

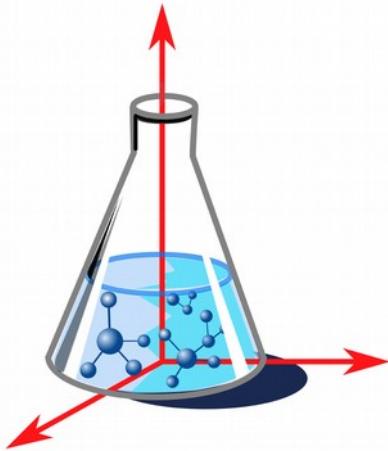


ChemFlow - From 2D chemical libraries to protein-ligand binding *free energies* in a few clicks

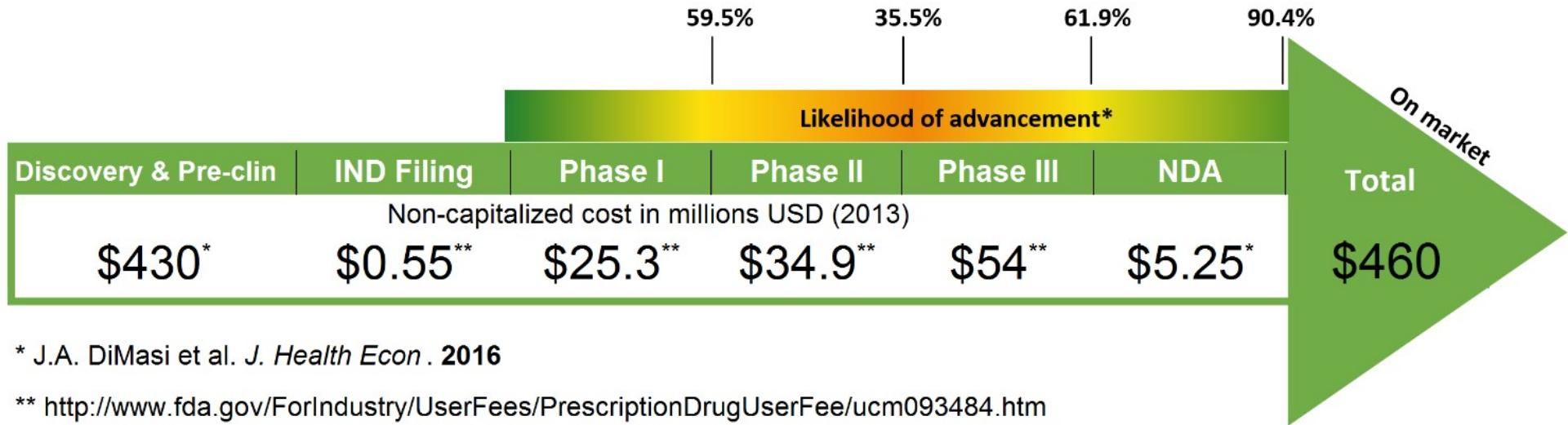
go Enry B. Gomes, PhD | dgomes@pq.cnpq.br

Departamento de Ciências da Computação

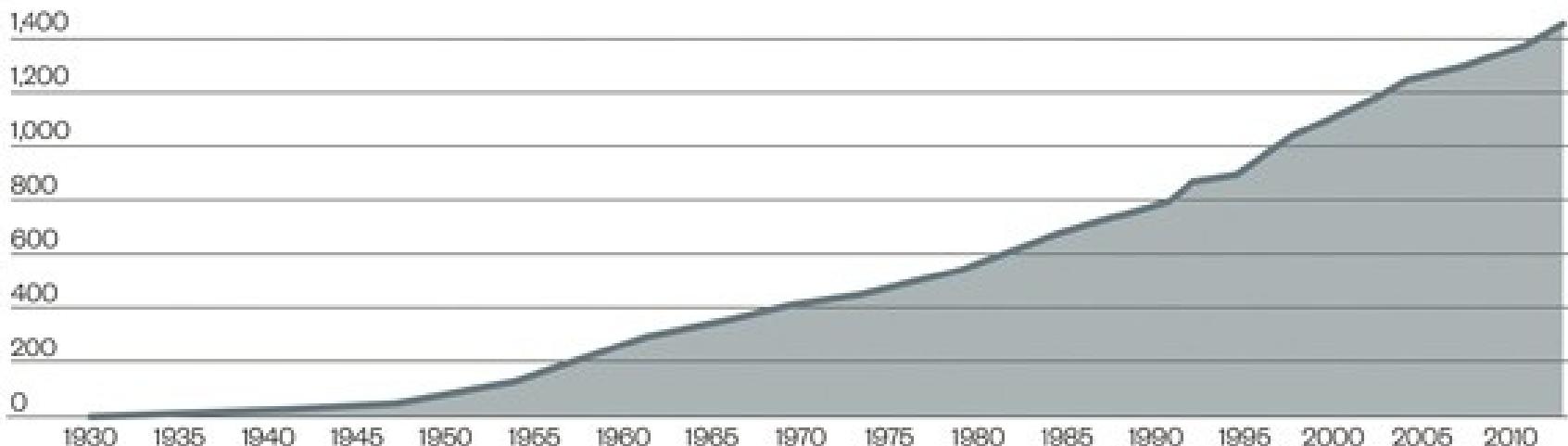
Universidade Federal de Juiz de Fora



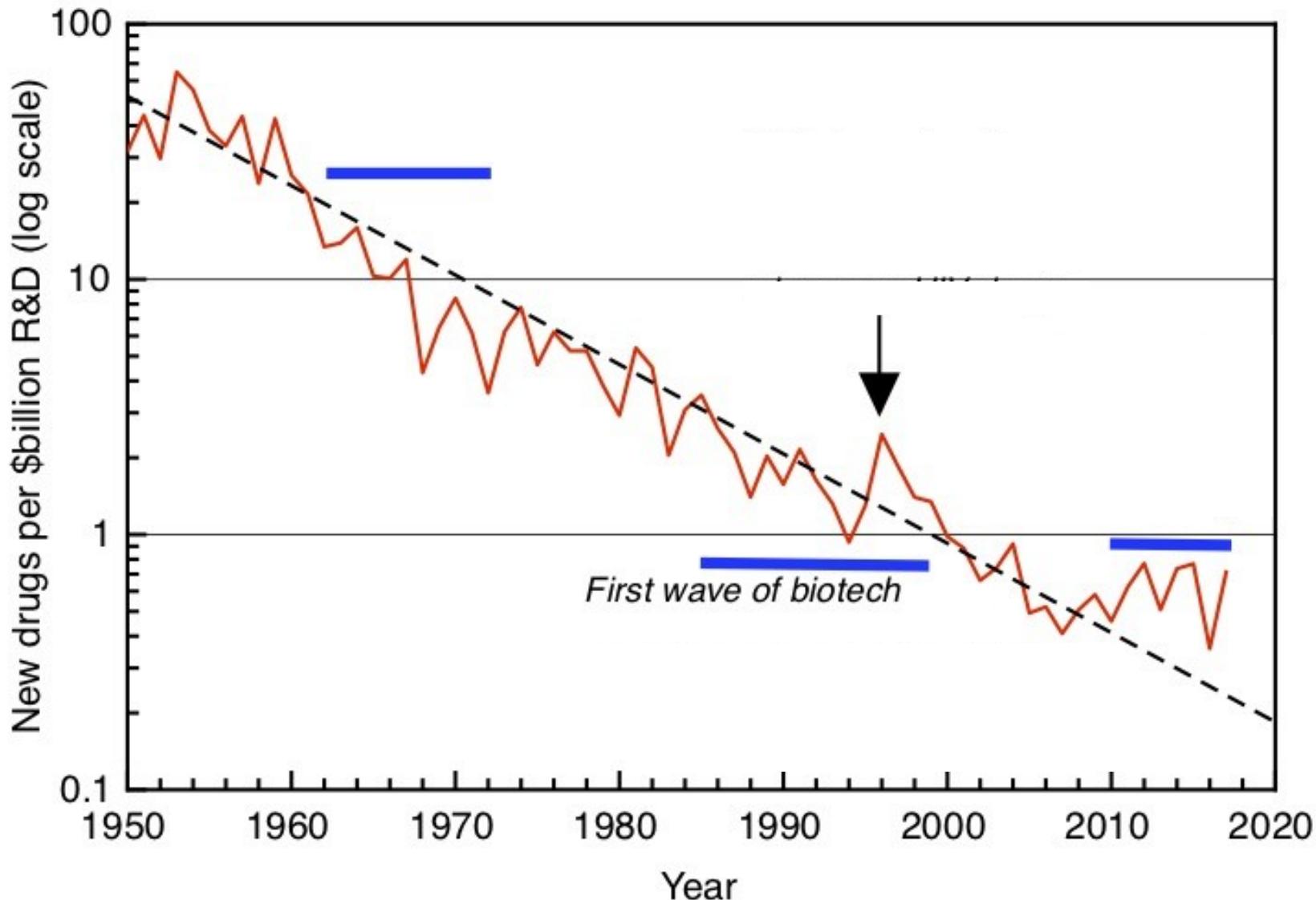
The Cost of Drug Discovery



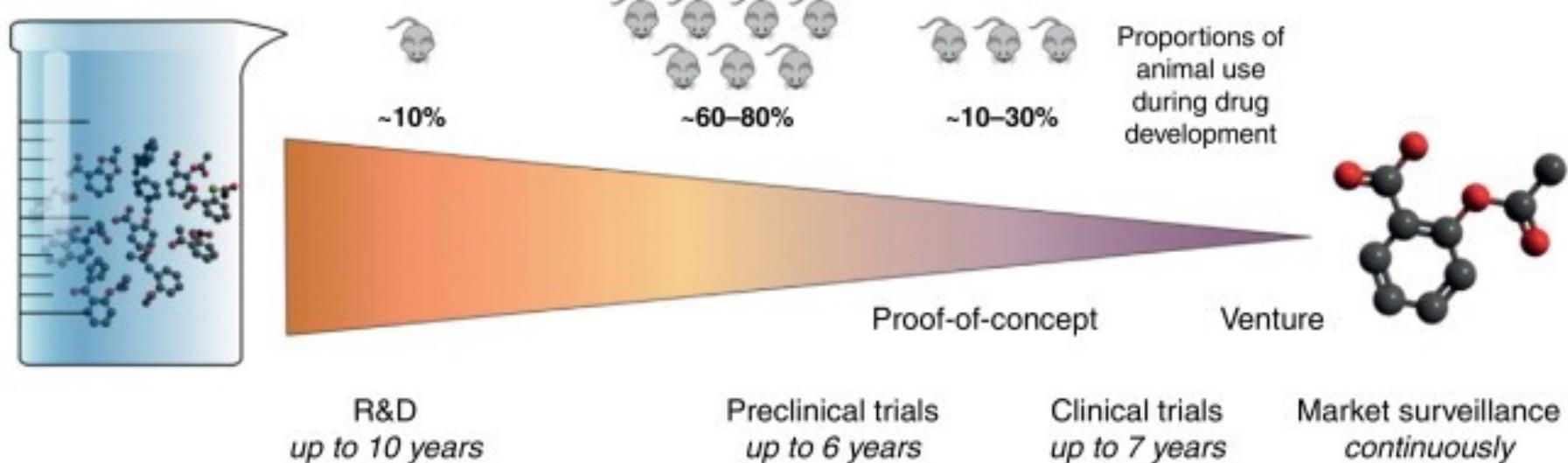
1600 cumulative NMEs



Efficiency of R&D in pharma industry



Reduce Animal Use



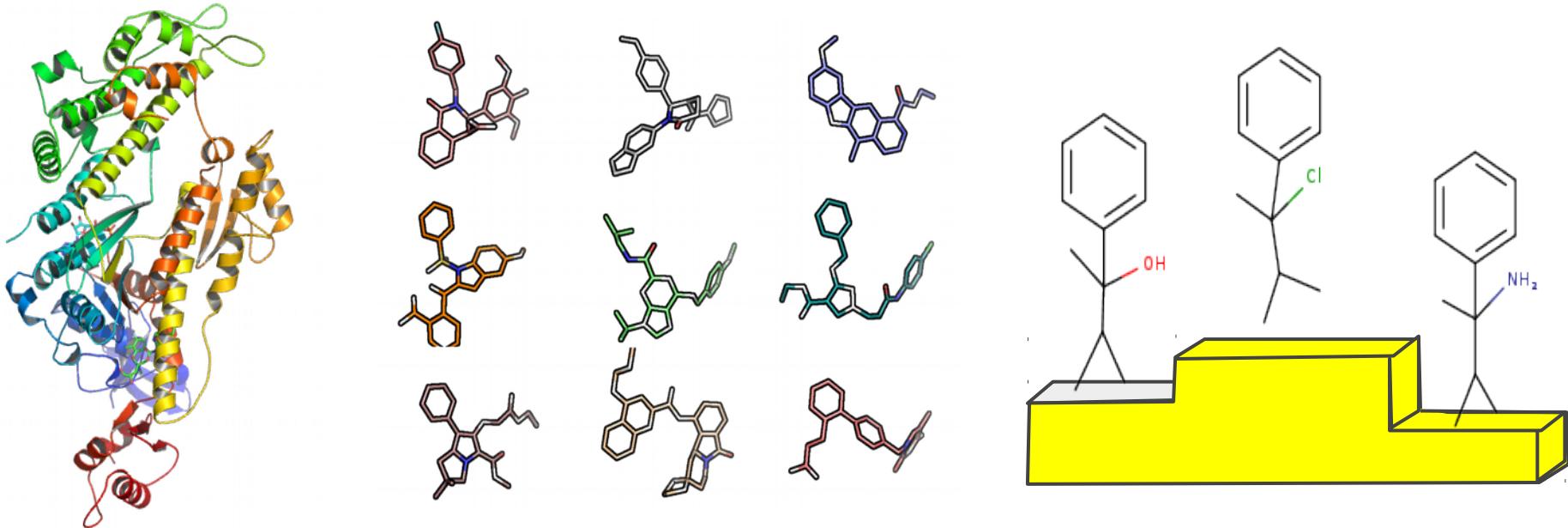
Animal numbers for the European Union 2008:

9.100.000 sacrifices

•••••
RENAMA

Compound Library for Virtual Screening Chemoinformatics

You're here



ChemFlow: From Smiles to “Free Energies”

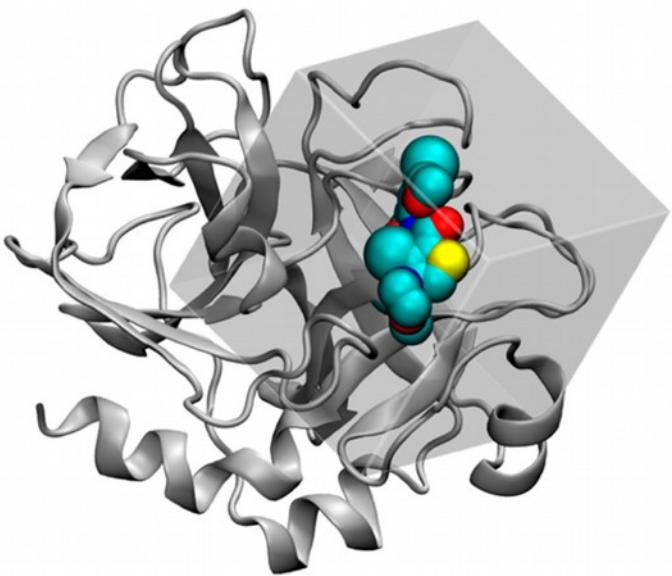
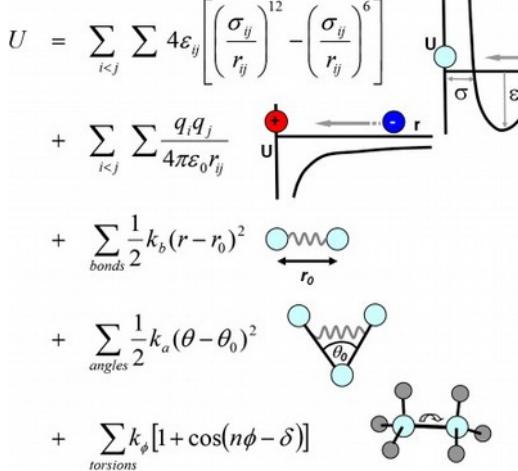


Image: Adv Drug
Del Rev

Image: 10.3390/molecules21111604

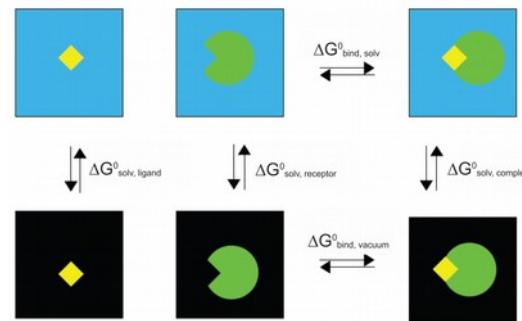
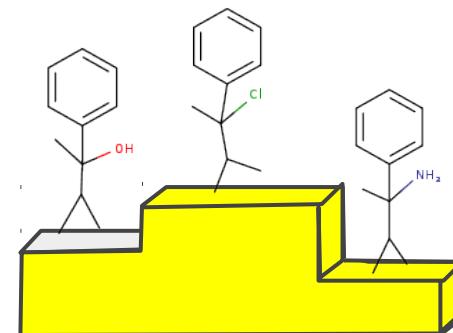


Image: ambermd.org



Main methods for predicting “ ΔG ” & computational demands

Quantitative Structure Activity

Relationship (QSAR)

$$\Delta G = k_0 + \sum k_i X_i \quad \Delta G = F(X) \quad (X \text{ is a descriptor})$$



Linear Interaction Energy (LIE)

$$\Delta G = \alpha \Delta \langle V_{t \text{ env}}^{\nu dW} \rangle + \beta \Delta \langle V_{t \text{ env}}^{\text{Elec}} \rangle$$

MM - Poisson-Boltzmann Surface area (MM-PBSA)

$$\Delta G = \langle \Delta G_{\text{Gas}} \rangle \langle \Delta G_{\text{PBSA}}^{\text{Desolv}} \rangle - T \Delta S$$

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamic integration (TI)

Non-equilibrium statistical mechanics

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

$$\Delta G = \int_0^1 \langle \frac{\partial V}{\partial \lambda} \rangle_\lambda d\lambda$$

Sample Design of a Compound Library for Virtual Screening

nature
International journal of science

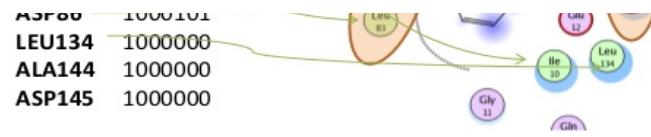
Article | Published: 06 February 2019

Ultra-large library docking for discovering new chemotypes

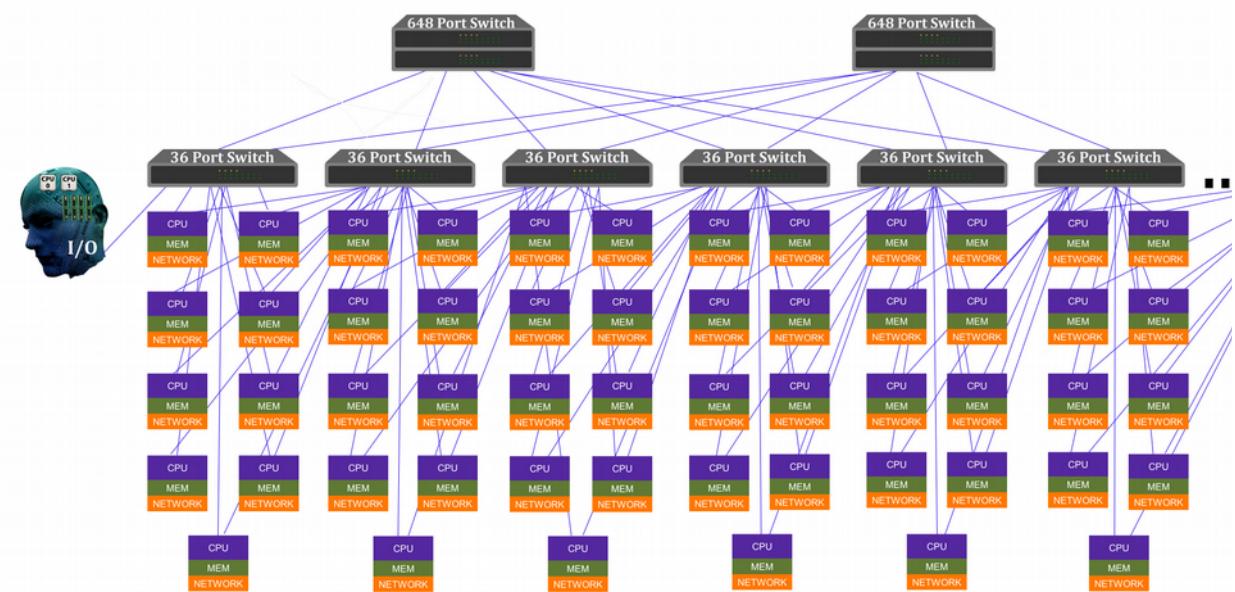
Jiankun Lyu, Sheng Wang, Trent E. Balias, Isha Singh, Anat Levit, Yurii S. Moroz, Matthew J. O'Meara, 8
Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, Andrey A. Tolmachev, Brian K. Shoichet ✉, Bryan
L. Roth ✉ & John J. Irwin ✉

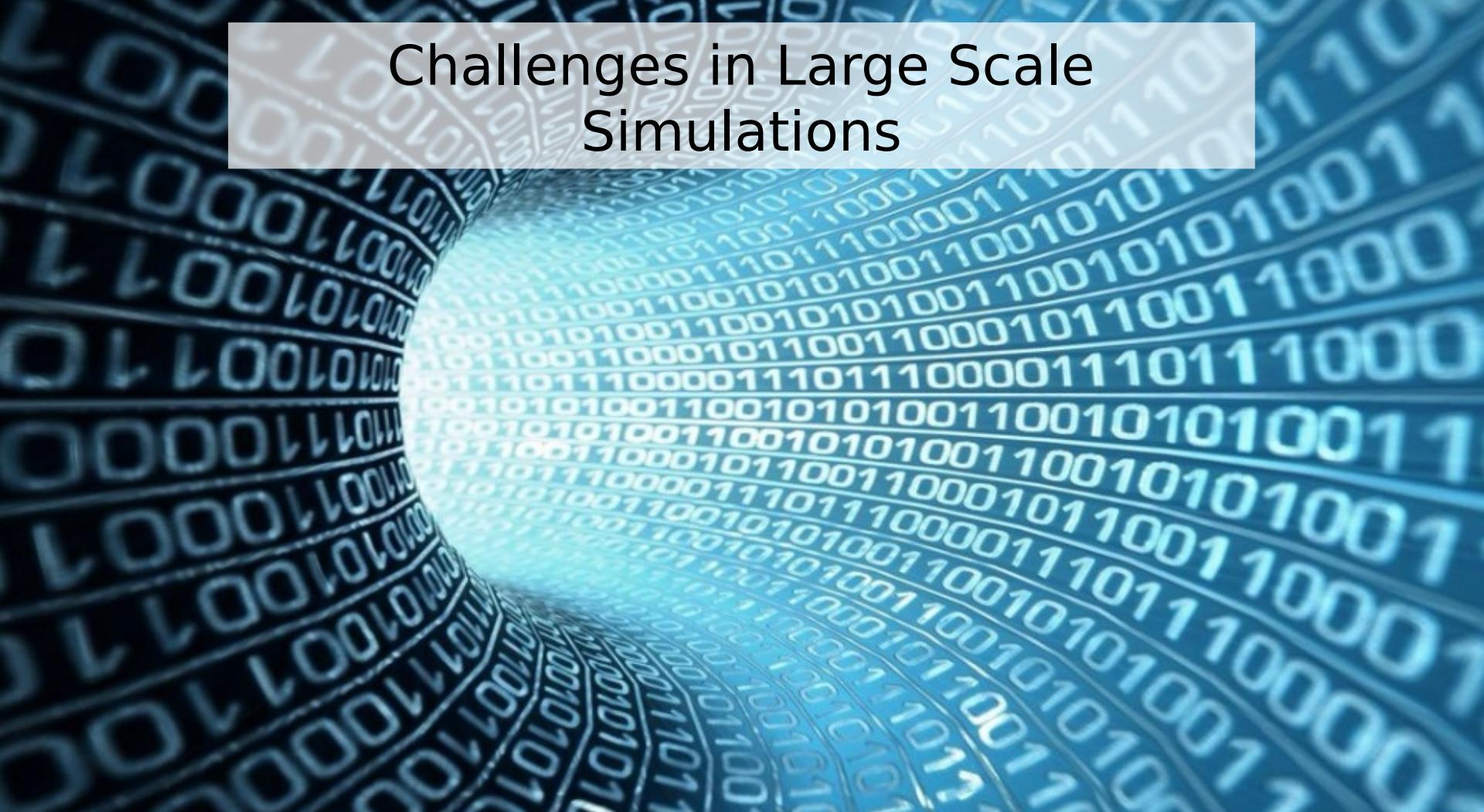
Nature **566**, 224–229 (2019) | Download Citation ↓

32k Accesses | **26** Citations | **276** Altmetric | Metrics ➞



Computational burden



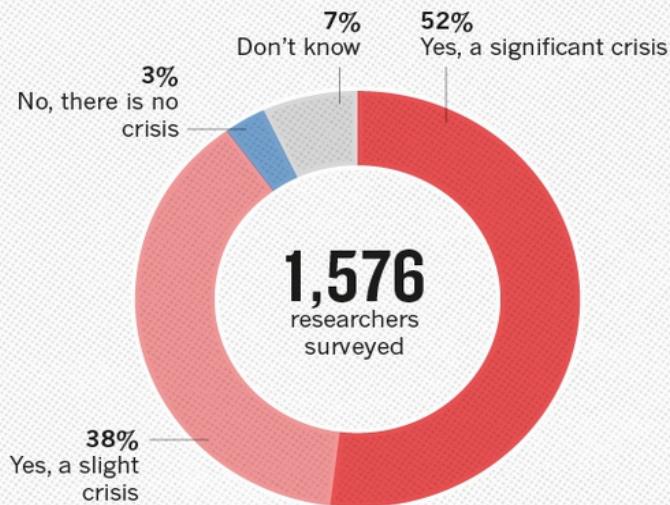


Challenges in Large Scale Simulations

- Complex workflows
- Heterogeneous components
- **Large data volumes and data rates**
- Data distribution, & transformations

No more excuses for non-reproducible methods

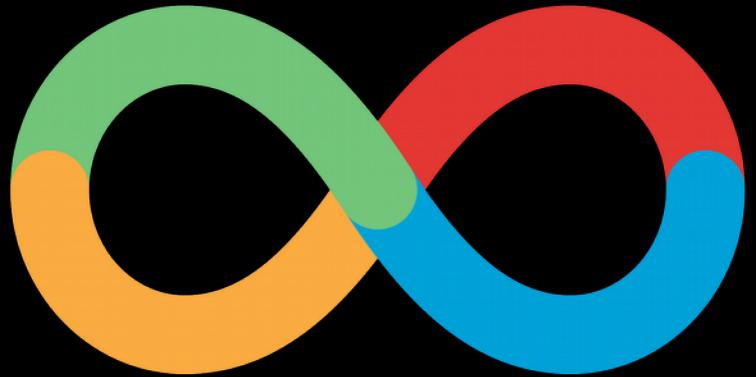
IS THERE A REPRODUCIBILITY CRISIS?



©nature







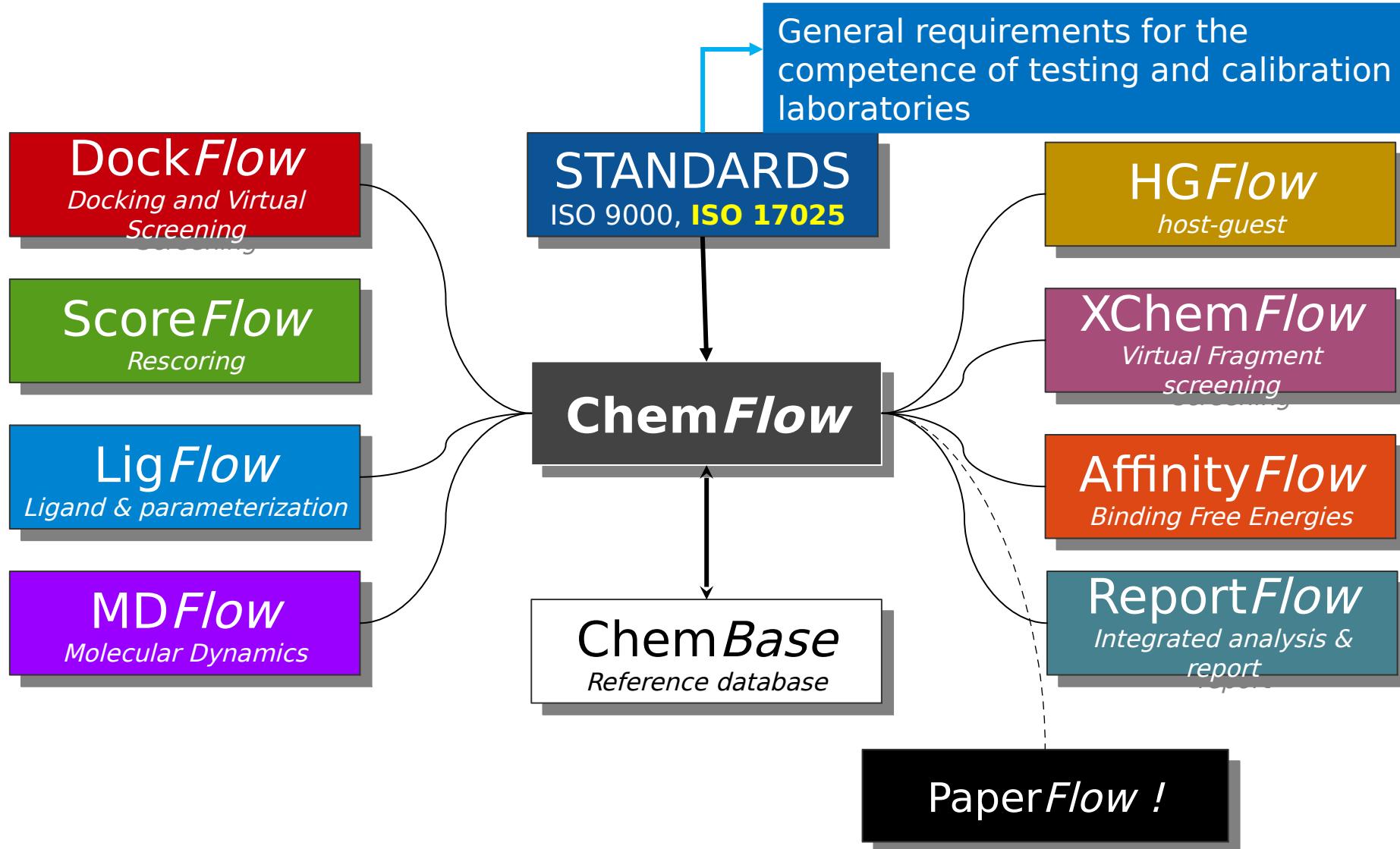
ChemFlow® WWDC 2018

Computational Chemistry is great
again !!!

•Diego Gomes | dgomes@pq.cnpq.br

•Université de Strasbourg / Instituto Nacional de Metrologia Qualidade e
Tecnologia

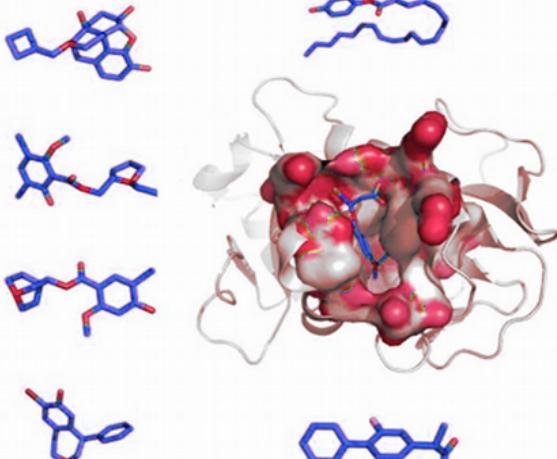
Ecosystem for computational chemistry



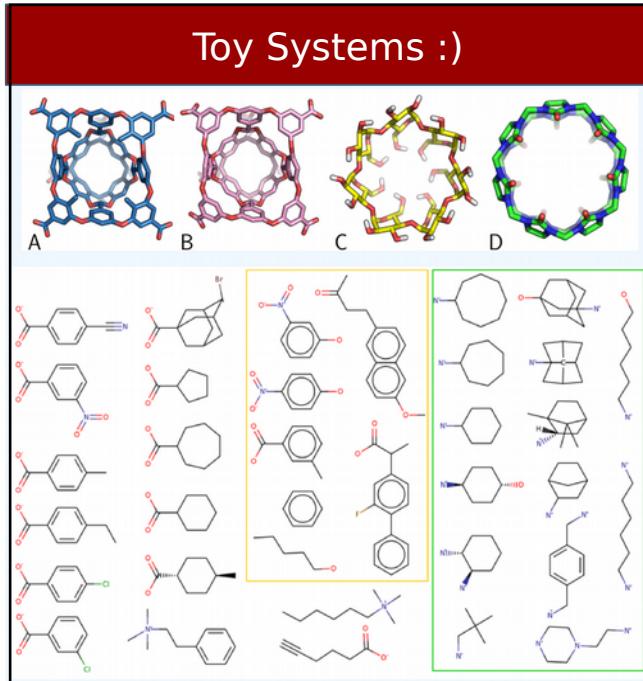
ChemFlow Scenarios

Current applications

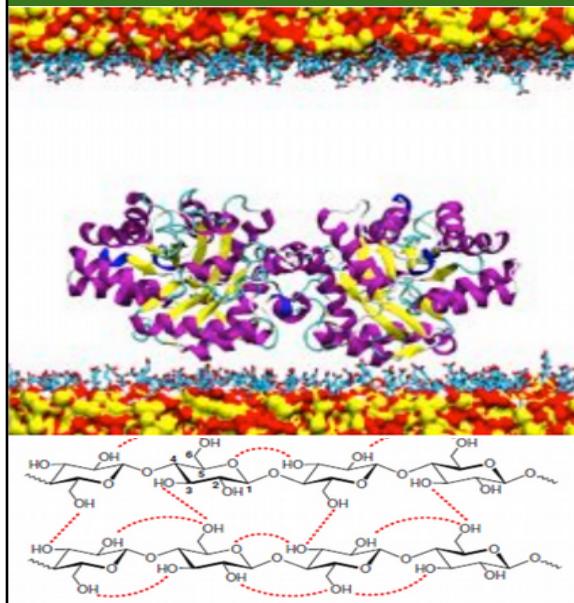
Drug Discovery & Design



Toy Systems :)



Design of Protein, Solvent...



Gomes, et al, J of Computer-Aided Molecular Design 2017

Hernandez, Hernandez, Gomes, Pascutti PLOS NTD 2018

Vieira, Gomes et al Adv Microb Biotec 2019

Hernandez, Hernandez , Gomes, Pascutti PLOS ONE 2019

Montalvo-Acosta, Gomes, et al, J. Phys. Chem. B, 2018

Oliveira, Costa, Gomes, (brazmedchem 2019)

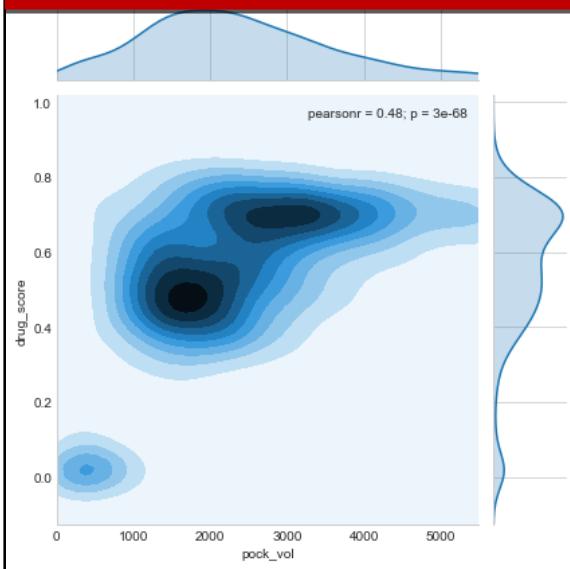
Fonseca,, Pascutti, **Gomes**, (In preparation)

Stoque, Sant'Anna, **Gomes**, (In preparation)

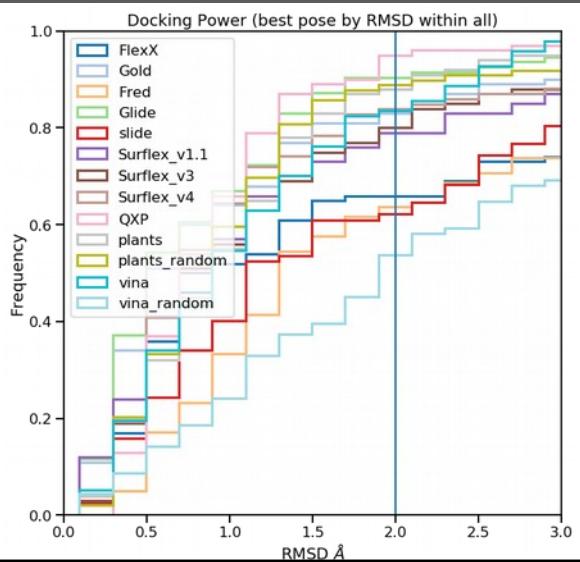
ChemFlow Scenarios

Current applications

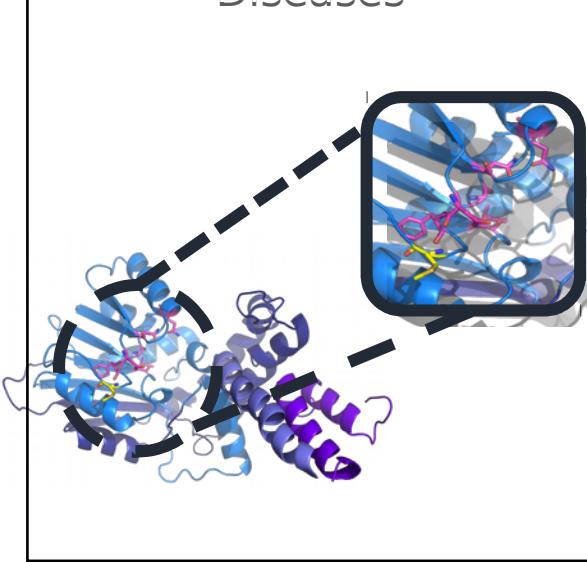
Pocket drugability



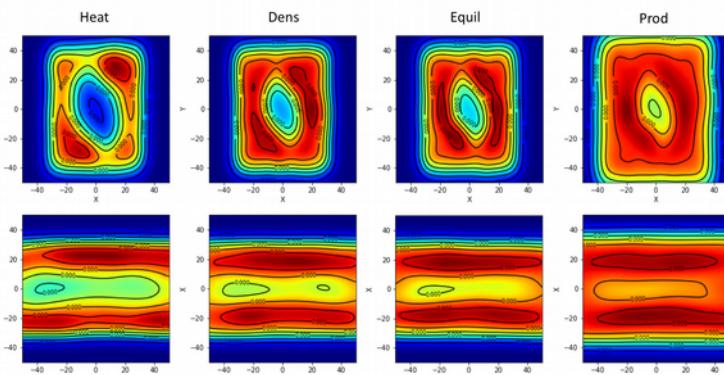
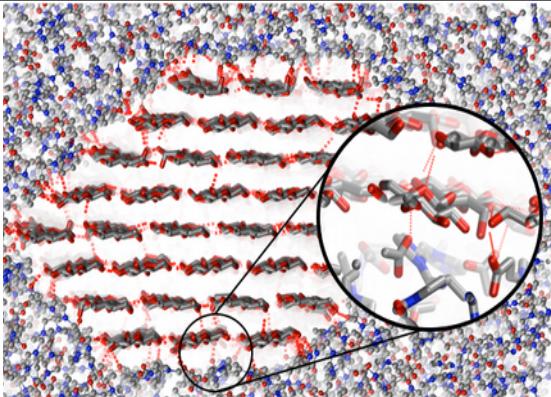
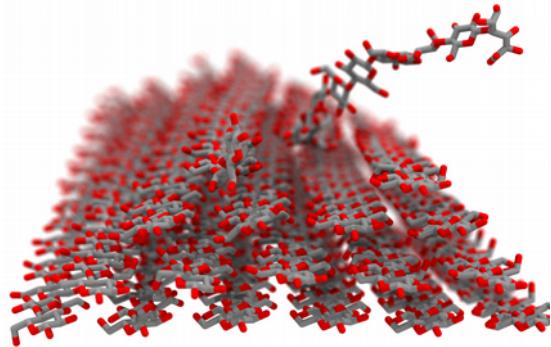
Protocol Validation



Drugs for Neglected Diseases

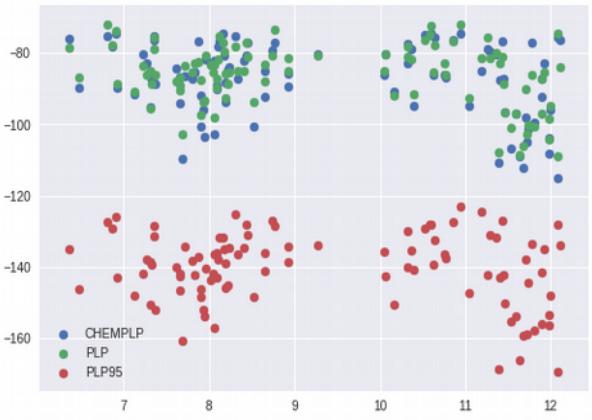


Improving biofuel production

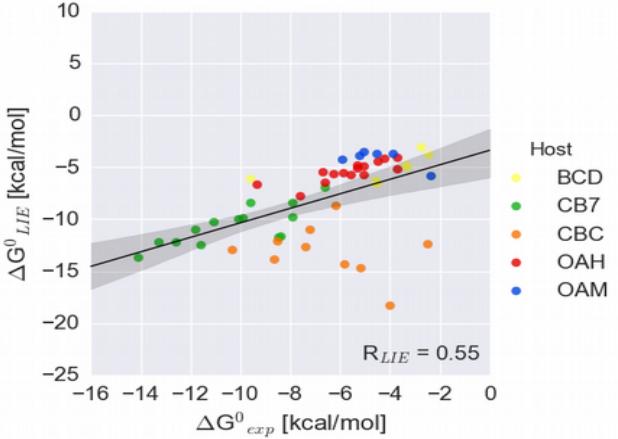


ReportFlow: An integrated analysis and reporting Notebook

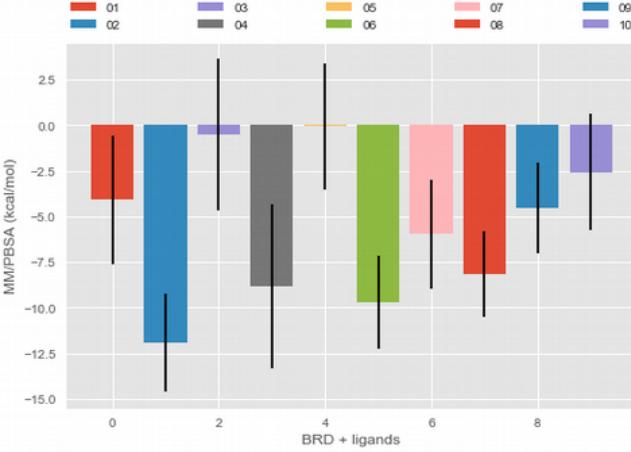
Compare docking campaigns, programs or energy functions



Compare methods



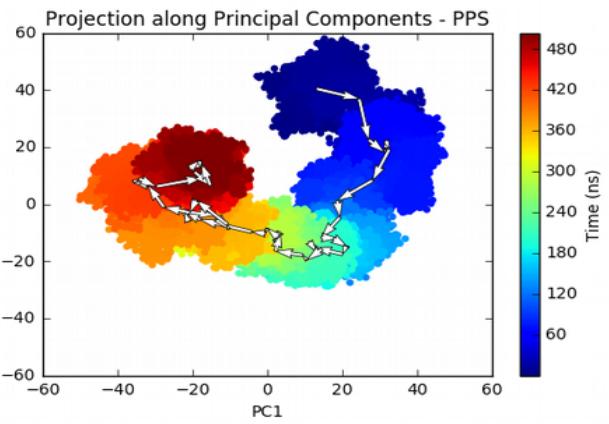
Compare binding energies



Comprehensive tables

PLANTS	ITERATION	RMSD	CHEMPLP	PLP	PLP95	ADT
-115.249	1	12.080773	-115.257	-109.0960	-169.263	-9.8537
-112.386	2	11.683341	-112.388	-106.1230	-159.160	-8.8748
-110.697	3	11.392603	-110.693	-107.7870	-168.639	-9.1044
-109.870	4	7.689499	-109.864	-102.9500	-160.784	-8.8748
-107.656	5	11.628859	-109.097	-108.7580	-165.954	-9.5625
-106.650	6	11.528613	-106.652	-101.1210	-155.193	-7.9395
-106.615	7	11.982867	-108.143	-104.2290	-156.259	-8.6876
-103.925	8	11.722159	-104.896	-102.8330	-158.884	-8.3565
-103.776	9	11.975485	-103.801	-98.3867	-153.541	-8.5412
-103.374	10	7.940215	-103.396	-93.3450	-153.786	-7.7961

Quick Results



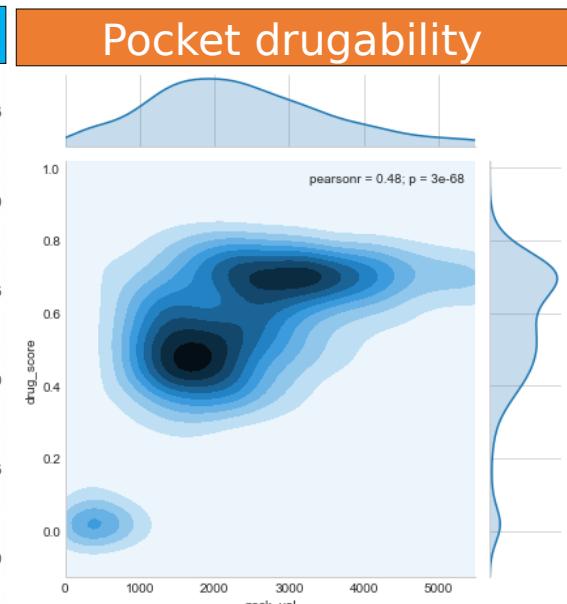
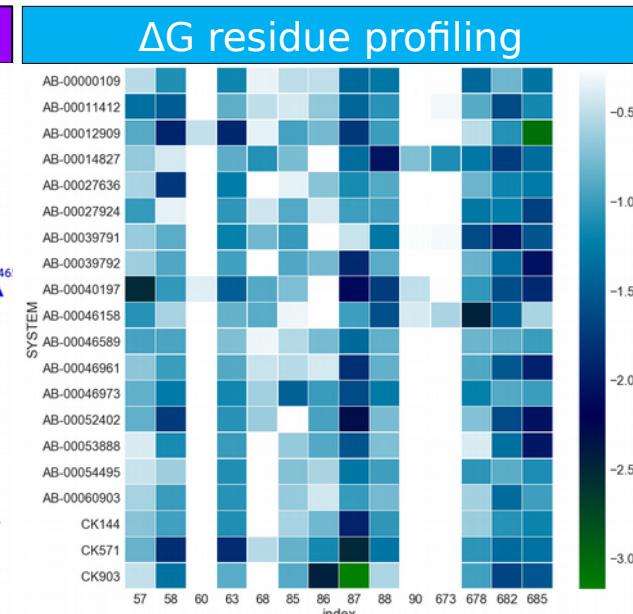
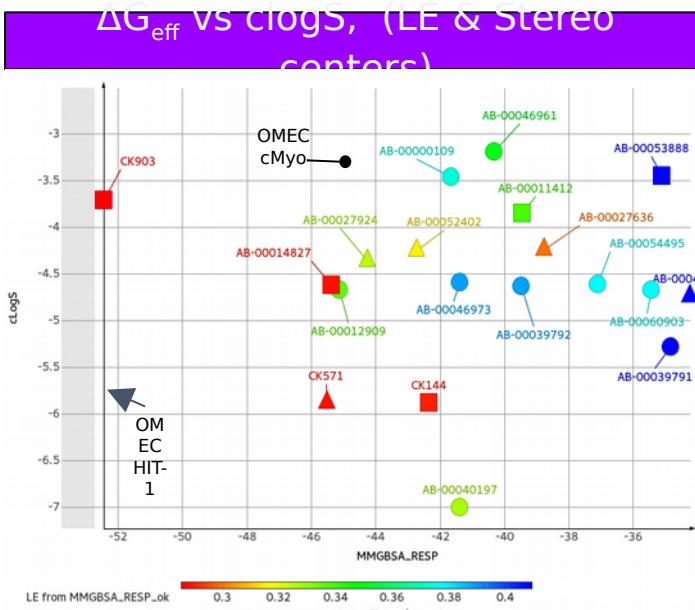
Interactive analysis

1 LIE vs MMPBSA Analysis Notebook

In [3]:	#Load Stuff									
In [7]:	#Read the data lie = pd.read_csv('LIE.csv') lie.columns = ['Complex', 'Group', 'exp', 'model', 'Host'] mmpbsa = pd.read_csv('MMPBSA1A.csv') mmpbsa.columns = ['Complex', 'Group', 'exp', 'model', 'Host'] energies = pd.concat([lie,mmpbsa],keys=['LIE','MMPBSA'],axis=1)									
Out[7]:	LIE MMPBSA									
Complex	Group	exp	model	Host	Complex	Group	exp	model	Host	
0	Oah-o01	2	-3.73	-4.08	OAH	cb7-c01	1	-9.90	-8.4000	CB7
1	Oah-o02	2	-5.90	-5.50	OAH	cb7-c02	1	-9.60	-7.0100	CB7
2	Oah-o03	2	-6.28	-5.63	OAH	cb7-c03	1	-6.60	-7.8500	CB7
3	Oah-o04	2	-6.72	-5.47	OAH	cb7-c04	1	-8.40	-7.7600	CB7
4	Oah-o05	2	-5.30	-4.83	OAH	cb7-c05	1	-8.50	-16.0400	CB7
5	Oah-o06	2	-5.60	-5.71	OAH	cb7-c06	1	-7.90	-8.9300	CB7
6	Oah-o07	2	-7.60	-7.78	OAH	cb7-c07	1	-10.10	-12.9600	CB7

ReportFlow

An integrated analysis and reporting Notebook



100% Free and Open
Source

ChemBase - Community service

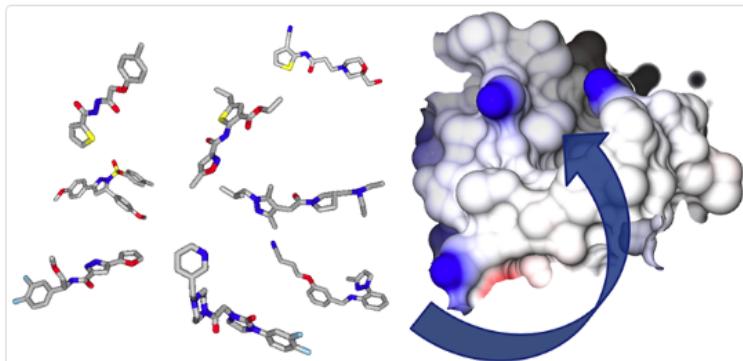
ChemBase

The ChemFlow database

ChemBase - The ChemFlow database

The ChemFlow Reference Simulation database (ChemBase) is an ongoing project whose aim is to provide well-documented simulation results for a variety of systems and from various simulation techniques. The results contained here are usually generated in-house at Laboratoire d'Ingénierie des Fonctions Moléculaires of the Université de Strasbourg - France, when certain criteria are satisfied, may also include results from provided from outside collaborators.

ChemBase is (probably going to be) a Web-based database with structural and dynamics analysis of small-molecule properties and molecular dynamics simulations.

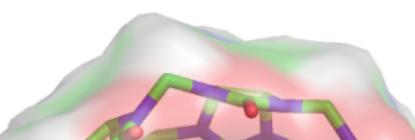


Chimiotèque Nationale du CNRS

ChemBase provides structures drug-like molecules from the Chimiotèque Nationale du CNRS, together computational predictions physico-chemical properties, ready-to-dock 3D structures and ready-to-simulate topologies and parameters. Parametrization is achieved for the AMBER/GAFF(1,2) forcefields both with AM1-BCC and RESP charges.

Host-Guest Systems

Host-Guest (HG) systems are ideally placed models of molecular recognition in solution with broad industrial



Chem**Flow** - Community service



Quimioteca Nacional

Início Cadastrar Moléculas Quimioteca Sobre C

Contato Acesso

br Qn Quimioteca National	B R A S I L	92 U Uranium 238.02891	2 8 18 32 21 9 2	53 I Iodine 126.90447	2 8 18 18 18 7 2	25 Mn Manganese 54.938045	2 8 13 2 2	53 I Iodine 126.90447	2 8 18 18 18 7 2	8 O Oxygen 15.9994	2 6 2	52 Te Tellurium 127.6	2 8 18 18 6 2	20 Ca Calcium 40.078	2 2 2 2 2 2 2
11 Na Sodium 22.98976928	2 8 1 2	6 C Carbon 12.0107	2 4 2 2	53 I Iodine 126.90447	2 8 18 18 18 7 2	8 O Oxygen 15.9994	2 6 2	11 Na Sodium 22.98976928	2 8 1 2	116 Lv Livermorium [293]	2 8 1 2 2 2 2				

Bem vindo à Quimioteca Nacional do Brasil

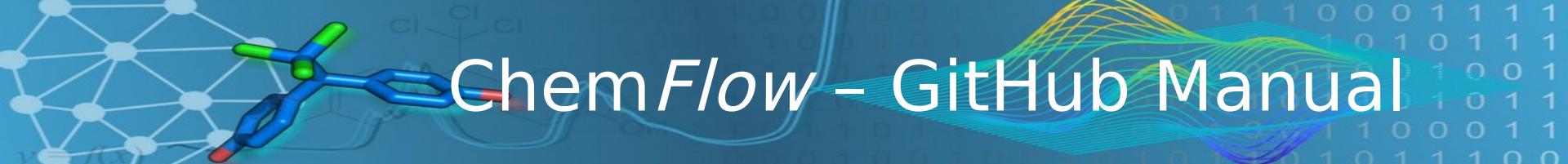
A Quimioteca Nacional é uma ferramenta que foi projetada para servir à comunidade científica com a missão de valorizar os produtos de síntese e extração, obtidos por grupos de pesquisa brasileiros. A força dessa iniciativa será a habilidade de conectar biólogos e químicos para o desenvolvimento de novos "hits" seu uso será um reaproveitamento direto dos recursos despendidos pelas agências de fomento no desenvolvimento dos compostos.

Financiamento

A Quimioteca Nacional é financiada principalmente pelo Edital CAPES Biologia Computacional implementado na parceria do INMETRO, UFRJ e UFMG, expandida para a UFJF e UNISTRA



ChemFlow - GitHub Manual



ChemFlow

Search docs

CONTENTS:

- ChemFlow
- Overview
- Workflows
- Features
- HPC Run
- Install
- User Manual

Tutorial

- ChemFlow - alpha-Thrombin
 - Provided files
 - LigFlow
 - DockFlow
 - ScoreFlow
 - Step 6.1: Run LigFlow to compute RESP charges.
 - Step 6.2: Run ScoreFlow to rescore the previous docking poses (best 3 for each ligand)
 - Step 7: Postprocess the results
 - Advanced
- To run DockFlow and ScoreFlow on a super computer

Contributing

Credits

Docs » Tutorial

[View page source](#)

Tutorial

ChemFlow - alpha-Thrombin

Copy the file "ChemFlow_tutorial_a-thrombin.tar.gz" present in ChemFlow/tutorial/ to place you want to run the tutorial.

Now extract with

```
tar xvfz ChemFlow_tutorial_a-thrombin.tar.gz
```

Now go to the folder and start playing :)

```
cd ChemFlow_tutorial_a-thrombin/
```

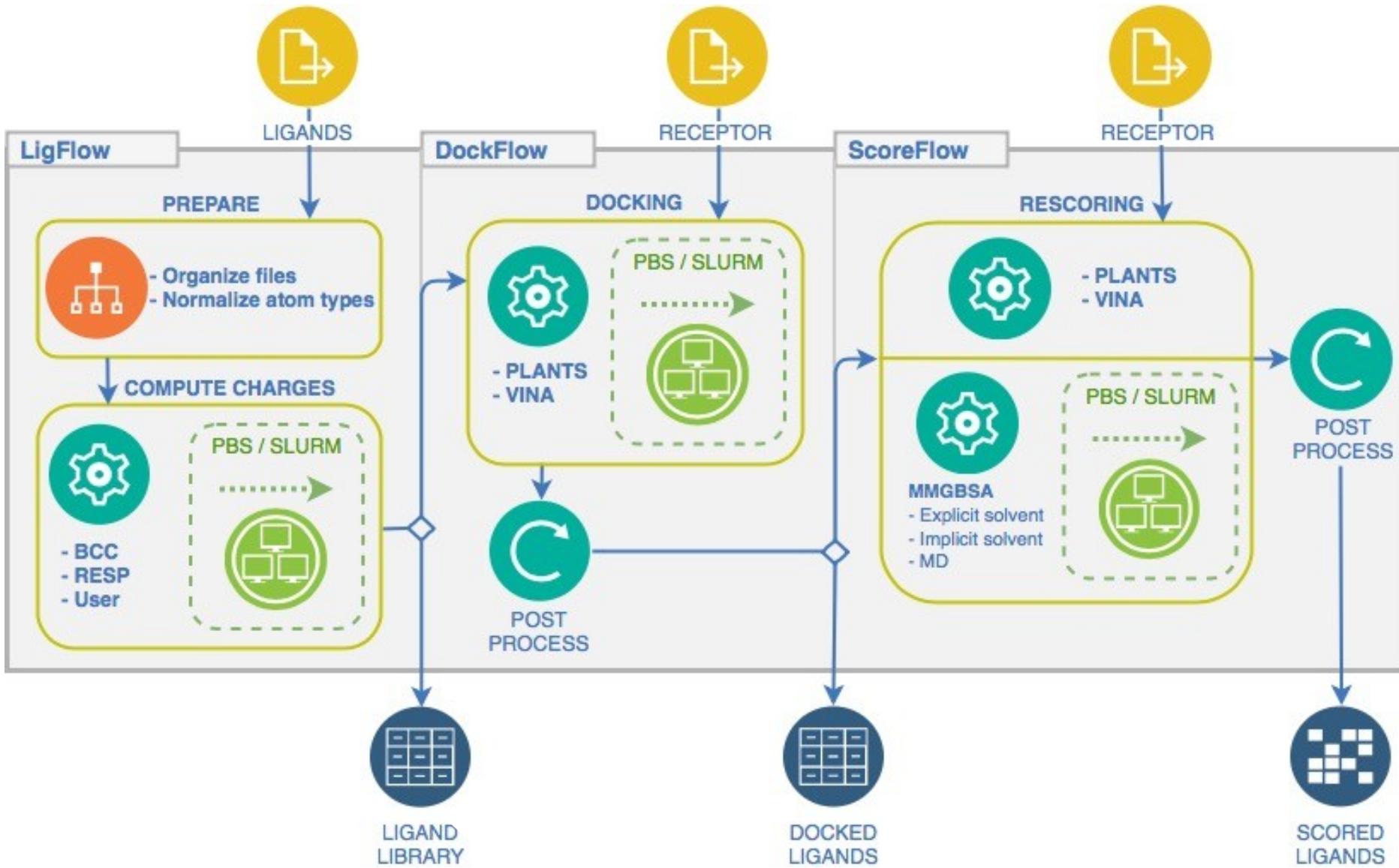
Provided files

1DWC.pdb	Original PDB
receptor.pdb	Receptor prepared with pdb4amber and -reduce.
receptor.mol2	Receptor prepared using SPORES.
reference_ligand.pdb	Ligand from 1DWC crystal structure.
reference_ligand.mol2	converted with openbabel.
ligands.smi	b1-b7 ligands.
ligands_crystal.smi	1D3D 1D3P 1D3Q 1D3T 1DWB 1DWC 1DWD
decoys.smi	decoys for a-thrombin, from DUD-E

LigFlow

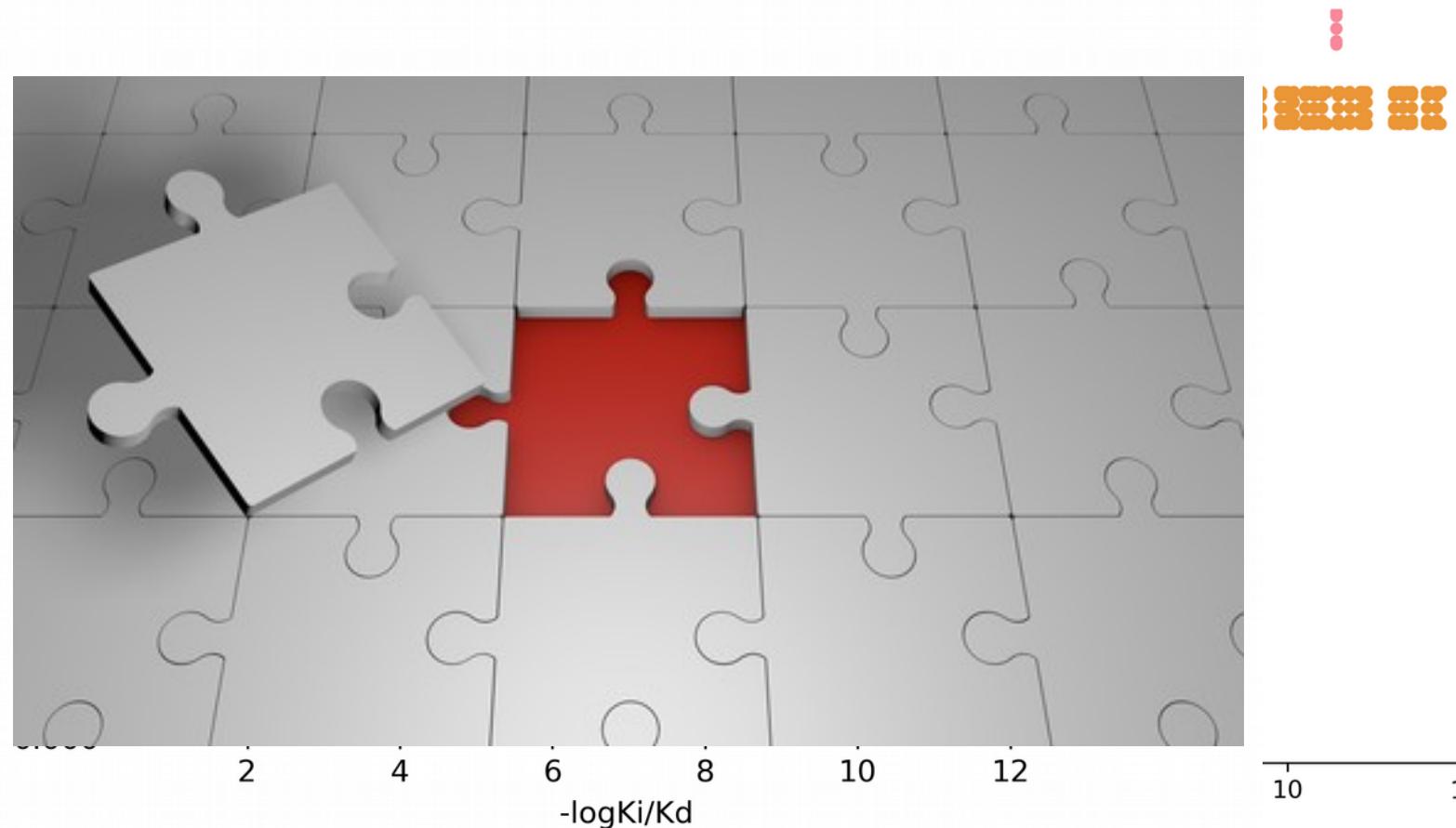
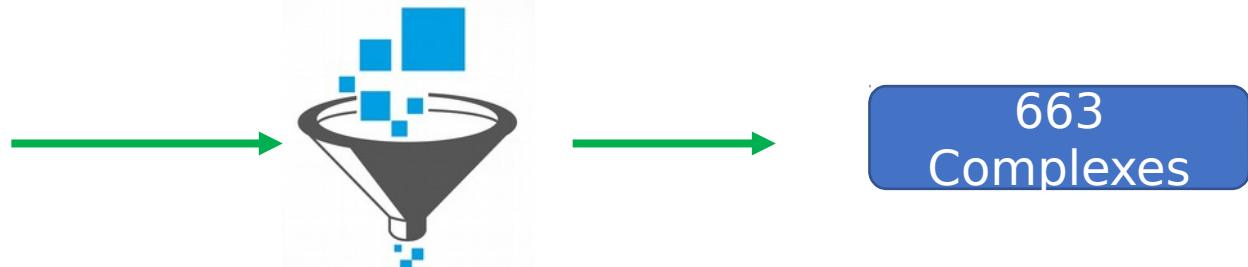
ChemFlow...

Summary for receptor-ligand systems



ChemFlow benchmark - Dataset

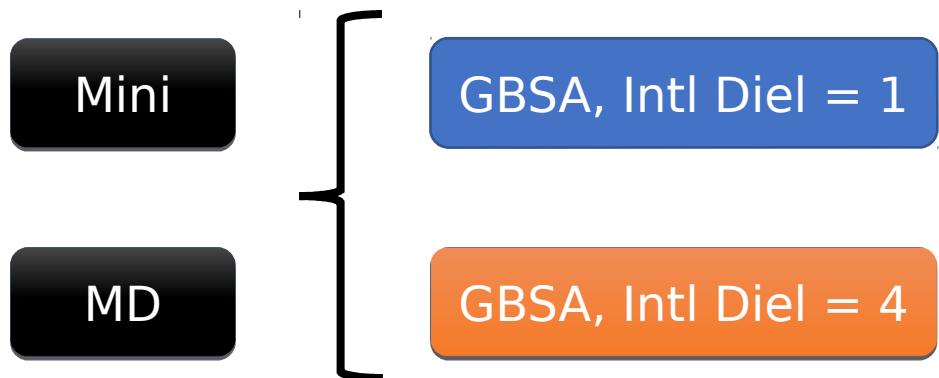
 NOVARTIS
858 complexes



Benchmark

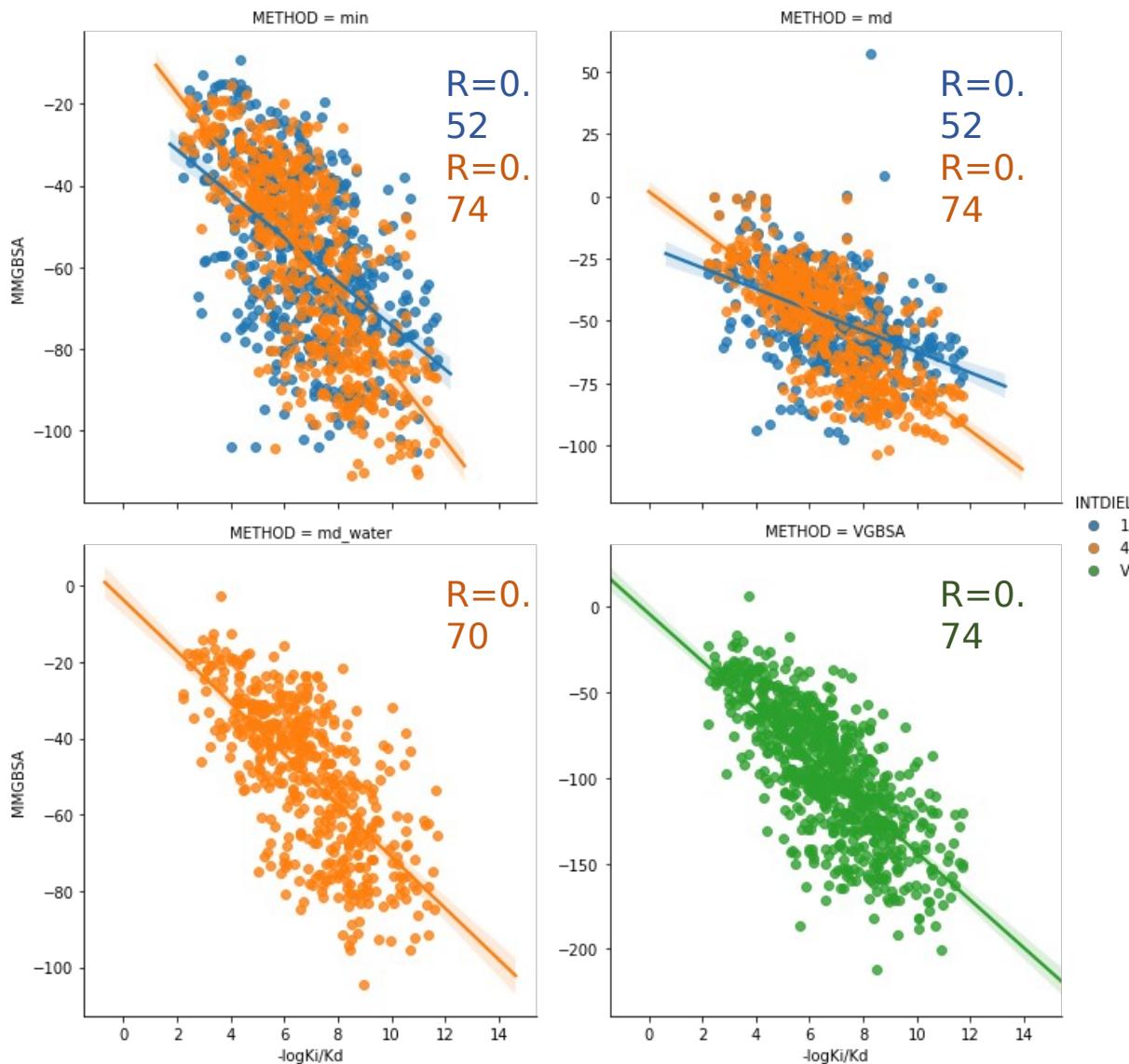
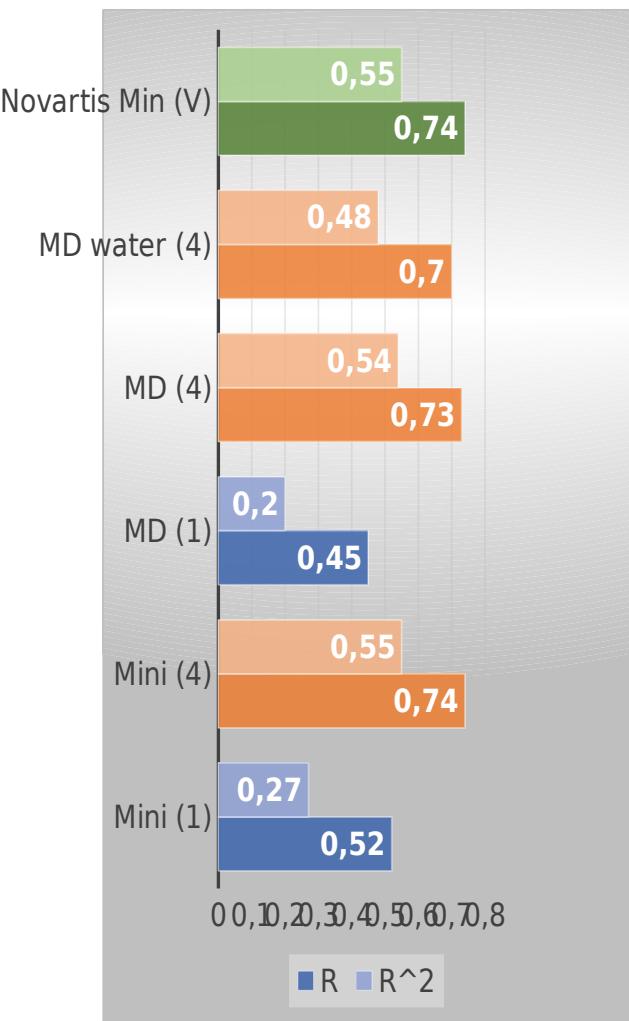


powered by
ChemFlow
663 Complexes



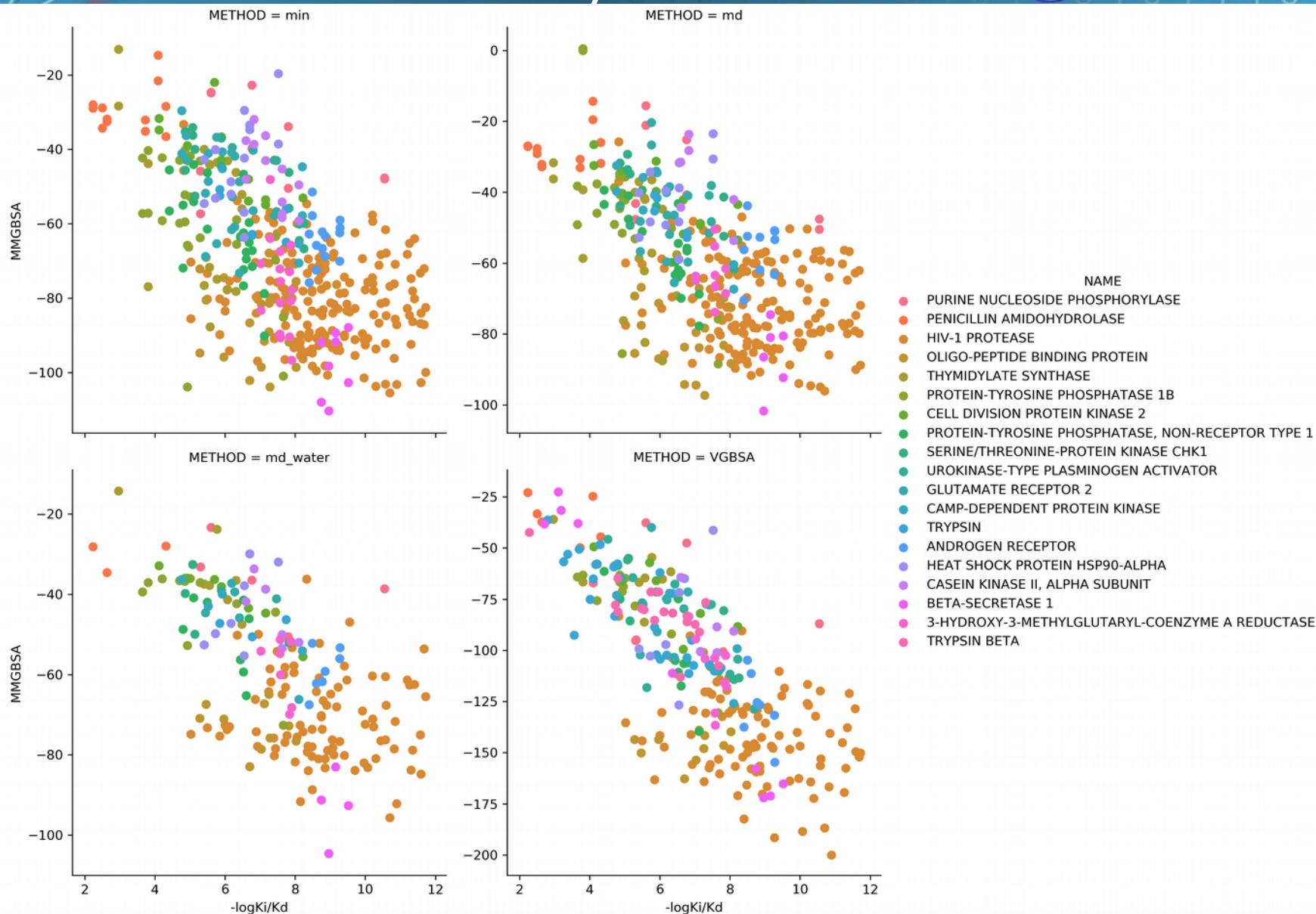


ChemFlow versus Novartis





ChemFlow versus Novartis Family issues

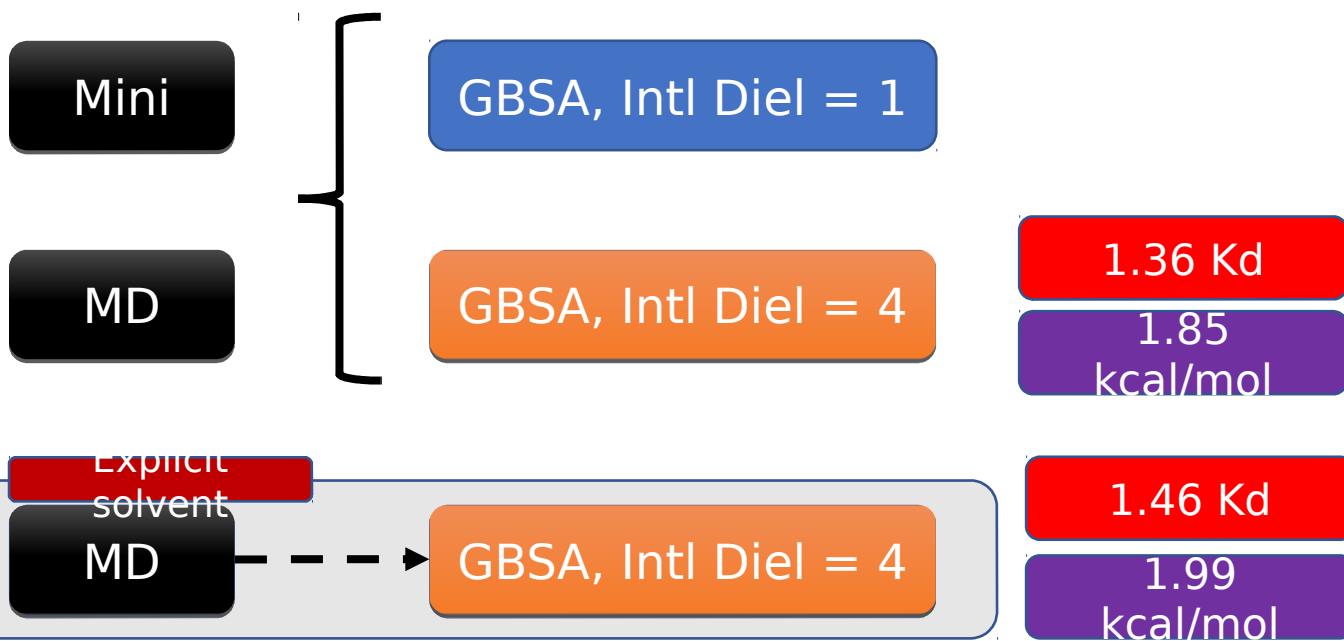


ChemFlow benchmark - Prediction quality

 NOVARTIS
663 complexes



powered by
ChemFlow
663 complexes



Smooth Muscle Myosin: Prioritization of modulating compounds

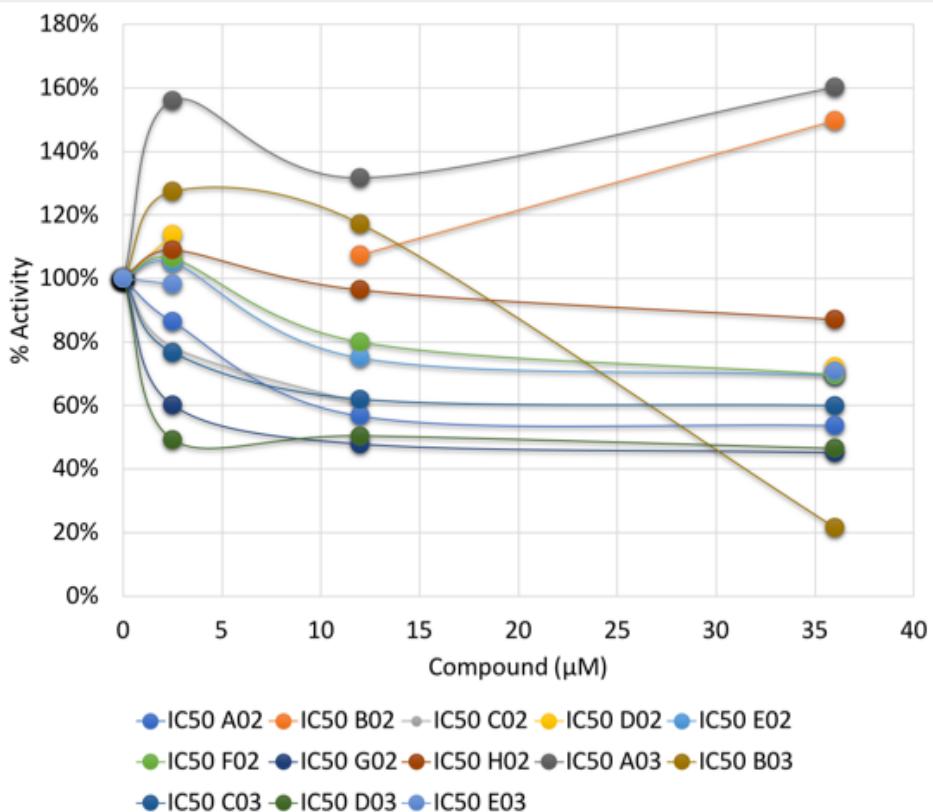


plate	compound	IC50 μ M
D03	AB-0004	2
G02	AB-0004	5
A02	AB-0002	20
C02	AB-0000	20
B03	AB-0005	30
C03	AB-0002	30
D02	AB-0001	100
E02	AB-0001	100
E03	AB-0006	100
F02	AB-0004	100
A03	AB-0003	weak activator
B02	AB-0003	weak activator
H02	AB-0004	not inhibiting

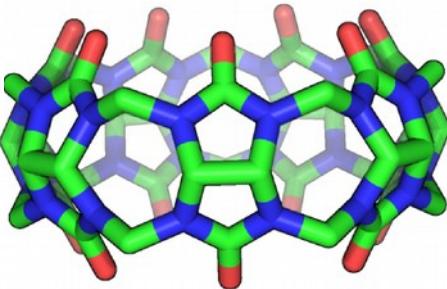
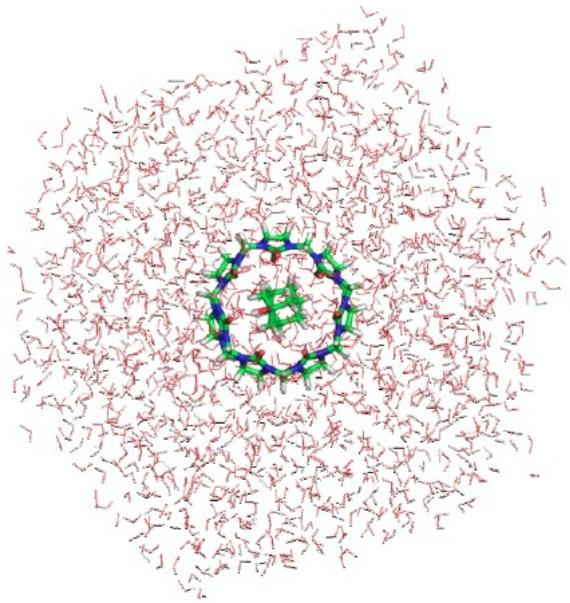
Confidential



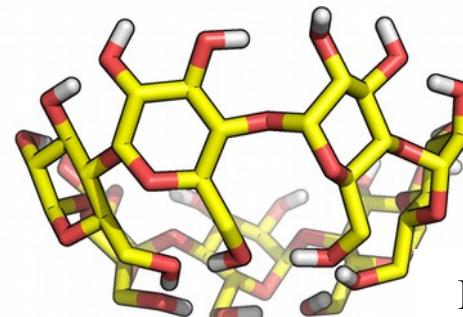
Linear Interaction Energy (LIE) Molecular Cages



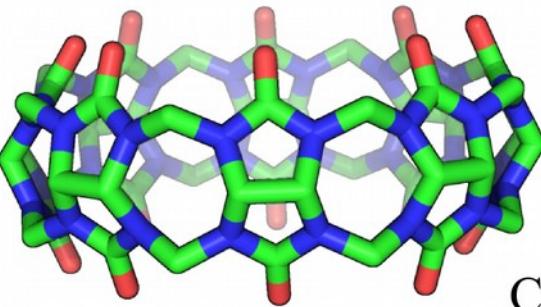
Drug Design Data Resource



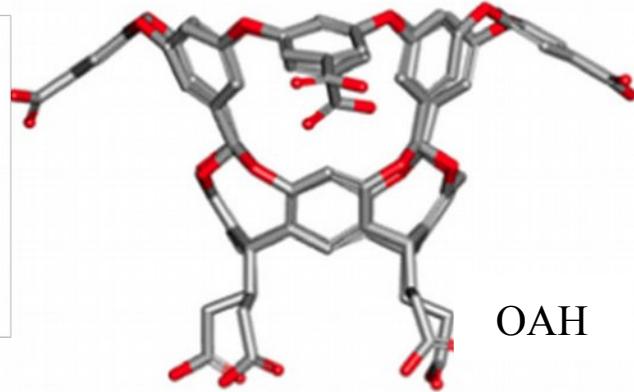
CB7



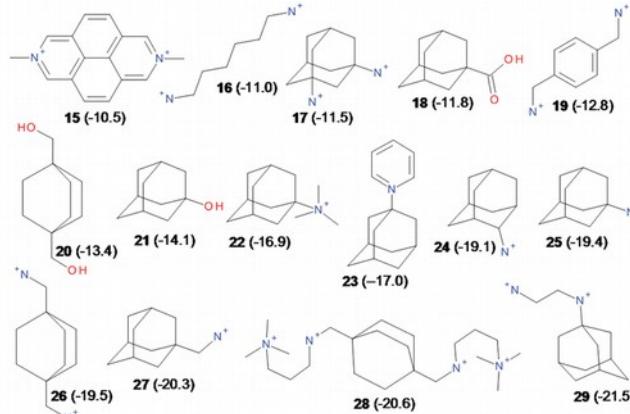
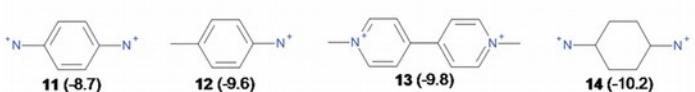
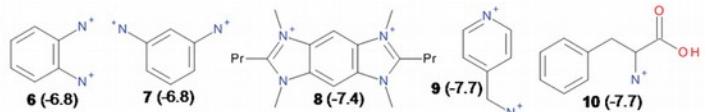
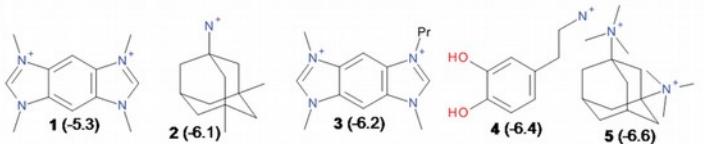
BCD



CB8

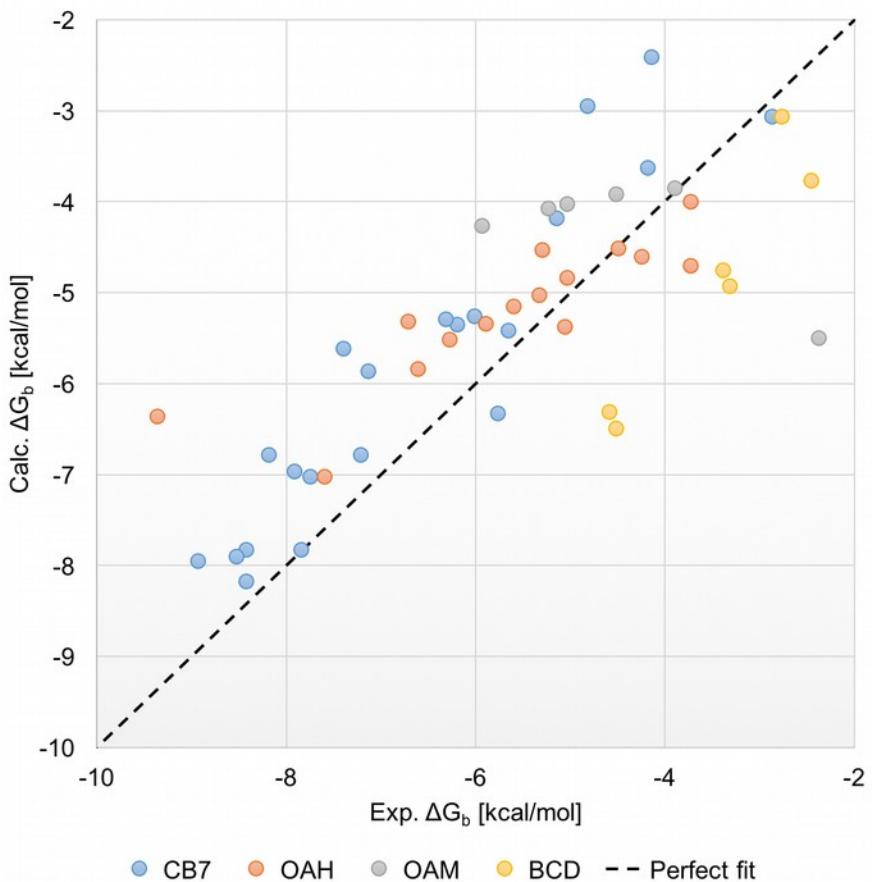


OAH



SAMPL4-6

Linear Interaction Energy (LIE) Molecular Cages



GAFF LIE model

$$\alpha = 0.43$$

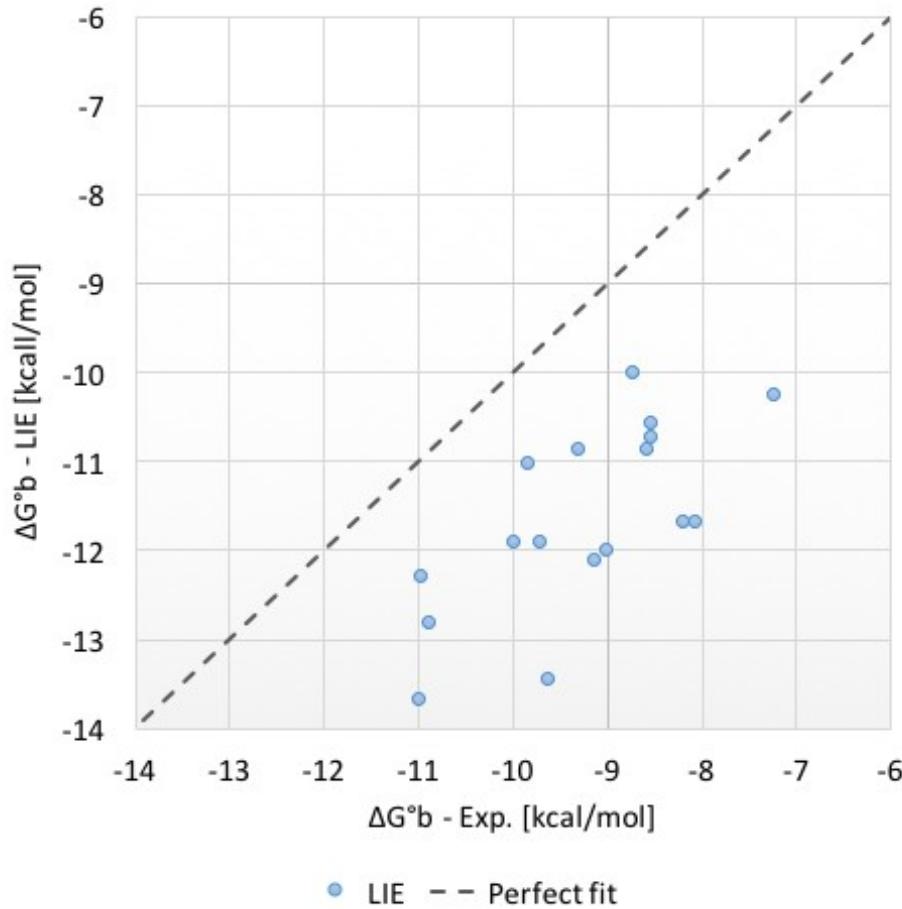
$$\beta = 0.20$$

Host Family	RMSE	R	Slope
Overall	1.08	0.81	0.72
CB7	1.17	0.92	1.05
OAH	0.66	0.88	0.56
OAM	1.06	0.98	0.20
BCD	1.48	0.95	1.47

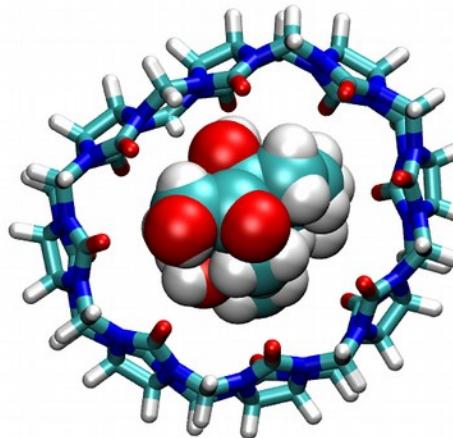
Best results from SAMPL4-5 challenges			
Host Family (Method)	RMSE	R	Slope
CB7 (SIE)	2.60	0.77	0.19
OAH/OAM (APR)	1.60	0.90	1.40
Best results from the Hydrophobe challenge			
CB7 (APR)	5.05	0.92	2.60
CB7 (QM/RRHO)	1.94	0.99	0.547

Linear Interaction Energy (LIE) Molecular Cages

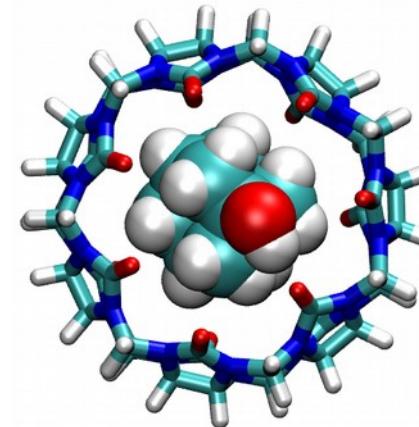
A Systematic or Model
error ???



	RMSE	R	Slope
LIE Model	2.45	0.69	0.68



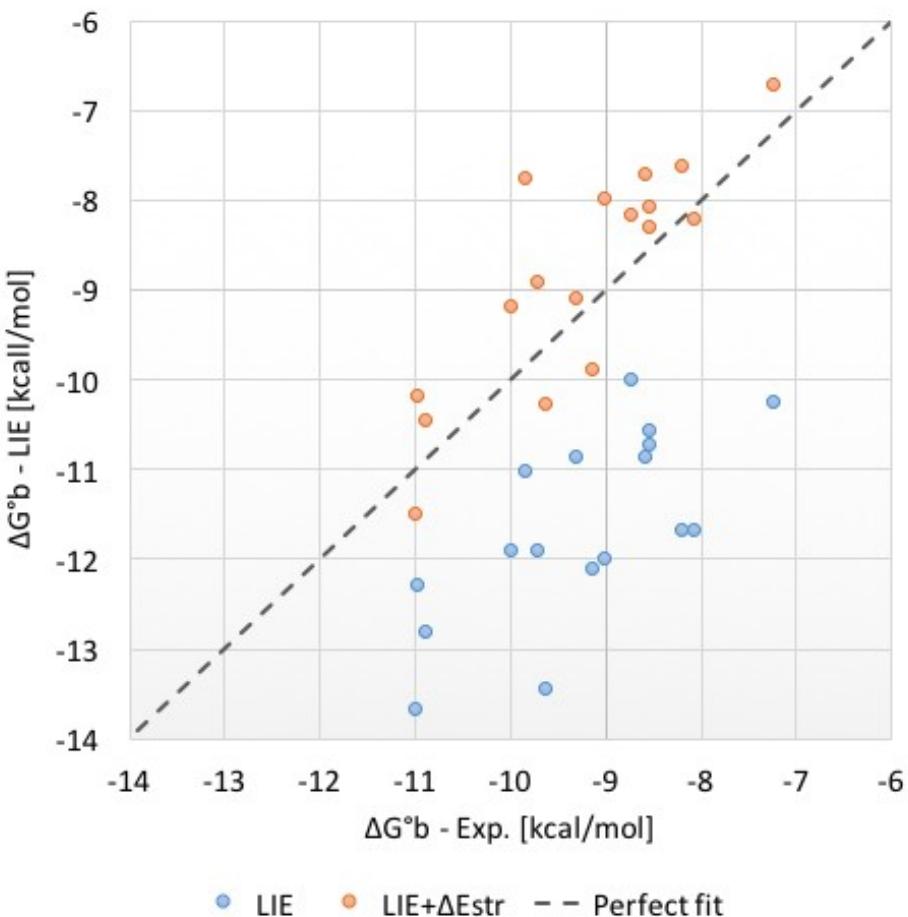
CB8 -
Prednisolone
RMSD=0.4~0.
6 Å



CB7 - 1-
Adamanthol
RMSD=0.1~0.
3 Å

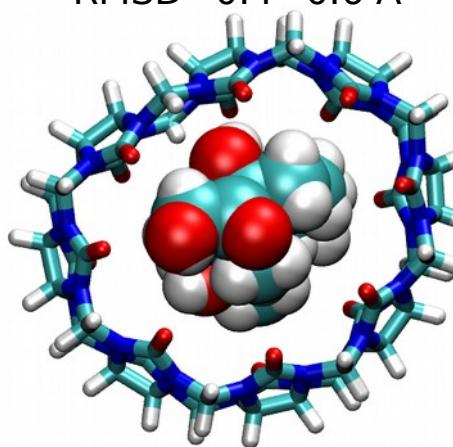
Linear Interaction Energy (LIE) Molecular Cages

$$\Delta G_b^\circ = \beta [\langle U_{L-s}^{elec} \rangle_b - \langle U_{L-s}^{elec} \rangle_{ub}] + \alpha [\langle U_{L-s}^{vdw} \rangle_b - \langle U_{L-s}^{vdw} \rangle_{ub}] + \boxed{\Delta E_{str}}$$

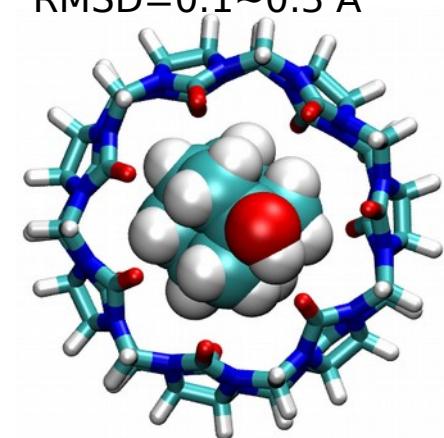


	RMSE	R	Slope
LIE model	2.53	0.69	0.68
LIE+ ΔE_{str} model	0.81	0.82	0.77

CB8 -
Prednisolone
RMSD=0.4~0.6 Å

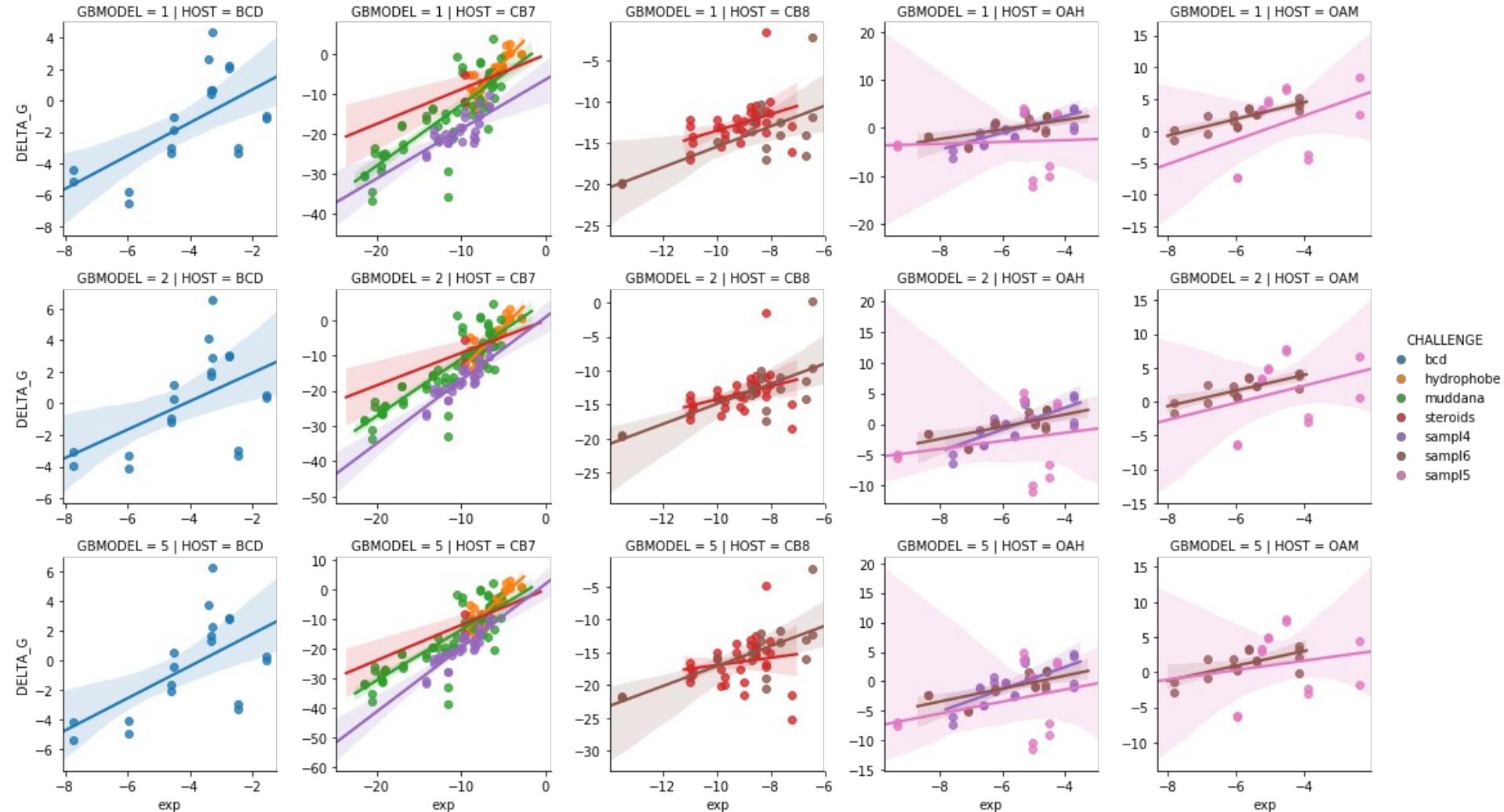


CB7 - 1-
Adamanthol
RMSD=0.1~0.3 Å

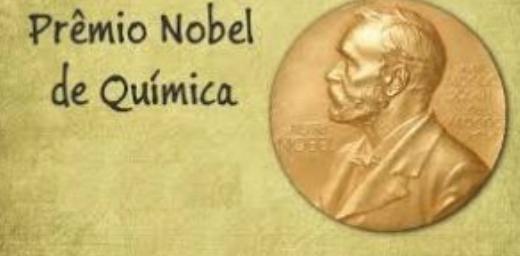


Linear Interaction Energy (LIE) Molecular Cages

ChemFlow + ReportFlow enables systematic evaluation of parameters



Thanks to ChemFlow Team & Advisors



2016

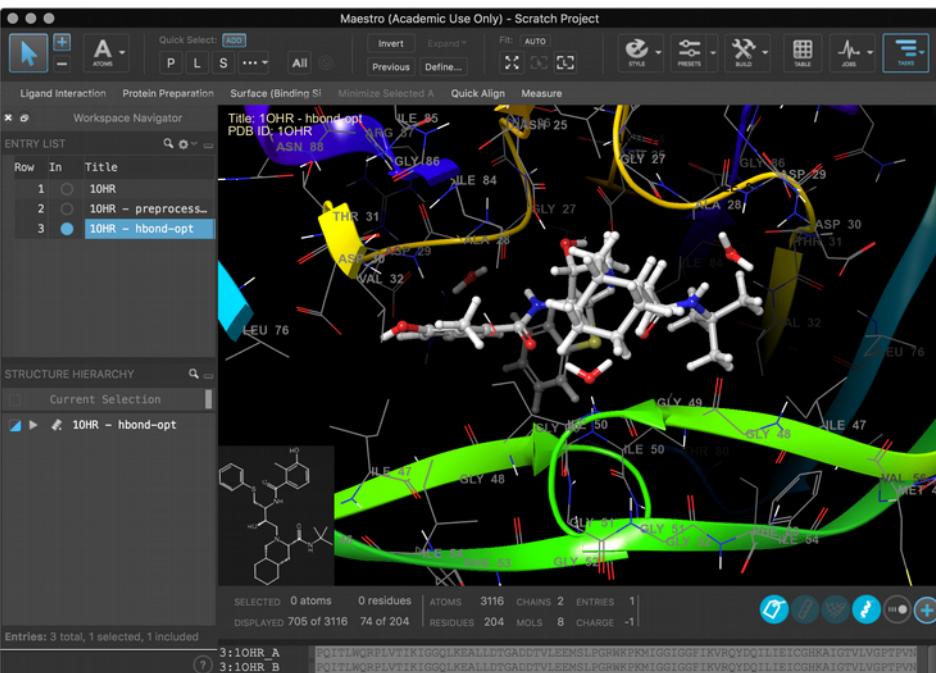
1987

2013



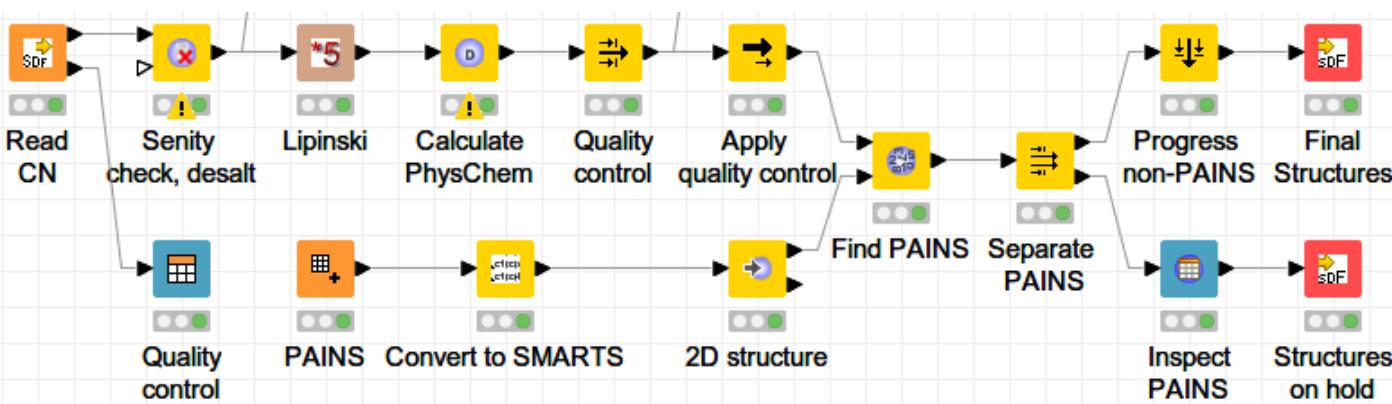
Workflows

WorkFlow Nutella

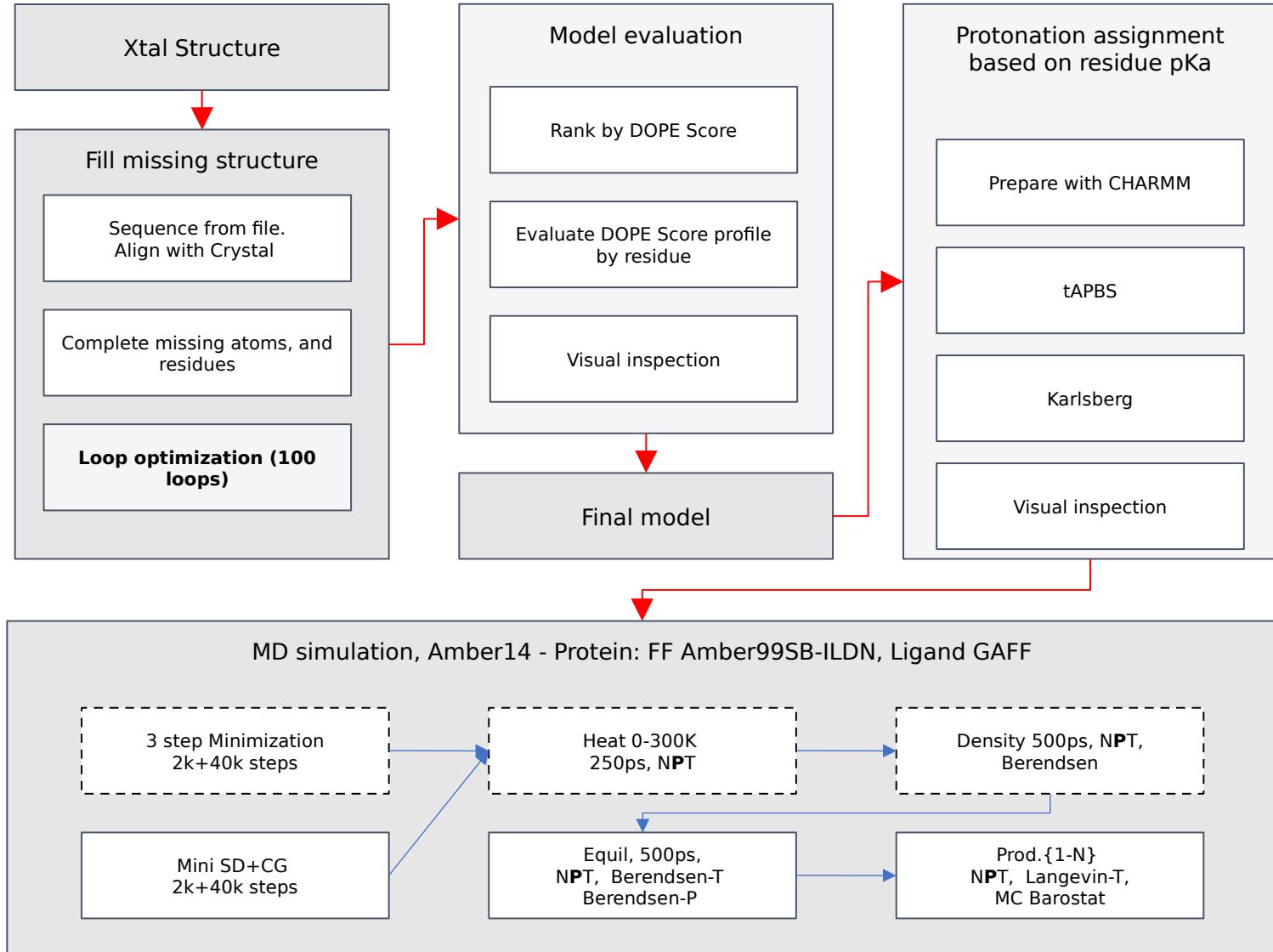


WorkFlow Raíz

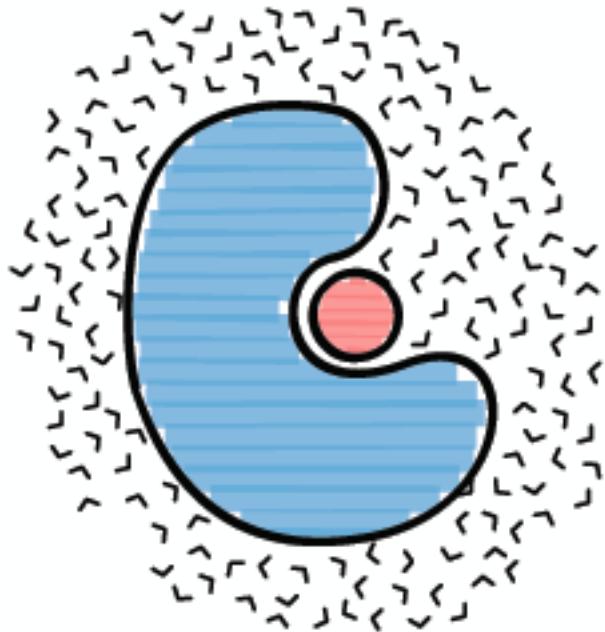
2979875	pri2015gp	SMM2_AB-00052350	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n444
2979814	pri2015gp	SMM2_AB-00012354	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979834	pri2015gp	SMM2_AB-00020176	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979854	pri2015gp	SMM2_AB-00023101-1	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979874	pri2015gp	SMM2_AB-00023101-2	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979895	pri2015gp	SMM2_AB-00023101-3	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979915	pri2015gp	SMM2_AB-00023823	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979935	pri2015gp	SMM2_AB-00025644	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979955	pri2015gp	SMM2_AB-00026534-1	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n443
2979975	pri2015gp	SMM2_AB-00027409	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n444
2979995	pri2015gp	SMM2_AB-00027413	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n444
2980015	pri2015gp	SMM2_AB-00029523	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n444
2980035	pri2015gp	SMM2_AB-00037702	_MD_prod.1	dbarreto	R	1:04:16	1	hpc-n444
2979754	pri2015gp	SMM2_AB-00009697	_MD_prod.1	dbarreto	R	1:04:29	1	hpc-n448
2979774	pri2015gp	SMM2_AB-00010406	_MD_prod.1	dbarreto	R	1:04:29	1	hpc-n448
2979794	pri2015gp	SMM2_AB-00012233	_MD_prod.1	dbarreto	R	1:04:29	1	hpc-n448
2979654	pri2015gp	SMM2_AB-00001506	_MD_prod.1	dbarreto	R	1:04:40	1	hpc-n447
2979674	pri2015gp	SMM2_AB-00007267	_MD_prod.1	dbarreto	R	1:04:40	1	hpc-n447
2979694	pri2015gp	SMM2_AB-00007268-2	_MD_prod.1	dbarreto	R	1:04:40	1	hpc-n447
2979714	pri2015gp	SMM2_AB-00007268-7	_MD_prod.1	dbarreto	R	1:04:40	1	hpc-n447
2979734	pri2015gp	SMM2_AB-00007268-8	_MD_prod.1	dbarreto	R	1:04:40	1	hpc-n448



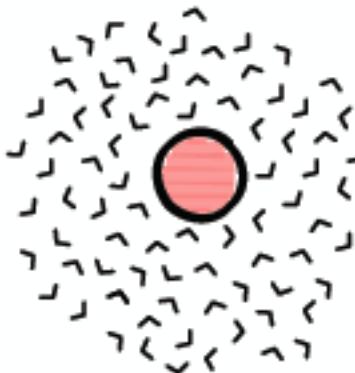
MD simulation workflow



Linear Interaction Energy (LIE)



Solution



$$\begin{aligned}\alpha &= 0.16 \\ \beta &= 0.50\end{aligned}$$

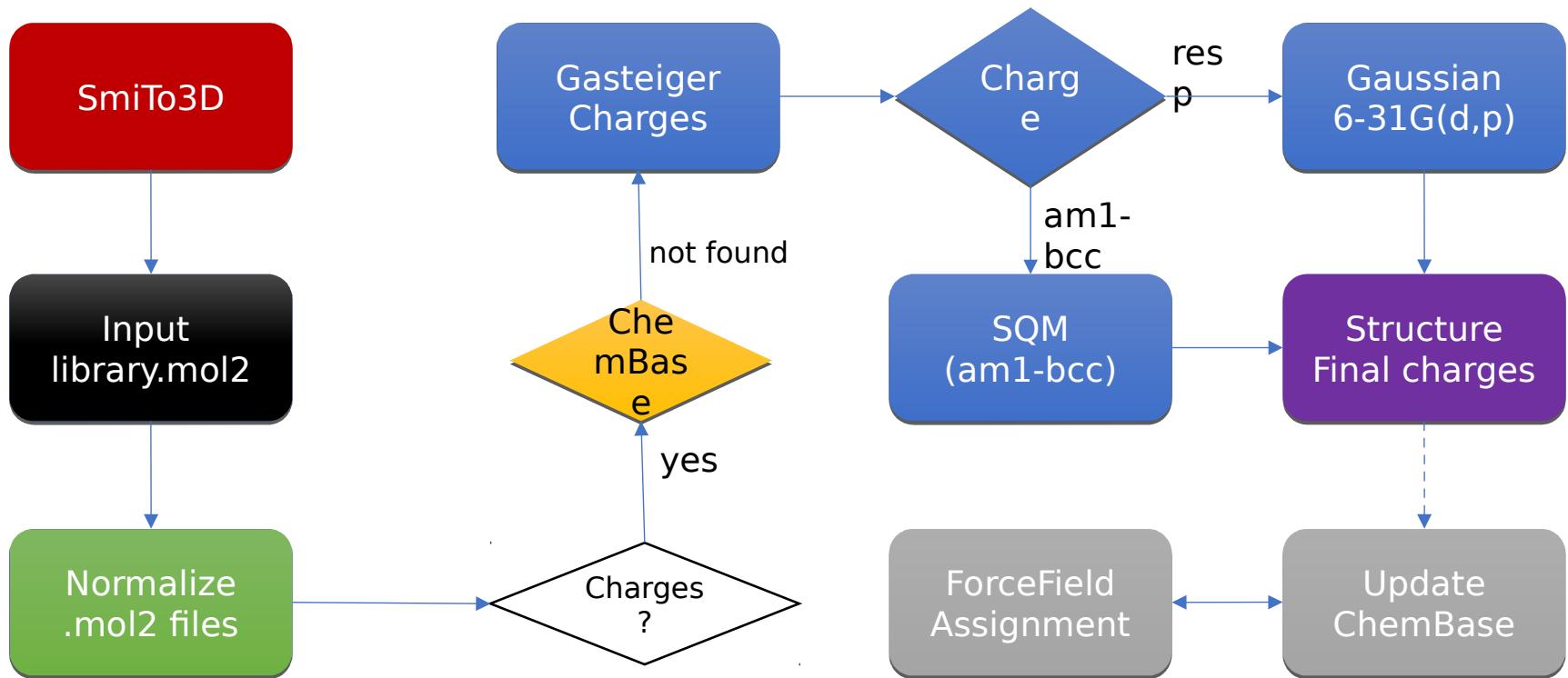
$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{vdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{\text{Elec}} \rangle$$

$$K_{eq} = \frac{Z_L(\text{site})}{Z_L(\text{bulk})}$$

Considerations:

- Infinite Dilution
- Small ligand with respect to the receptor
- Negligible conformational change of the receptor upon binding

LigFlow - Compound handling and parameterization



Smi - > mol2

- 1D to 3D
- Protonate pH 7
- Optimize (ETKDG)

Normalize .mol2 files

- Split .mol2
- **Adjust atomtypes**
- Deduplicate (docking confs)

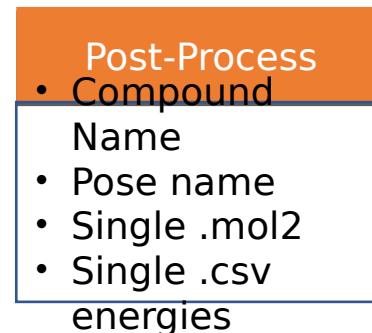
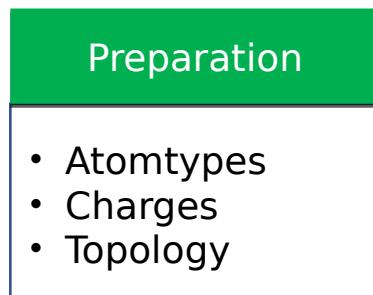
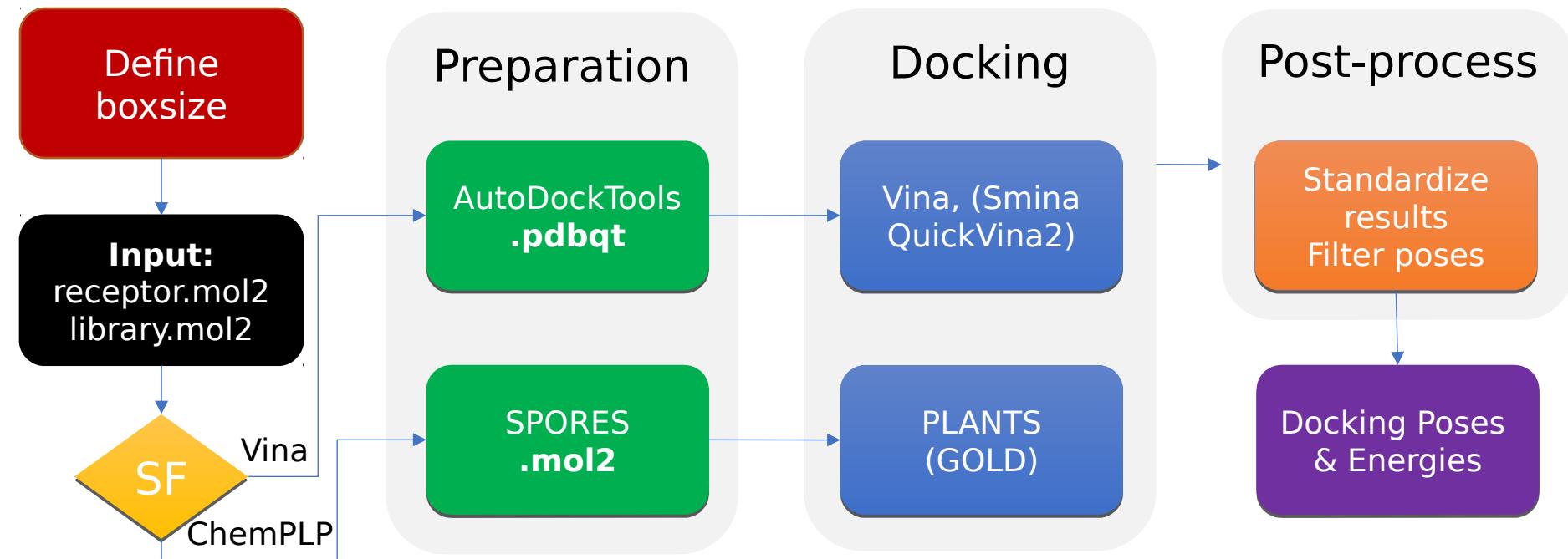
Charge/QM

- AM1-BCC & RESP
- Opt geometries
- Store checkpoint

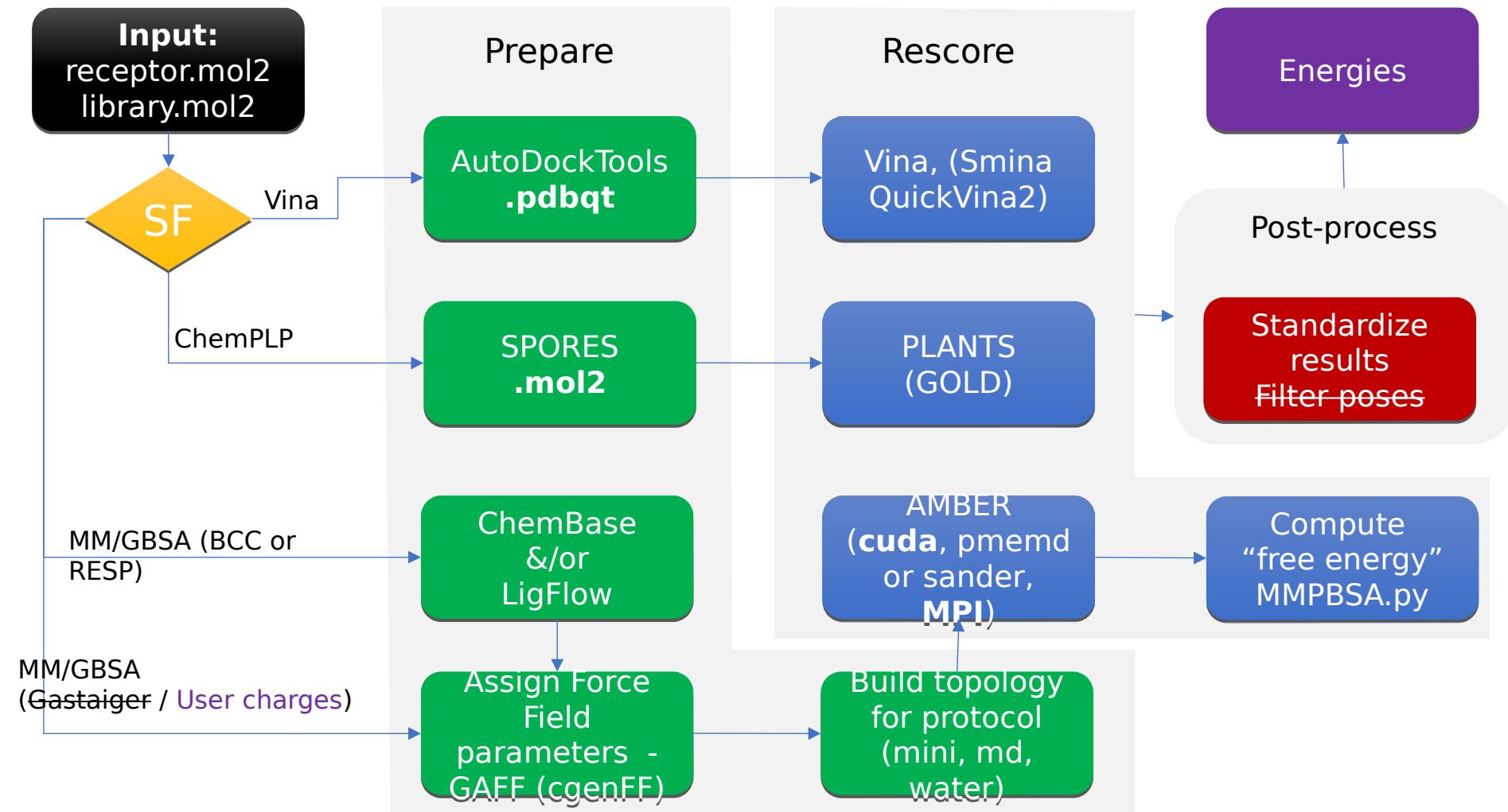
ChemBase

- Prevents expensive charge re-calculation

DockFlow - Docking



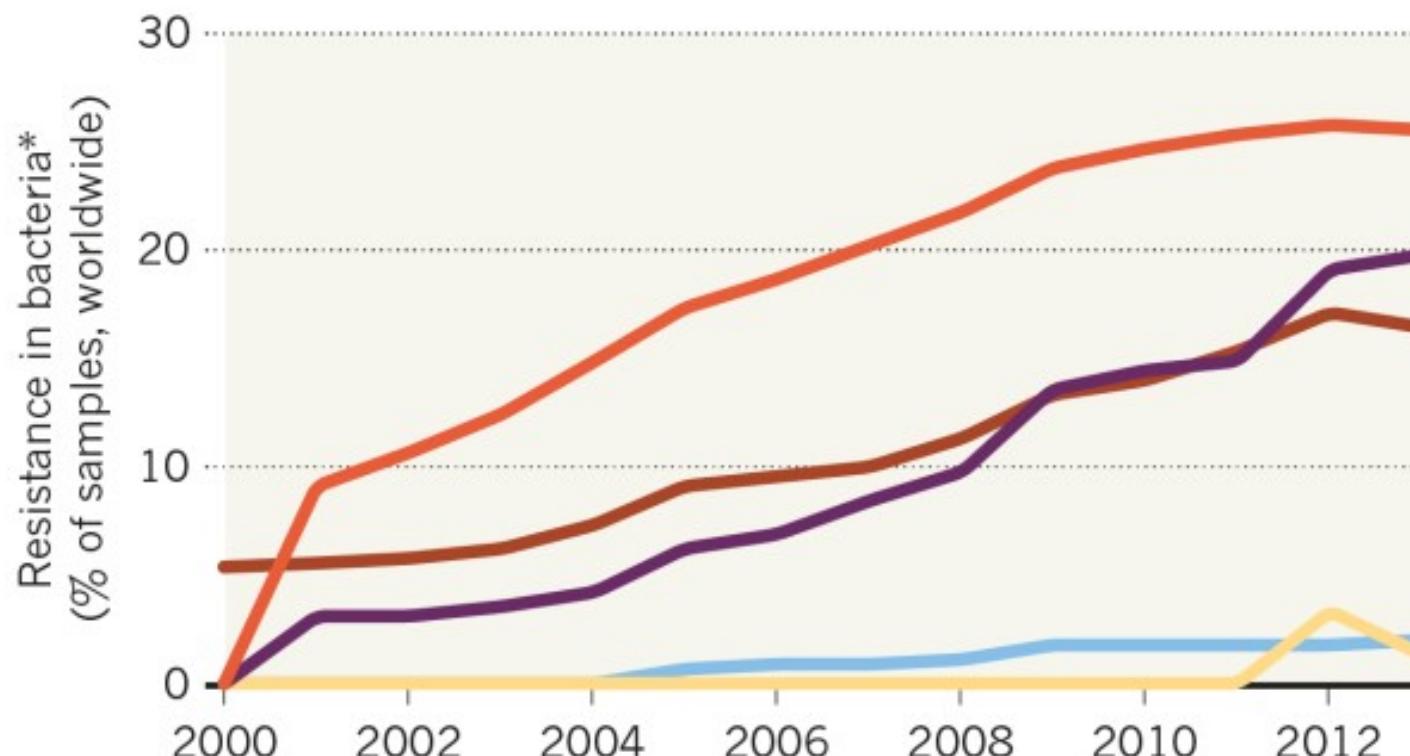
ScoreFlow – Rescoring complexes



The Spread of Antibiotic Resistance

An increasing proportion of bacteria display resistance to common antibiotics.

— Fluoroquinolones — Cephalosporins (3rd gen) — Aminoglycosides
— Carbapenems — Polymyxins



*Enterobacteriace, including *Escherichia coli*, *Klebsiella pneumonia*, *Enterobacter* and *Salmonella*

©nature

Drug Discovery process

