Current Development in Non-Equilibrium Switching: The Mapper Feature

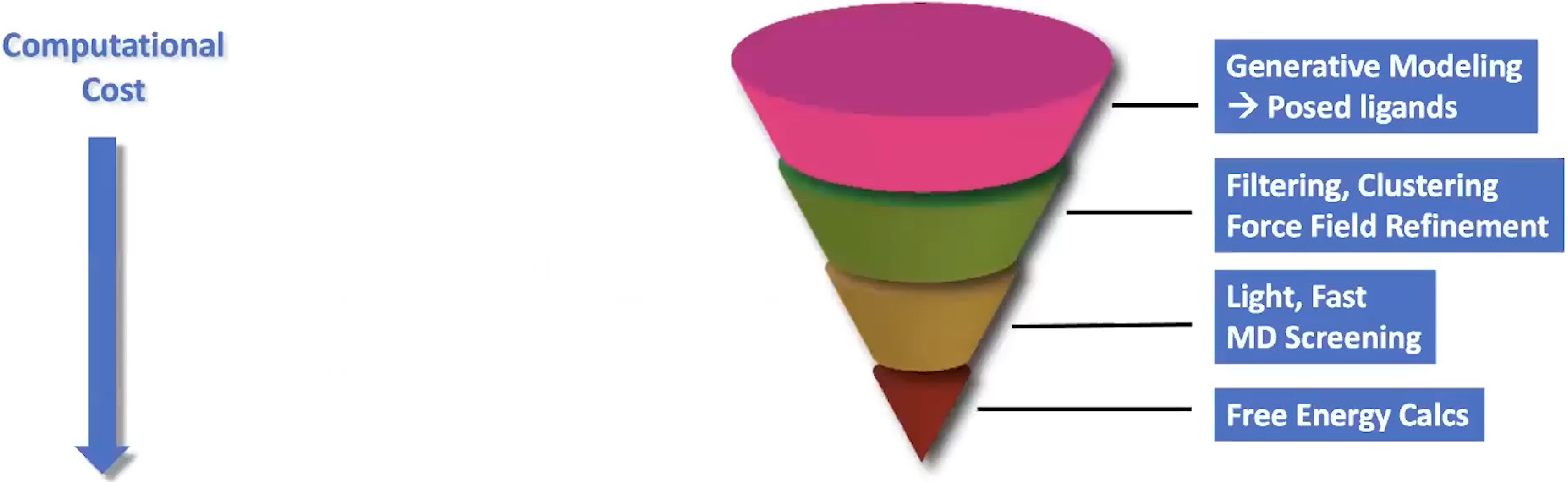
Presented by Gaetano Calabró, PhD
Senior Scientific Software Developer
Thursday, June 30, 2022

OpenEye SCIENTIFIC

Overview

- Introduction and Background
- The OE Mapper Features
- The OE Mapper Validation
- Conclusions

MD in Structure-Based Lead Optimization



- Heavier MD methods staged to offer more value later in triaging

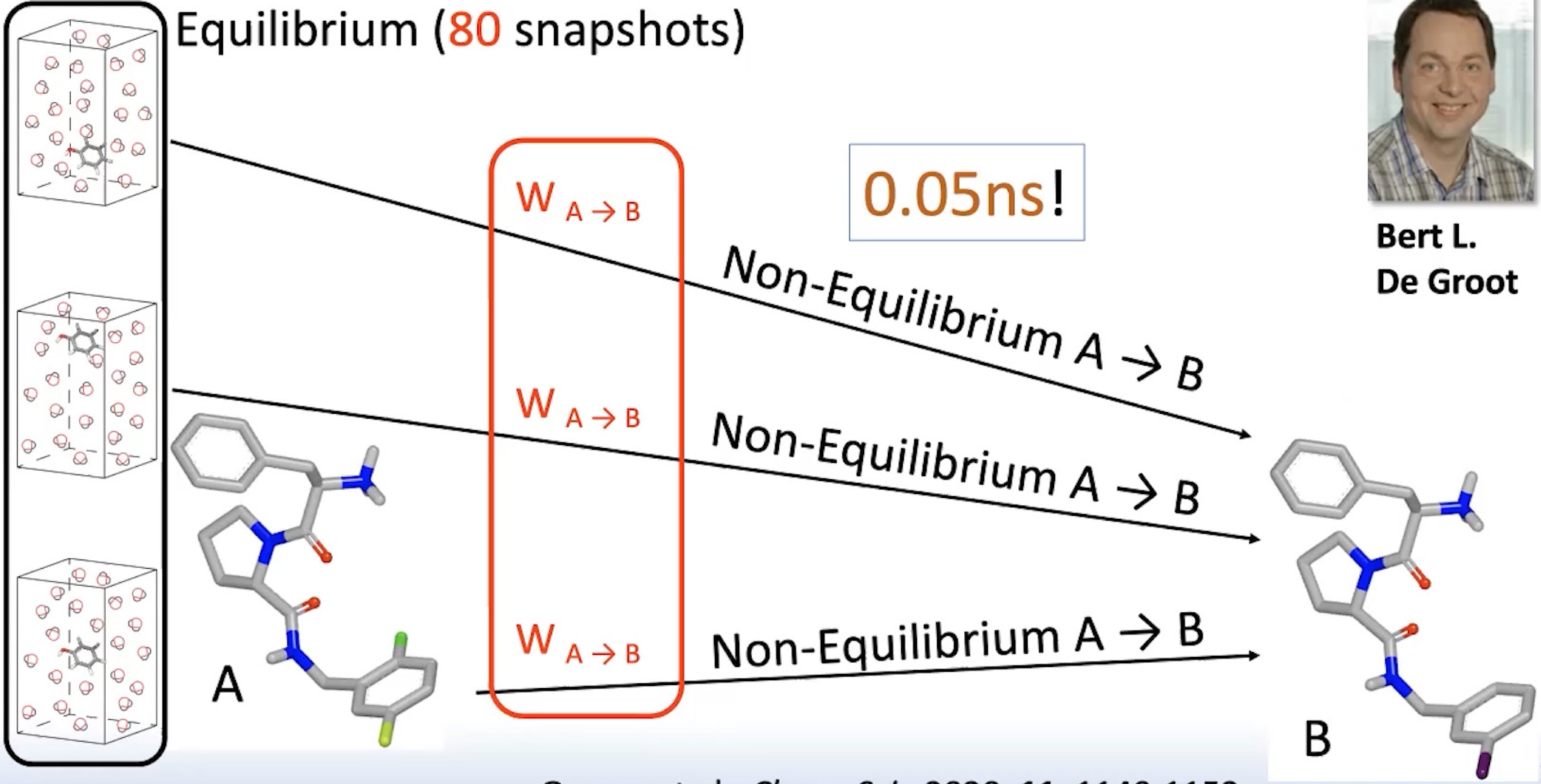
OpenEye SCIENTIFIC

MD in Structure-Based Lead Optimization

- Heavier MD methods staged to offer more value later in triaging

OpenEye SCIENTIFIC

RBFE Methods NES (Non-Equilibrium Switching)

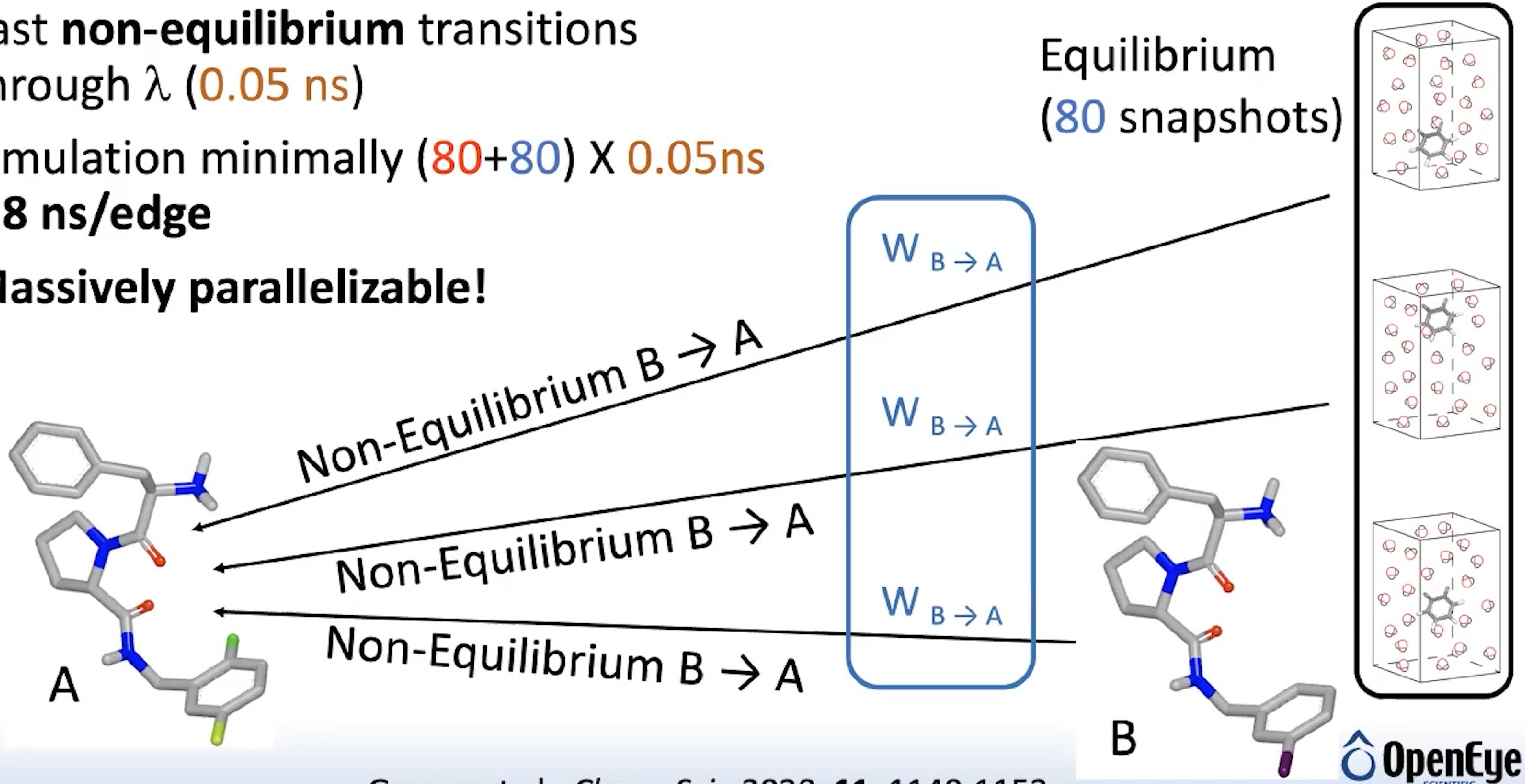


Bert L. De Groot   Vytautas Gapsys

Gapsys et al., *Chem. Sci.*, 2020, **11**, 1140-1152

RBFE Methods NES (Non-Equilibrium Switching)

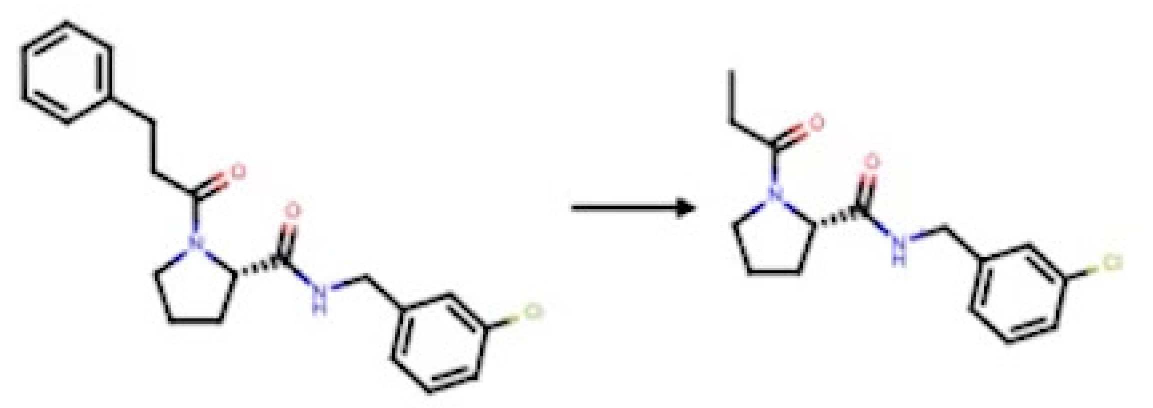
- Fast **non-equilibrium** transitions through λ (0.05 ns)
- Simulation minimally (80+80) X 0.05ns **= 8 ns/edge**
- **Massively parallelizable!**



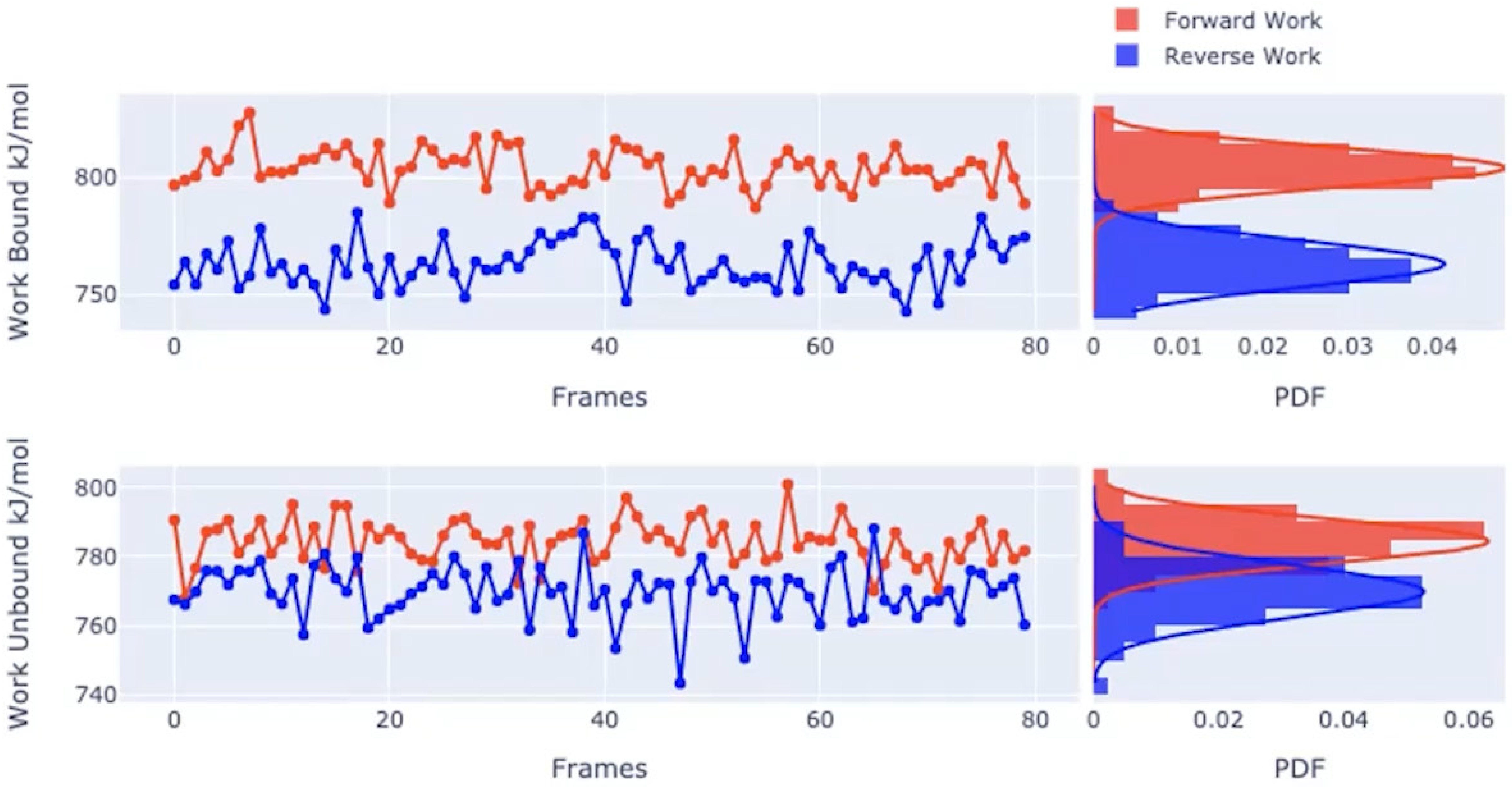
Gapsys et al., *Chem. Sci.*, 2020, **11**, 1140-1152

RBFE Methods NES (Non-Equilibrium Switching)

3K to 3B

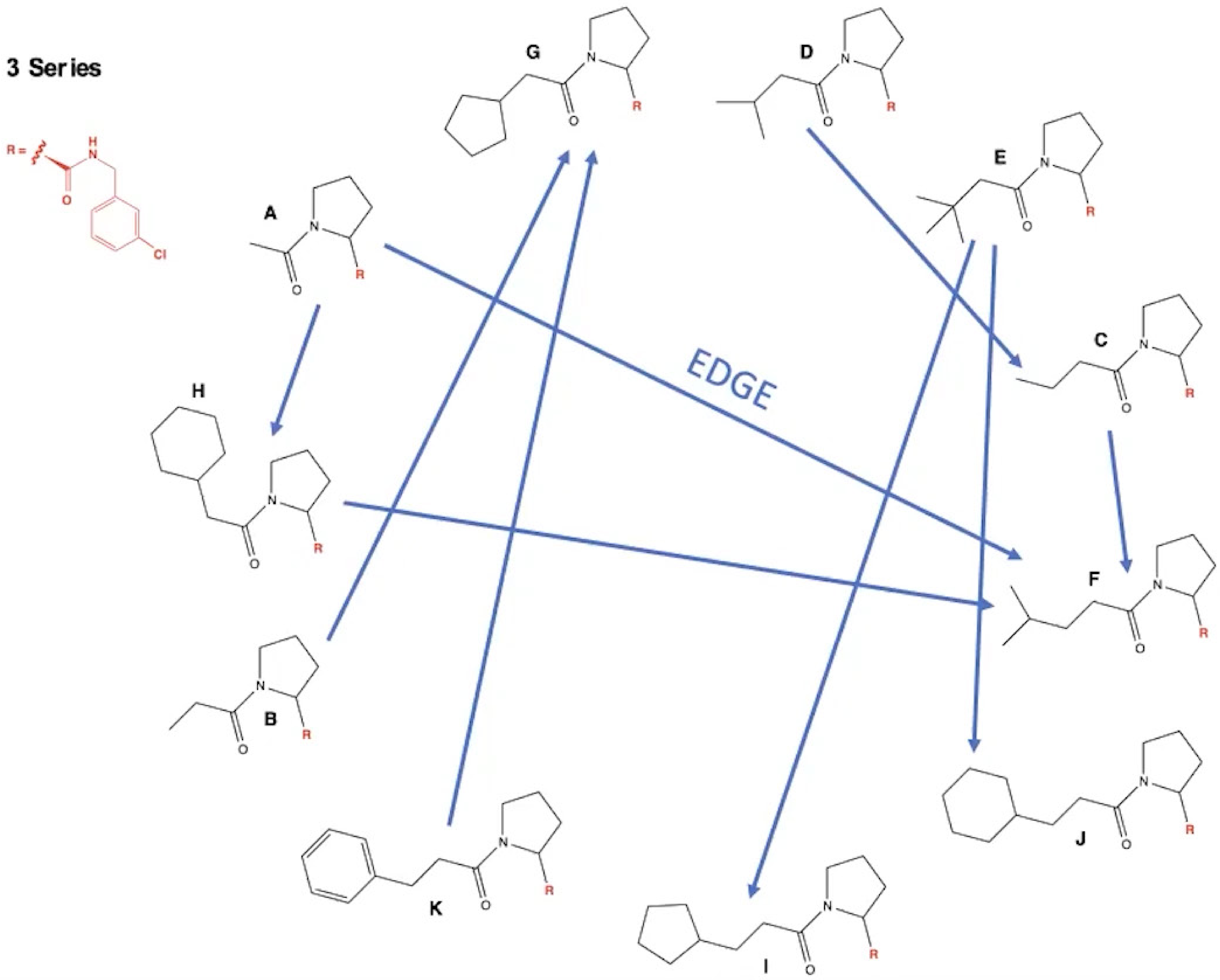


BAR score:
8.30 ± 1.02 kJ/mol



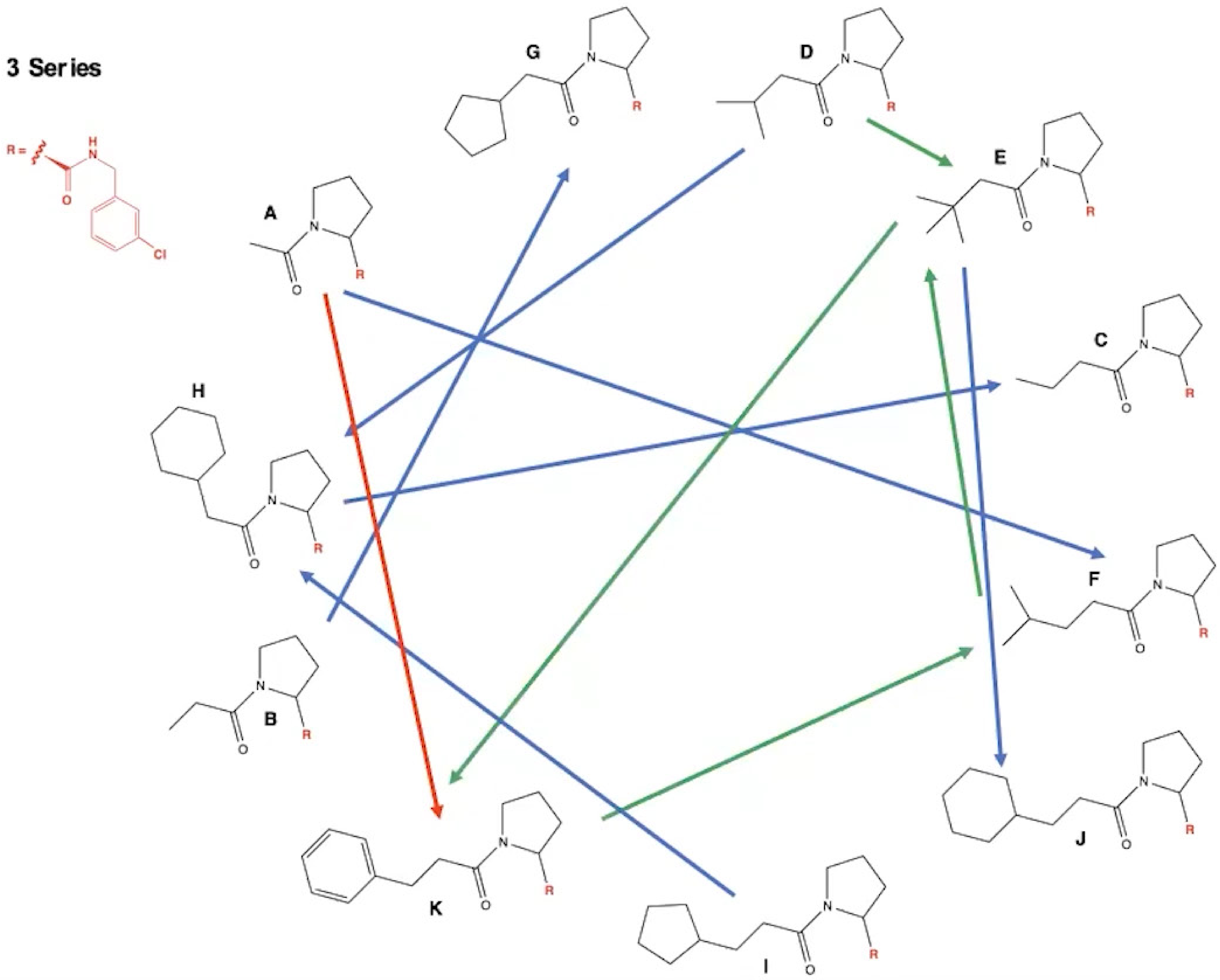
BAR: $$\sum_{i=1}^{n_f} \frac{1}{1+\exp\left(\ln \frac{n_f}{n_r} + \beta(w_i - \Delta G)\right)} = \sum_{j=1}^{n_r} \frac{1}{1+\exp\left(\ln \frac{n_r}{n_f} + \beta(w_j - \Delta G)\right)}$$

Why do we need a mapper?



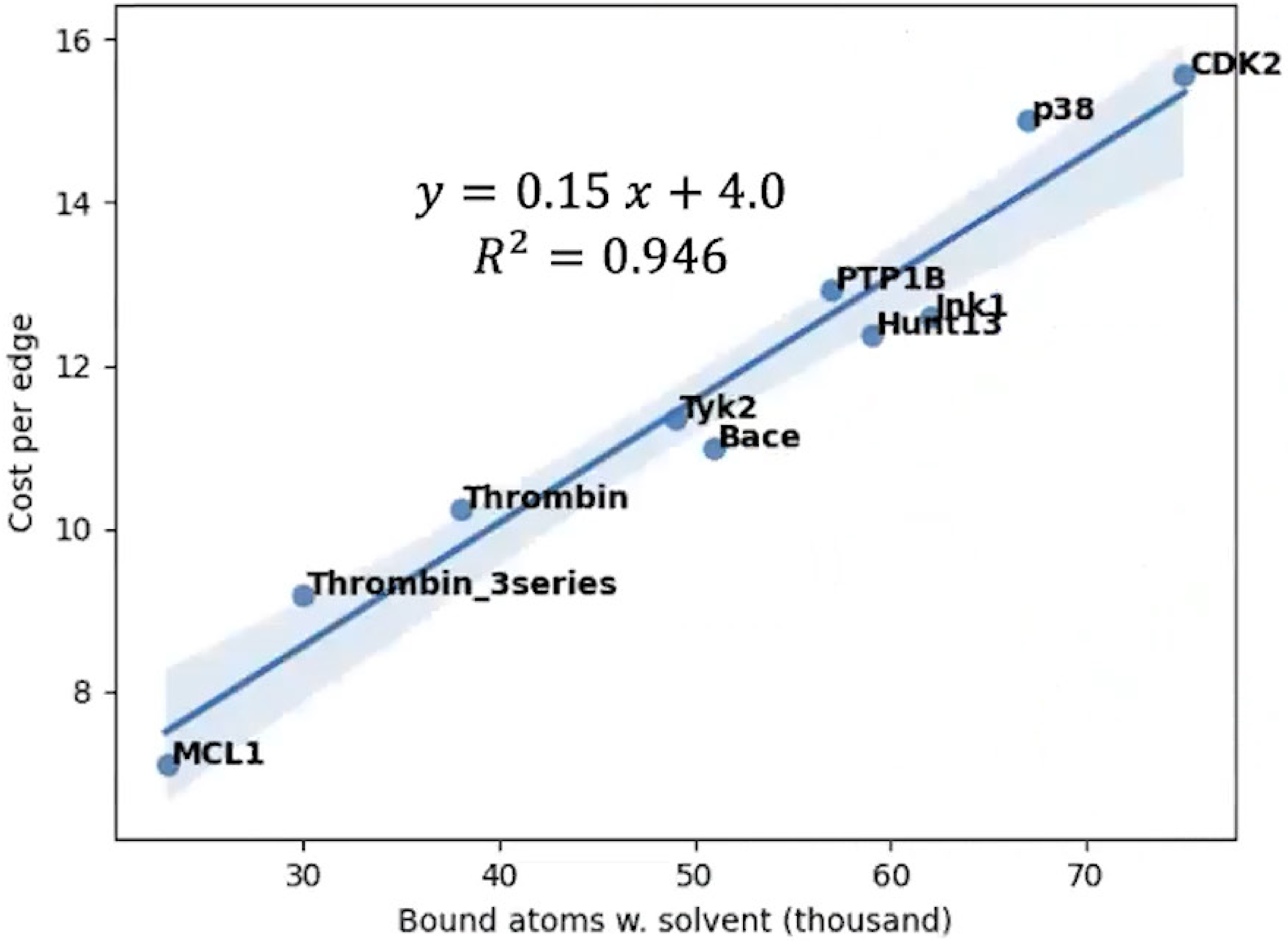
OpenEye SCIENTIFIC

Why do we need a mapper?



Why do we need a mapper?

- In general, given N compounds N(N-1)/2 possible edges



11 Thrombin inhibitors (55 edges) ~ $550
32 Bace inhibitors (496 edges) ~ $5500

Why do we need a mapper?

- The mapper should avoid “difficult” edges
- Cycles closure help in the Affinity prediction introducing redundances
- To expensive running all the edges

The OE Mapper

- **The Mapper goal** is to produce a set of edges where the transformed pair of compounds are “similar”: the RBFE edge calculation is likely to be successful and accurate

- The OE Mapper is mainly based on LOMAP(*)
    - LOMAP uses the chemical graph only (MCS)

(*) Liu, Y. Wu, et al, (2013), “Lead Optimization Mapper: Automating free energy calculations for lead optimization”, J. Comput. Aided Mol. Design, 27(9):755-770

The OE Mapper floe report

The Mapper Score

(With OpenEye variations)

- The similarity Matrix Scoring

$$\begin{array}{c|ccccc} & C_1 & C_2 & \dots & & C_n \\ C_1 & 1 & S_{1,2} & S_{1,3} & \dots & S_{1,n} \\ C_2 & S_{2,1} & 1 & S_{2,3} & \dots & S_{2,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ & \dots & \dots & \dots & \dots & \dots \\ C_n & S_{n,1} & S_{n,2} & S_{n,3} & \dots & 1 \end{array}$$

$$S_{ij} \in [0,1]$$

$$S_{ij} = S_{ji}$$

$$S_{i,j} = \prod_{k=1}^{M} R_k(i,j)$$

- R_k Charge Based
  - If C_i and C_j same charge 1 else 0
- R_k MCSS Based (GMX biased)
  - (a) $H(mcs_{hw} - ths)$
  - (b) $e^{-\beta(Nhw_i + Nhw_j - 2mcs_{hw})}$
- R_k ROCS Based
  - Shape and Color

The Mapper Score

- The similarity Matrix Scoring

(With OpenEye variations)

$$
\begin{array}{c} \\ C_1 \\ C_2 \\ \dots \\ \\ C_n \end{array}
\begin{array}{c}
\begin{array}{ccccc} C_1 & C_2 & \dots & & C_n \end{array} \\
\begin{pmatrix}
1 & S_{1,2} & S_{1,3} & \dots & S_{1,n} \\
S_{2,1} & 1 & S_{2,3} & \dots & S_{2,n} \\
\dots & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots \\
S_{n,1} & S_{n,2} & S_{n,3} & \dots & 1
\end{pmatrix}
\end{array}
\quad S_{ij} \in [0,1] \quad S_{ij} = S_{ji}
$$

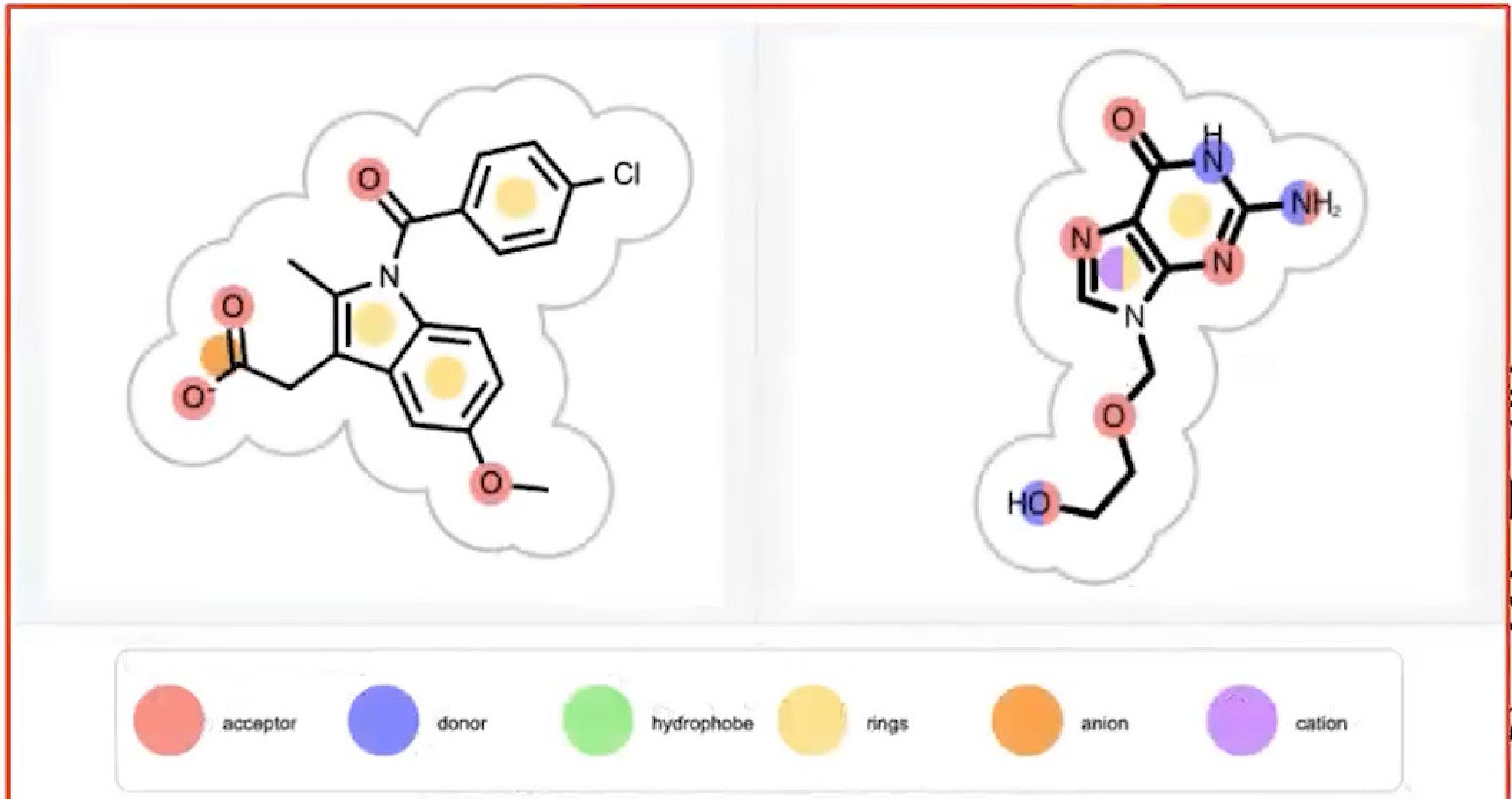


Figure 2. Color-coding of chemical features

ge Based

nd C_j same charge 1 else 0

S Based (GMX biased)

$$H(mcs_{hw} - ths)$$

$$e^{-\beta(Nhw_i + Nhw_j - 2mcs_{hw})}$$

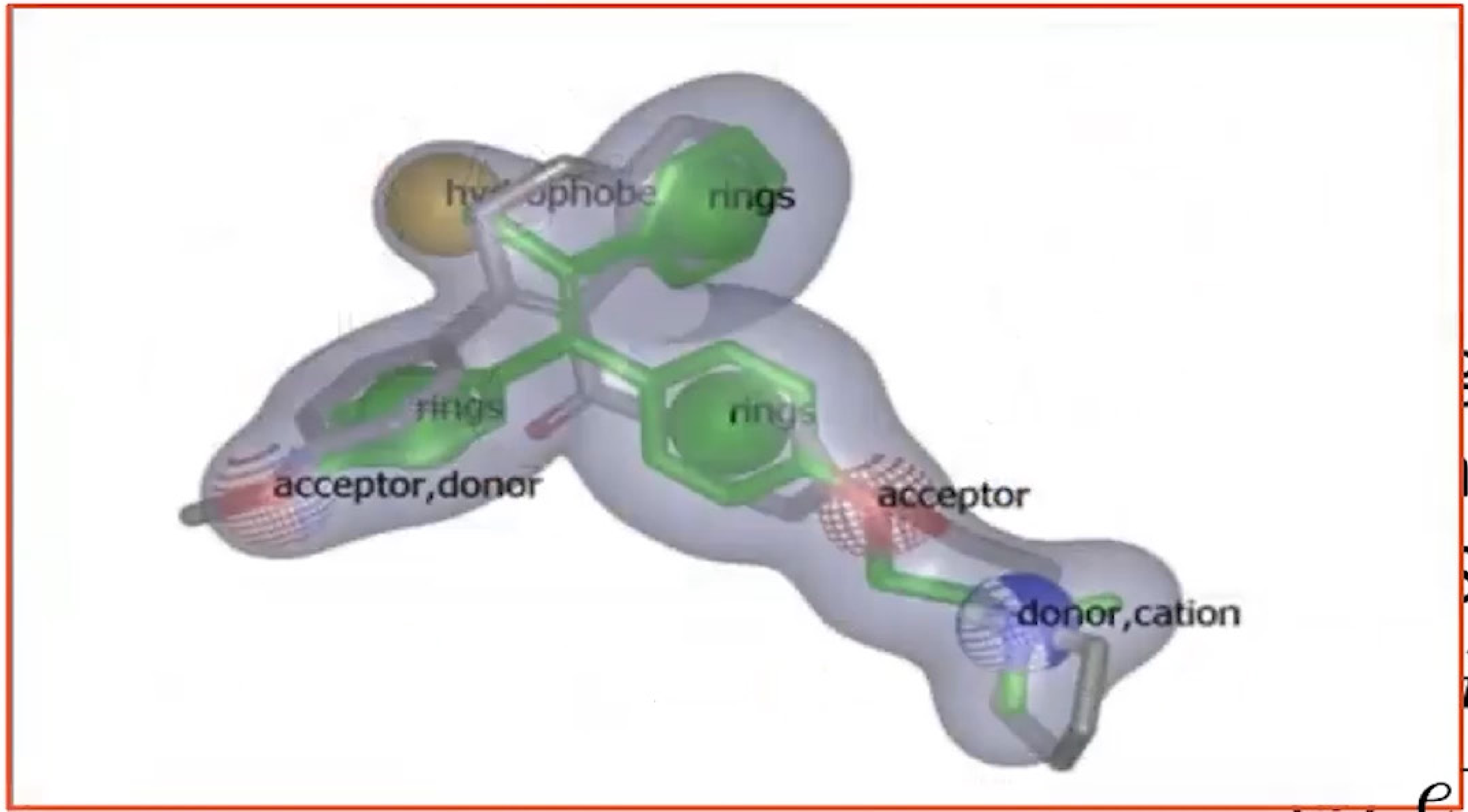
- R_k ROCS Based
  - Shape and Color

The Mapper Score

- The similarity Matrix Scoring

(With OpenEye variations)

$$
\begin{array}{c} \\ C_1 \\ C_2 \\ \dots \\ \\ C_n \end{array}
\begin{array}{c}
\begin{array}{ccccc} C_1 & C_2 & \dots & & C_n \end{array} \\
\begin{pmatrix}
1 & S_{1,2} & S_{1,3} & \dots & S_{1,n} \\
S_{2,1} & 1 & S_{2,3} & \dots & S_{2,n} \\
\dots & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots \\
S_{n,1} & S_{n,2} & S_{n,3} & \dots & 1
\end{pmatrix}
\end{array}
\quad S_{ij} \in [0,1] \quad S_{ij} = S_{ji}
$$



ge Based

nd C_j same charge 1 else 0

S Based (GMX biased)

$H(mcs_{hw} - ths)$

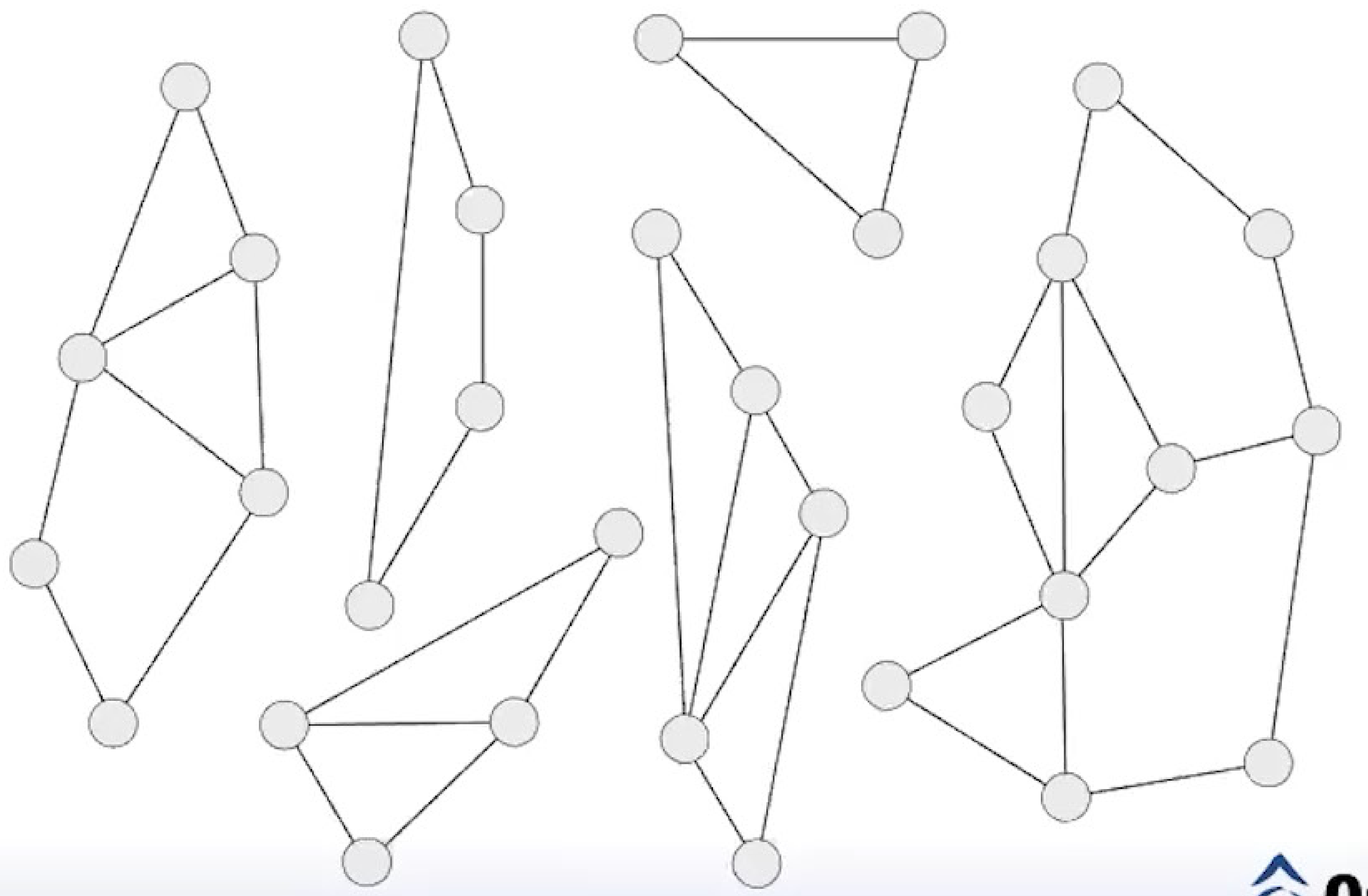
$e^{-\beta(Nhw_i + Nhw_j - 2mcs_{hw})}$

- R_k ROCS Based
  - Shape and Color

The Mapper Graph

- Building the graph
  1. Create edges where $S_{ij} \geq Cut_{Off} > 0$

$$\begin{pmatrix} 1 & S_{1,2} & S_{1,3} & \dots & S_{1,n} \\ S_{2,1} & 1 & S_{2,3} & \dots & S_{2,n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ S_{n,1} & S_{n,2} & S_{n,3} & \dots & 1 \end{pmatrix}$$

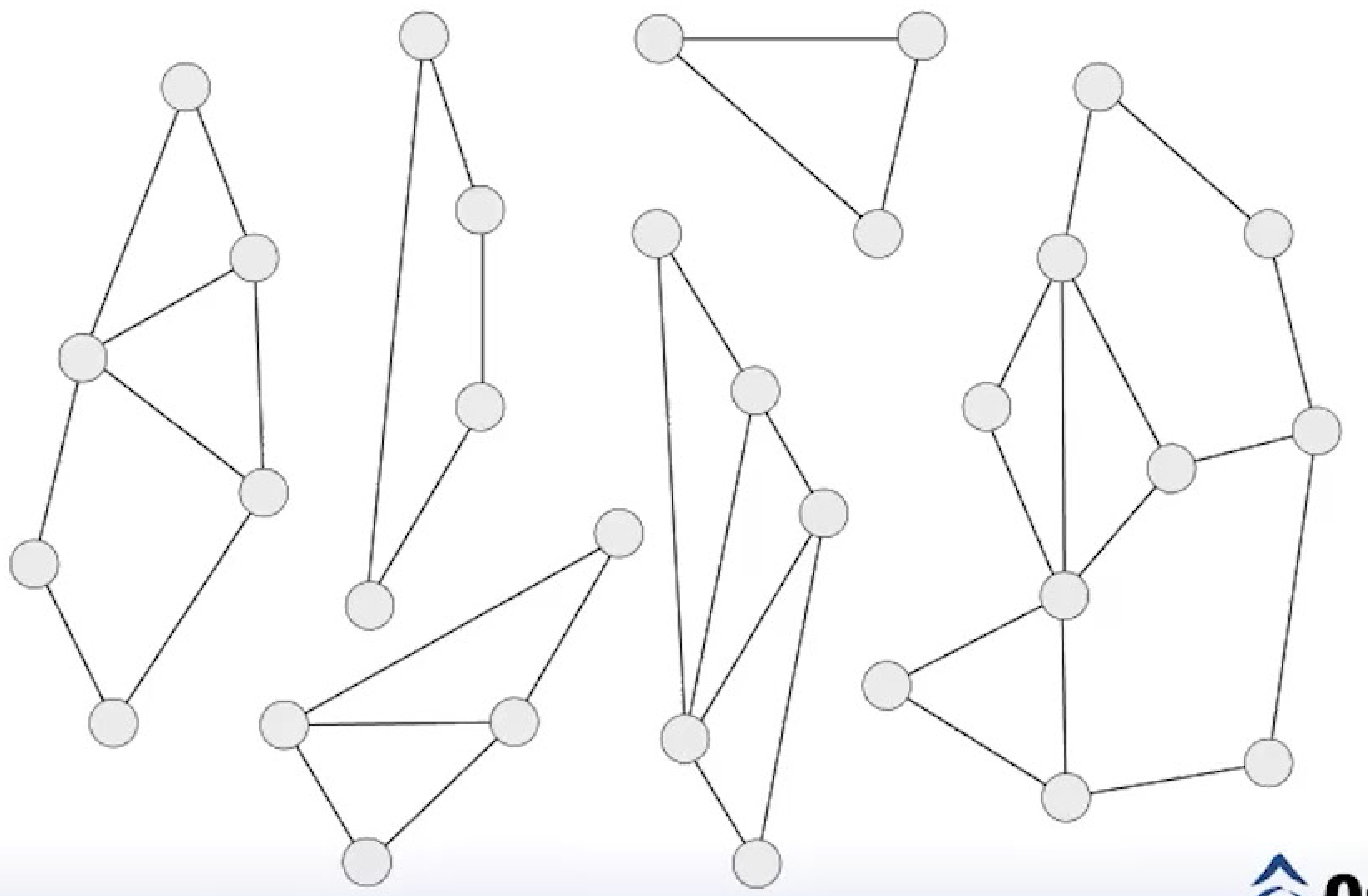


OpenEye SCIENTIFIC

The Mapper Graph

- Building the graph
  - For each one of the cluster minimize:
    - Cycles and MAXDIST constraints

$$\begin{pmatrix} 1 & S_{1,2} & S_{1,3} & \ldots & S_{1,n} \\ S_{2,1} & 1 & S_{2,3} & \ldots & S_{2,n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ S_{n,1} & S_{n,2} & S_{n,3} & \ldots & 1 \end{pmatrix}$$

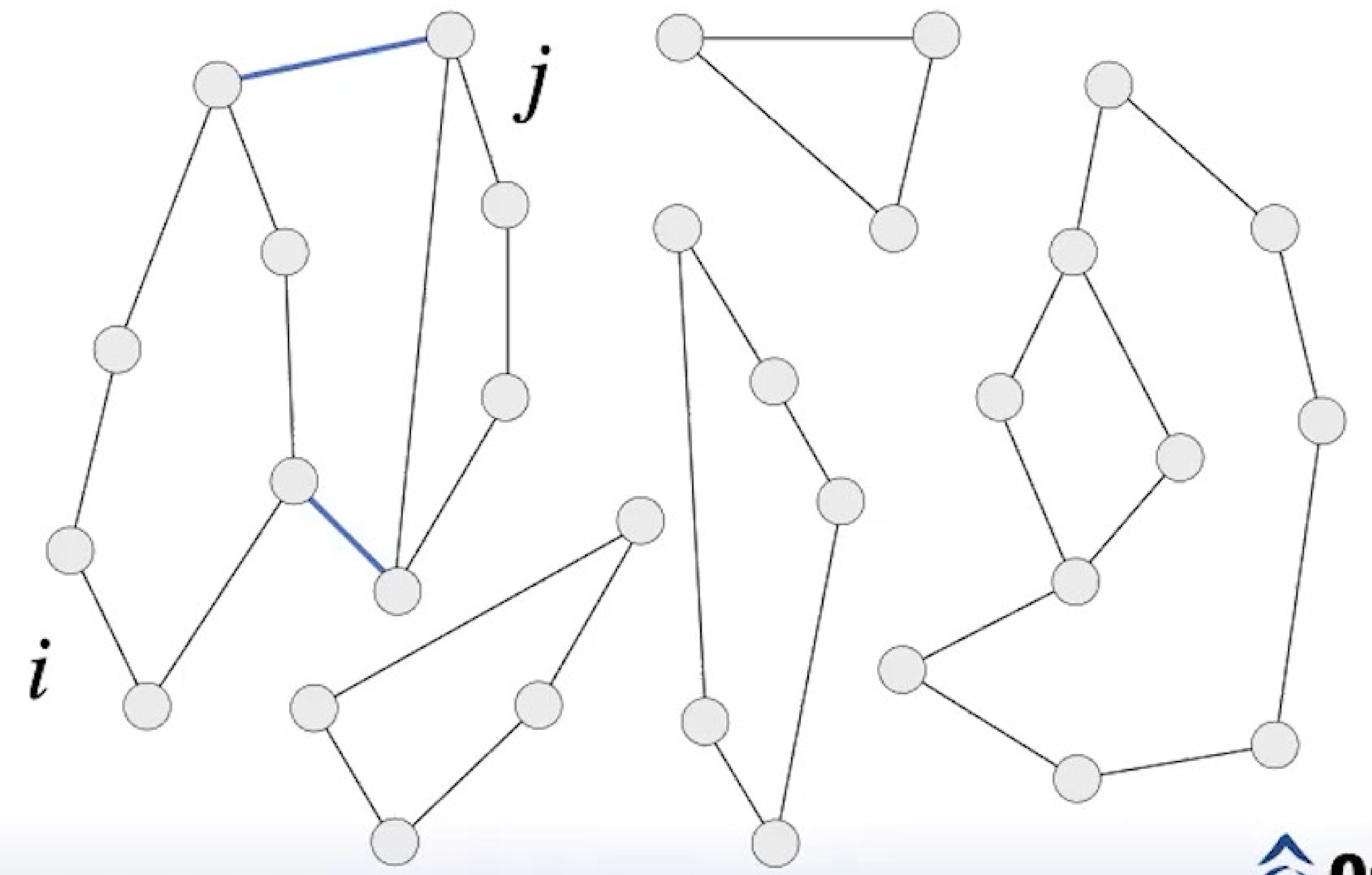


OpenEye SCIENTIFIC

The Mapper Graph

- Building the graph
  - Connect the Subgraphs:
    - (a) $\forall Cl_i, Cl_j \; e_{ij} \mid S_{i,j} = \max\limits_{S_{Ci,Cj}}$
    - (b) $\forall Cl_h, Cl_k \; e_{hk} \mid S_{h,k} = \max\limits_{S_{Ch,Ck}-e_{ij}}$

$$\begin{pmatrix} 1 & S_{1,2} & S_{1,3} & \dots & S_{1,n} \\ S_{2,1} & 1 & S_{2,3} & \dots & S_{2,n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ S_{n,1} & S_{n,2} & S_{n,3} & \dots & 1 \end{pmatrix}$$



NES Protocol mainly followed Gapsys et al.*

(With OpenEye variations)

- GROMACS 2020
- OpenFF 2.0 (Sage) with Amber ff14
- Equilibrium runs done separately
  - Bound and unbound ligand
  - 1X 6 ns, no clustering
  - No NES knowledge embedded
- NES runs: 80 frames with 50ps switching per frame
  - OpenEye alchemical chimeric A/B ligands
  - $\Delta\Delta G$ correlations symmetrized around A→B | B→A
- Schrodinger JACS ‘15 datasets: 8 targets
- Hunt ‘13 Bace dataset and Calabro Thrombin 3-series dataset

*Gapsys et al., *Chem. Sci.*, 2020, **11**, 1140-1152

OpenEye SCIENTIFIC

NES Protocol mainly followed Gapsys et al.*

(With OpenEye variations)

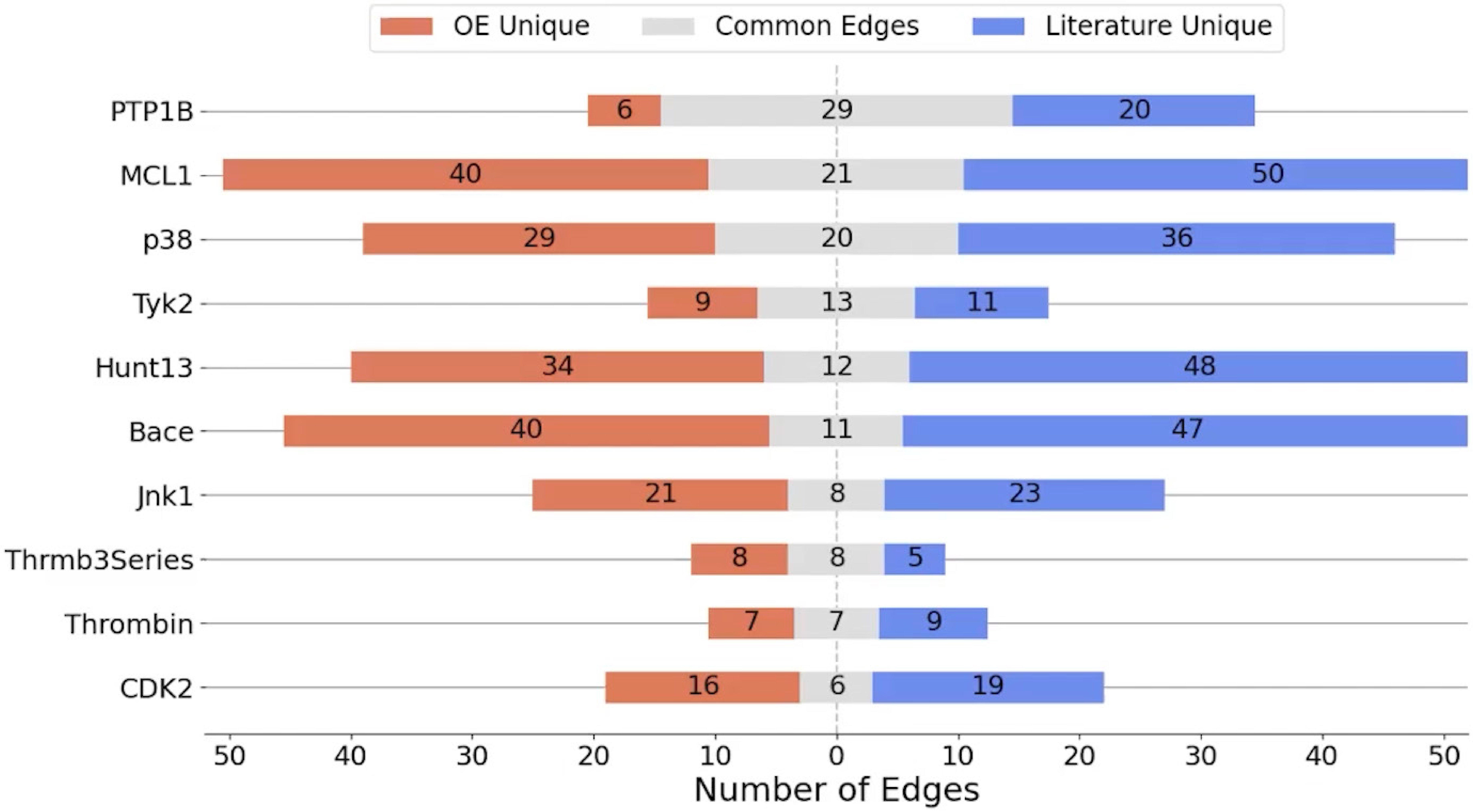
- GROMACS 2020
- OpenFF 2.0 (Sage) with Amber ff14
- Equilibrium runs done separately
  - Bound and unbound ligand
  - 1X 6 ns, no clustering
  - No NES knowledge embedded
- NES runs: 80 frames with 50ps swi
  - OpenEye alchemical chimeric A/B liga
  - ΔΔG correlations symmetrized around A→B | B→A
- Schrodinger JACS '15 datasets: 8 targets
- Hunt '13 Bace dataset and Calabro Thrombin 3-series dataset

Compare:
- **OE Mapper**
- **Literature maps (FEP+)**

OpenEye SCIENTIFIC

*Gapsys et al., *Chem. Sci.*, 2020, **11**, 1140-1152

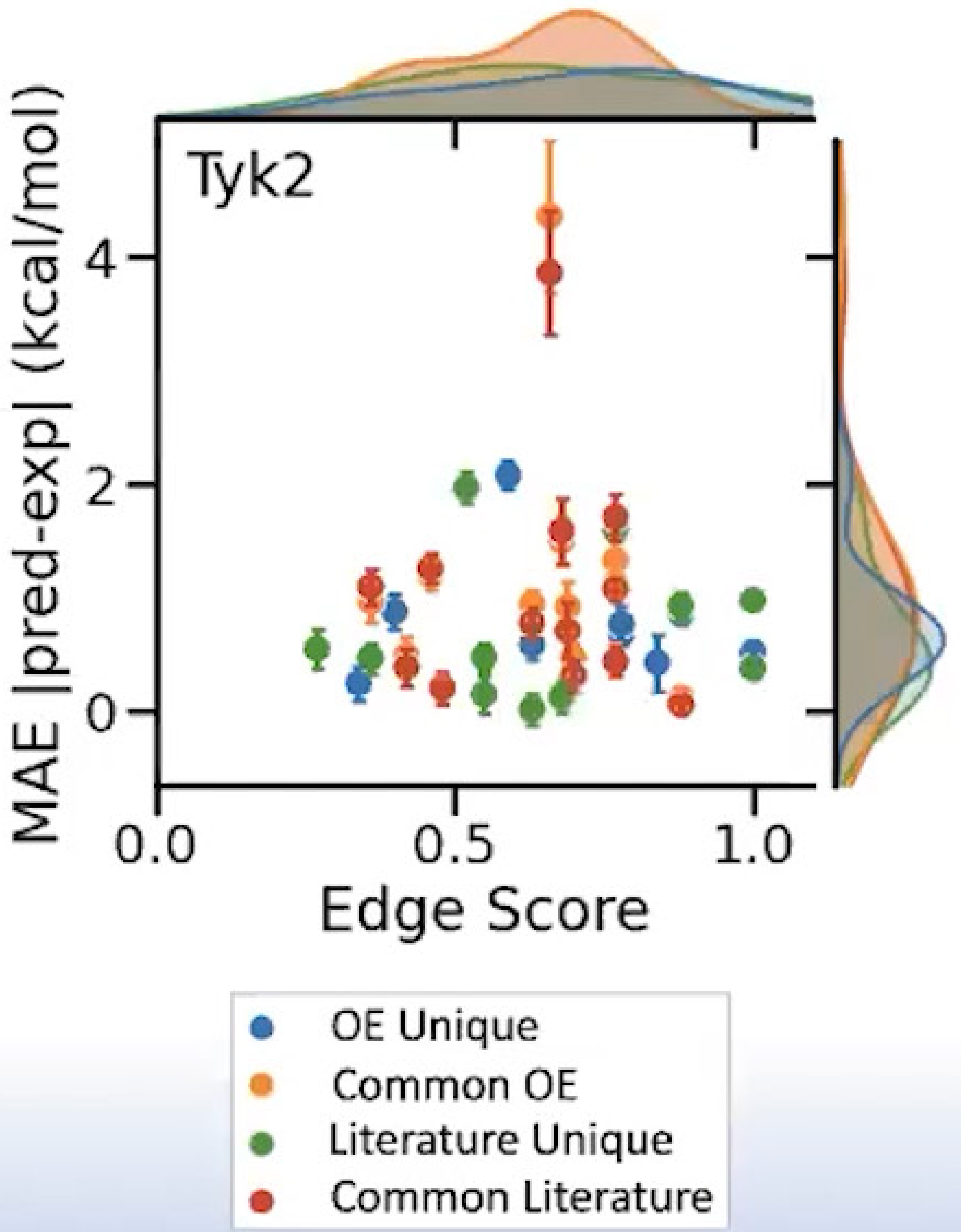
Common Edges Diverging Diagram



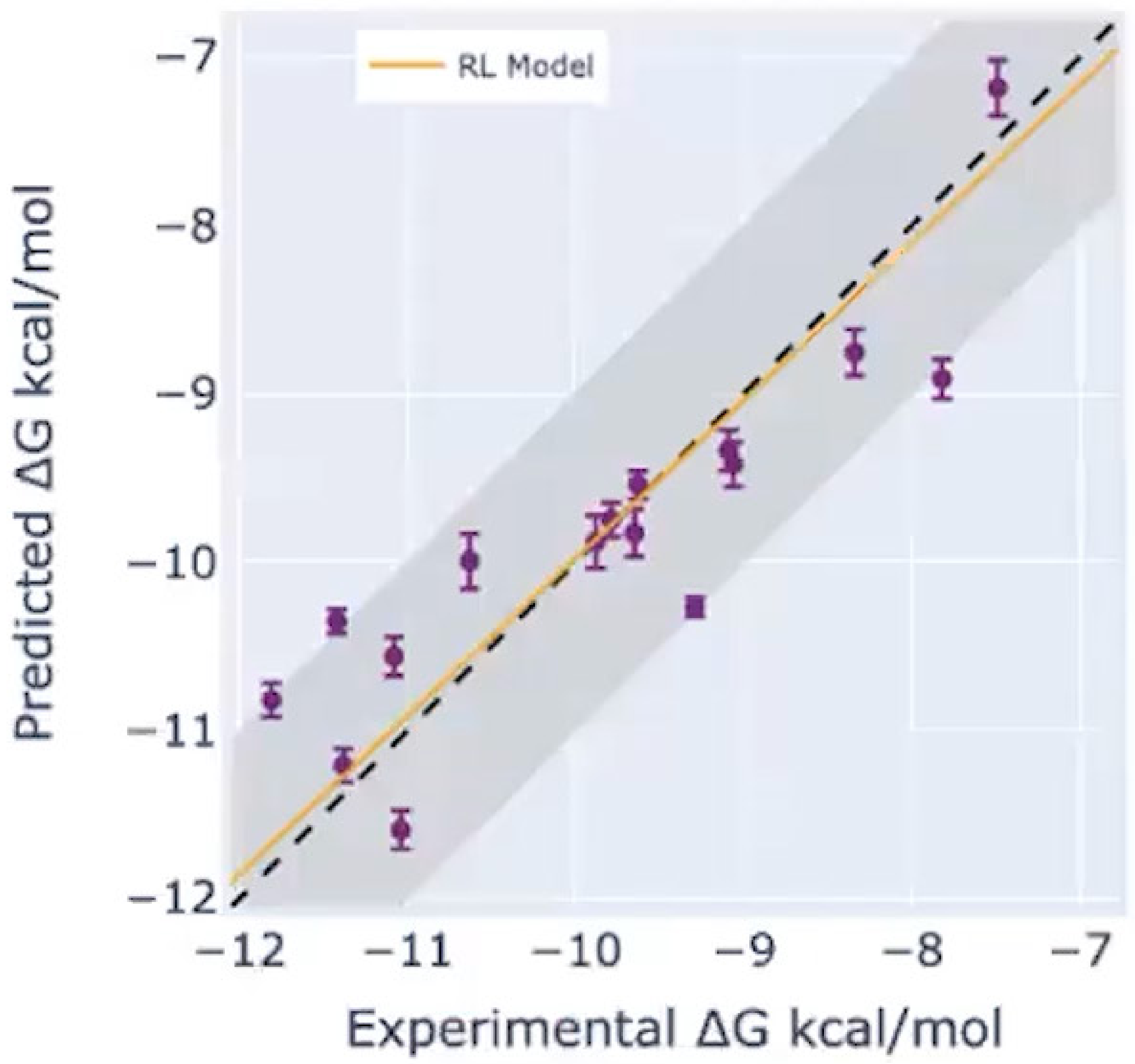
OpenEye SCIENTIFIC

Tyk2

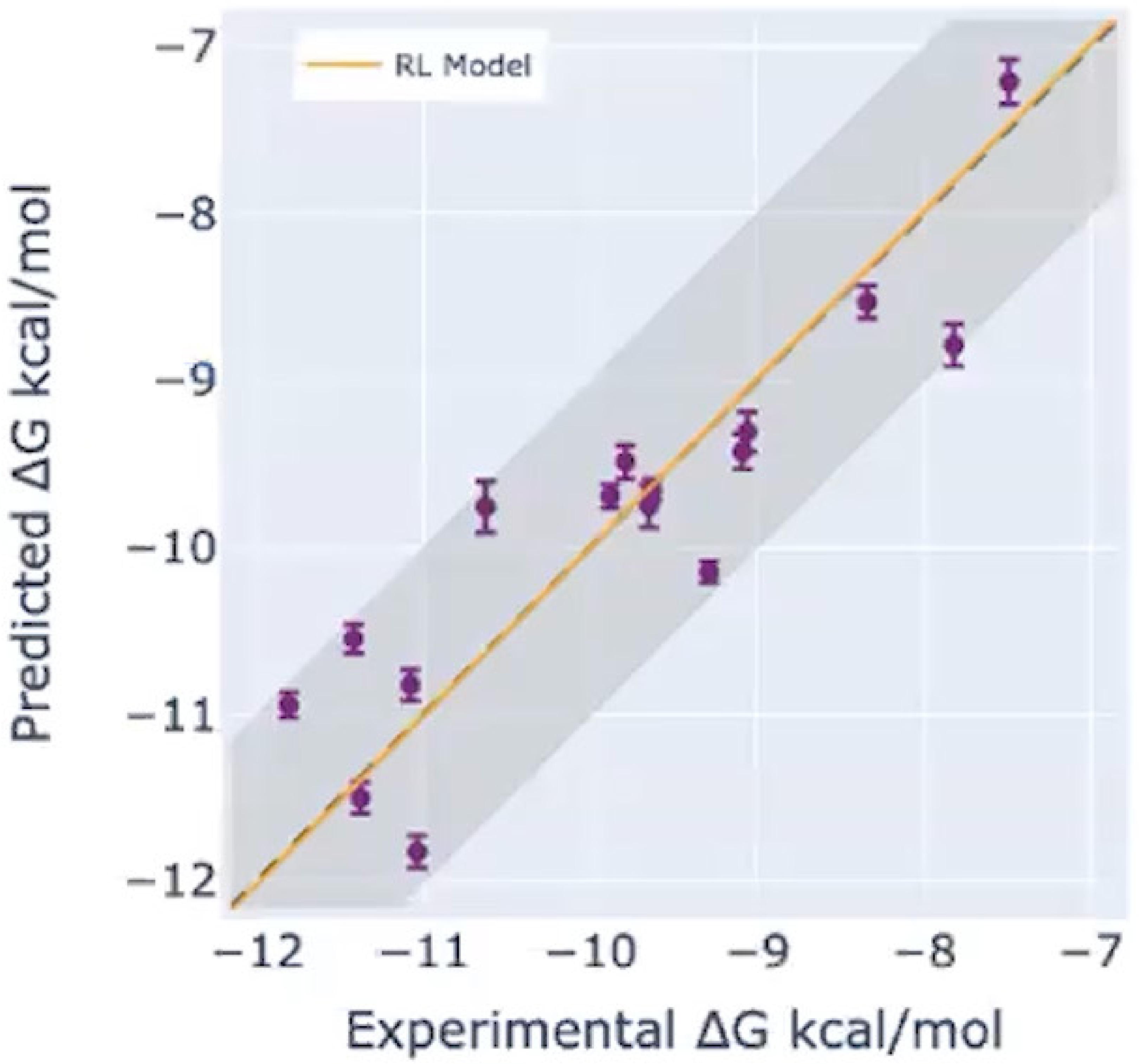
Metric	OE Mapper	Literature Mapper
Ligands	16	
Edges	22	24



OE Mapper



Literature Mapper



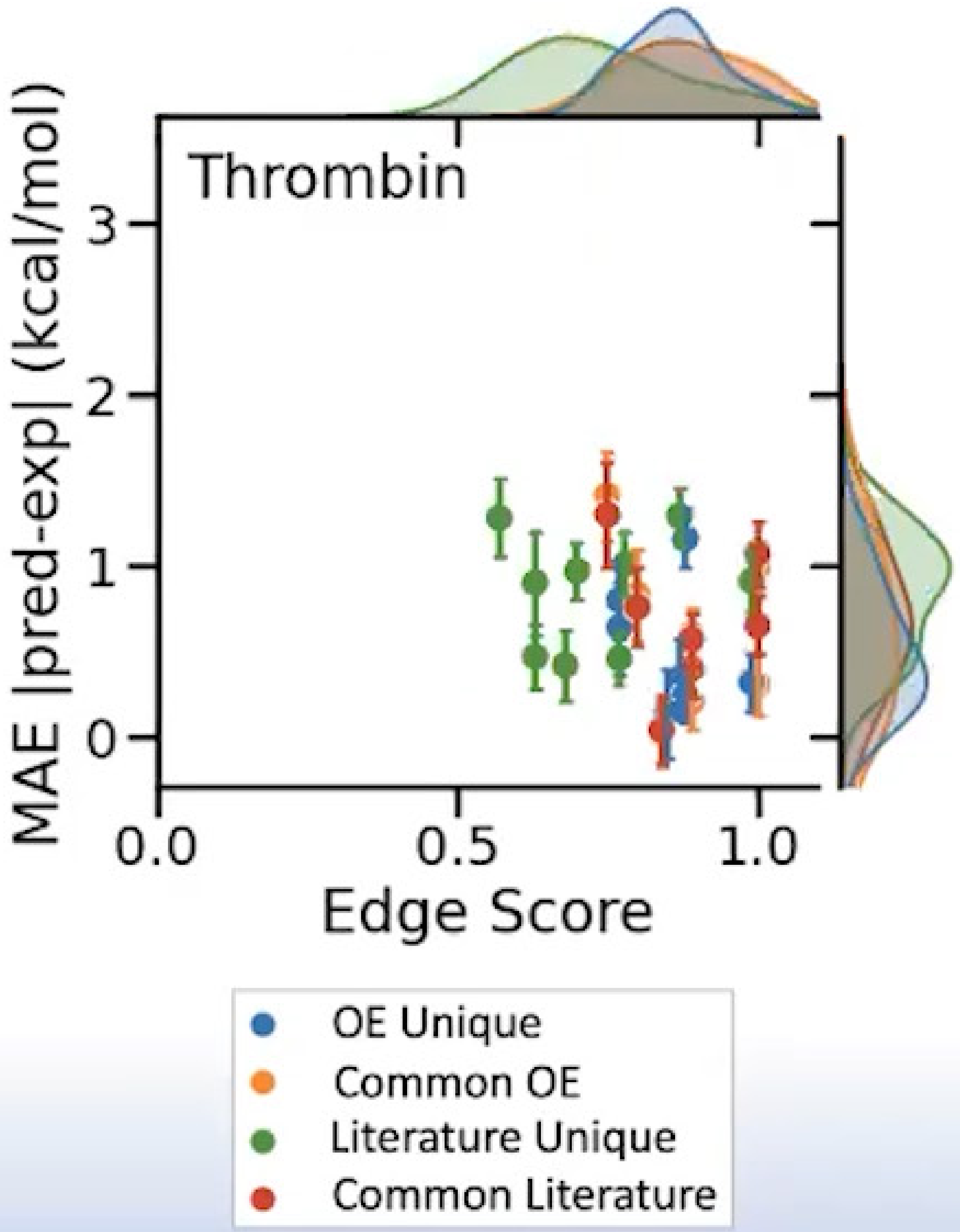
Metric	OE Mapper	Literature Mapper
Pearson's r^2	0.783 ± 0.075	0.804 ± 0.074
Kendall's τ	0.750 ± 0.102	0.733 ± 0.114
MAE[a]	0.477 ± 0.092	0.454 ± 0.088
RMAE[b]	0.456 ± 0.091	0.434 ± 0.097

[a]Mean Absolute Error in kcal/mol.
[b]MAE divided by the Mean Absolute Deviation of Experimental ΔG.

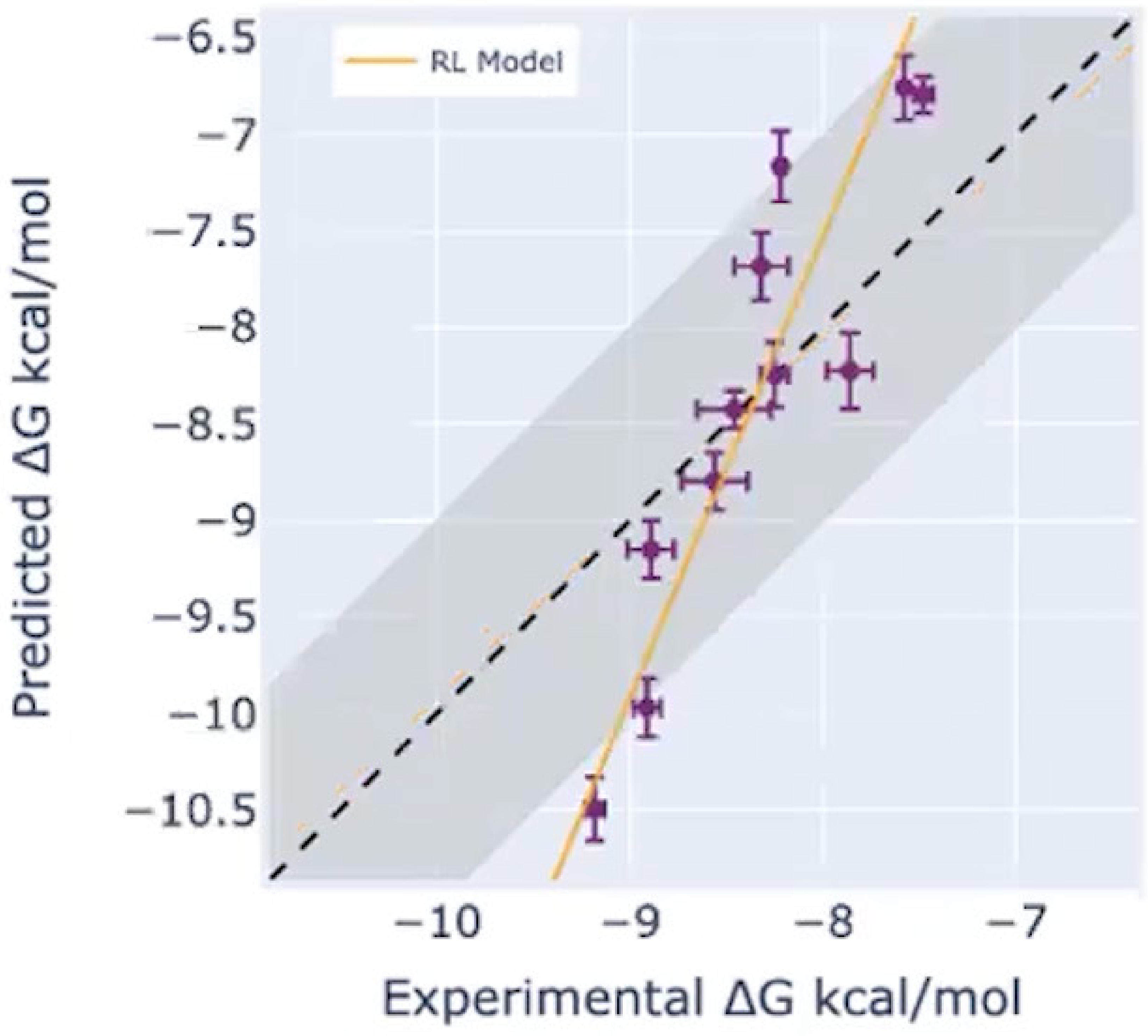
OpenEye SCIENTIFIC

Thrombin

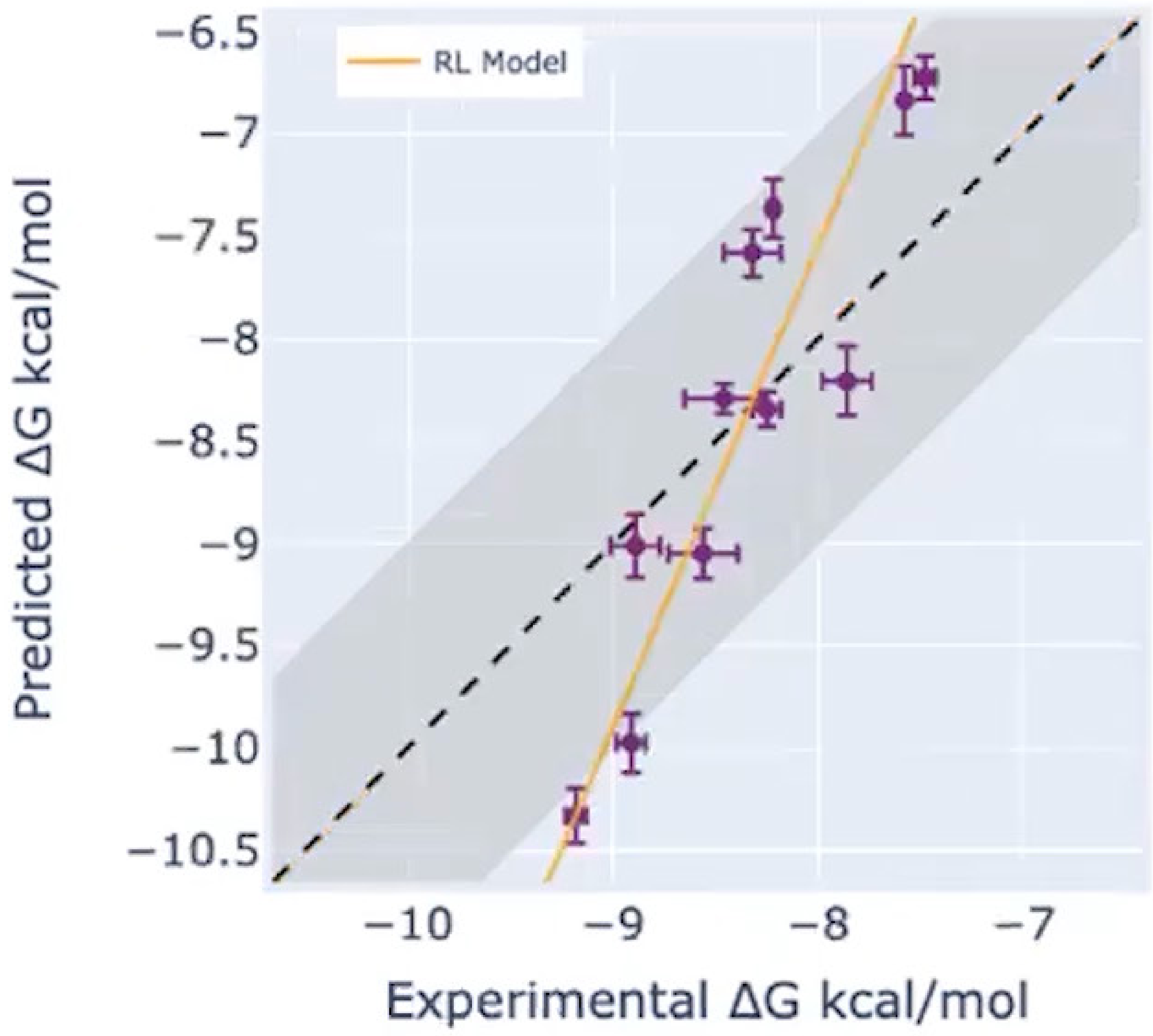
Metric	OE Mapper	Literature Mapper
Ligands	11	
Edges	14	16



OE Mapper



Literature Mapper



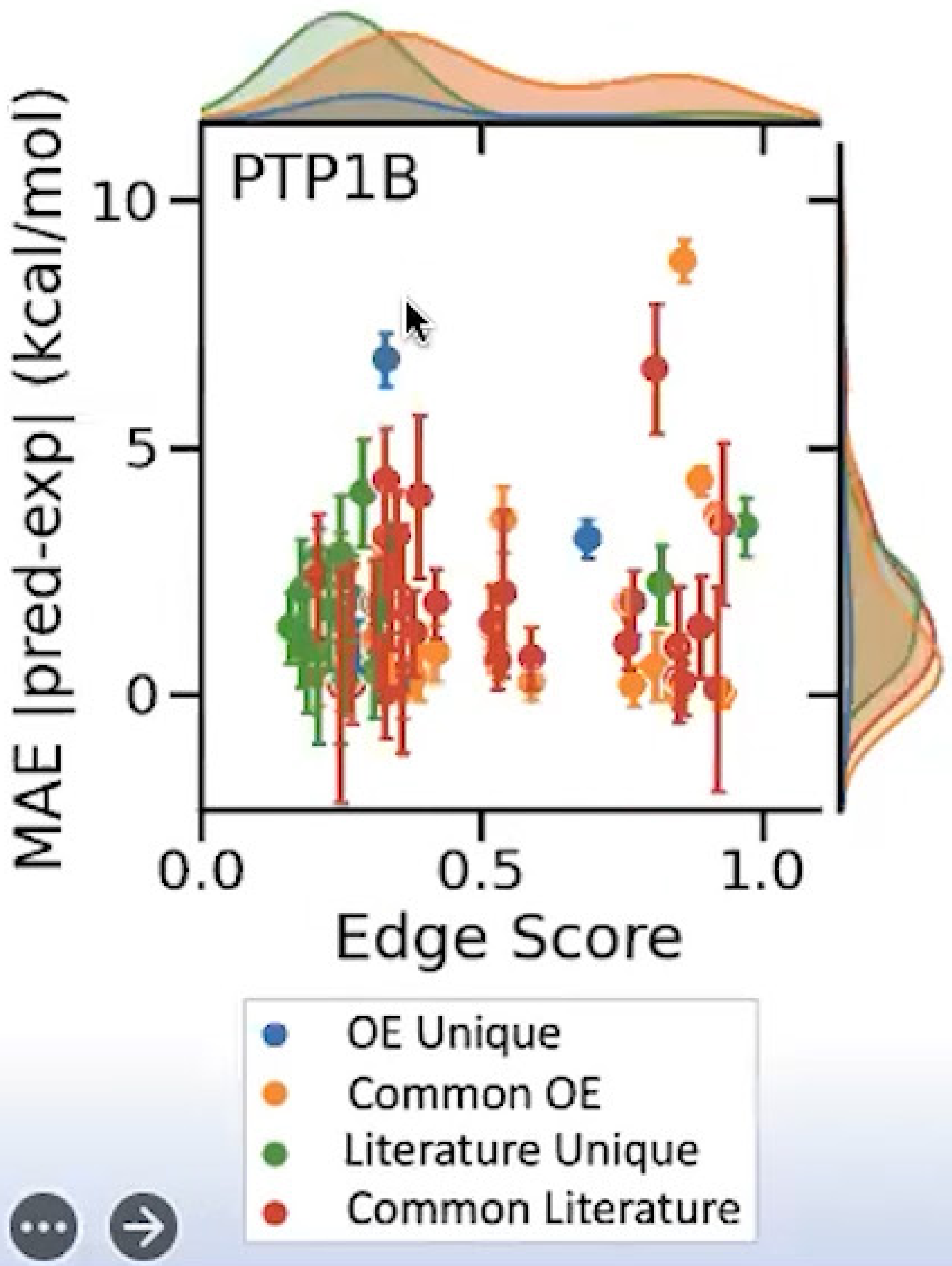
Metric	OE Mapper	Literature Mapper
Pearson's r^2	0.827 ± 0.113	0.824 ± 0.114
Kendall's τ	0.855 ± 0.119	0.818 ± 0.127
MAE[a]	0.588 ± 0.125	0.594 ± 0.113
RMAE[b]	1.399 ± 0.445	1.412 ± 0.381

[a]Mean Absolute Error in kcal/mol.
[b]MAE divided by the Mean Absolute Deviation of Experimental ΔG.

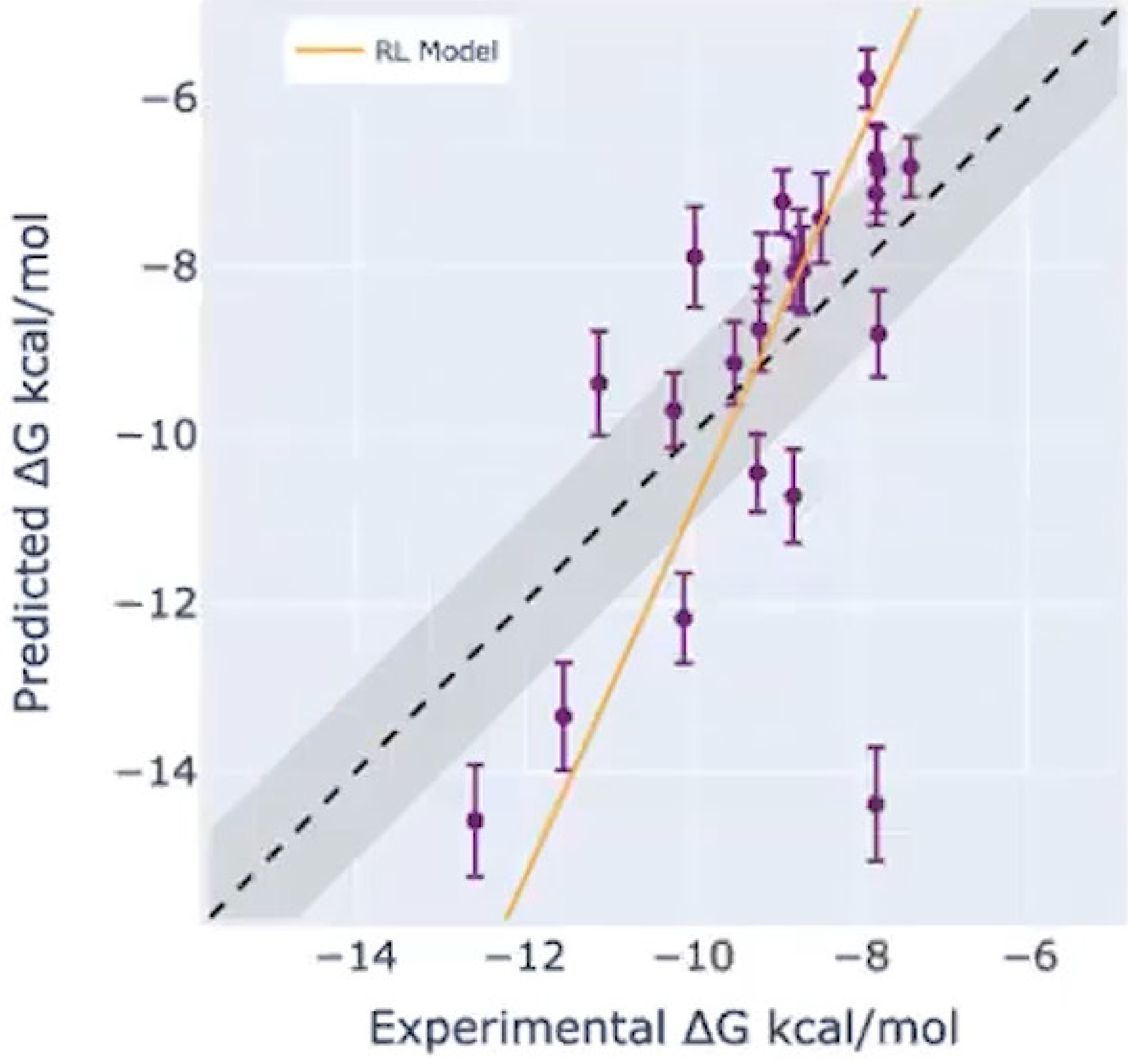
OpenEye SCIENTIFIC

PTP1B

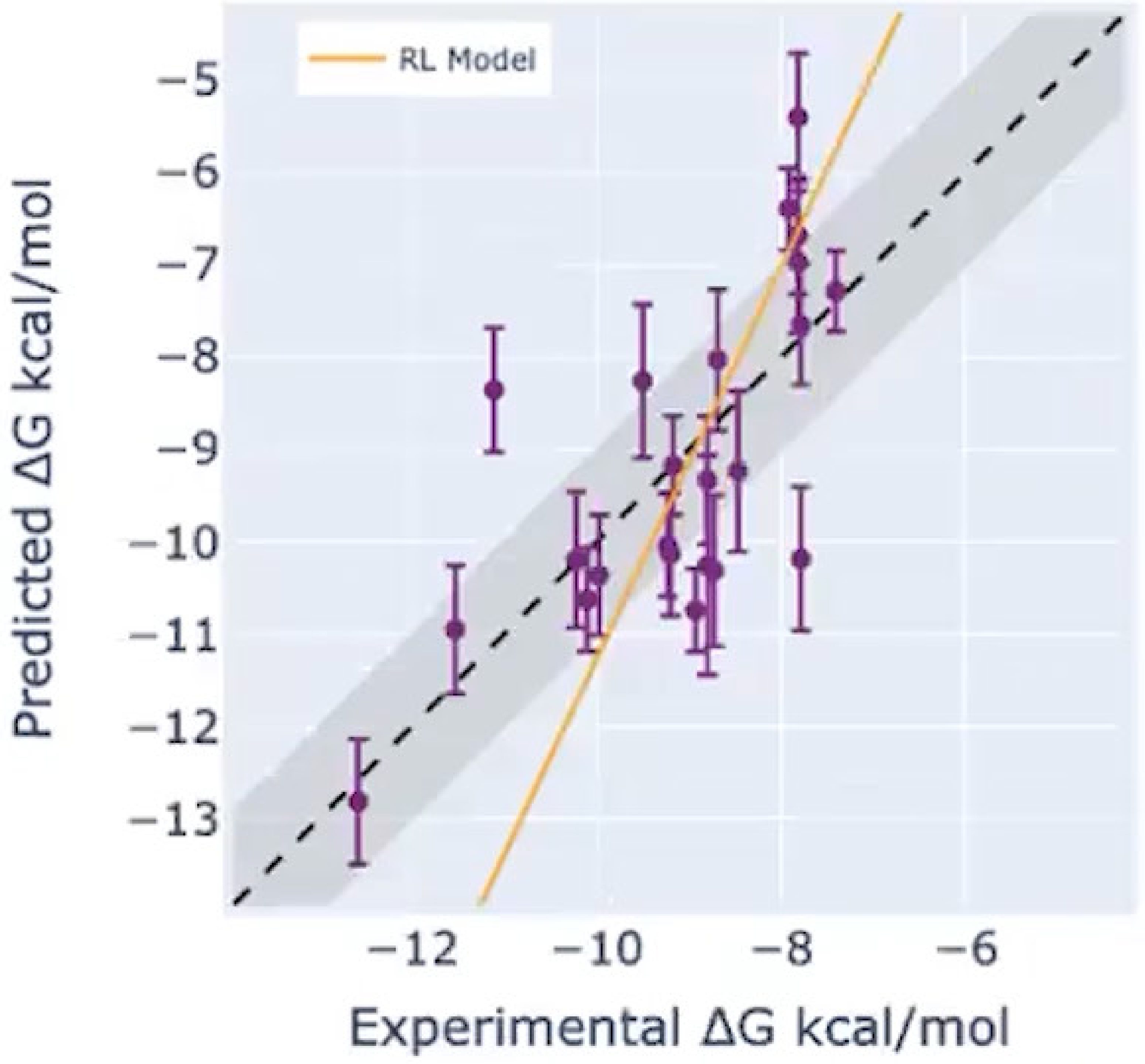
Metric	OE Mapper	Literature Mapper
Ligands	23	
Edges	35	49



OE Mapper



Literature Mapper



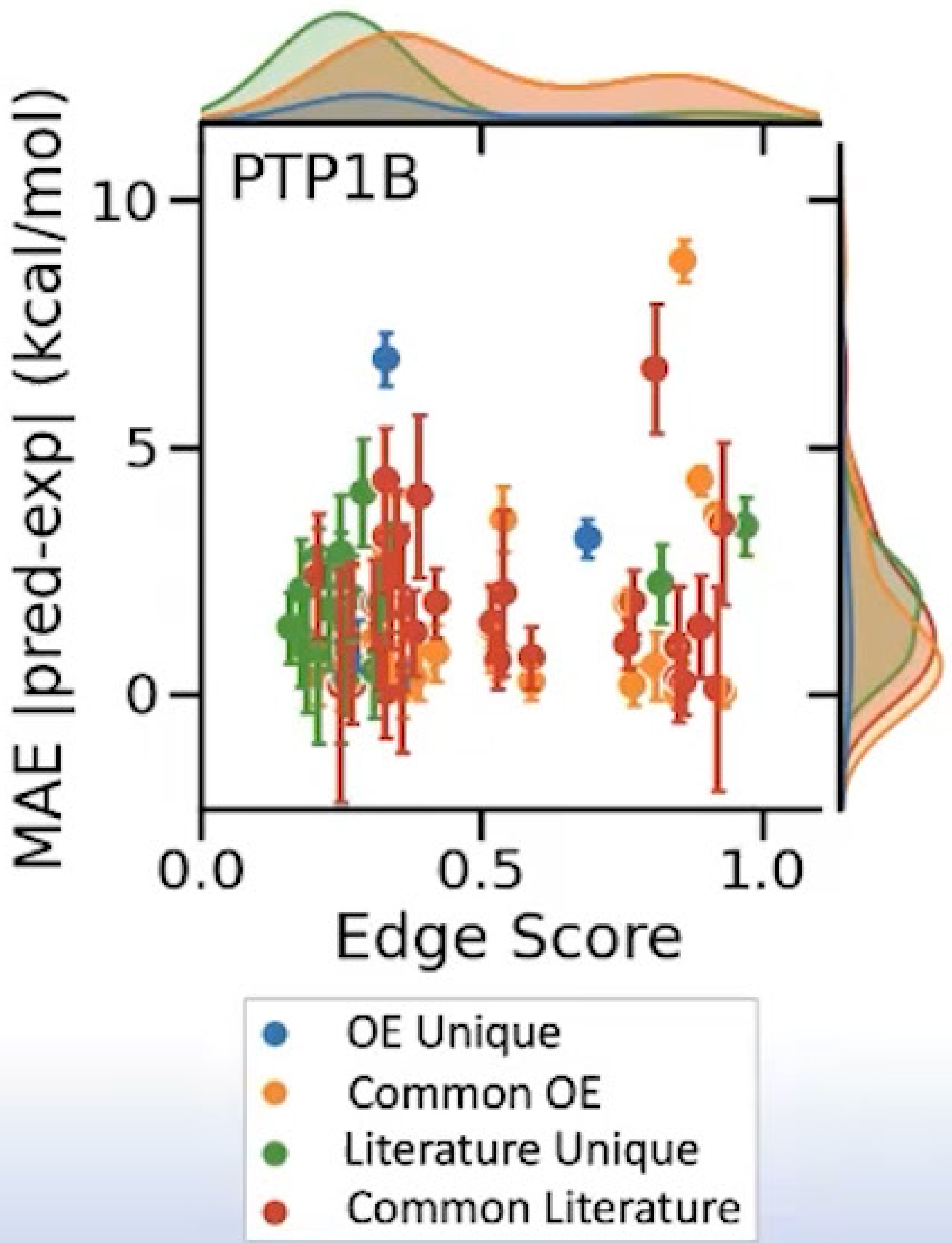
Metric	OE Mapper	Literature Mapper
Pearson's r^2	0.385 ± 0.248	0.481 ± 0.170
Kendall's τ	0.503 ± 0.147	0.487 ± 0.133
MAE[a]	1.445 ± 0.251	0.995 ± 0.157
RMAE[b]	1.437 ± 0.373	0.989 ± 0.284

[a]Mean Absolute Error in kcal/mol.
[b]MAE divided by the Mean Absolute Deviation of Experimental ΔG.

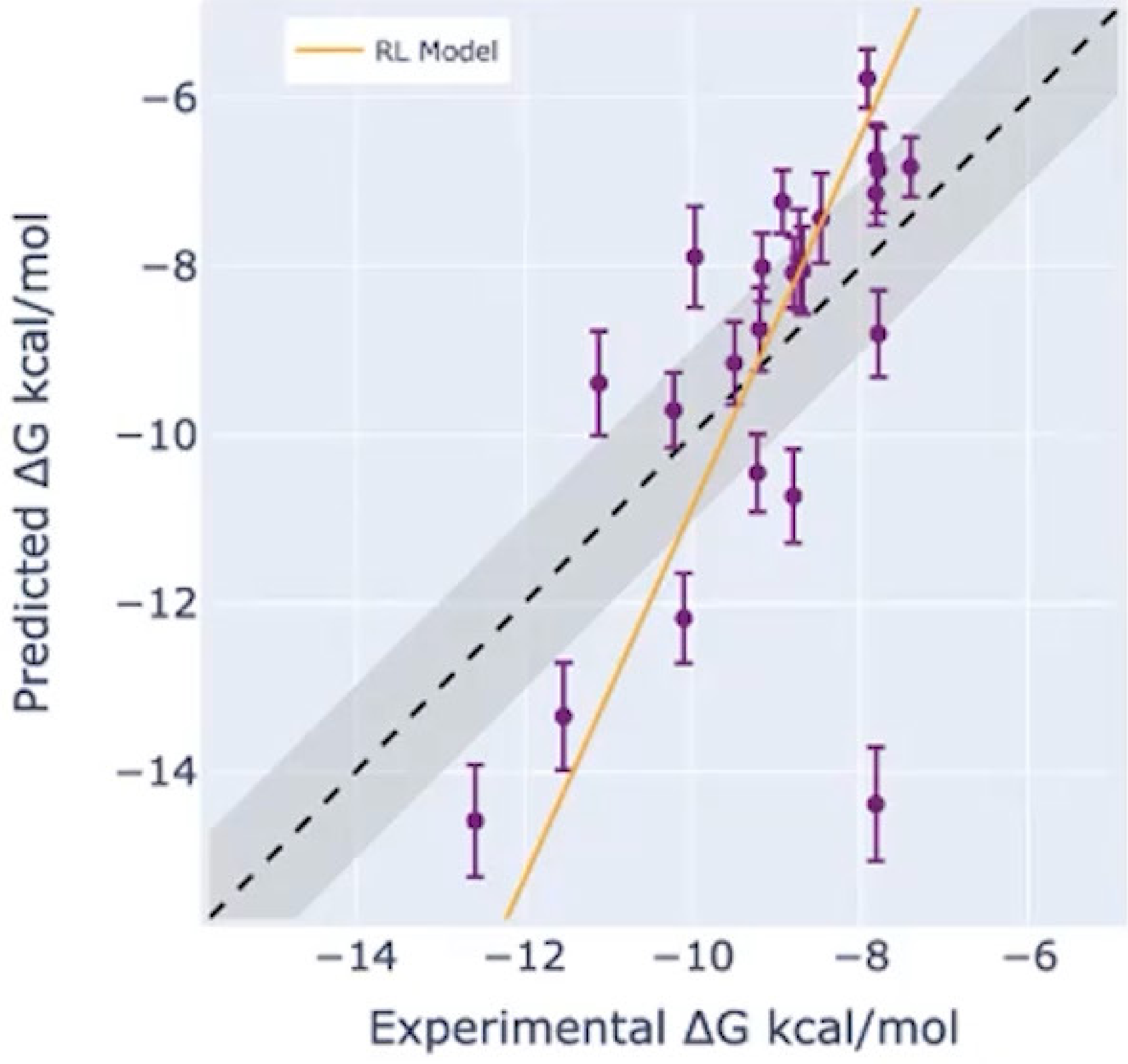
OpenEye SCIENTIFIC

PTP1B

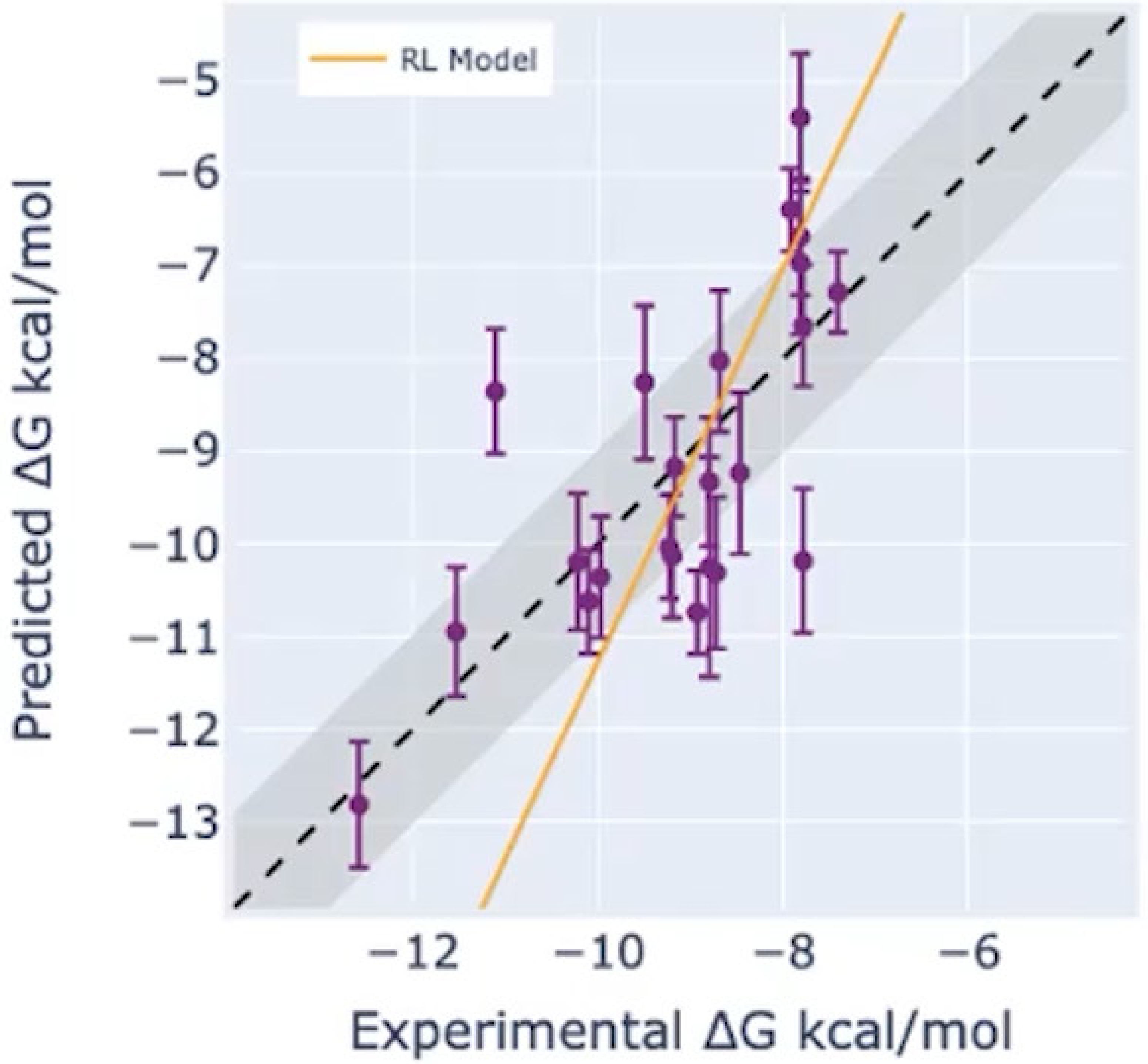
Metric	OE Mapper	Literature Mapper
Ligands	23	
Edges	35	49



OE Mapper



Literature Mapper

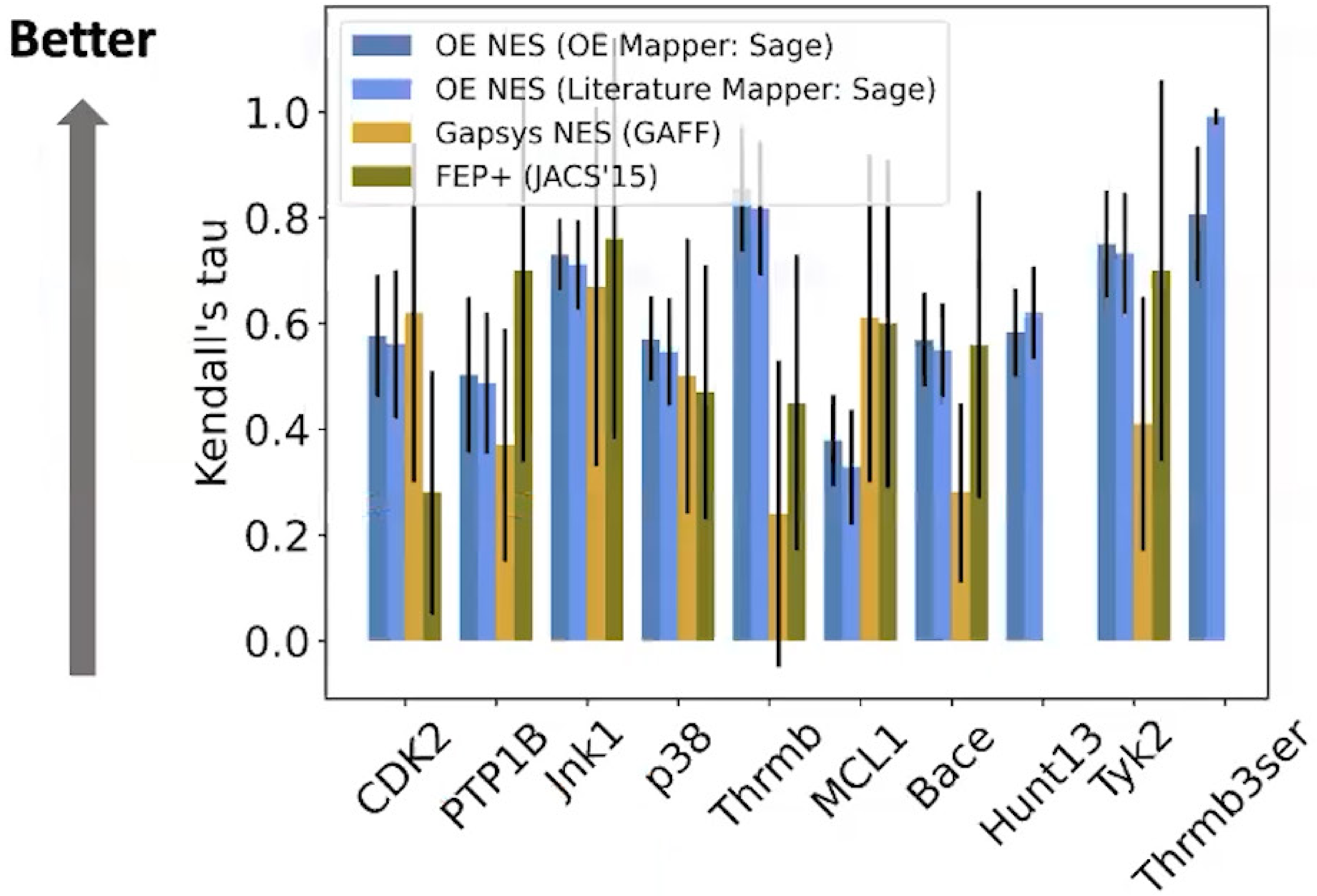


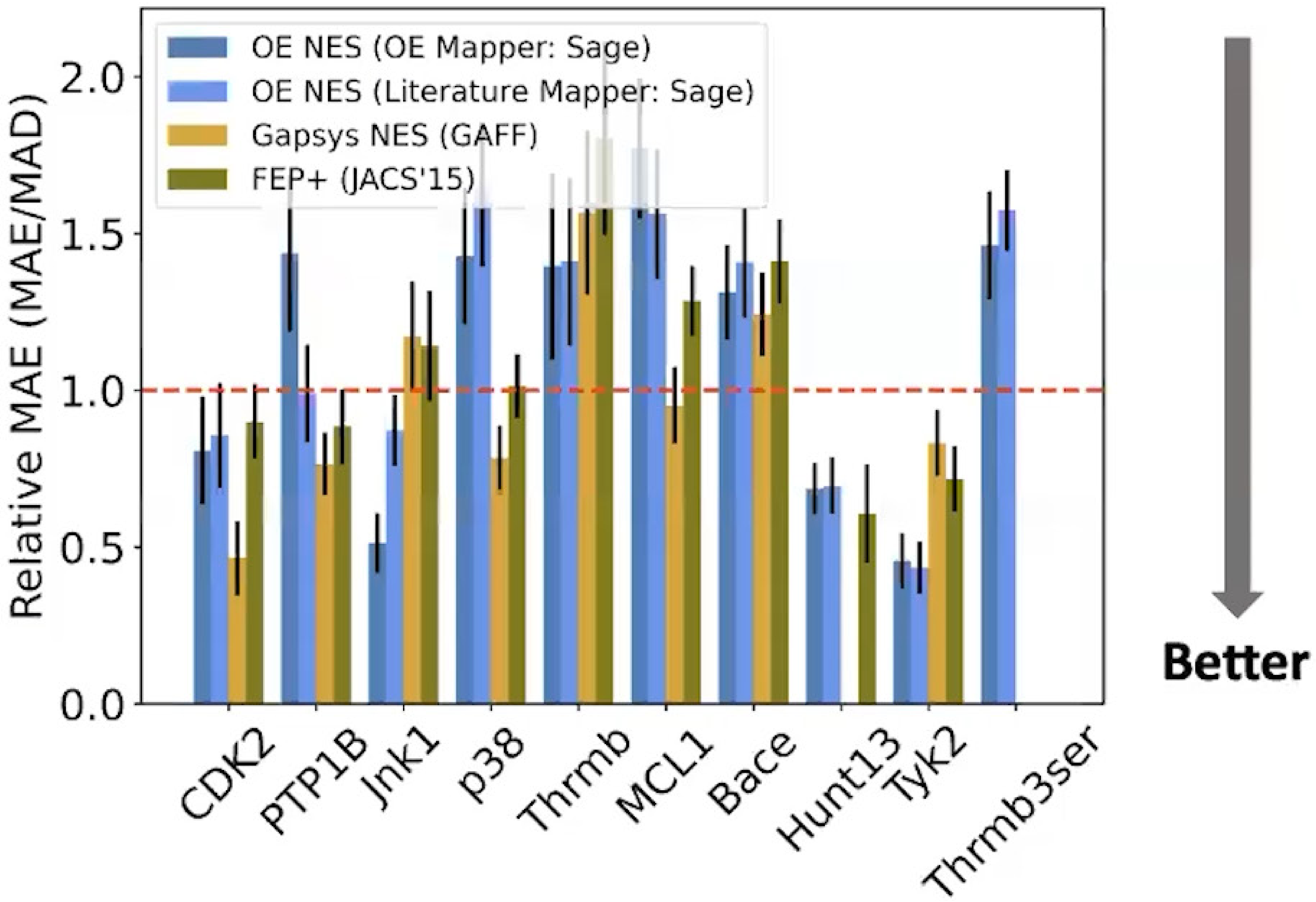
Metric	OE Mapper	Literature Mapper
Pearson's r²	0.385 ± 0.248	0.481 ± 0.170
Kendall's τ	0.503 ± 0.147	0.487 ± 0.133
MAE[a]	1.445 ± 0.251	0.995 ± 0.157
RMAE[b]	1.437 ± 0.373	0.989 ± 0.284

[a]Mean Absolute Error in kcal/mol.
[b]MAE divided by the Mean Absolute Deviation of Experimental ΔG.

OpenEye SCIENTIFIC

Direct Predictions of ΔG : 9 Datasets





- OE NES has comparable accuracy to literature RBFE benchmarks

Conclusions

- The starting OE Mapper implementation performs as expected compared to literature maps
- Still the edge scoring sometimes does not reflect the accuracy of the calculation

Plans: include equilibrium information in the mapper scoring

Acknowledgements

- Christopher Bayly
- Agnes Huang
- Hyesu Jang
- Geoff Skillman


Thank You!
OpenEye
SCIENTIFIC

For more information, please contact:

sales@eyesopen.com | info@eyesopen.com

www.openeye.inc

+1-505-473-7385

OpenEye SCIENTIFIC


Click Here
To Explore Our
On-Demand Webinar Library
OpenEye
SCIENTIFIC