Thermodynamic Integration with Enhanced Sampling (TIES)

A. P. Bhati, S. Wan, D. W. Wright and P. V. Coveney
agastya.bhati.14@ucl.ac.uk

Centre for Computational Science
Department of Chemistry
University College London

16th May 2017
Credits due

- Peter V. Coveney
- Shunzhou Wan
- David W. Wright
- UCL Overseas Research Scholarship
- Inlaks Shivdasani Foundation
Overview

- Brief introduction to the binding affinity and methods to calculate it
- Importance of being able to predict binding affinities reliably
- Issues with the traditional *in silico* approaches
- Ensemble simulation based approach: TIES
- Success story of TIES: Case studies
Overview

- Brief introduction to the binding affinity and methods to calculate it
- Importance of being able to predict binding affinities reliably
- Issues with the traditional *in silico* approaches
- Ensemble simulation based approach: TIES
- Success story of TIES: Case studies
Overview

- Brief introduction to the binding affinity and methods to calculate it
- Importance of being able to predict binding affinities reliably
- Issues with the traditional *in silico* approaches
- Ensemble simulation based approach: TIES
- Success story of TIES: Case studies
Overview

- Brief introduction to the binding affinity and methods to calculate it
- Importance of being able to predict binding affinities reliably
- Issues with the traditional in silico approaches
- Ensemble simulation based approach: TIES
- Success story of TIES: Case studies
Overview

- Brief introduction to the binding affinity and methods to calculate it
- Importance of being able to predict binding affinities reliably
- Issues with the traditional \textit{in silico} approaches
- Ensemble simulation based approach: TIES
- Success story of TIES: Case studies
An experiment is usually made on a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations.
Molecular Dynamics (MD) Simulations

An experiment is usually made on a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations.

In MD simulation, macroscopic properties corresponding to experimental observables are defined in terms of ensemble averages.¹

¹Coveney & Wan, PCCP, 2016, 18, 30236-30240, DOI: 10.1039/C6CP02349E
Molecular Dynamics (MD) Simulations

An experiment is usually made on a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations.

In MD simulation, macroscopic properties corresponding to experimental observables are defined in terms of ensemble averages.\(^1\)

Free energy is such a measurement

\(^1\)Coveney & Wan, PCCP, 2016, 18, 30236-30240, DOI: 10.1039/C6CP02349E
Binding Affinity

How does it help us?

Ligand binding driven by changes in the Gibbs free energy. The more negative the $\Delta G$, the stronger the binding.
Binding Affinity

How does it help us?

- Ligand binding driven by changes in the Gibbs free energy
- The more negative the $\Delta G$ the stronger the binding
In silico free energy calculation methods

- Docking methods
- Linear Interaction method
- MMPBSA+NMODE
- Thermodynamic integration
- Free energy perturbation (EXP, BAR, MBAR)
Thermodynamic Integration

Hybrid potential function:

\[ H(\lambda) = \lambda H_A + (1 - \lambda) H_B \]

The coupling parameter, \( \lambda \), defines the progress of a system along the path, B to A, as \( \lambda \) is changed from 0 to 1.

\[ \Delta G = \int_{0}^{1} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \]

Alchemical mutation from B (left) to A (right)
Relative Binding Affinity

Relative binding affinity calculations with alchemical mutation: make use of thermodynamic cycle to calculate binding free energy differences

\[
\Delta G_{\text{binding}} = \Delta G_{\text{binding ligand}}^2 - \Delta G_{\text{binding ligand}}^1 = \Delta G_{\text{alch ligand}} - \Delta G_{\text{alch complex}}
\]
Relative Binding Affinity

Relative binding affinity calculations with alchemical mutation: make use of thermodynamic cycle to calculate binding free energy differences

\[
\Delta G^{\text{binding}}_{\text{ligand}} = \Delta G^{\text{binding}}_{\text{ligand}}^{2} - \Delta G^{\text{binding}}_{\text{ligand}}^{1} = \Delta G^{\text{alch}}_{\text{ligand}} - \Delta G^{\text{alch}}_{\text{complex}}
\]
Relative Binding Affinity

Relative binding affinity calculations with alchemical mutation: make use of thermodynamic cycle to calculate binding free energy differences

\[ \Delta \Delta G_{binding} = \Delta G_{binding_{ligand2}} - \Delta G_{binding_{ligand1}} = \Delta G_{alch_{ligand}} - \Delta G_{alch_{complex}} \]
Application of binding affinity prediction

Drug designing: Lead optimisation
Application of binding affinity prediction

Drug designing: Lead optimisation
Searching for a needle in a haystack

www.phrm.co.uk
Application of binding affinity prediction

Drug designing: Lead optimisation
Searching for a needle in a haystack
Application of binding affinity prediction

Drug designing: Lead optimisation
Searching for a needle in a haystack

High-Throughput Screening (HTS)

- HTS can test thousands of compounds per day
- Cost of HTS is substantial: $1-10 per compound
Application of binding affinity prediction

Drug designing: Lead optimisation
Searching for a needle in a haystack

Virtual Screening:
Systematic computer-based prediction of binding affinity of compounds to proteins
Issues with the available *in silico* methods

**Theories exist** - Then predictions are possible, and in principle, we should be able to apply existing methods in drug screening domains
Issues with the available *in silico* methods

**Theories exist** - Then predictions are possible, and in principle, we should be able to apply existing methods in drug screening domains.
Single vs Ensemble MD simulations

The binding free energy and potential energy derivative can vary widely (up to 12 kcal/mol) between two single simulations. **Single simulation: not reproducible, unscientific!**

L1Q-LI9 ligand transformation bound to CDK2\(^2\)

Drug-HIV1 protease\(^3\)

---

1. Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979
2. Wright, Hall, Kenway, Jha & Coveney, JCTC, (2014), DOI: 10.1021/ct4007037
Single vs Ensemble MD simulations

The binding free energy and potential energy derivative can vary widely (up to 12 kcal/mol) between two single simulations.

Single simulation: not reproducible, unscientific!

L1Q-LI9 ligand transformation bound to CDK2

The energy/energy derivatives from ensemble simulations follow well defined Gaussian distributions.

1Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979
2Wright, Hall, Kenway, Jha & Coveney, JCTC, (2014), DOI: 10.1021/ct4007037

Drug-HIV1 protease
Thermodynamic Integration with Enhanced Sampling (TIES)

Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979
Thermodynamic Integration with Enhanced Sampling (TIES)

Binding Affinity Calculator (BAC) is a software toolkit which automates the implementation of TIES (and ESMACS) methods for binding affinity calculations.

\[ \text{Preparation} \]

\[ \lambda \]

<table>
<thead>
<tr>
<th>0.0</th>
<th>0.05</th>
<th>0.1,...,0.9</th>
<th>0.95</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>5</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>...</td>
<td>5</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>...</td>
<td>5</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>...</td>
<td>5</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>...</td>
<td>5</td>
<td>1</td>
<td>...</td>
</tr>
</tbody>
</table>

\[ \text{Equilibration} \]

NAMD, 96*13*5 cores, 2.5 hours

\[ \text{Production} \]

NAMD, 96*13*5 cores, 5 hours

\[ \text{Statistical Analyses} \]

Desktops, <10 minutes

\[ ^1 \text{Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979} \]
TIES: Convergence of errors

Variation of error with different parameters

L1Q-LI9 ligand transformation bound to CDK2

1 Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979
TIES: Biomolecular systems studied

Crystal structures of the protein from the Protein Data Bank

CDK2
MCL1
PTP1B
Thrombin
TYK2
TIES predictions

\[ \Delta \Delta G_{\text{TIES}} \text{ (kcal/mol)} \]

\[ \Delta \Delta G_{\text{experimental}} \text{ (kcal/mol)} \]

RMSE = 0.9 kcal/mol
MAE = 0.7 kcal/mol
\( r_p = 0.84 \)
\( r_s = 0.85 \)

\(^1\)Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979
TIES predictions

1Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979
TIES reproducibility

Reproducibility of TIES predictions for transformation between ligands bound to CDK2

\[ \Delta \Delta G_{\text{calc}} (\text{kcal/mol}) \]
\[ \Delta \Delta G_{\text{exp}} (\text{kcal/mol}) \]

\[ ^1 \text{Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979} \]
Thrombin S1 pocket: water inflow

Captured successfully by TIES

*Bhati, Wan, Wright & Coveney, JCTC, (2017), DOI: 10.1021/acs.jctc.6b00979*
Blind study with Pfizer: Tropomysin receptor kinase A

The TIES study gives the same ranking of the binding free energies as the experimental data for the compounds studied

- Pearson correlation of 0.88
- 10 out of the 14 pairs directional agreements
- A better agreement might be achieved when the error bars of the experimental $\Delta \Delta G$ are also taken into account

---

1Wan, Bhati, Skerratt, Omoto, Shanmugasundaram, Bagal, Coveney, JCIM (2017), DOI: 10.1021/acs.jcim.6b00780
Blind study with Pfizer: Tropomysin receptor kinase A

The TIES study gives the same ranking of the binding free energies as the experimental data for the compounds studied

- Pearson correlation of 0.88
- 10 out of the 14 pairs directional agreements
- A better agreement might be achieved when the error bars of the experimental $\Delta\Delta G$ are also taken into account

\[ r = 0.88 \]

---

1Wan, Bhati, Skerratt, Omoto, Shanmugasundaram, Bagal, Coveney, JCIM (2017), DOI: 10.1021/acs.jcim.6b00780
Blind study with Pfizer: Tropomysin receptor kinase A

The TIES study gives the same ranking of the binding free energies as the experimental data for the compounds studied

- Pearson correlation of 0.88
- 10 out of the 14 pairs directional agreements
- A better agreement might be achieved when the error bars of the experimental $\Delta \Delta G$ are also taken into account

---

1Wan, Bhati, Skerratt, Omoto, Shanmugasundaram, Bagal, Coveney, JCIM (2017), DOI: 10.1021/acs.jcim.6b00780
Blind study with GSK: Bromodomain 4

\[ r_s = 0.92 \]

\[ \Delta G_{\text{TIES}} \text{ (kcal/mol)} \]

\[ \Delta G_{\text{exp}} \text{ (kcal/mol)} \]

---

\(^1\) Wan, Bhati, Zasada, Wall, Green, Bamborough, Coveney, JCTC (2017), DOI: 10.1021/acs.jctc.6b00794
Excellent scalability

- TIES workflow is perfectly scalable
- Giant run on PRACE’s Tier0 SuperMUC

LRZ press release

London Science Museum blog post
Conclusions

- The traditional *in silico* methods to calculate binding affinity are not reliable, and hence, not widely applicable in pharmaceutical domain.
Conclusions

- The traditional *in silico* methods to calculate binding affinity are not reliable, and hence, not widely applicable in pharmaceutical domain.
- Ensemble simulation based method is the way out.
Conclusions

- The traditional *in silico* methods to calculate binding affinity are not reliable, and hence, not widely applicable in pharmaceutical domain.
- Ensemble simulation based method is the way out.
- TIES substantially improves the accuracy, precision and reliability of calculated relative binding affinities.
Conclusions

- The traditional *in silico* methods to calculate binding affinity are not reliable, and hence, not widely applicable in pharmaceutical domain.
- Ensemble simulation based method is the way out.
- TIES substantially improves the accuracy, precision and reliability of calculated relative binding affinities.
- Ligand-protein binding affinity predictions made with TIES are reproducible.