Exploring druggability with the JEDI collective variable

Joan Clark-Nicolas

PLUMED users meeting

25th – 27th May 2017

SISSA Trieste (Italy)

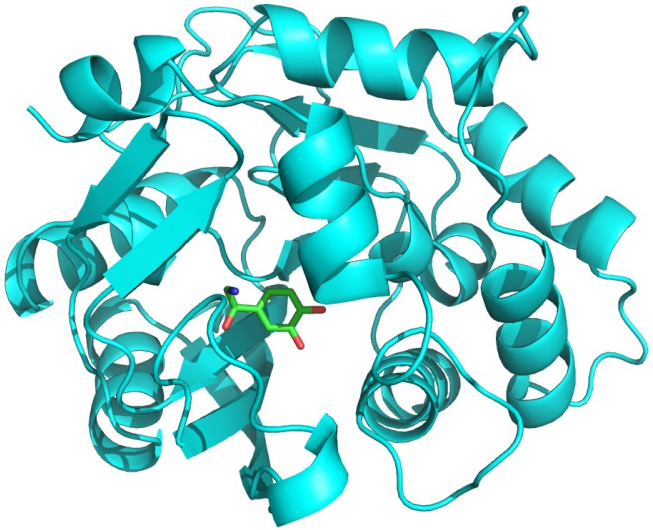
Contents

Part 1: Protein Druggability

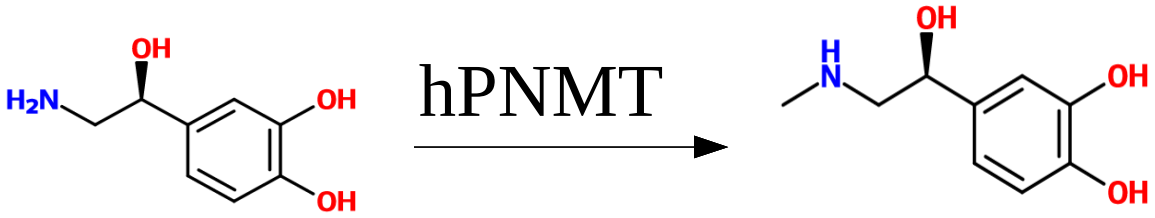
What are proteins and diseases?
And how are drugs related to them?

How are diseases related to proteins?

"From a simplified perspective, proteins can be considered as very complex machines that perform a function inside a cell..."



Noradrenaline Adrenaline

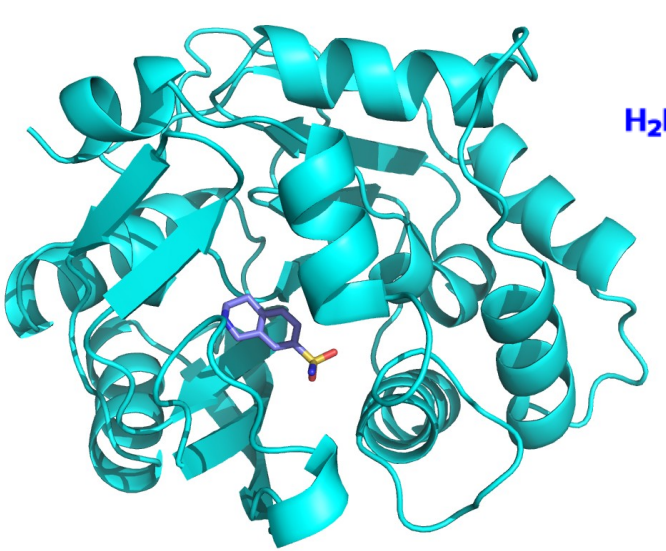


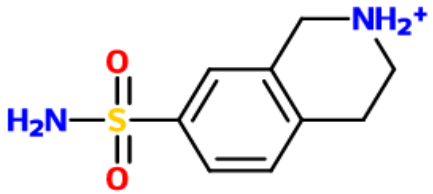
Noradrenaline Adrenaline

"...and diseases can be considered the result of a protein not performing its function properly"

How do drugs cure diseases?

"From a simplified perspective, proteins can be considered as very complex machines that perform a function inside a cell..."



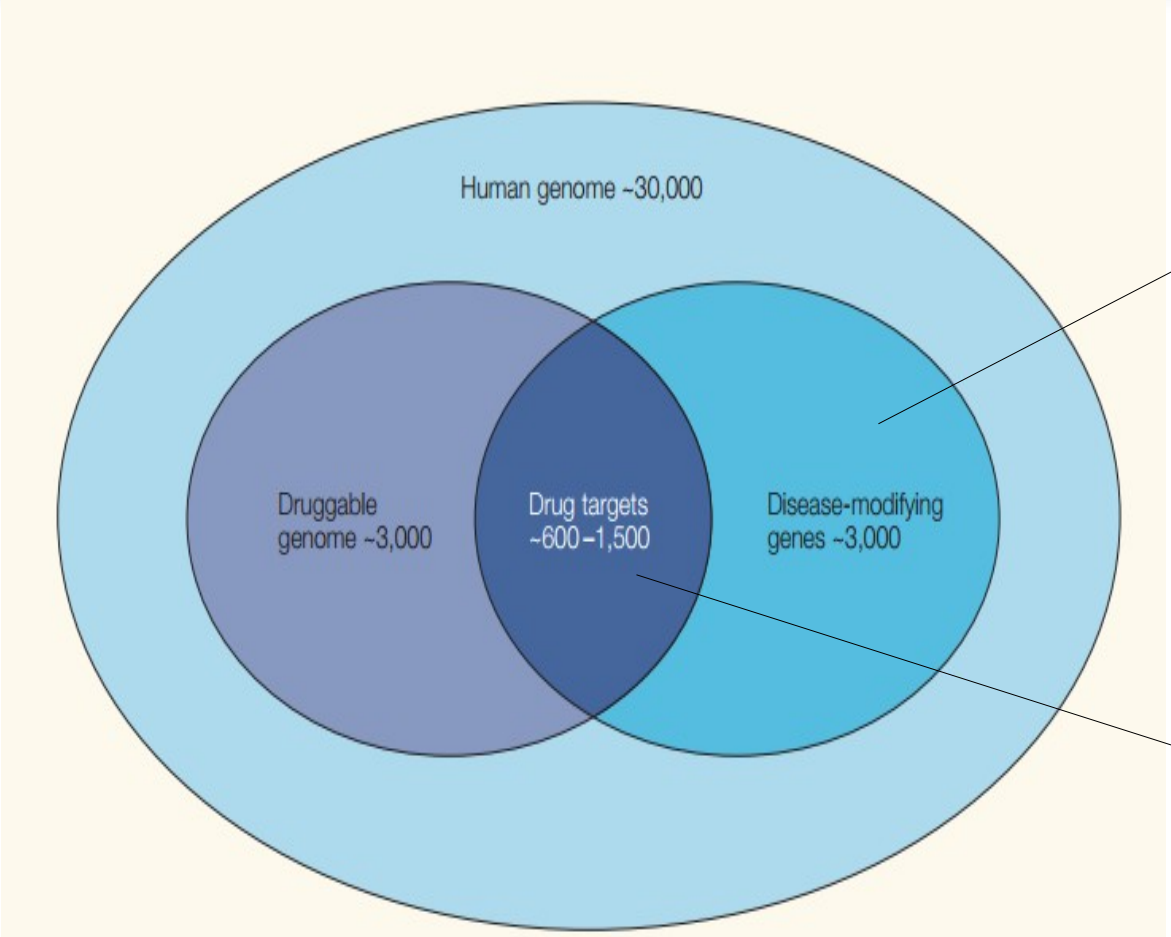


hPNMT → **X**

"...and diseases can be considered the result of a protein not performing its function properly."

*"A drug is a molecule whose role is **usually** to correct this malfunction by binding to a target protein and modulating its activity"*

Not all proteins can bind drugs...



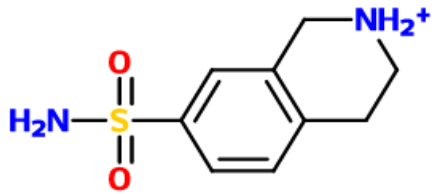
Hopkins & Groom, Nat. Rev. Drug. Disc. 2002

~10% human genome is involved in diseases

~20% - 50% of human genes involved in diseases code proteins targeted by drugs

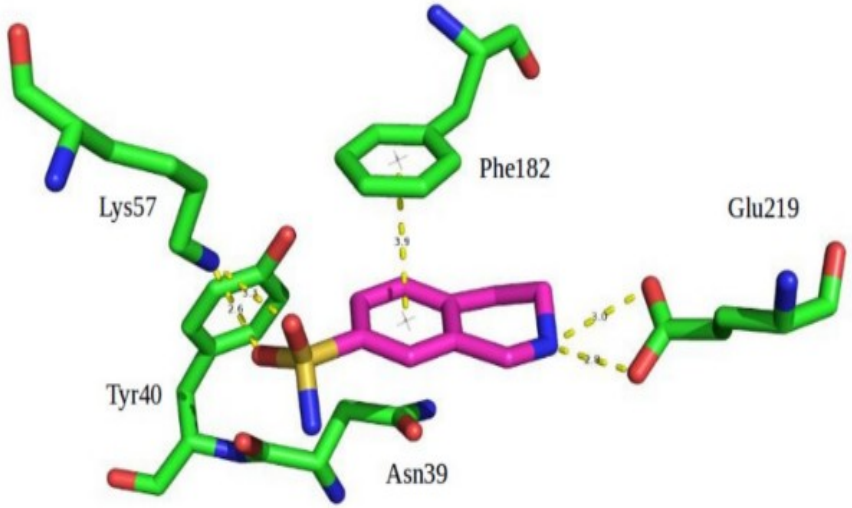
Druggability: ***the ability of a protein to have its activity modulated by the non-covalent binding of a small molecule***

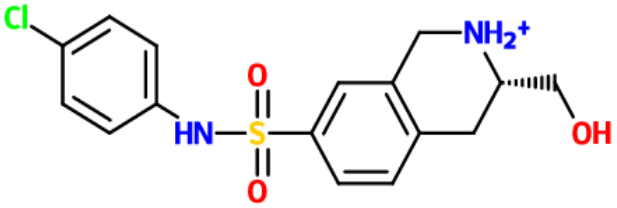
Can protein flexibility help drug design?



Inhibitor SKF
Volume = 258 Å^3
Ki= 580 nM

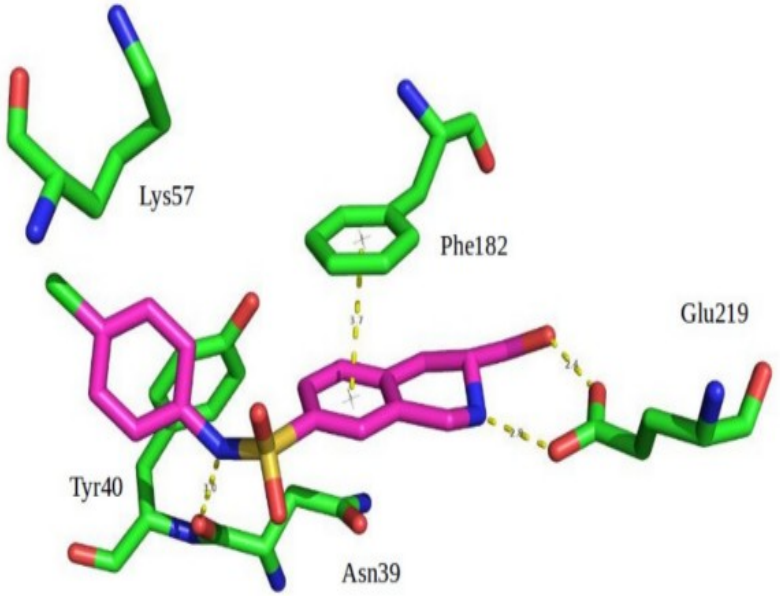
Pocket Volume: 304 Å^3



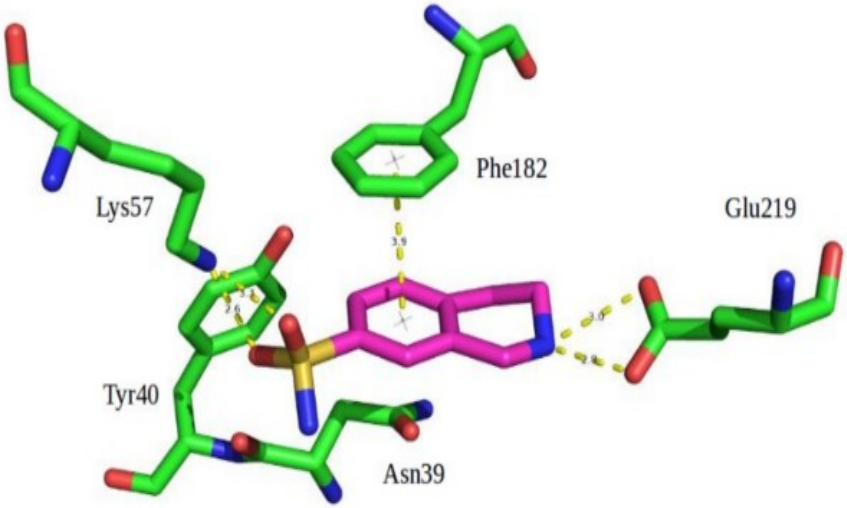


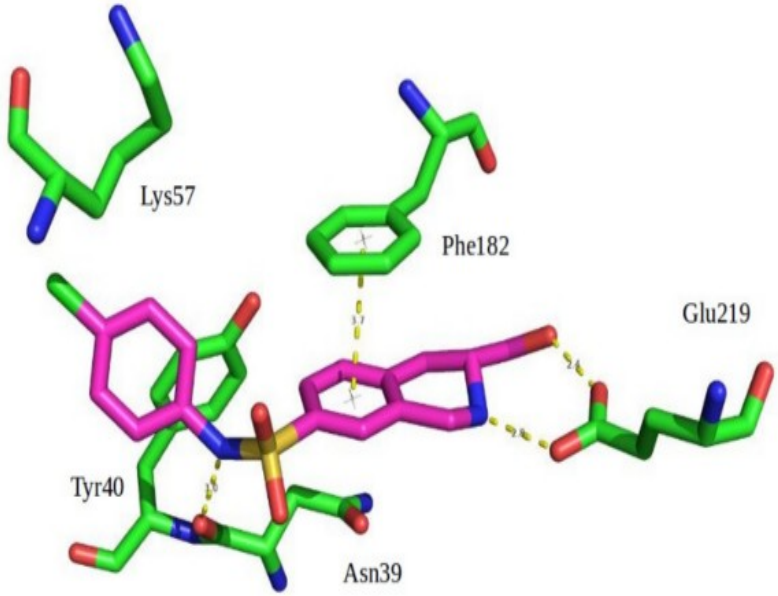
Inhibitor F83
Volume = 422 Å^3
Ki= 63 nM

Pocket Volume: 545 Å^3



How to find more potent ligands?





Druggability: *the ability of a protein to BIND a small molecule*

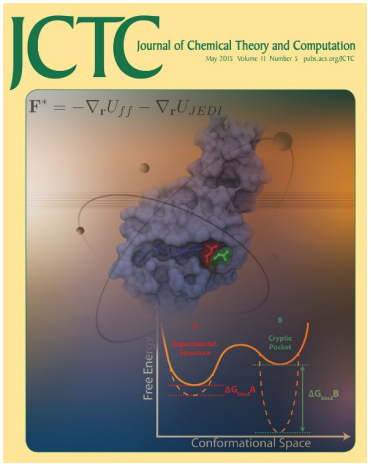
JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

A Collective Variable for the Rapid Exploration of Protein Druggability

Rémi Cuchillo, Kevin Pinto-Gil, and Julien Michel*



Quantification of Protein Druggability: JEDI

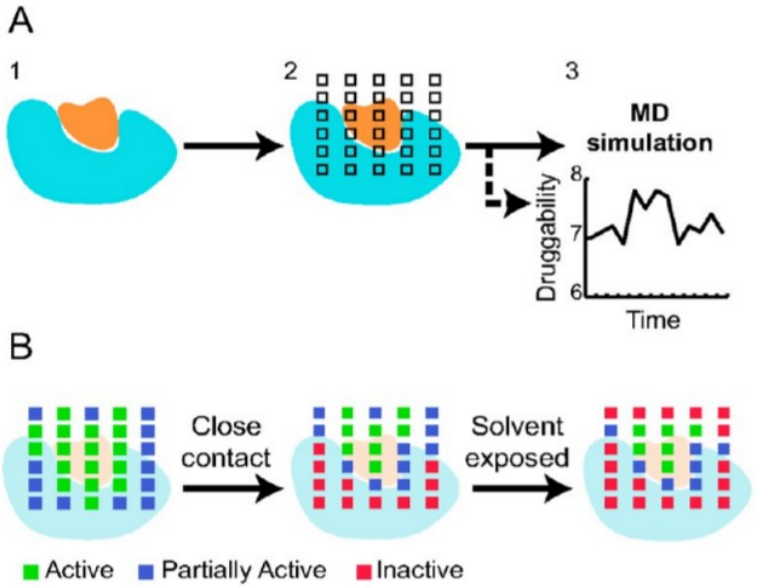
JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

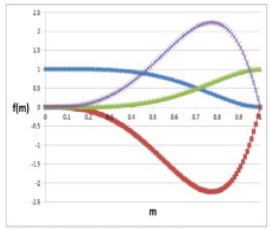
A Collective Variable for the Rapid Exploration of Protein Druggability

Rémi Cuchillo, Kevin Pinto-Gil, and Julien Michel*



$$a_i = S^{off}_{BS_i}(1.0, BS_{min}, \Delta BS)\, S^{on}_{mind_i}(1.0, CC_{mind}, \Delta CC) S^{on}_{exposure_i}(1.0, E_{min}, \Delta E)$$

Kernel Functions



Michel, J., JEDI Theory, Feb. 2016

Activity calculation

$$a_i = S^{off}_{BS_i}(1.0, BS_{min}, \Delta BS)\, S^{on}_{mind_i}(1.0, CC_{mind}, \Delta CC) S^{on}_{exposure_i}(1.0, E_{min}, \Delta E)$$

Exposure of grid point i to the

JEDI score

$$JEDI = V_{druglike}(\alpha V_a + \beta H_a + \gamma)$$

Symbol	Definition	Value
α	PLS derived volume coefficient	5.31
β	PLS derived hydrophobicity coefficient	24.29
γ	PLS derived intercept	-13.39
Δ_g	grid spacing	0.15 nm
BS_{min}	Minimum distance to *ligand group* from which the lig_i value starts to decrease	0.2 nm
ΔBS	distance interval over which lig_i decreases to 0	0.6 nm
CC_{mind}	distance below which a grid point is fully in close contact with the *binding site group*	0.15 nm
ΔCC	distance interval over which a grid point is in partial contact with the *binding site group*	0.15 nm
E_{min}	minimum exposure value from which a grid point is considered to be partially exposed to the *binding site group*	10.0
ΔE	interval over which a grid point becomes fully exposed to the *binding site group*	20.0
$CC2_{min}$	minimum distance below which a grid point is overlapping the binding site group	0.15 nm
$\Delta CC2$	distance interval over which a grid point is in partial contact with the binding site group	0.14 nm

Symbol	Definition	Value
GP_{min}	distance above which a grid point k is considered neighbor of grid point i	0.25 nm
GP_{max}	distance below which a grid point k is considered neighbor of grid point i	0.35 nm
d_{hydro}	distance below which a grid point i is in contact with a binding site atom (for hydrophobicity calculation)	0.40 nm
Δd_{hydro}	distance interval over which a grid point i is in partial contact with a binding site atom (for hydrophobicity calculation)	0.05 nm
V_{max}	volume below which $V_{druglike}$ is equal to 1	0.5 nm^3
ΔV_{max}	volume interval over which $V_{druglike}$ goes from 1 to 0	0.050 nm^3
V_{min}	volume below which $V_{druglike}$ is equal to 0	0.0 nm^3
ΔV_{min}	volume interval over which $V_{druglike}$ goes from 0 to 1	0.050 nm^3

$$V = \sum_{i=1}^{N} a_i V_g$$

$$V_g = spacing^3$$

$$V_a = \frac{V}{V_{max}}$$

PAR $V_{druglike} = S^{off}_{V}(1.0, V_{max}, \Delta V_{max}) S^{on}_{V}(1.0, V_{min}, \Delta V_{min})$

at the same time as

Penalize pockets that are too big

Penalize pockets that are too small

Quantification of Protein Druggability: JEDI

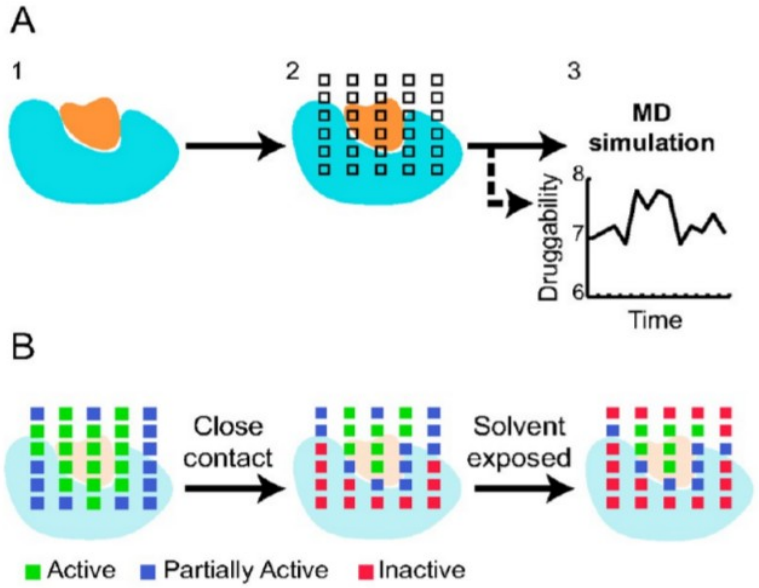
JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

A Collective Variable for the Rapid Exploration of Protein Druggability

Rémi Cuchillo, Kevin Pinto-Gil, and Julien Michel*



$$a_i = S^{off}_{BS_i}(1.0, BS_{min}, \Delta BS)\; S^{on}_{mind_i}(1.0, CC_{mind}, \Delta CC) S^{on}_{exposure_i}(1.0, E_{min}, \Delta E)$$

$$JEDI = F\left(a_{gridpoints}\right)$$

$$a_i = F\left(\vec{r}_{gridpoints}, \vec{r}_{atoms}\right)$$

$$JEDI = F\left(\vec{r}_{gridpoints}, \vec{r}_{atoms}\right)$$

CONTINUOUS FUNCTION **CONTINUOUS DERIVATIVES**

JEDI vs Experimental druggability

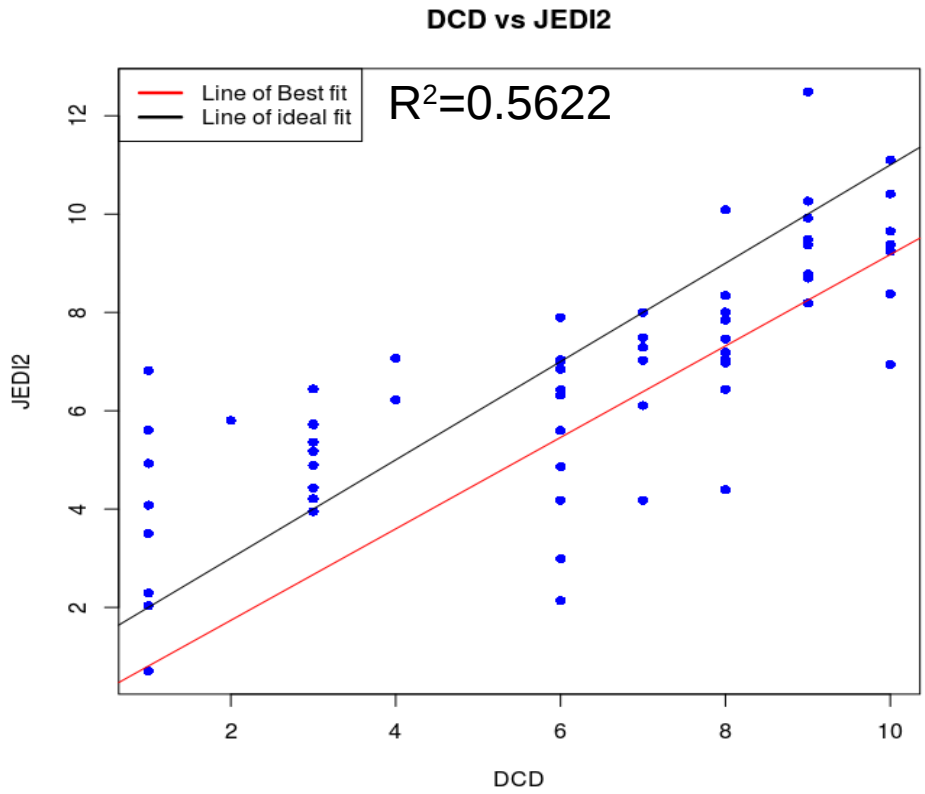
Druggable Cavity Directory — Public Repository for Collaborative Dataset Creation

Non redundant dataset

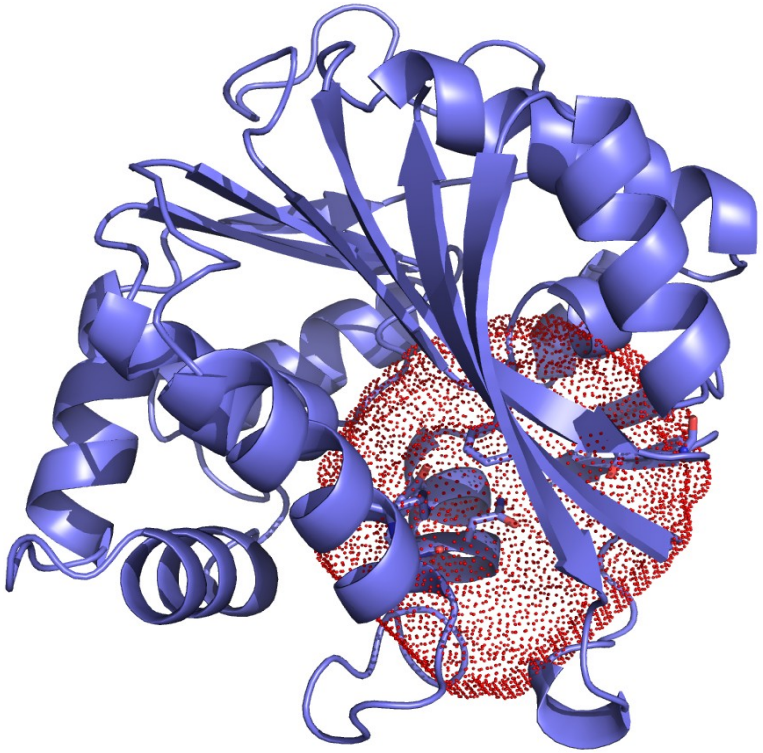
Proteins chosen and scored manually

Discrete scores

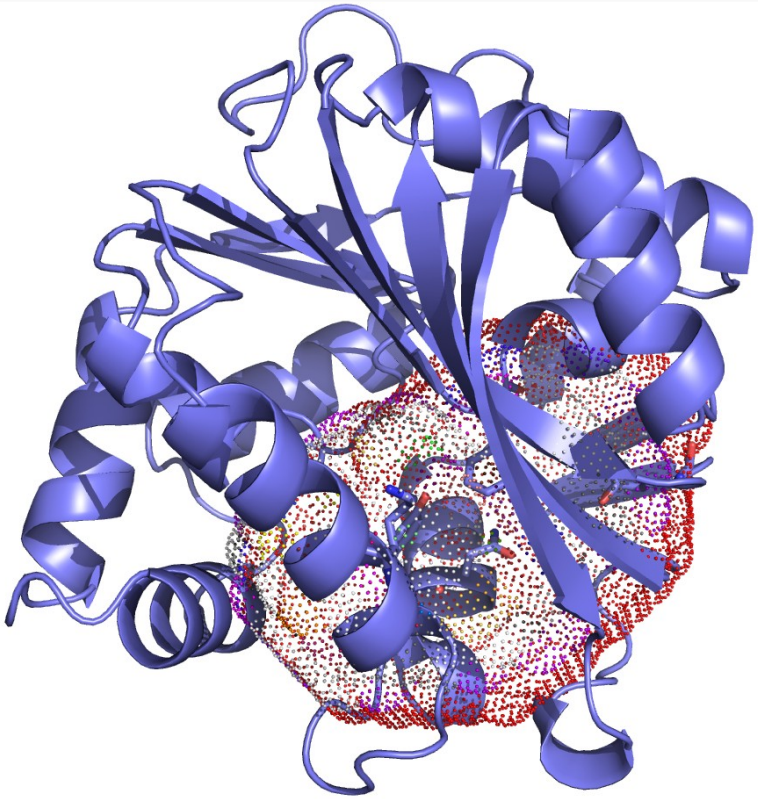
1-4: Non druggable (few or no known ligands)
5-7: Difficult (some known ligands)
8-10: Druggable (many known ligands)



Can JEDI distinguish hPNMT conformations?



Pocket	Small
LIGAND	SKF (580nM)
JEDI	7.3219 (8.17*)
Vd	1
Va	0.61497
Ha	0.71826



Pocket	Big
LIGAND	F83 (63 nM)
JEDI	9.8087 (8.98*)
Vd	1
Va	0.82106
Ha	0.77558

Part 2: Dynamics

Can we predict the binding mode of the big ligand (F83) if knowing only the structure bound to the small one (SKF)?

Druggability vs time

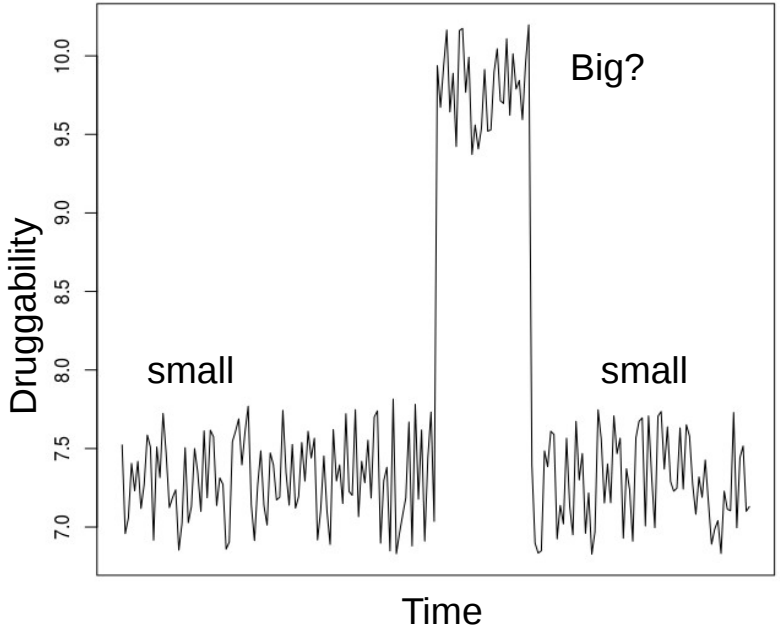
JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

A Collective Variable for the Rapid Exploration of Protein Druggability

Rémi Cuchillo, Kevin Pinto-Gil, and Julien Michel*

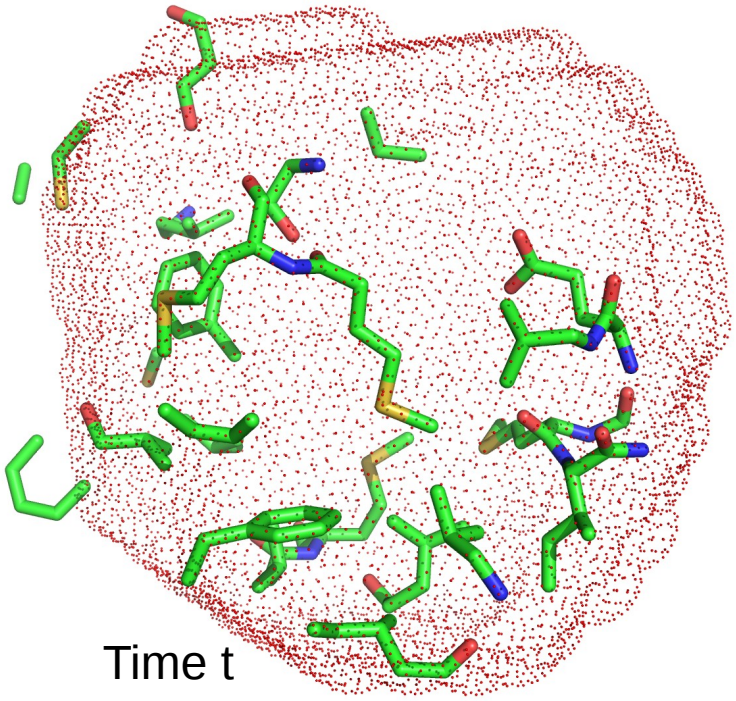
PLUMED

GROMACS
FAST. FLEXIBLE. FREE.



Grid Update

If the atoms move, we need to updat the grid according to that movement!



Time t

1) Calculate binding site COM_0 at time t_0

2) Move atoms

3) calculate binding site COM at time $t=t_0+dt$

4) Center binding site at time t at the COM at time t_0 (not shown)

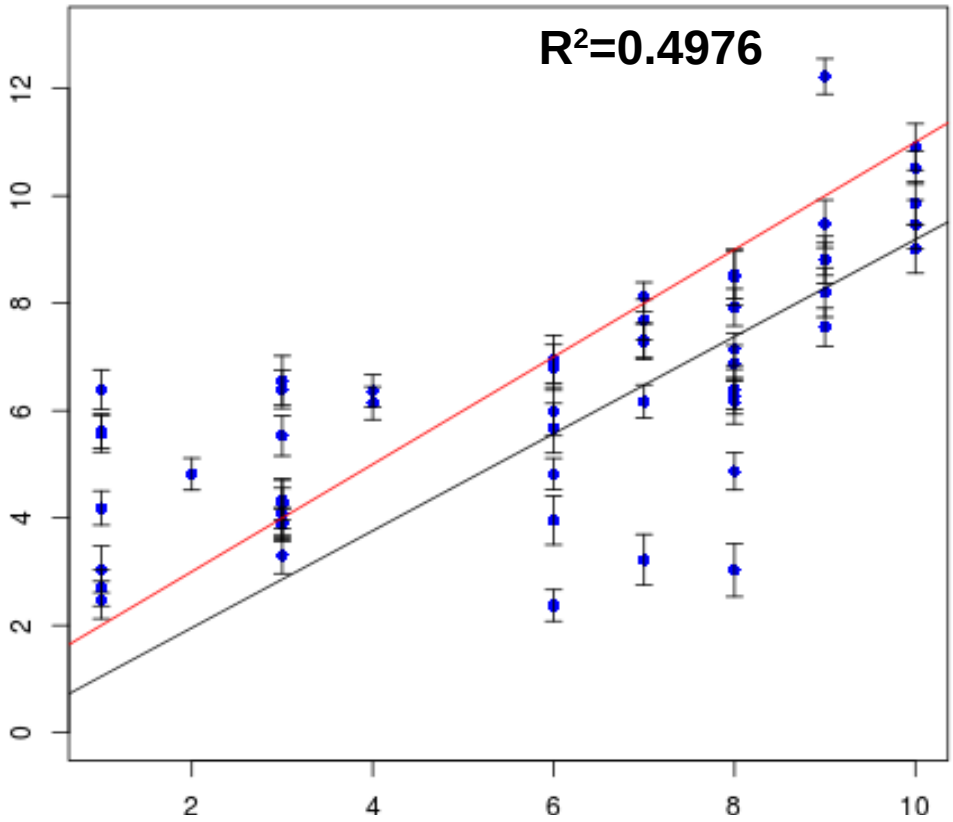
5) Calculate the rotation matrix R that fits the translated binding site at time t to the binding site at time t_0 with the lowest RMSD

Behaviour of the JEDI scores in MD

MD Setup details

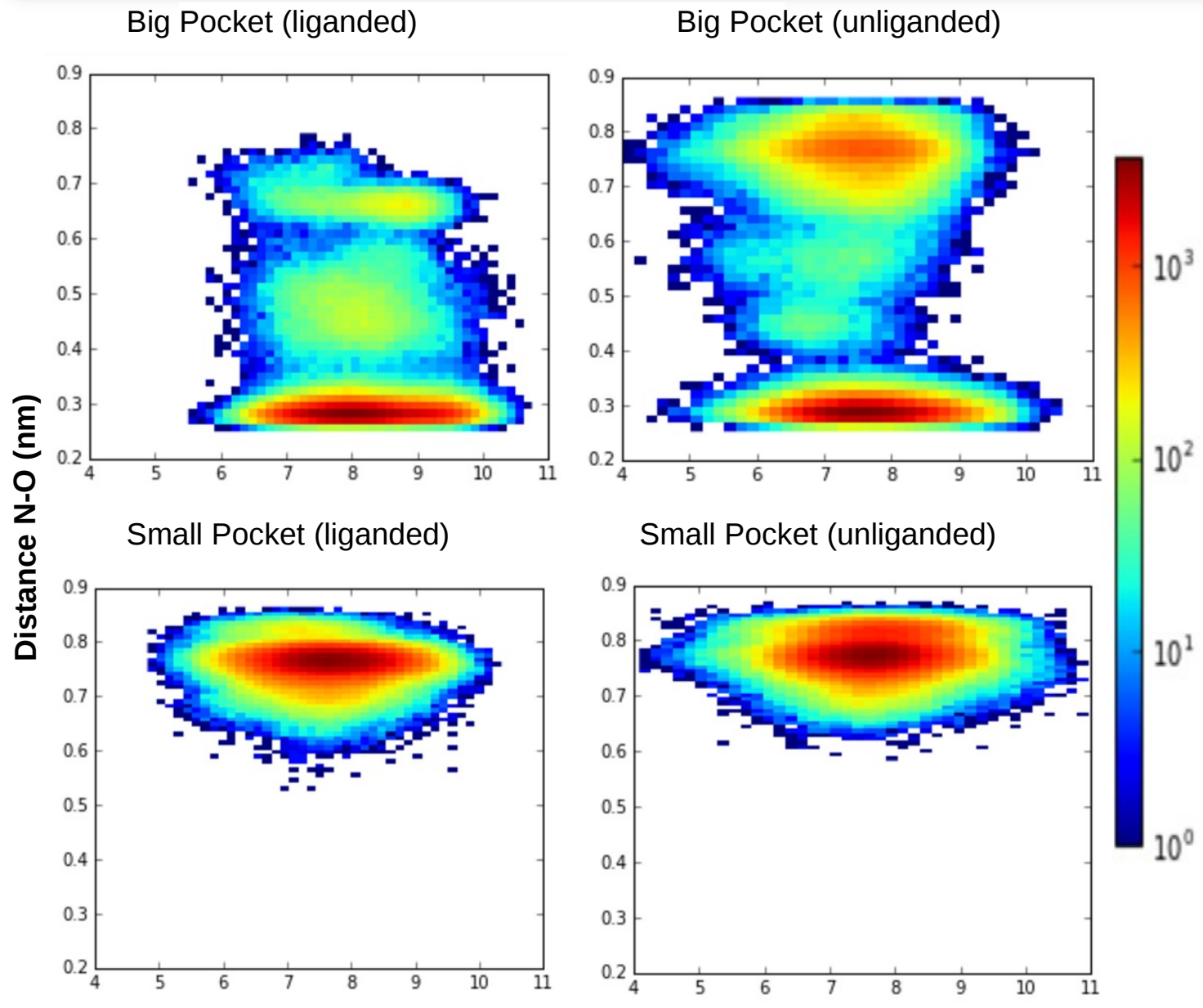
- 1 ns MD at 300 K
- dt=0.002 ps
- Explicit solvent (TIP3P)
- Ligand cavities filled with water
- Saving *JEDI* every 1000 steps
- GROMACS 5.1[4] + PLUMED 2.2b[5]

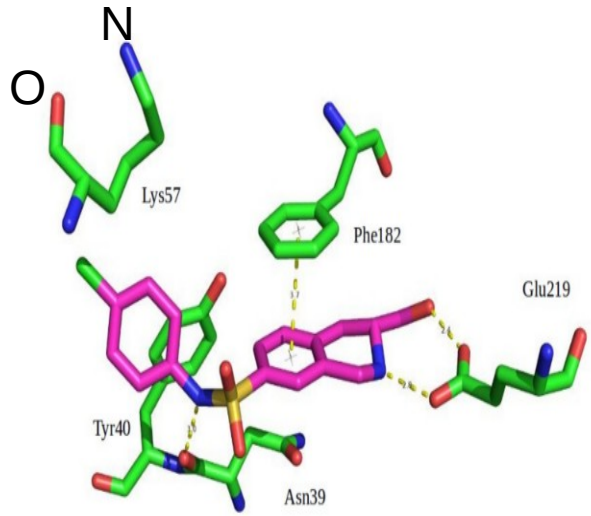
https://github.com/michellab/plumed2
https://github.com/michellab/jedi-utilities

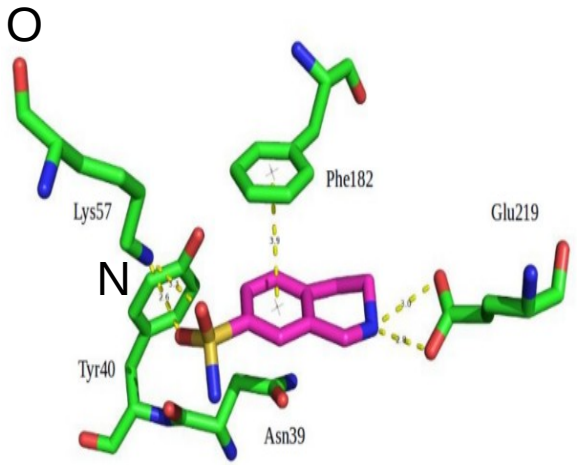


Short simulations don't show big fluctuations of the JEDI score and the correlation does not decrease significantly

Sampling of hPNMT conformations







Part 3: Biased MD

How to make hPNMT sample the conformation that binds the big ligand (F83) from the one that binds the small one (SKF)?

Biasing Potential and Force

CONTINUOUS FUNCTION

$$JEDI = F\left(\vec{r}_{gridpoints}, \vec{r}_{atoms}\right)$$

$$U_{JEDI} = F\left(JEDI\right) = F\left(\vec{r}_{gridpoints}, \vec{r}_{atoms}\right)$$

$$U_{system} = U_{FF} + U_{JEDI}$$

DIFFERENTIABLE FUNCTION

$$\frac{d\left(JEDI\right)}{dx_{atom_i}} = \left(\ldots\right)$$

$$\vec{F}_{x_i} = -\frac{d\left(U_{system}\right)}{dx_i} = -\left[\frac{d\left(U_{FF}\right)}{dx_i} + \frac{d\left(U_{JEDI}\right)}{dx_i}\right]$$

"Classic" MD force

JEDI bias force

Biasing Potential and Force

HARMONIC POTENTIAL

$$U_{JEDI} = k\left(JEDI - JEDI_0\right)^2$$

$$\vec{F}_{JEDI_i} = -\frac{d\left[k\left(JEDI - JEDI_0\right)^2\right]}{dx_i} = -2k\left(JEDI - JEDI_0\right)\frac{d\left(JEDI\right)}{dx_i}$$

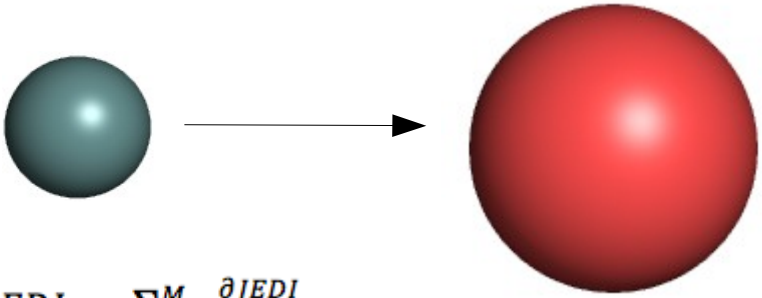
IN COMPARISON TO MD FORCES WE WANT JEDI FORCES TO BE:

- Slowly varying absolute value (difference in force norm between steps)
- Slowly varying in direction (angle between force vectors of consecutive steps)

Removal of net force and torque

Grid point i pushes atom j...

...but atom j does not push back



$$F_{ij}+F_{ji}\neq 0$$

$$\nabla JEDI_x = \sum_{j=1}^{M} \frac{\partial JEDI}{\partial x_{p_j}}$$

$$\nabla JEDI_y = \sum_{j=1}^{M} \frac{\partial JEDI}{\partial y_{p_j}}$$

$$\nabla JEDI_y = \sum_{j=1}^{M} \frac{\partial JEDI}{\partial z_{p_j}}$$

$$\nabla TorJEDI_x = \sum_{j=1}^{M} y_{p_j} \frac{\partial JEDI}{\partial z_{p_j}} - z_{p_j} \frac{\partial JEDI}{\partial y_{p_j}}$$

$$\nabla TorJEDI_y = \sum_{j=1}^{M} z_{p_j} \frac{\partial JEDI}{\partial x_{p_j}} - x_{p_j} \frac{\partial JEDI}{\partial z_{p_j}}$$

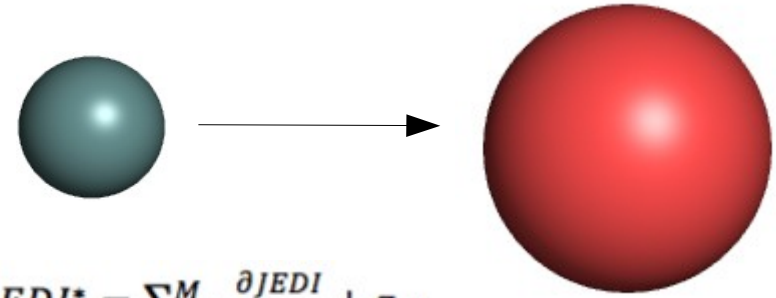
$$\nabla TorJEDI_z = \sum_{j=1}^{M} x_{p_j} \frac{\partial JEDI}{\partial y_{p_j}} - y_{p_j} \frac{\partial JEDI}{\partial x_{p_j}}$$

$$\begin{pmatrix} \\ \\ \\ \\ \\ \\ \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Removal of net force and torque

Grid point i pushes atom j...

...but atom j does not push back



$$F_{ij}+F_{ji}\neq 0$$

$$\begin{pmatrix} \nabla JEDI_x^* = \sum_{j=1}^{M} \frac{\partial JEDI}{\partial x_{p_j}} + z_{3j} \\ \nabla JEDI_y^* = \sum_{j=1}^{M} \frac{\partial JEDI}{\partial y_{p_j}} + z_{3j+1} \\ \nabla JEDI_z^* = \sum_{j=1}^{M} \frac{\partial JEDI}{\partial z_{p_j}} + z_{3j+2} \\ \nabla TorJEDI_x^* = \sum_{j=1}^{M} y_{p_j}(\frac{\partial JEDI}{\partial z_{p_j}} + z_{3j+2}) - z_{p_j}(\frac{\partial JEDI}{\partial y_{p_j}} + z_{3j+1}) \\ \nabla TorJEDI_y^* = \sum_{j=1}^{M} z_{p_j}(\frac{\partial JEDI}{\partial x_{p_j}} + z_{3j}) - x_{p_j}(\frac{\partial JEDI}{\partial z_{p_j}} + z_{3j+2}) \\ \nabla TorJEDI_z^* = \sum_{j=1}^{M} x_{p_j}(\frac{\partial JEDI}{\partial y_{p_j}} + z_{3j+1}) - y_{p_j}(\frac{\partial JEDI}{\partial x_{p_j}} + z_{3j}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Undetermined Linear Equation system: $Ax=y$

Removal of net force and torque

1) Calculate The Moore-Penrose pseudo inverse matrix $\mathbf{A}^{+} = \mathbf{A}^{t}(\mathbf{A}\mathbf{A}^{t})^{-1}$

2) Solving the linear equation system $\boldsymbol{z} = \mathbf{A}^{+}\boldsymbol{y}$

The solution of 2) minimizes the sum:

$$\|\boldsymbol{z}\| = \sum_{i=1}^{3M} z_i^2$$

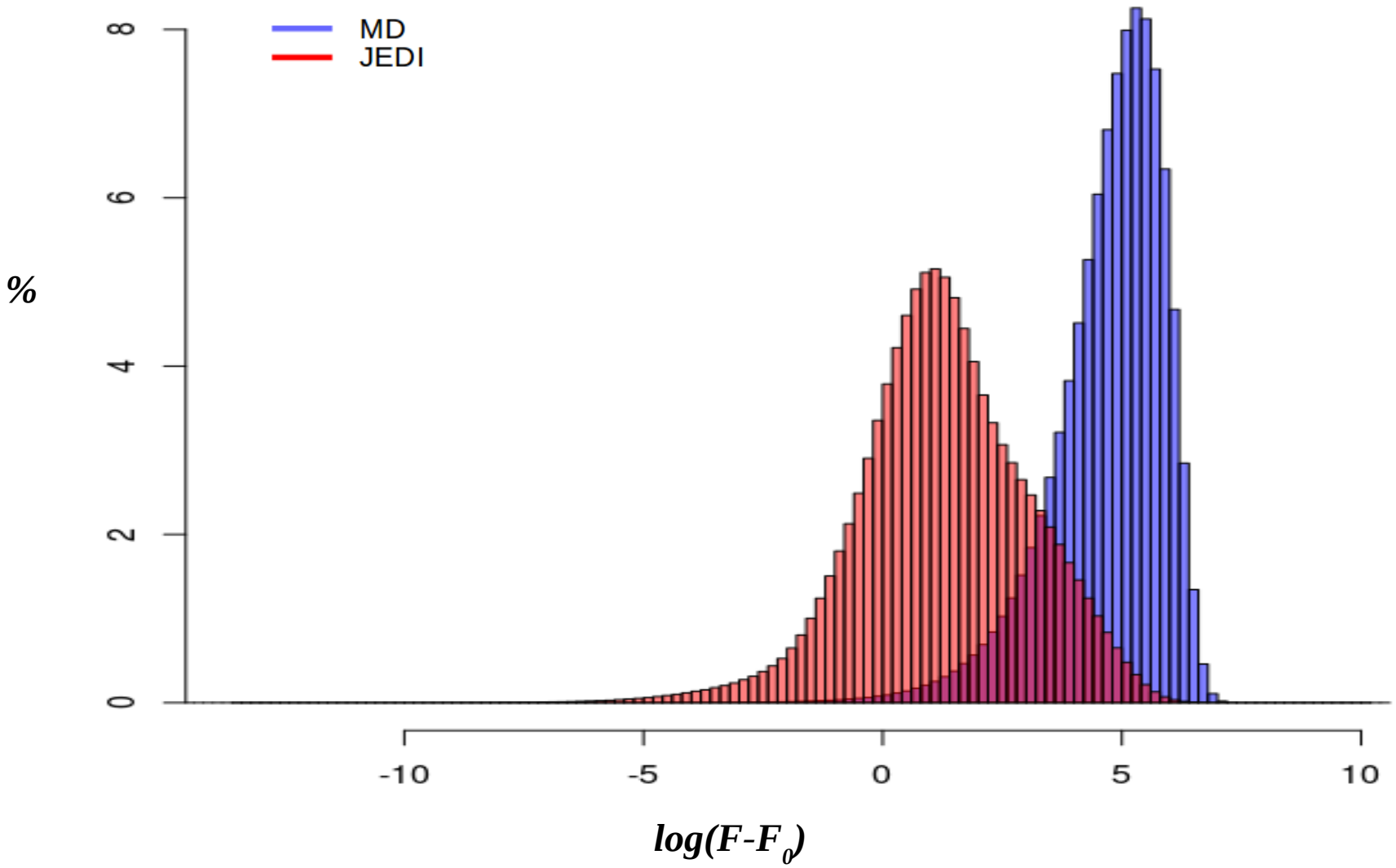
$$\frac{\partial JEDI^*}{\partial x_{p_j}} = \frac{\partial JEDI}{\partial x_{p_j}} + z_{3j}$$

$$\frac{\partial JEDI^*}{\partial y_{p_j}} = \frac{\partial JEDI}{\partial y_{p_j}} + z_{3j+1}$$

$$\frac{\partial JEDI^*}{\partial z_{p_j}} = \frac{\partial JEDI}{\partial z_{p_j}} + z_{3j+2}$$

Behaviour of the JEDI forces

Change in the absolute value of the force between consecutive steps

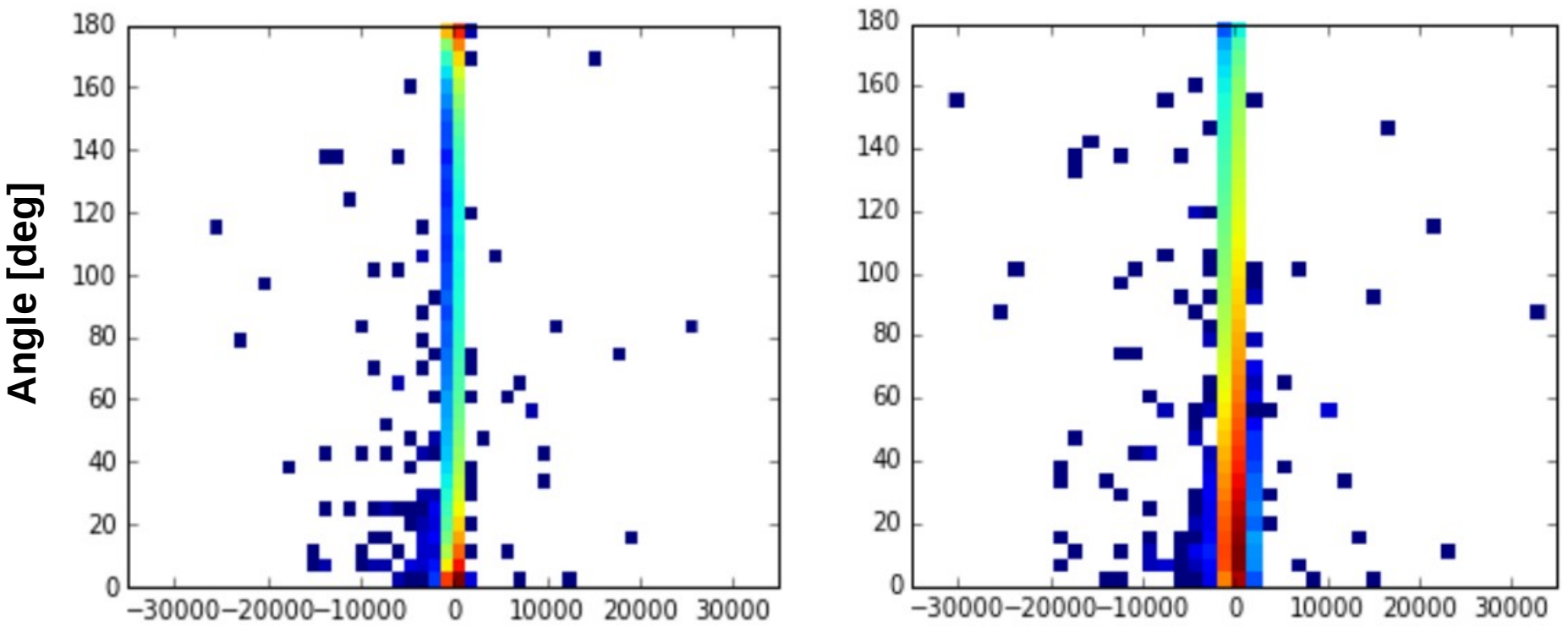


Behaviour of the JEDI forces

Change in the direction of the force between consecutive steps

JEDI Forces

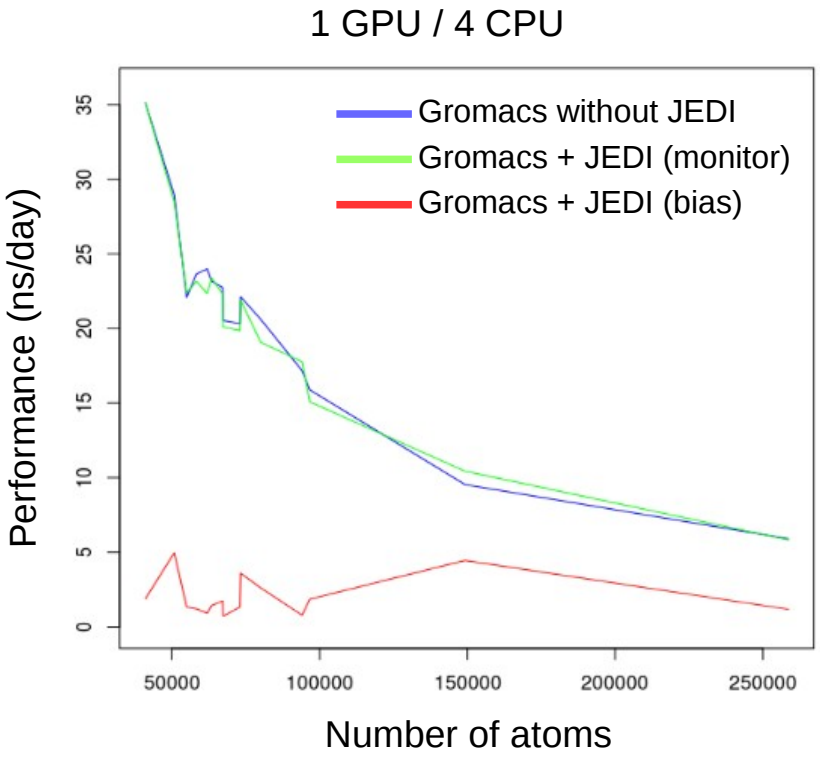
MD Forces



Part 4: Performance

Can high druggability conformations be sampled faster using JEDI?

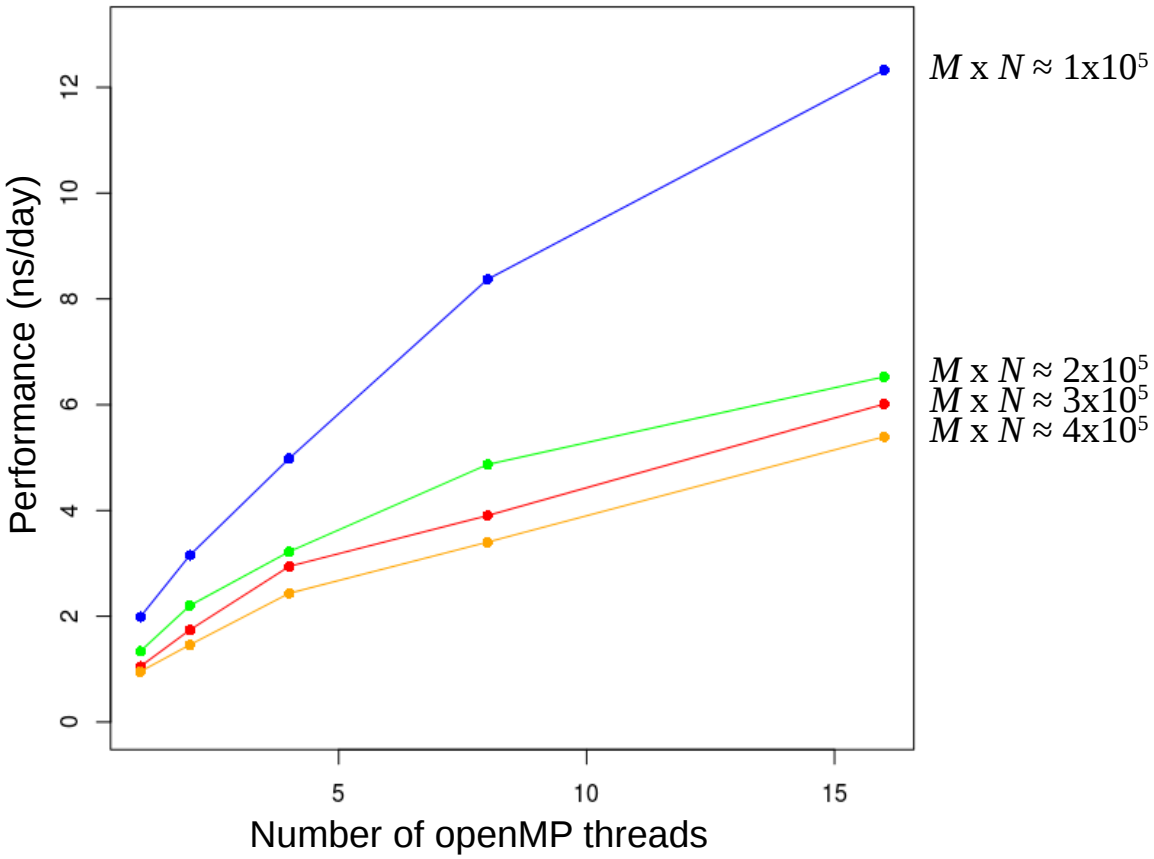
Serial performance




```
for atom in binding_site:
    do_stuff()
    for point in grid:
        if point too_far_from atom:
            skip_point()
        endif
        do_more_stuff()
        for neighbour in neighbours[point]:
            do_even_more_stuff()
        endfor
    endfor
    still_more_stuff()
    for point in grid:
        just_a_bit_more()
    endfor
    last_things_to_do()
endfor
```

$$maxIter = M\left[\sum_{i=0}^{N}(1 + neighbours[i])\right]$$

Parallel performance using openMP




```
#pragma omp parallel for
for atom in binding_site:
        do_stuff()
    for point in grid:
        if point too_far_from atom:
            skip_point()
        endif
        do_more_stuff()
        for neighbour in neighbours[point]:
            do_even_more_stuff()
        endfor
    endfor
    still_more_stuff()
    for point in grid:
        just_a_bit_more()
    endfor
    last_things_to_do()
endfor
```

Up to 3x-6x increase in speed using openMP

$$maxIter = M\left[\sum_{i=0}^{N}(1 + neighbours[i])\right]$$

Part 5: Conclusions and Outlook

How to make JEDI still better? What to do with it then?

Conclusions

- JEDI shows a proper correlation with manually assigned druggability scores and allows to distinguish two conformations of the binding site of hPNMT
- The JEDI score is stable and no big fluctuations are seen in short MD simulations
- Different conformations of hPNMT can show similar JEDI scores in long MD simulations
- The biasing forces are stable in short biased simulations
- The performance of the code is still to be improved

Further improvement

JCTC Journal of Chemical Theory and Computation

Article

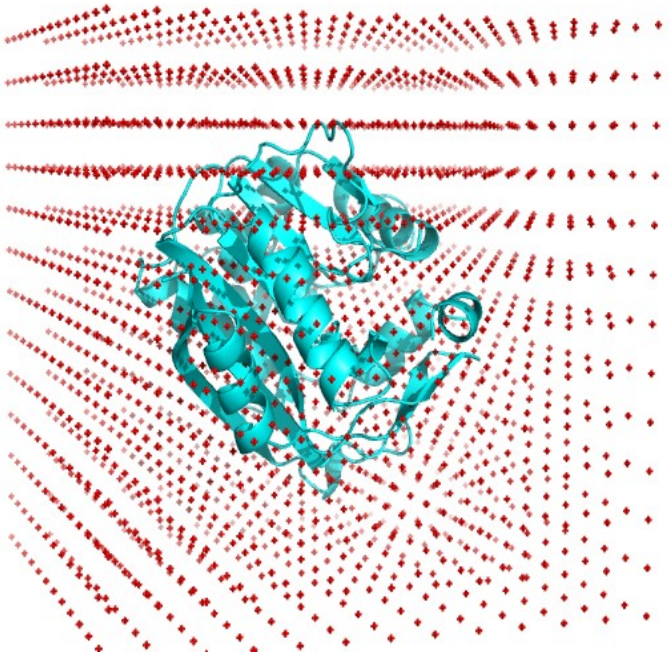
pubs.acs.org/JCTC

Accurate Multiple Time Step in Biased Molecular Simulations

Marco Jacopo Ferrarotti, Sandro Bottaro, Andrea Pérez-Villa, and Giovanni Bussi*

Scuola Internazionale Superiore di Studi Avanzati (SISSA), via Bonomea 265, 34136 Trieste, Italy

Evaluation of forces every n steps instead of 1 to boost performance



Overlapping a big grid onto the whole protein and clustering the grid points in order to detect the opening of transient pockets

Application to hPNMT

- Perform JEDI-biased simulations of hPNMT to search for the desired conformational change
- Perform free energy calculations to determine the population of the conformations with higher JEDI scores
- Perform docking calculations of known ligands in order to verify the results

Acknowledgements

Michel research group

Dr. Julien Michel

Royal Society University Research Fellow
Room 263
School of Chemistry
University of Edinburgh
Joseph Black Building
King's Buildings
David Brewster Road
Edinburgh
EH9 3FJ
United Kingdom

Antonia Mey

Degree:
PhD. Physics, University of Nottingham, 2013

Keywords:
Free energy calculations, Molecular simulation workflows

Jordi Juárez Jiménez

Degree:
PhD. Biomedicine, University of Barcelona, 2014

Keywords:
Cyclophilins inhibition, Docking and scoring, Protein dynamics

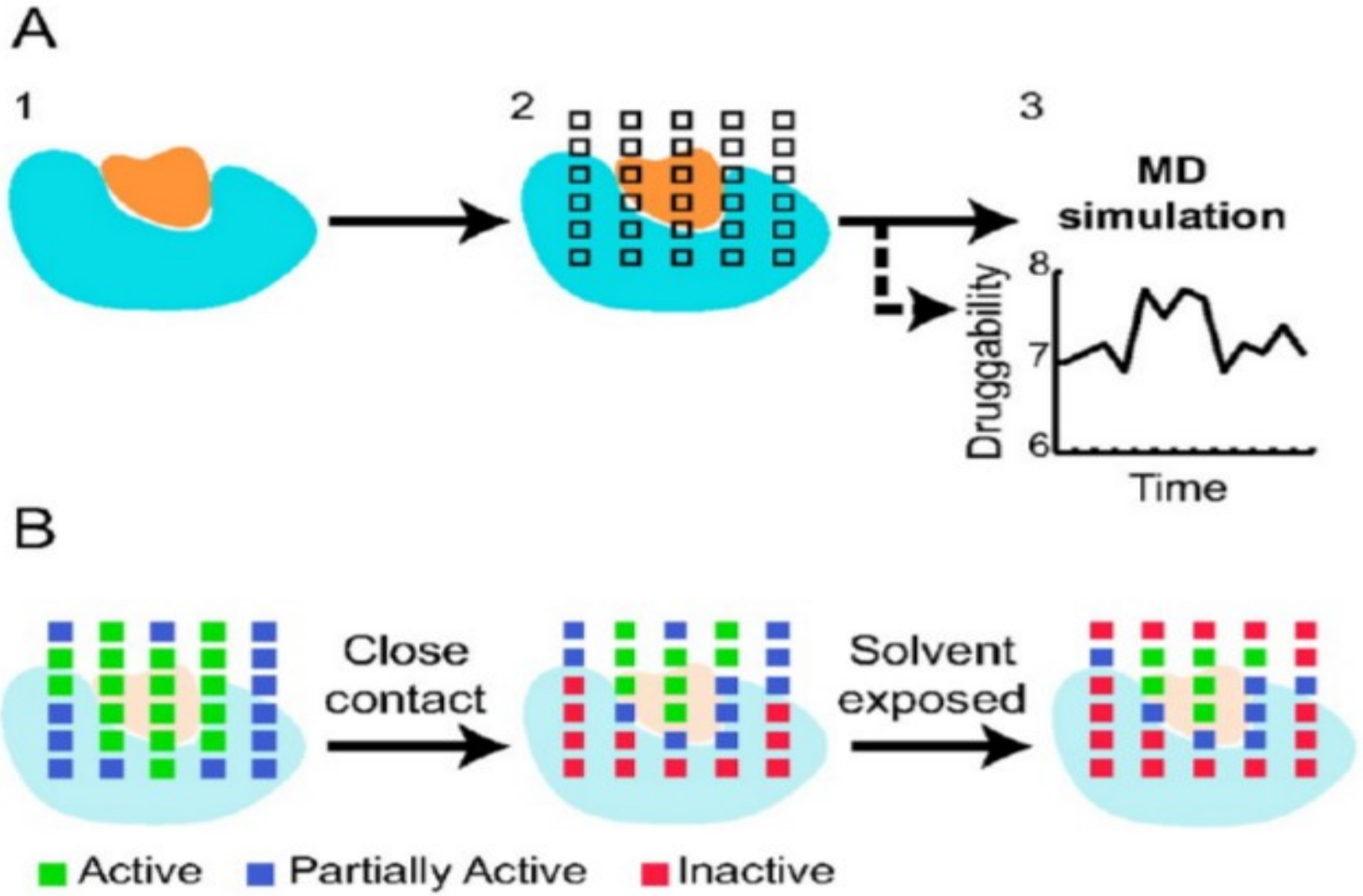
Funding

European Research Council

Acknowledgements

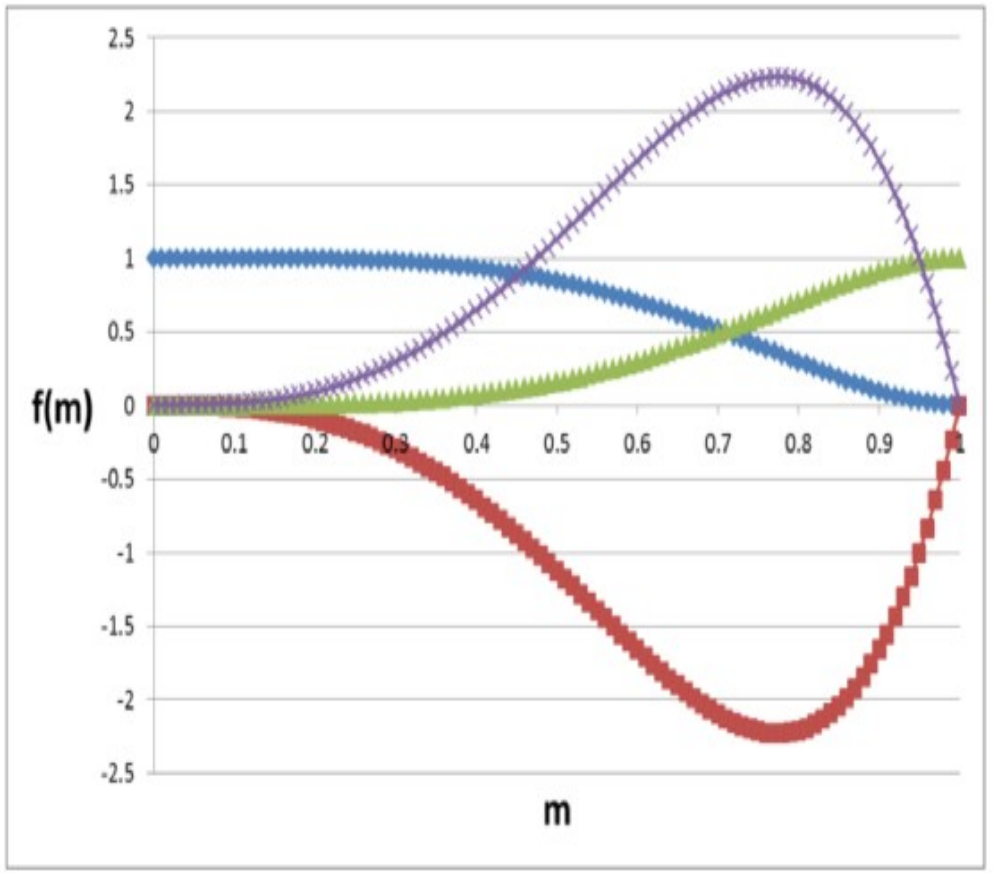
PLUMED

What is JEDI?



Cuchillo *et al*, JCTC 2015

Kernel Functions



$$S_v^{on}(k, v_{min}, \Delta) = \begin{cases} 0 \; if \; m < 0 \\ k[1 - (1 - m^2)^2(1 + 2m^2)] \; if \; 0 \leq m \leq 1 \\ k \; if \; m > 1 \end{cases}$$

$$S_v^{off}(k, v_{min}, \Delta) = \begin{cases} k \; if \; m < 0 \\ k[(1 - m^2)^2(1 + 2m^2)] \; if \; 0 \leq m \leq 1 \\ 0 \; if \; m > 1 \end{cases}$$

$$m = \frac{v - v_{min}}{\Delta}$$

Michel, J., JEDI Theory, Feb. 2016

Penalize the descriptors with values too big or too low regarding those of the benchmark set

Activity calculation

$$a_i = S^{off}_{BS_i}(1.0, BS_{min}, \Delta BS)\, S^{on}_{mind_i}(1.0, CC_{mind}, \Delta CC) S^{on}_{exposure_i}(1.0, E_{min}, \Delta E)$$

Distance of grid point *i* to a known ligand
(1 if not defined)

Distance of grid point *i* to binding site atoms

Exposure of grid point *i* to the solvent

Hydrophobicity Descriptor

$$h_i = \frac{apolar_i}{contacts_i}$$

$$contacts_i = \sum_{j=1}^{M} S^{off}_{\|\boldsymbol{r}_{ij,t}\|}(a_i, d_{hydro}, \Delta d_{hydro})$$

$$apolar_i = \sum_{j=1}^{M} I_{apolar}(j) S^{off}_{\|\boldsymbol{r}_{ij,t}\|}(a_i, d_{hydro}, \Delta d_{hydro})$$

$$I_{apolar}(j) \begin{cases} 1 \; if \; j \in apolar \; group \\ 0 \; if \; j \in polar \; group \end{cases}$$

$$H_a = \sum_{i=1}^{N} \frac{h_i a_i}{\sum_{i=1}^{N} a_i}$$

Apolar and polar atoms are defined at the same time as the grid, before the simulation

Active Volume Descriptor

$$V = \sum_{i=1}^{N} a_i V_g$$

$$V_g = spacing^3$$ → **ONLY FOR EVENLY SPACED GRIDS!!!!**

$$V_a = \frac{V}{V_{max}}$$ → **PARAMETER**

Druglike Volume Descriptor

$$V_{druglike} = S_V^{off}(1.0, V_{max}, \Delta V_{max}) S_V^{on}(1.0, V_{min}, \Delta V_{min})$$

Penalize pockets that are too big

Penalize pockets that are too small

JEDI score

$$JEDI = V_{druglike}(\alpha V_a + \beta H_a + \gamma)$$

Symbol	Definition	Value
α	PLS derived volume coefficient	5.31
β	PLS derived hydrophobicity coefficient	24.29
γ	PLS derived intercept	-13.39
Δ_g	grid spacing	0.15 nm
BS_{min}	Minimum distance to *ligand group* from which the lig_i value starts to decrease	0.2 nm
ΔBS	distance interval over which lig_i decreases to 0	0.6 nm
CC_{mind}	distance below which a grid point is fully in close contact with the *binding site group*	0.15 nm
ΔCC	distance interval over which a grid point is in partial contact with the *binding site group*	0.15 nm
E_{min}	minimum exposure value from which a grid point is considered to be partially exposed to the *binding site group*	10.0
ΔE	interval over which a grid point becomes fully exposed to the *binding site group*	20.0
$CC2_{min}$	minimum distance below which a grid point is overlapping the binding site group	0.15 nm
$\Delta CC2$	distance interval over which a grid point is in partial contact with the binding site group	0.14 nm

Symbol	Definition	Value
GP_{min}	distance above which a grid point k is considered neighbor of grid point i	0.25 nm
GP_{max}	distance below which a grid point k is considered neighbor of grid point i	0.35 nm
d_{hydro}	distance below which a grid point i is in contact with a binding site atom (for hydrophobicity calculation)	0.40 nm
Δd_{hydro}	distance interval over which a grid point i is in partial contact with a binding site atom (for hydrophobicity calculation)	0.05 nm
V_{max}	volume below which $V_{druglike}$ is equal to 1	0.5 nm^3
ΔV_{max}	volume interval over which $V_{druglike}$ goes from 1 to 0	0.050 nm^3
V_{min}	volume below which $V_{druglike}$ is equal to 0	0.0 nm^3
ΔV_{min}	volume interval over which $V_{druglike}$ goes from 0 to 1	0.050 nm^3

JEDI version 1

- Written in C
- Parameters hard-coded
- Apolar and polar atoms definition hard coded
- Works in implicit solvent
- Gromacs 4.5.5
- PLUMED 1.3

JEDI version 2

- Written in C++
- Parameters supplied by *jedi.params*
- Grid, apolar and polar atoms file generated with a preprocessor *jedi_setup.py*
- Handles explicit solvent and PBC
- Gromacs 5.1.0
- PLUMED 2.2