
**Couladouros lab**
**Anna Kapela**
**Maria Ouzouni**

**University of Thrace**
**Agianian lab**
**Dr. Maria Pavlaki**

**University of Ioannina**
**Christoforidis lab (cell-free assays)**
**Alexandra Papafotika**
**Dr. Vasiliki Lazani**

ARIS

American Association for Cancer Research

# Links to the movies I ‘ve shown

**Villin headpiece protein folding**
https://www.youtube.com/watch?v=sD6vyfTtE4U

**A basic introduction to proteins & drugs**
https://www.youtube.com/watch?v=u49k72rUdyc

**How Does a Drug Molecule Find its Target Binding Site?**
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3221467/
(link to .avi file at the end of the article will d/l the movie)

**Simulation of the Wild Type PI3Kα protein**
http://journals.plos.org/ploscompbiol/article/asset?unique&id=info:doi/10.1371/journal.pcbi.1003895.s031

**Simulation of the oncogenic H1047R mutant PI3Kα protein**
http://journals.plos.org/ploscompbiol/article/asset?unique&id=info:doi/10.1371/journal.pcbi.1003895.s032