BioExcel Centre of Excellence for Computational Biomolecular Research

pmx: free energy calculations

Vytautas Gapsys
vgapsys@mpibpc.mpg.de
BioExcel Partners 2019

BioExcel is funded by the European Union Horizon 2020 program under grant agreements 675728 and 823830.
BioExcel: Overview

- **SOLUTIONS**
  - SOFTWARE: Widely used, fast and scalable codes for integrative modelling and molecular simulations
  - CORE DEVELOPERS: The scientists who wrote the code and know it best work with us!
  - WORKFLOWS: User-friendly and efficient systems for workflow executions and data processing

- **SERVICES**
  - TRAINING: Webinars, “ask-me-anything” sessions, hands-on workshops for everyone from newbies to advanced users
  - CUSTOMISATION: Tailored solutions adapted to your needs
  - CONSULTANCY: Personalized support with software usage, tuning and scientific aspects of the research

- **ACADEMIA INDUSTRY**

 Vytautas Gapsys
 BioExcel/pmx/Alchemy
Molecular Simulations (GROMACS)
Free Energy Calculations (PMX)

PMX provides an automated framework for the introduction of amino acid mutations in proteins and thus removes some of the most laborious and time consuming steps in traditional methods for free energy calculations.
Integrative modelling (HADDOCK)

- Mutagenesis
- Cross-linking
- H/D exchange
- Bioinformatic predictions
- NMR titrations
- NMR crosssaturation
- Other sources e.g. SAXS, cryoEM
- NMR anisotropy data
- RDCs, para-restraints, diffusion anisotropy
Hybrid QM/MM calculations (CP2K)

Workflow of our **MiMiC interface** in combination with **CP2K** and classical **MD code**.
Usability and Automation

- Make ICT technologies easier to use by biomolecular researchers, both in academia and industry
- Devise efficient workflow environments with associated data integration
pmx: Free Energy Calculations

pmx Alchemy for Proteins, DNA and Ligands
Amino Acid Mutations

Vytautas Gapsys

BioExcel/pmx/Alchemy
Hybrid Structure/Topology

Vytautas Gapsys

BioExcel/pmX/Alchemistry
Application: Protein Thermostability
Application: Protein Thermostability

Wild Type (WT) \hspace{1cm} \text{Mutant (Mut)}

\begin{align*}
\Delta G^\text{WT Folded} &
\end{align*}

\begin{align*}
\Delta G^\text{Mut Folded} &
\end{align*}

\begin{align*}
\Delta G^\text{WT Unfolded} &
\end{align*}

\begin{align*}
\Delta G^\text{Mut Unfolded} &
\end{align*}

\begin{align*}
\Delta \Delta G^\text{Mutation Folding} &= \Delta G^\text{Mut Folding} - \Delta G^\text{WT Folded} = \Delta G^\text{Mutation Folded} - \Delta G^\text{Mutation Unfolded}
\end{align*}

Vytautas Gapsys

BioExcel/pmx/Alchemistry
In total:
- 119 mutations
- at 55 positions

In total:
- 119 mutations
- at 55 positions
Application: Force Field Combination

In total:
- 119 mutations
- at 55 positions

Vytautas Gapsys BioExcel/PMX/Alchemistry
Application: Drug Resistant Mutations

Application:
Large Scale Drug Resistant Mutation Scan

Aldeghi, Gapsys, de Groot, ACS Central Science, 2018
Aldeghi, Gapsys, de Groot, ACS Central Science, 2019
Application: Drug Resistant Mutations

Change in the ligand binding free energy upon an amino acid mutation

\[ \Delta\Delta G_{\text{mutation}} = \Delta G_2 - \Delta G_3 = \Delta G_4 - \Delta G_1 \]
Application: Drug Resistant Mutations

- 17 systems
- 134 mutations
- 27 ligands
The overall averaged accuracy in terms of RMSE reaches 1-1.2 kcal/mol
DNA Mutations

Vytautas Gapsys
DNA: Nucleic Acid Mutations

Amino acid mutations

Val → pmx → Val2Phe

DNA nucleotide mutations

G → pmx → G2T

Gapsys, de Groot, JCTC, 2017

Vytautas Gapsys

BioExcel/pmX/Alchemistry
DNA: Nucleic Acid Mutations

- A2G, A2C, A2T
- G2A, G2C, G2T
- C2A, C2G, C2T
- T2A, T2C, T2G
Application: Protein-DNA Binding

Gapsys, de Groot,

JCTC, 2017
Application: Protein-DNA Binding

Change in the Protein-DNA binding free energy upon nucleotide mutation

$$\Delta \Delta G_{\text{mutation}} = \Delta G_2 - \Delta G_3 = \Delta G_4 - \Delta G_1$$
Application: Protein-DNA Binding

- 16 systems
- 397 mutations
Application: Protein-DNA Binding

\[ \Delta \Delta G_{\text{calc}} = 0.84 \Delta \Delta G_{\text{exp}} + 0.47 \]

Number of mutations: 397
Number of systems: 16

Vytautas Gapsys
BioExcel/pmX/Alchemistry
Ligands

Ligand Modifications

ΔG₁

ΔG₂

ΔG₃

ΔG₄

Vytautas Gapsys

BioExcel/pmX/Alchemy
pmx Ligands

atoms_to_morph.py
Identifies atoms to be morphed

make_hybrid.py
Builds hybrid topology

build_mst_graph.py
Suggests ligand pairs

RDKit
Open-Source Cheminformatics and Machine Learning

Open source toolkit for cheminformatics

Vytautas Gapsys
BioExcel/pmX/Alchemistry
Protein-Ligand Binding

Application:
482 ligand modifications in protein-ligand binding

Gapsys, Perez-Benito, Aldeghi, Seeliger, van Vlijmen, Tresadern, de Groot,
under review

Vytautas Gapsys

BioExcel/pmx/Alchemistry
Protein-Ligand Complexes

- **11 systems**
- **482 modifications**

- PDE2: 21 ligands, 34 perturbations
- Galectin: 8 ligands, 8 perturbations
- cMet: 12 ligands, 25 perturbations
- BACE: 80 ligands, 144 perturbations (divided in 3 sets)
- JNK1: 21 ligands, 31 perturbations
- TYK2: 16 ligands, 24 perturbations
- MCL1: 42 ligands, 71 perturbations
- CDK2: 16 ligands, 25 perturbations
- Thrombin: 11 ligands, 16 perturbations
- PTP1b: 23 ligands, 49 perturbations
- P38: 34 ligands, 56 perturbations
Overall Results

All data sets: 482 mutations

- **FEP+**
- **pmx**

### AUE, k/mol

<table>
<thead>
<tr>
<th>Method</th>
<th>AUE</th>
<th>cor</th>
<th>ΔG values</th>
<th>Error</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maestro</strong></td>
<td>3.66 ± 0.19</td>
<td>0.69 ± 0.03</td>
<td>68.2 kcal/mol</td>
<td>0.71 ± 0.02</td>
<td>0.73 ± 0.02</td>
</tr>
<tr>
<td><strong>FEP+</strong></td>
<td>3.88 ± 0.15</td>
<td>0.61 ± 0.03</td>
<td>63.1 kcal/mol</td>
<td>0.72 ± 0.02</td>
<td>0.74 ± 0.02</td>
</tr>
<tr>
<td><strong>GAFF</strong></td>
<td>4.01 ± 0.18</td>
<td>0.64 ± 0.04</td>
<td>59.2 kcal/mol</td>
<td>0.73 ± 0.02</td>
<td>0.75 ± 0.02</td>
</tr>
<tr>
<td><strong>CGenFF</strong></td>
<td>3.64 ± 0.14</td>
<td>0.63 ± 0.03</td>
<td>66.2 kcal/mol</td>
<td>0.74 ± 0.02</td>
<td>0.76 ± 0.02</td>
</tr>
<tr>
<td><strong>Consensus:</strong></td>
<td><strong>GAFF+CGenFF</strong></td>
<td><strong>5.62 ± 0.16</strong></td>
<td><strong>69.4 kcal/mol</strong></td>
<td><strong>0.75 ± 0.02</strong></td>
<td><strong>0.77 ± 0.02</strong></td>
</tr>
</tbody>
</table>

### Pearson Correlation

- **Maestro**
- **FEP+**
- **GAFF**
- **CGenFF**
- **Consensus: GAFF+CGenFF**

### Additional Plots

- **FEP+ OPLS 3**
- **GAFF**
- **CGenFF**
- **Consensus: GAFF+CGenFF**
D3R: Free Energy Calculations
Cathepsin S

Elisée, Gapsys, Mele, Chaput, Selwa, de Groot, Iorga
in revision
D3R: Cathepsin S

Poses generated by Eddy Elisée based on the crystallographic poses released in the previous D3R-GC3
One ligand was selected as a reference.
D3R: Ligand Mapping

Redundancies were included.

Vytautas Gapsys

BioExcel/pmx/Alchemy
D3R: Calculation

Setup

- Non-equilibrium protocol
- 109 edges
- 3 repeats per edge
- 44 ns per repeat
- 2 force fields: Gaff 2.1, CGenFF 4.1
- Gromacs2018 + pmx
D3R: Results (RMSE)

- Free energy
- Structure Based
- Ligand Based

RMSEC, kcal/mol

Vytautas Gapsys
BioExcel/PMX/Alchemistry
D3R: Results (Pearson correlation)

Pearson's r

Free energy
Structure Based
Ligand Based

Vytautas Gapsys
BioExcel/px/Alchemistry
D3R: Results (Spearman correlation)
D3R: Results (Kendall correlation)

Vytautas Gapsys BioExcel/pmx/Alchemistry
pmx and free energies
Dr. Servaas Michielssens
Dr. Daniel Seeliger
Dr. Matteo Aldeghi
Dr. Yuriy Khalak
Professor Dr. Bert de Groot

Small molecule study
Dr. Laura Benitez
Dr. Gary Tresadern
Professor Dr. Herman van Vlijmen

D3R: Cathepsin S
Dr. Bogdan I. Iorga
Eddy Elisee
Dr. Nawel Mele
Dr. Ludovic Chaput
Dr. Edithe Selwa