Jenke Scheen

A data-driven approach to relative Free Energy Perturbation reliability predictions for alchemical free energy calculations in drug design

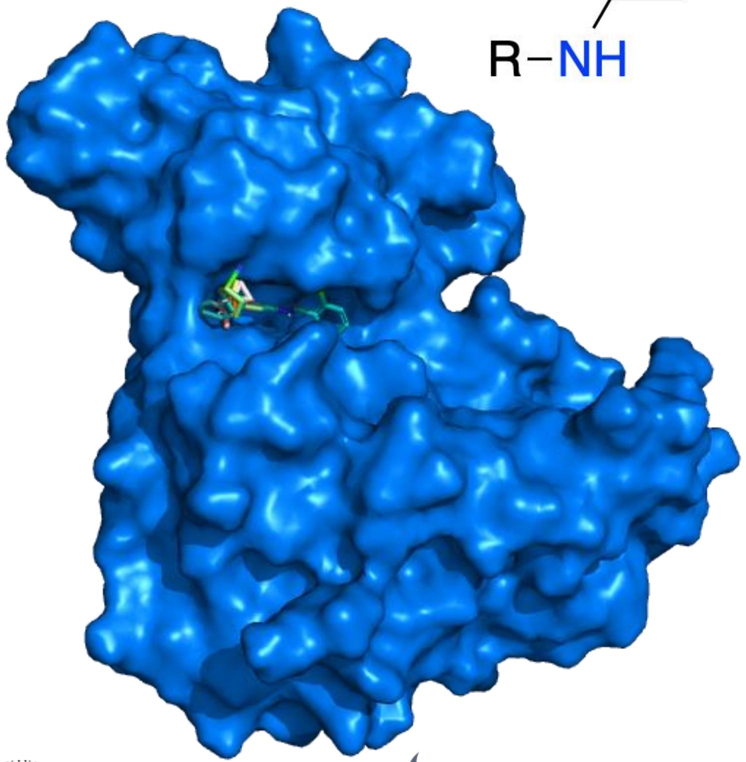
MGMS
Young Modellers' Forum
2021/22
Friday, 11th February, 2022

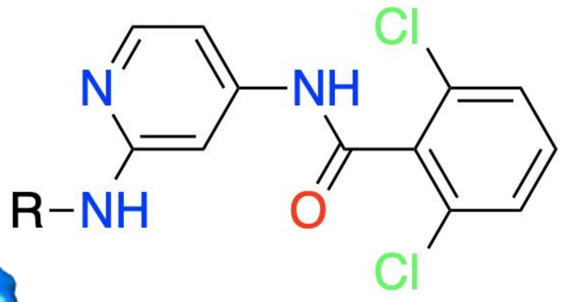
@Jenkescheen

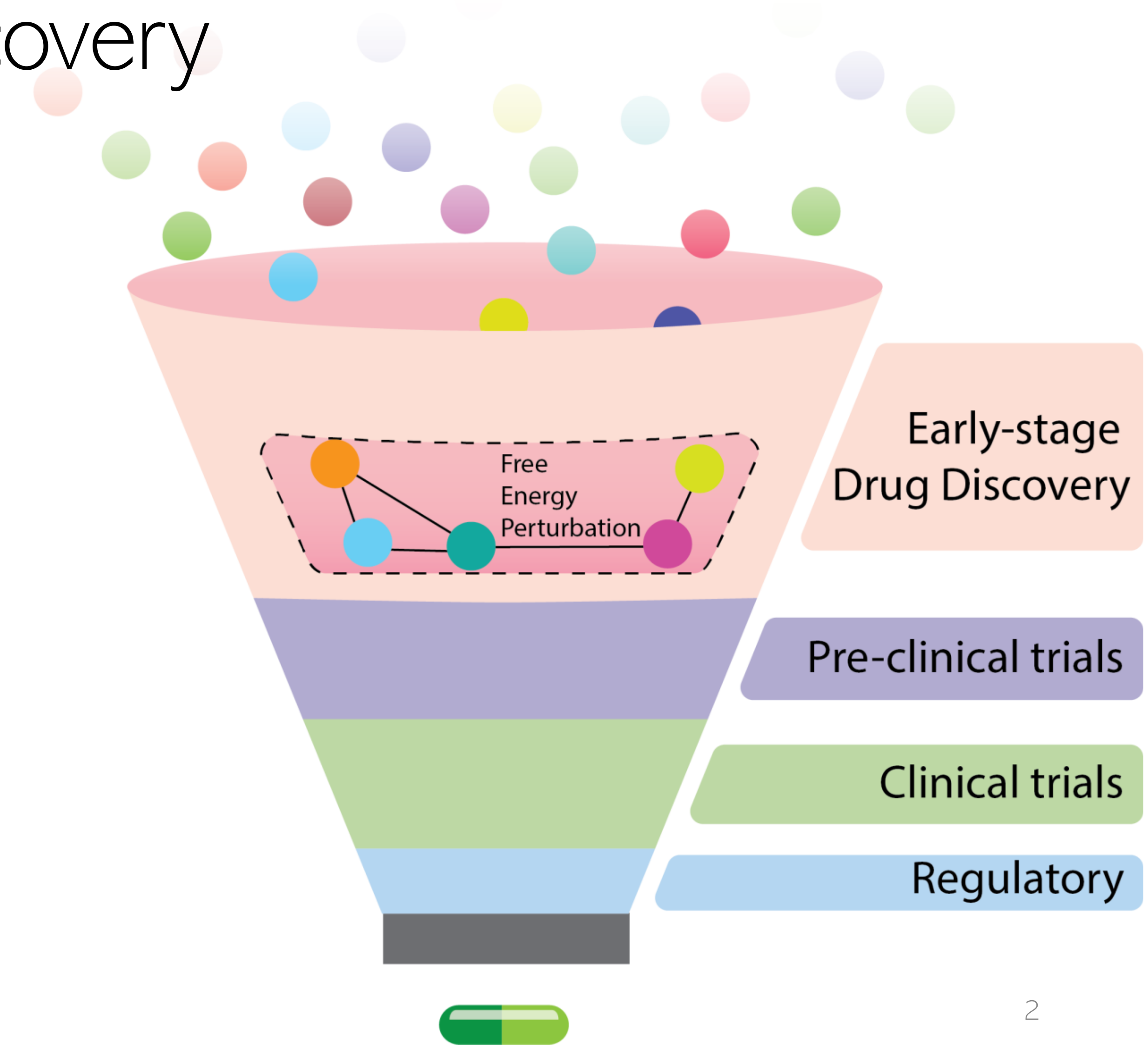
j.scheen@sms.ed.ac.uk

THE UNIVERSITY of EDINBURGH
cresset
EaStCHEM

Free Energy Perturbation (FEP) is a a helpful tool in supporting ligand optimization in early-stage drug discovery

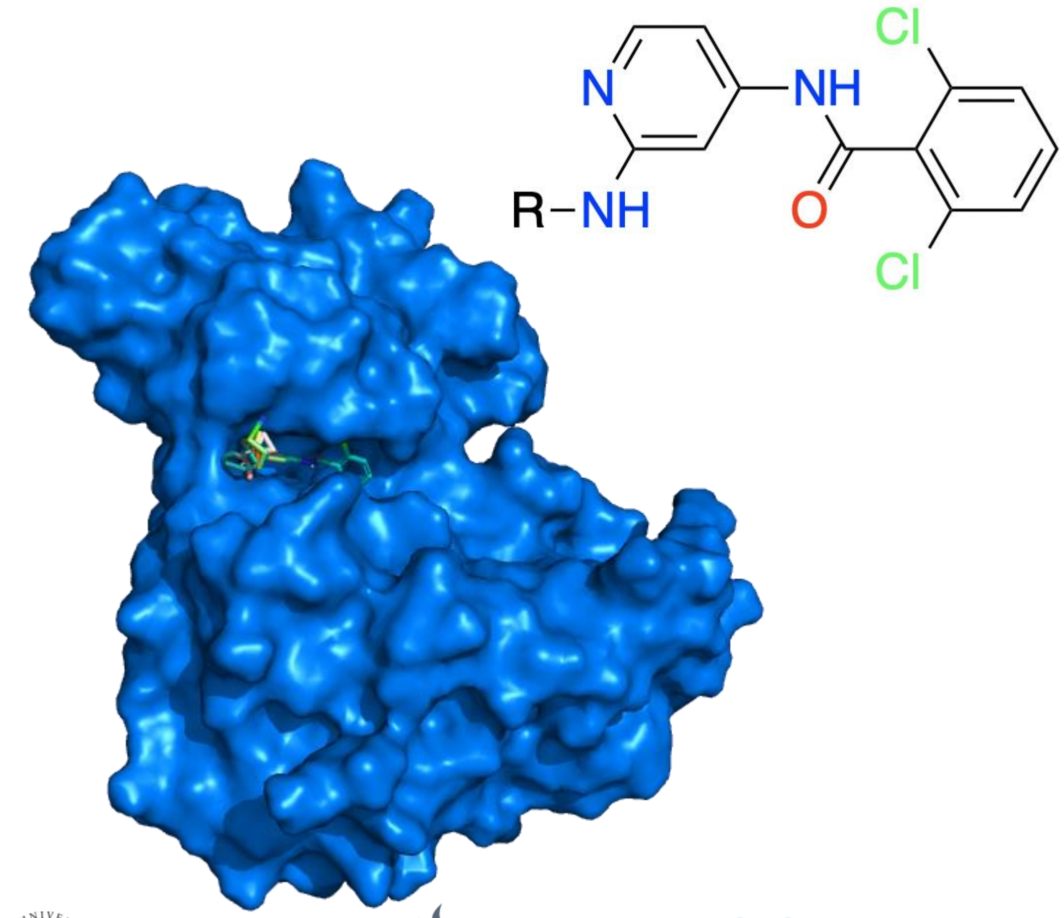


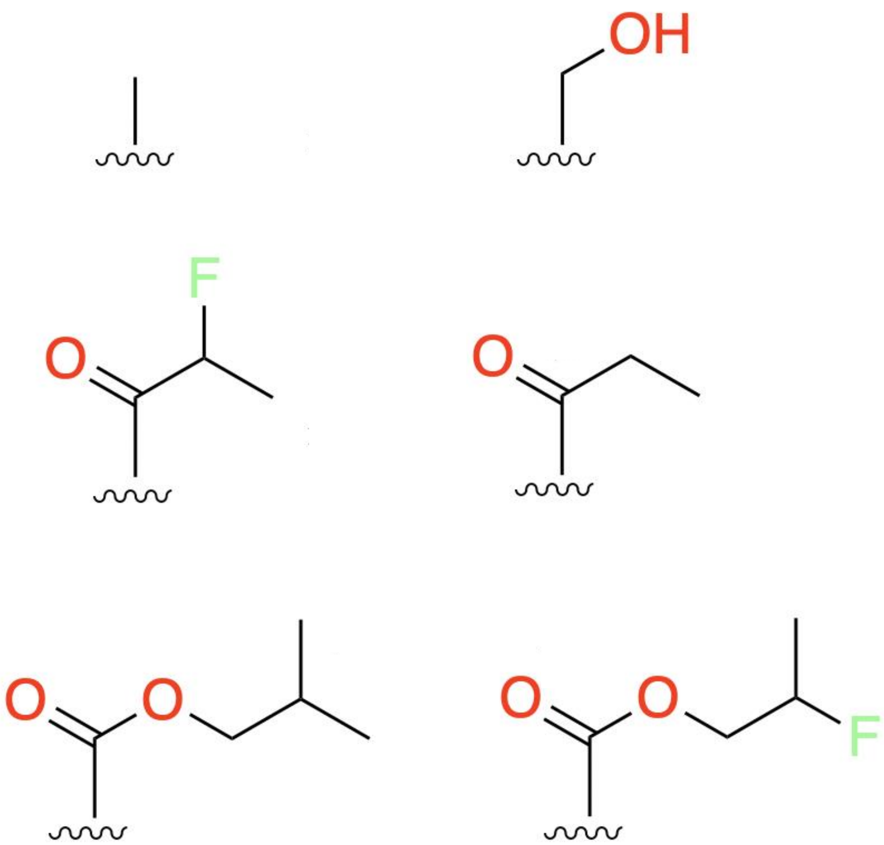




THE UNIVERSITY of EDINBURGH cresset EaStCHEM

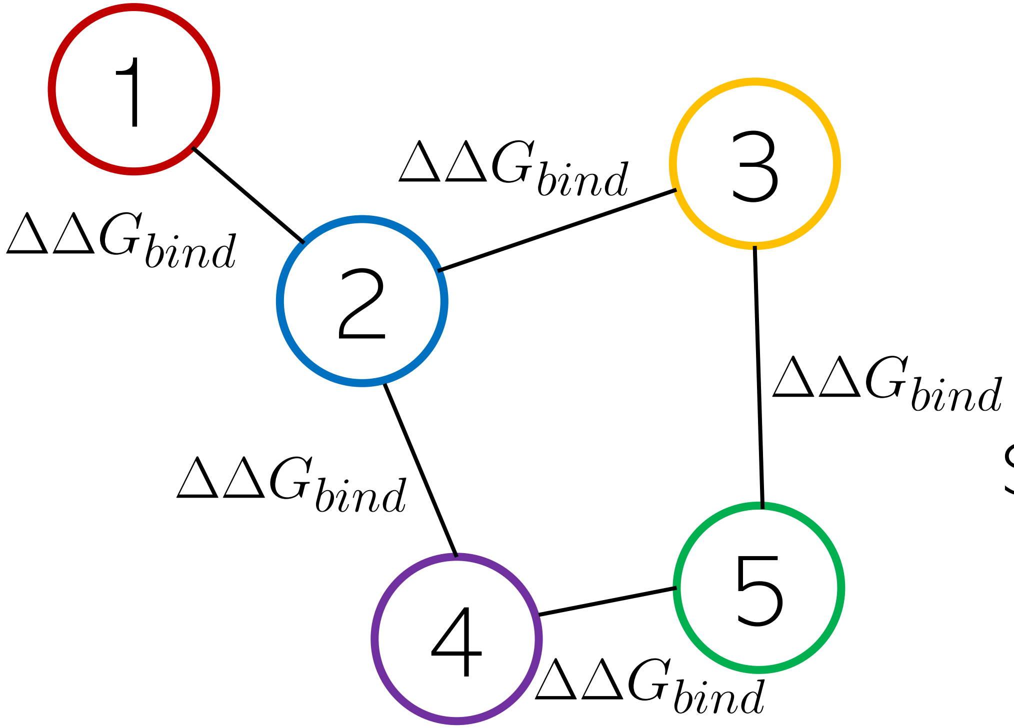
Free Energy Perturbation (FEP) can predict the relative binding free energy between ligands





THE UNIVERSITY of EDINBURGH cresset EaStCHEM

FEP calculations on ligand series require a graph of transformations through the series



Weighted Least Squares regression

$\Delta\Delta G_{bind,\ ligand\ 1} = ..$

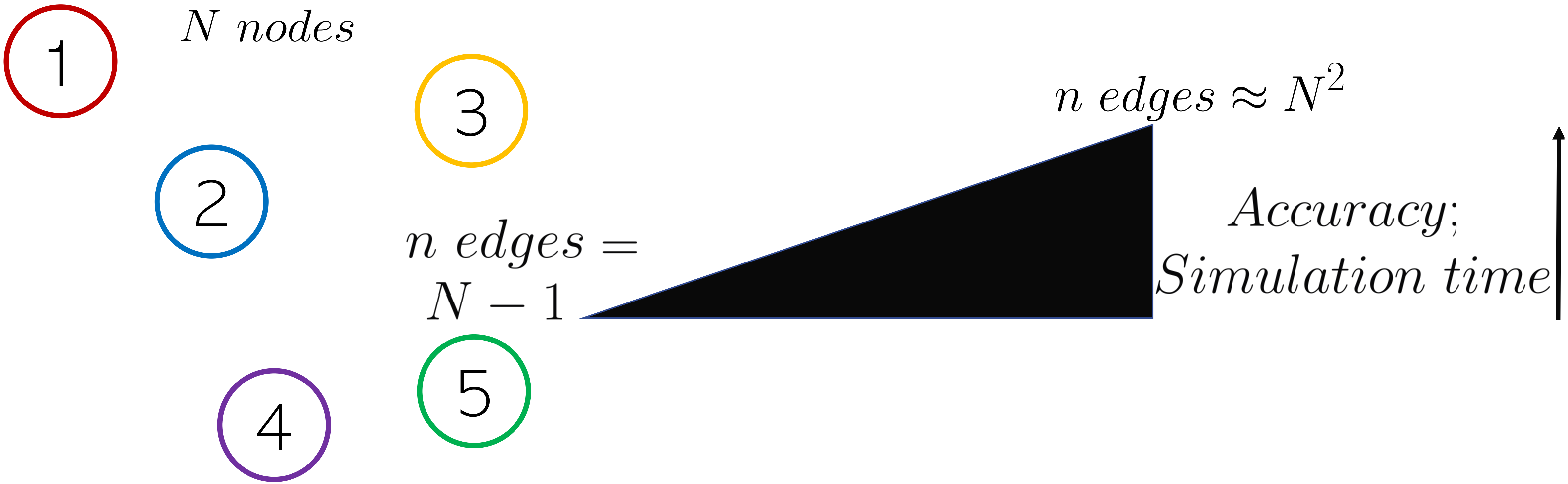
$\Delta\Delta G_{bind,\ ligand\ 2} = ..$

$\Delta\Delta G_{bind,\ ligand\ 3} = ..$

$\Delta\Delta G_{bind,\ ligand\ 4} = ..$

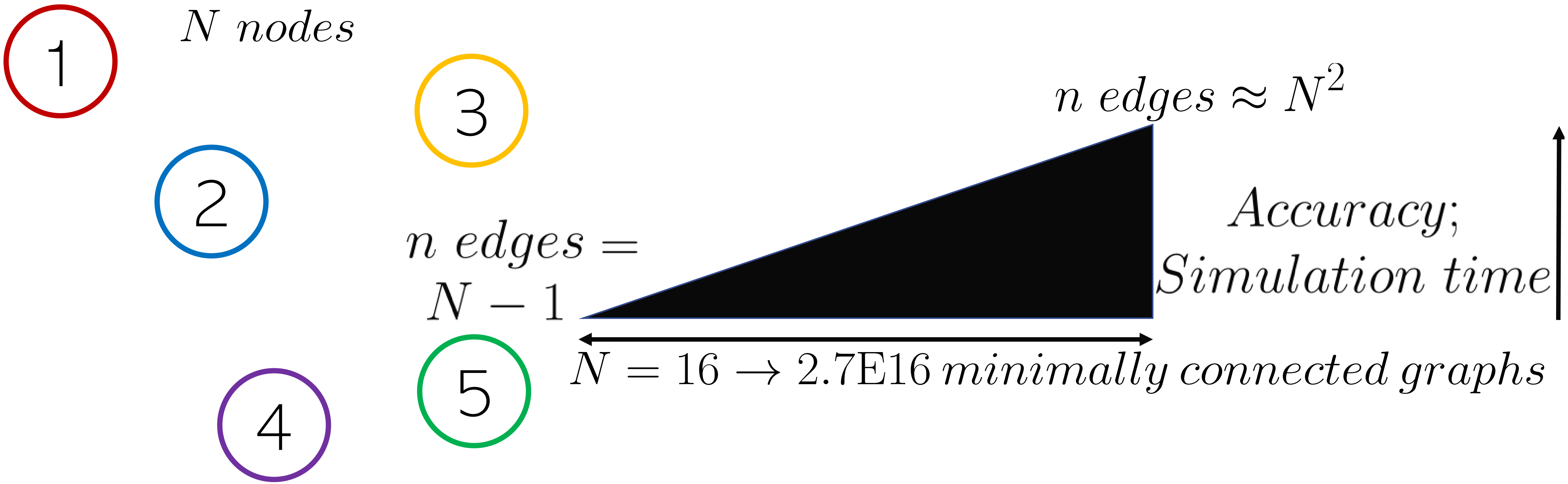
$\Delta\Delta G_{bind,\ ligand\ 5} = ..$

Problem statement – how do we pick a graph?

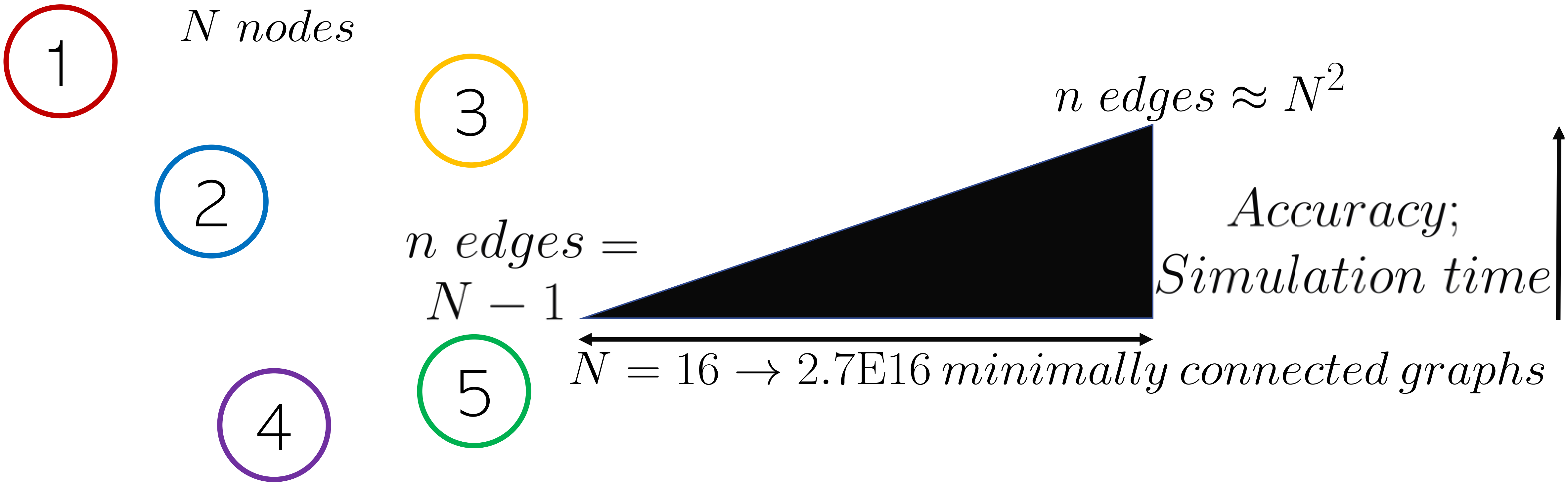


THE UNIVERSITY of EDINBURGH
cresset
EaStCHEM

Problem statement – how do we pick a graph?

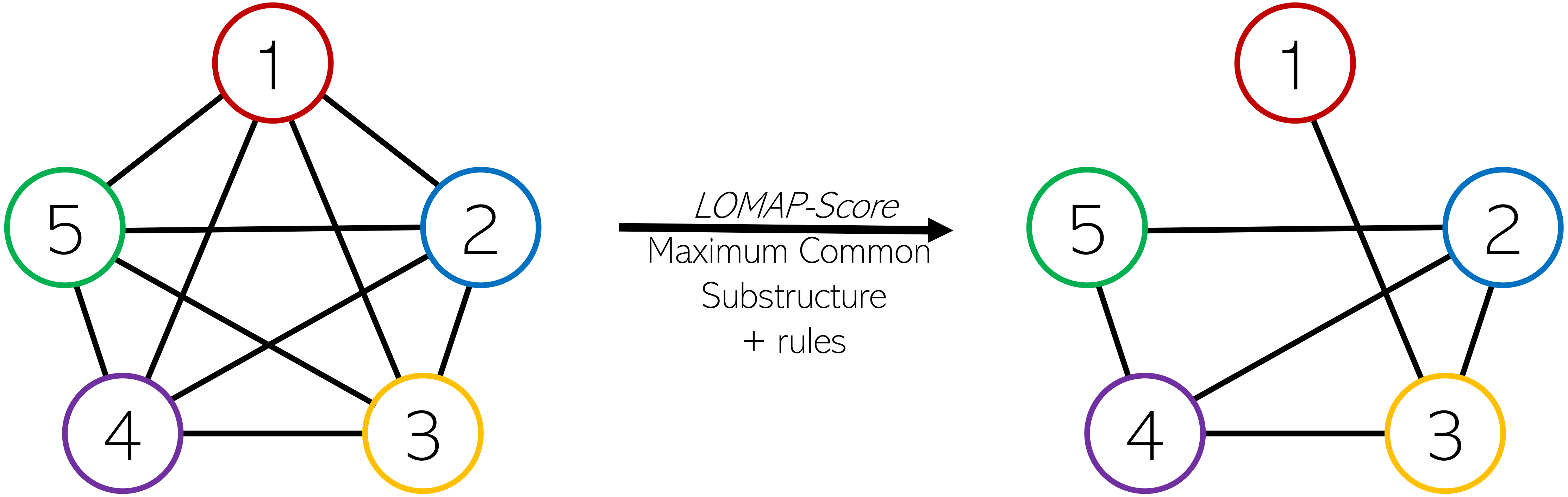


Problem statement – how do we pick a graph?



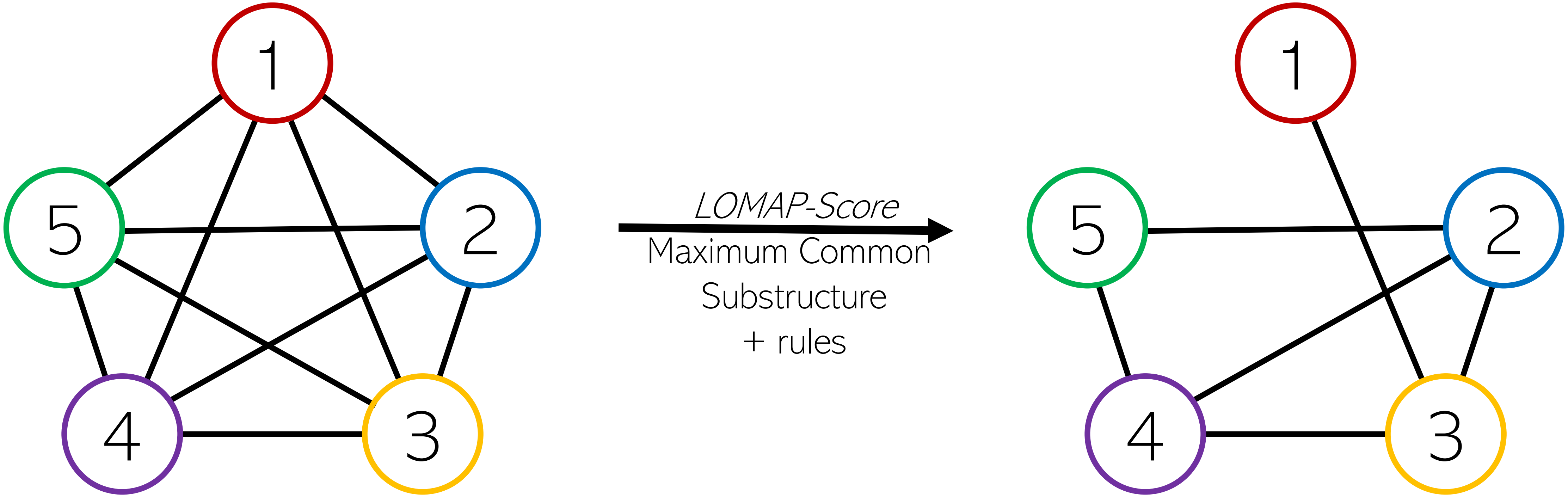
Goal: find a graph that minimises the number of edges (n), while maximising prediction accuracy (versus experiment).

State-of-the-art: LOMAP-Score depends on MCS and expert-derived rules



Shuai Liu et al, Lead optimization mapper: automating free energy calculations for lead optimization, J Comput Aided Mol Des . 2013 Sep;27(9):755-70. doi: 10.1007/s10822-013-9678-y. Epub 2013 Sep 26.

State-of-the-art: LOMAP-Score depends on MCS and expert-derived rules



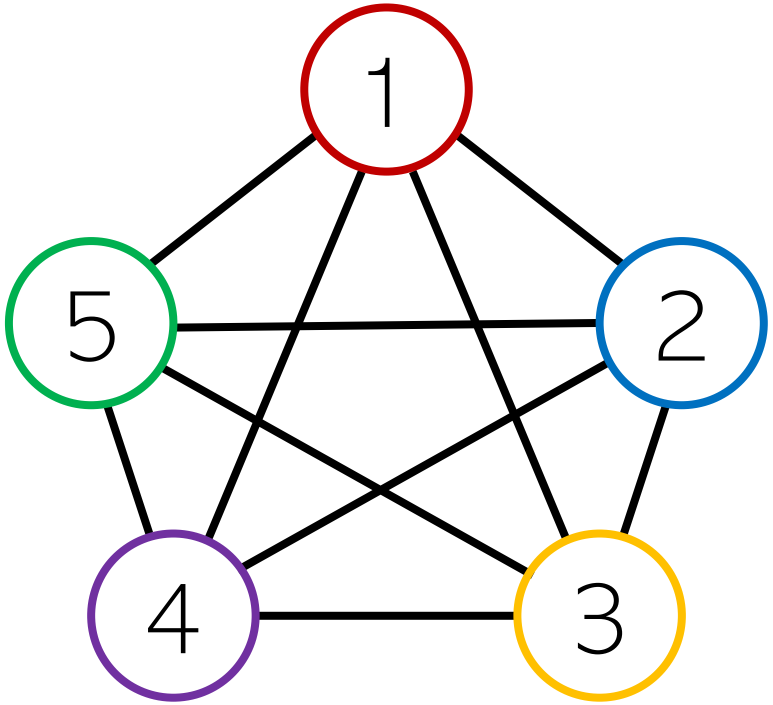
Extensively tweaked to perform adequately

Does not scale well to higher N (>50)

Limited transferability to other FEP codes

Shuai Liu et al, Lead optimization mapper: automating free energy calculations for lead optimization, J Comput Aided Mol Des . 2013 Sep;27(9):755-70. doi: 10.1007/s10822-013-9678-y. Epub 2013 Sep 26.

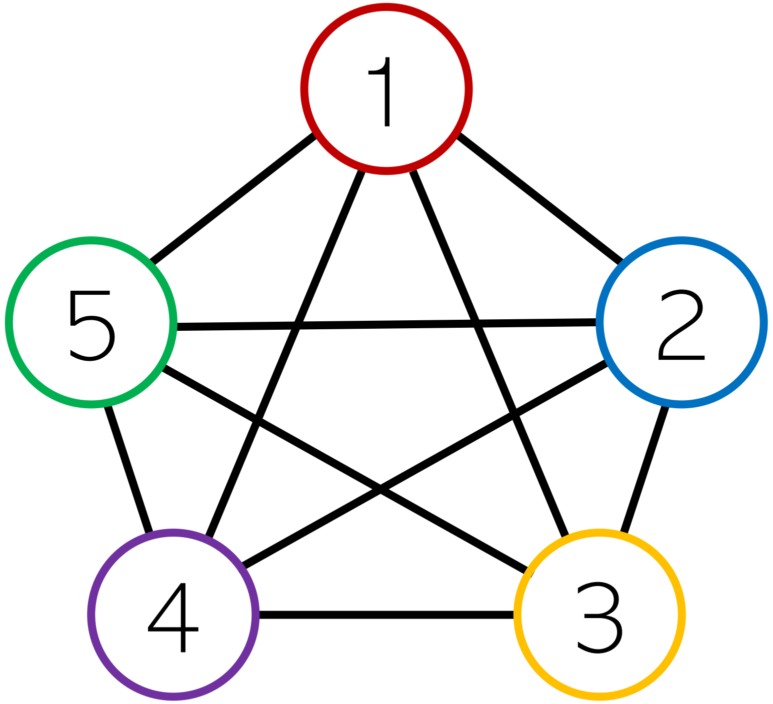
The ideal metric: $\Delta\Delta G_{offset}$



- Absolute differences between FEP predictions and experimental values

J Scheen et al, Hybrid Alchemical Free Energy/Machine-Learning Methodology for the Computation of Hydration Free Energies , J Chem Inf Model . 2020 Nov 23;60(11):5331-5339. doi: 10.1021/acs.jcim.0c00600.

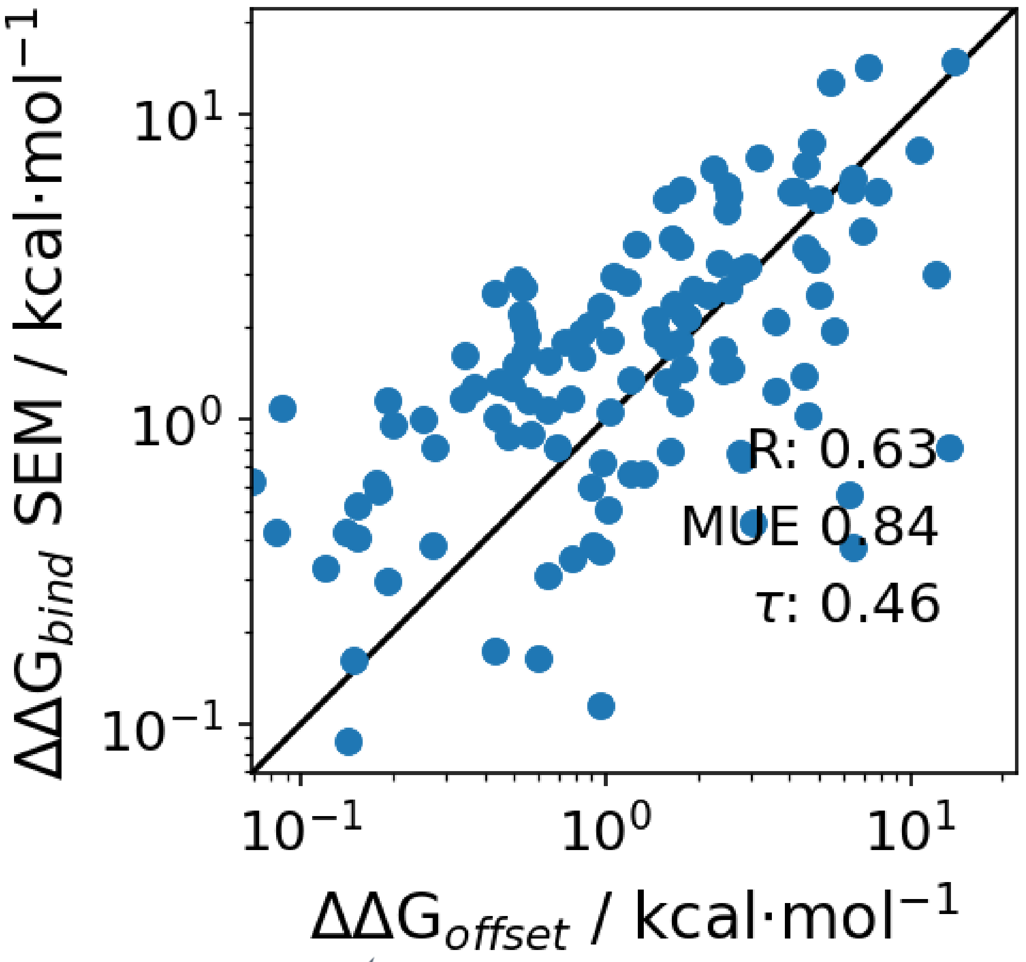
The ideal metric: $\Delta\Delta G_{offset}$



- Absolute differences between FEP predictions and experimental values
- Neither are known *a priori*
- Proposed graph has theoretically minimal FEP 'mistakes'

J Scheen et al, Hybrid Alchemical Free Energy/Machine-Learning Methodology for the Computation of Hydration Free Energies , J Chem Inf Model . 2020 Nov 23;60(11):5331-5339. doi: 10.1021/acs.jcim.0c00600.

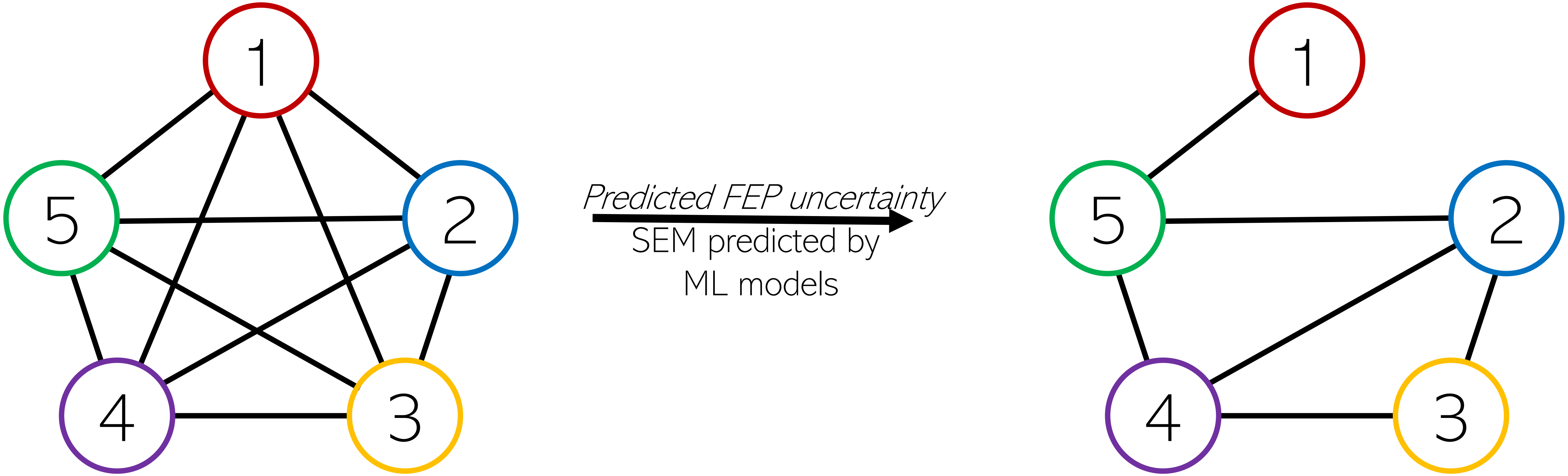
A more practical metric: FEP uncertainty



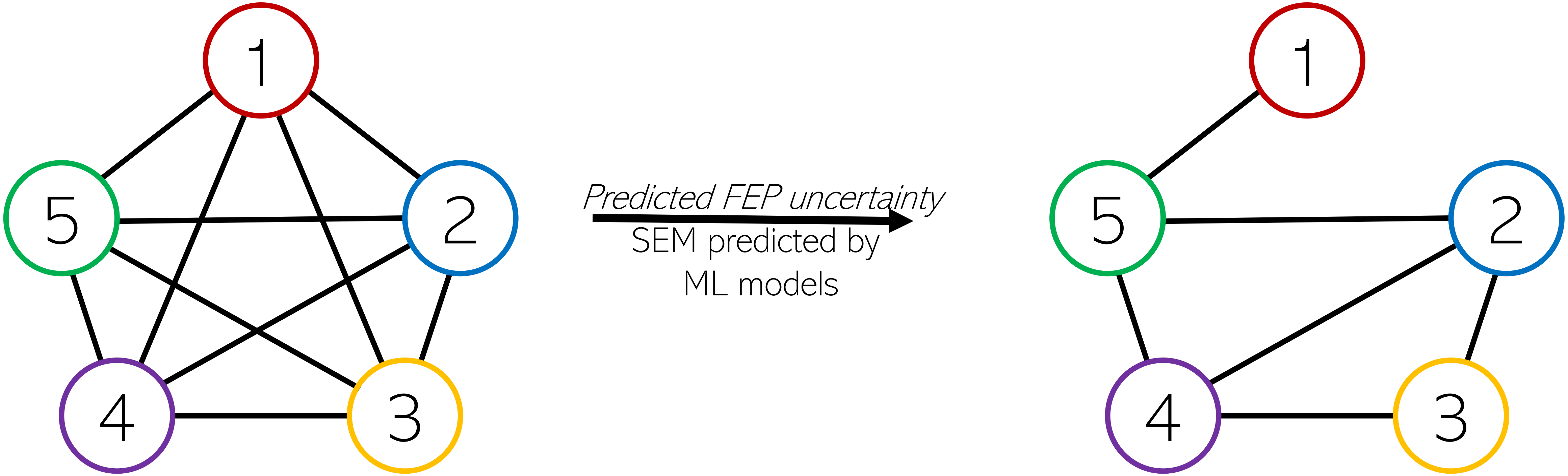
- FEP standard error of the mean (SEM) tends to be low when offset is low
- No experimental data required
- Still require FEP simulations to obtain SEMs

THE UNIVERSITY of EDINBURGH cresset EaStCHEM

This work: predict FEP uncertainty with machine-learning models



This work: predict FEP uncertainty with machine-learning models



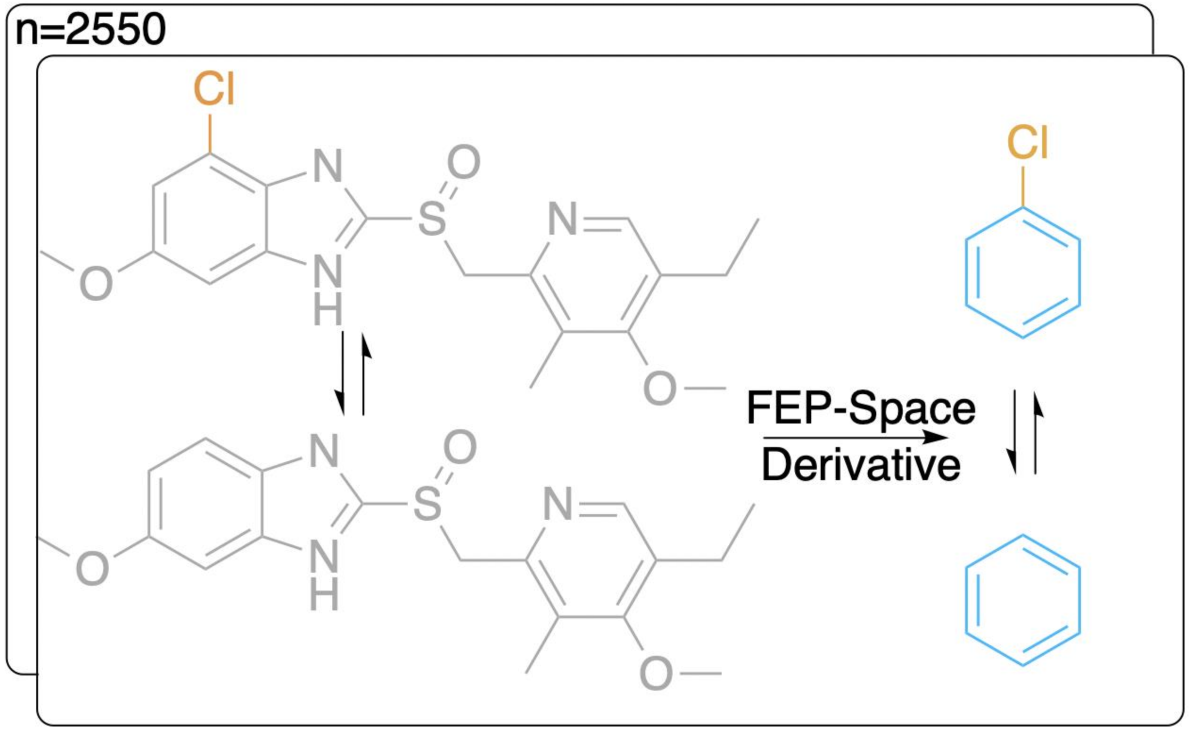
Can do a representative set of FEPs in advance

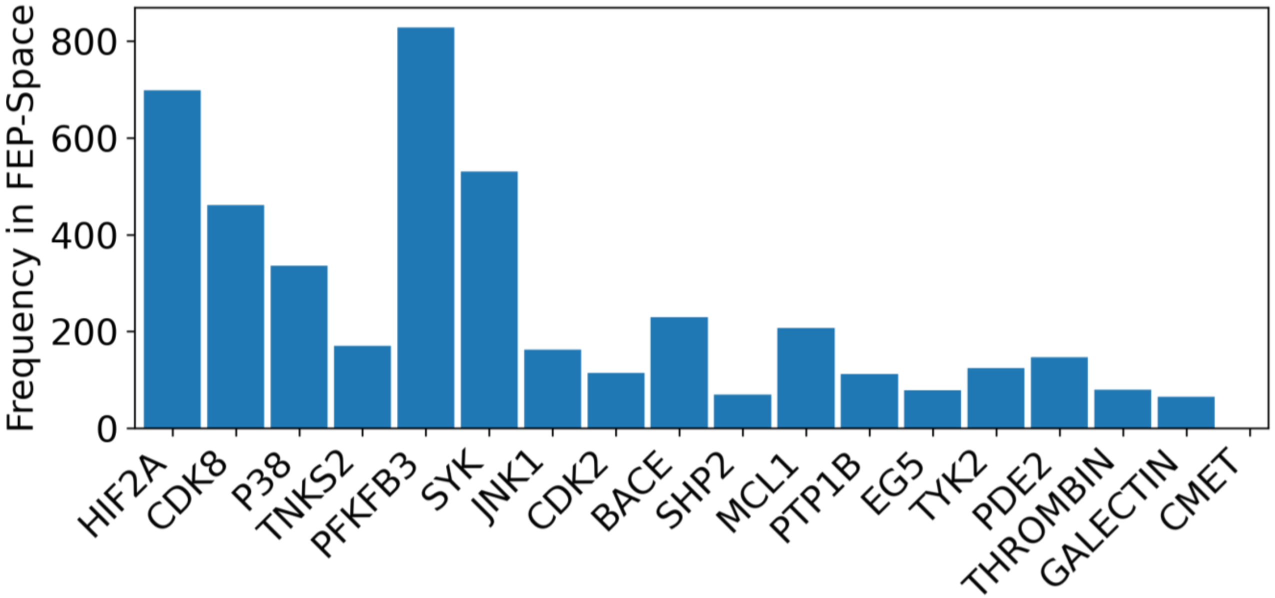
Fast, scalable estimation for large N

Does not require expert/subjective engineering

THE UNIVERSITY of EDINBURGH cresset EaStCHEM

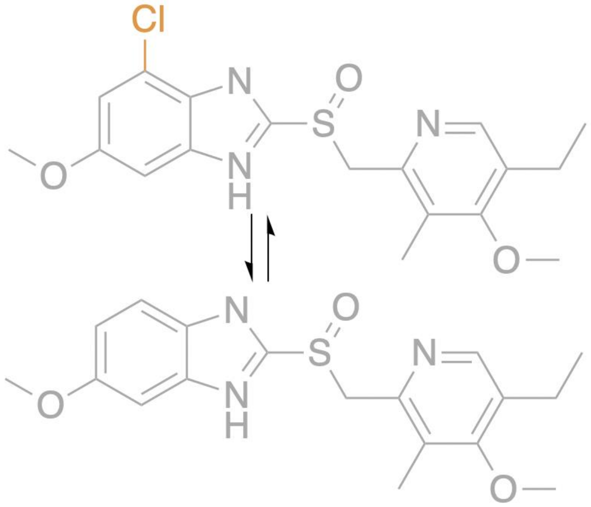
FEP-Space was created as a training set to represent all realistic FEP transformations

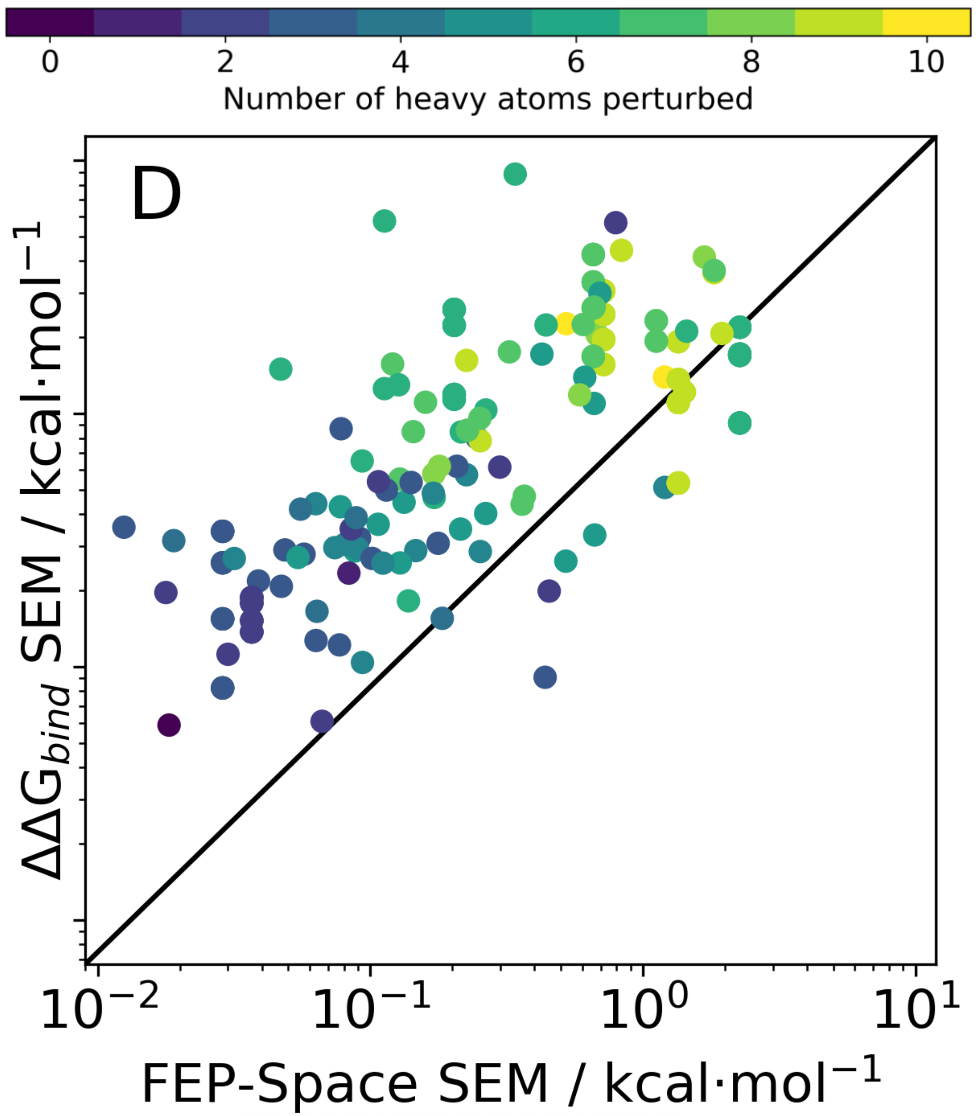


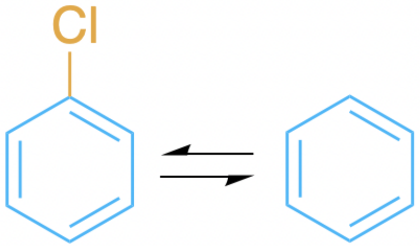


~25K GPUh

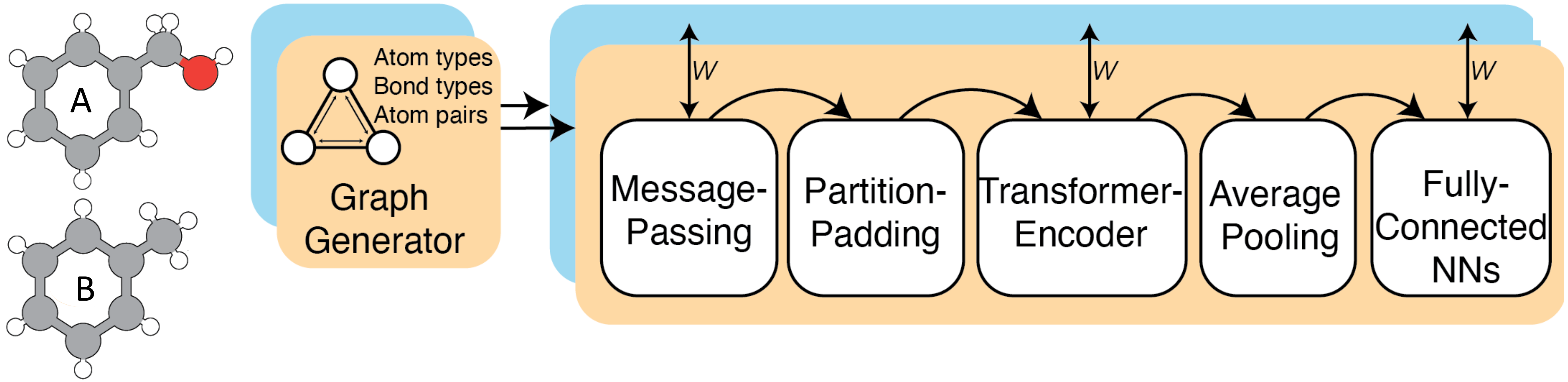
FEP-Space is a generalisable training domain at the cost of losing some correlation to its source SEM





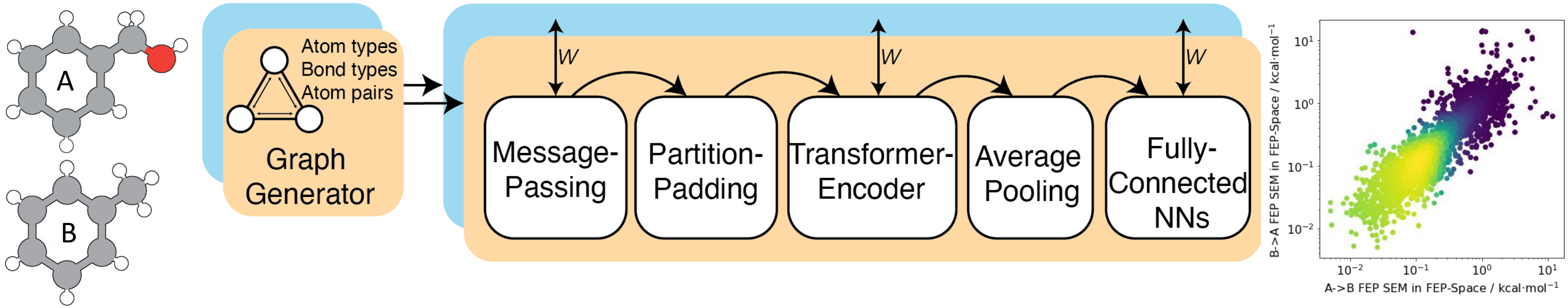


FEP-NN: a graph neural network that trains on FEP-Space ligand pairs



Davide Chicco Siamese Neural Networks: An Overview. Artificial Neural Networks. 2021 73--94. doi: 10.1007/978-1-0716-0826-5_3.

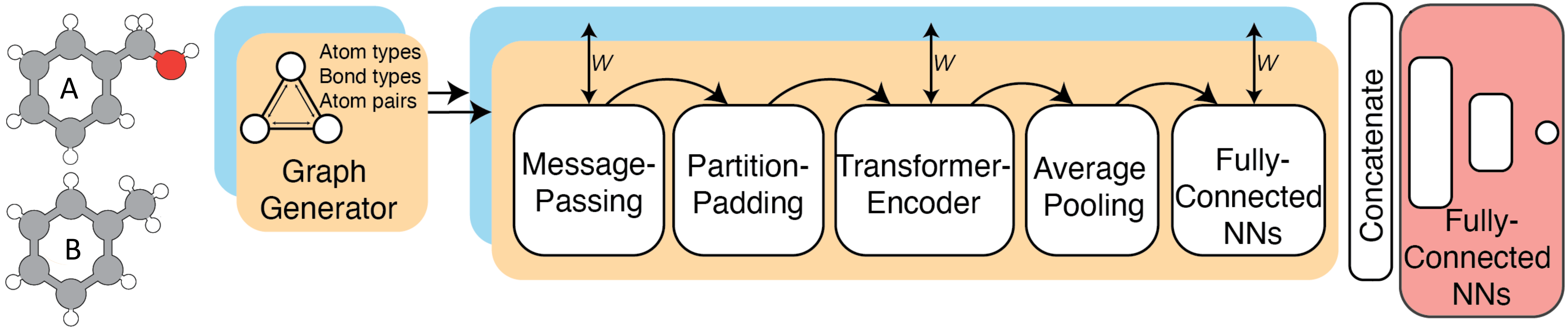
FEP-NN: a graph neural network that trains on FEP-Space ligand pairs



THE UNIVERSITY of EDINBURGH cresset EaStCHEM

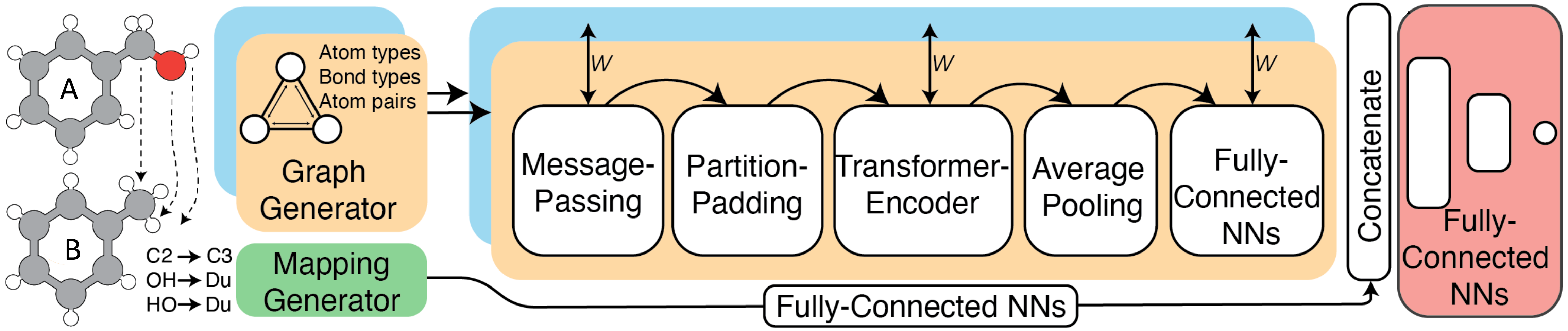
Davide Chicco Siamese Neural Networks: An Overview. Artificial Neural Networks. 2021 73--94. doi: 10.1007/978-1-0716-0826-5_3.

FEP-NN: a graph neural network that trains on FEP-Space ligand pairs



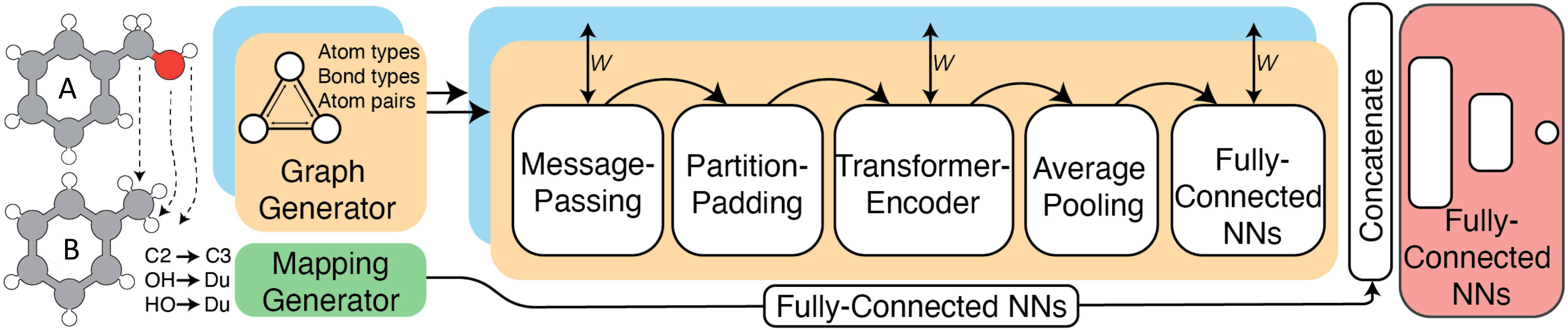
Davide Chicco Siamese Neural Networks: An Overview. Artificial Neural Networks. 2021 73--94. doi: 10.1007/978-1-0716-0826-5_3.

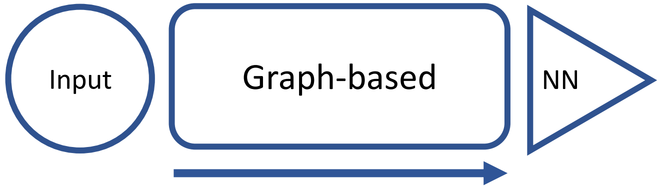
FEP-NN: a graph neural network that trains on FEP-Space ligand pairs



Davide Chicco Siamese Neural Networks: An Overview. Artificial Neural Networks. 2021 73--94. doi: 10.1007/978-1-0716-0826-5_3.

FEP-NN: a graph neural network that trains on FEP-Space ligand pairs

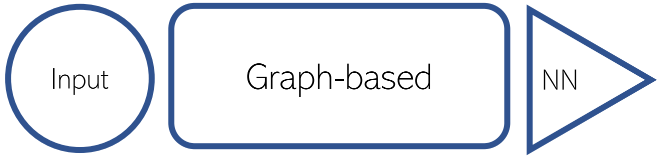


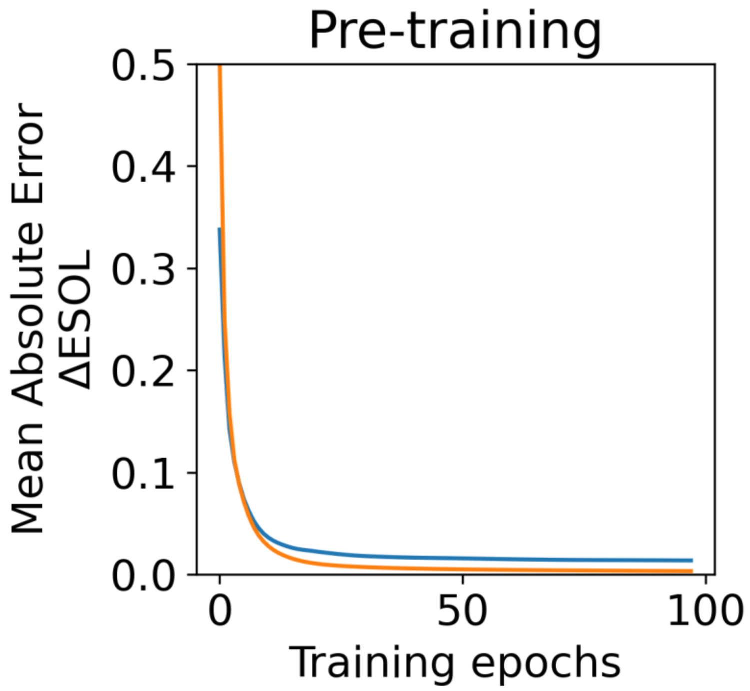


THE UNIVERSITY of EDINBURGH cresset EaStCHEM

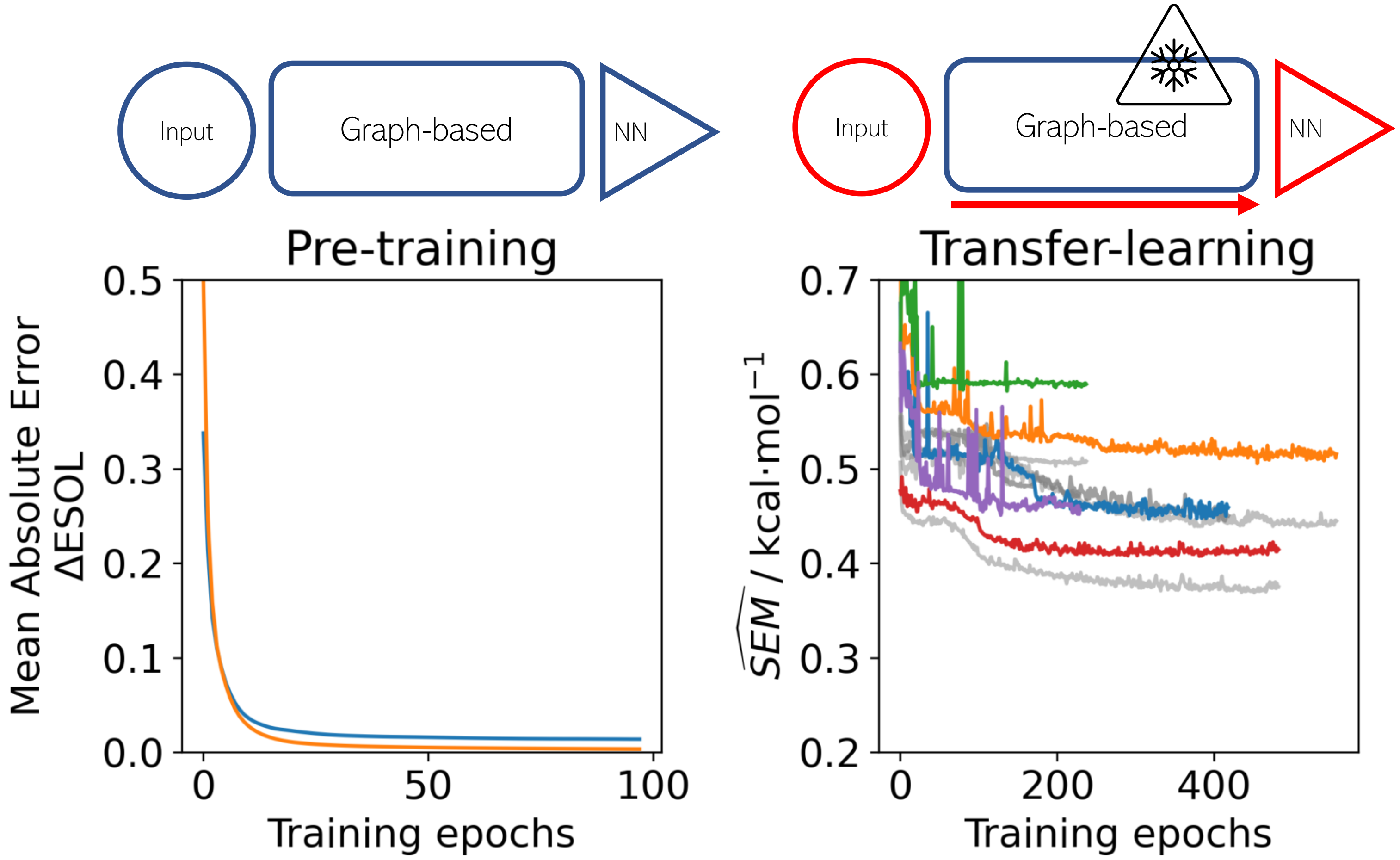
Davide Chicco Siamese Neural Networks: An Overview. Artificial Neural Networks. 2021 73--94. doi: 10.1007/978-1-0716-0826-5_3.

FEP-NN: training on FEP-space using a transfer-learning approach

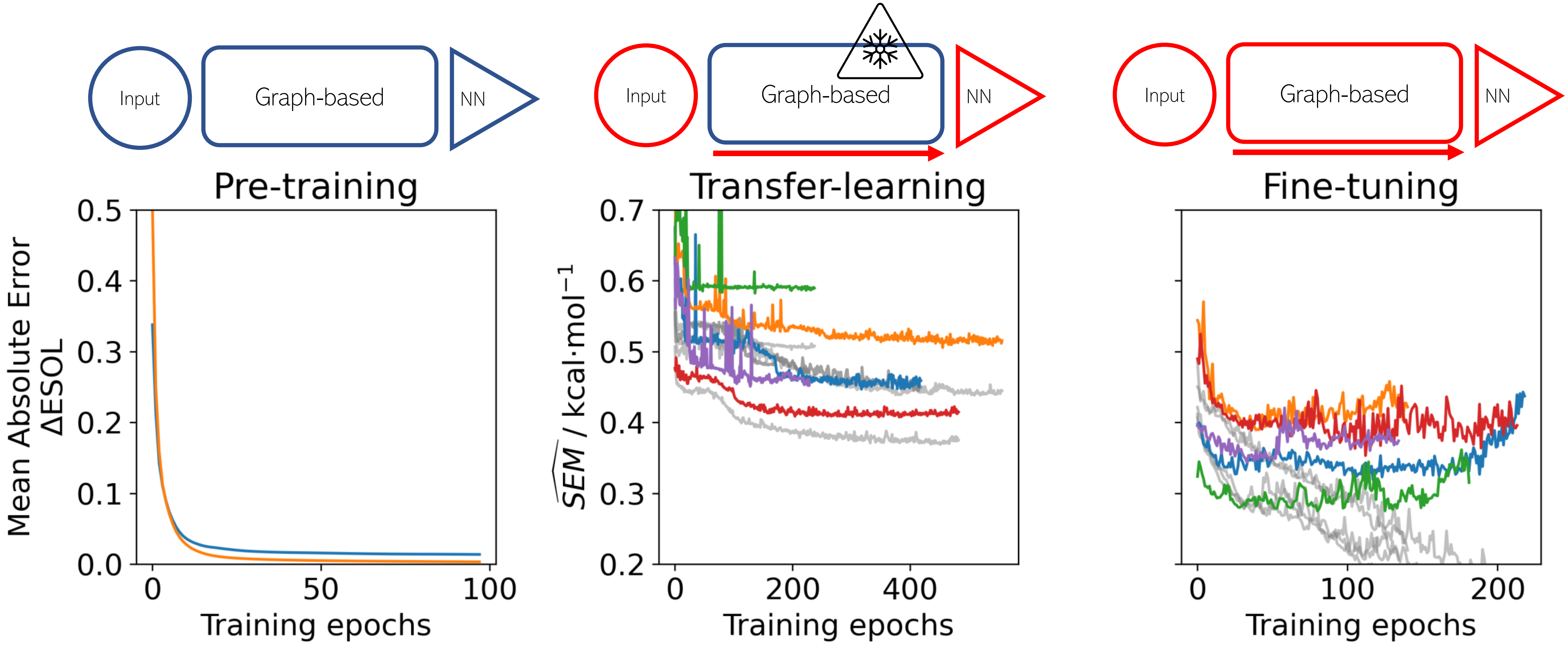




FEP-NN: training on FEP-space using a transfer-learning approach

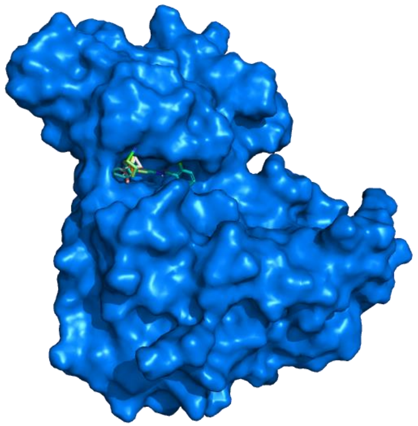


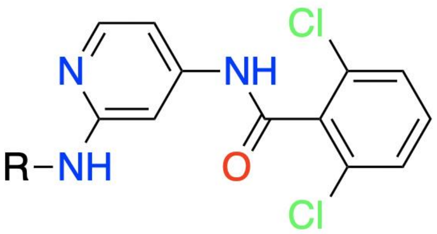
FEP-NN: training on FEP-space using a transfer-learning approach



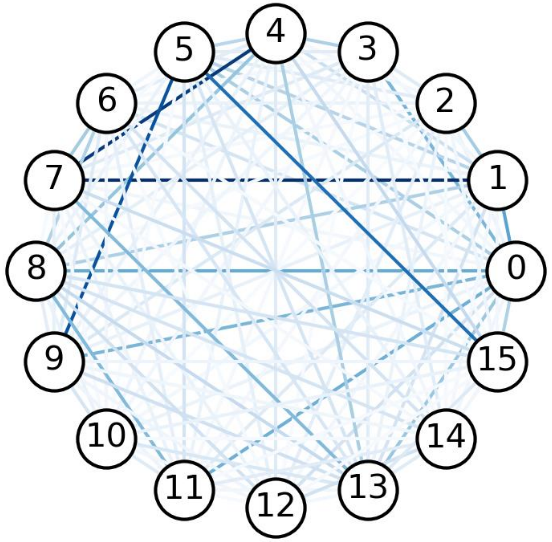
THE UNIVERSITY of EDINBURGH
cresset
EaStCHEM

Applying trained model to test set – TYK2 methodology

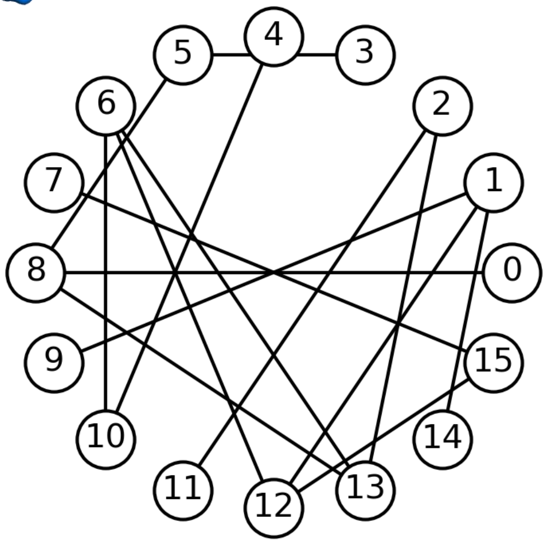




Fully-connected:



Generated graph:



Free Energy Predictions:

$$\Delta\Delta G_{bind,\ ligand\ 1} = ..$$

$$\Delta\Delta G_{bind,\ ligand\ 2} = ..$$

$$\Delta\Delta G_{bind,\ ligand\ 3} = ..$$

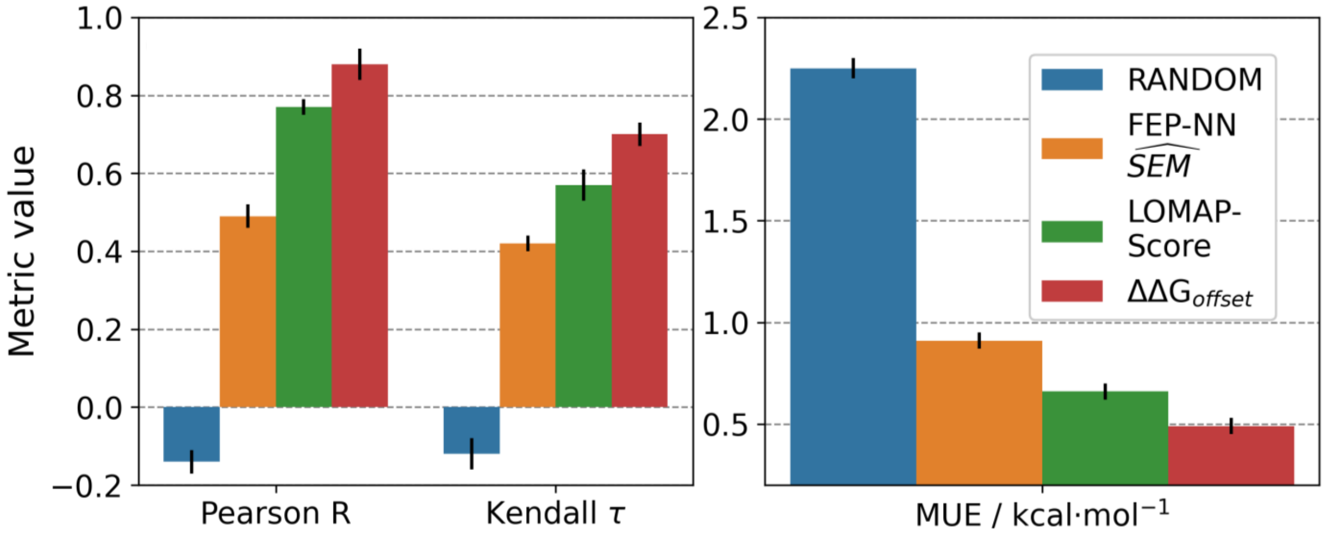
$$\Delta\Delta G_{bind,\ ligand\ 4} = ..$$

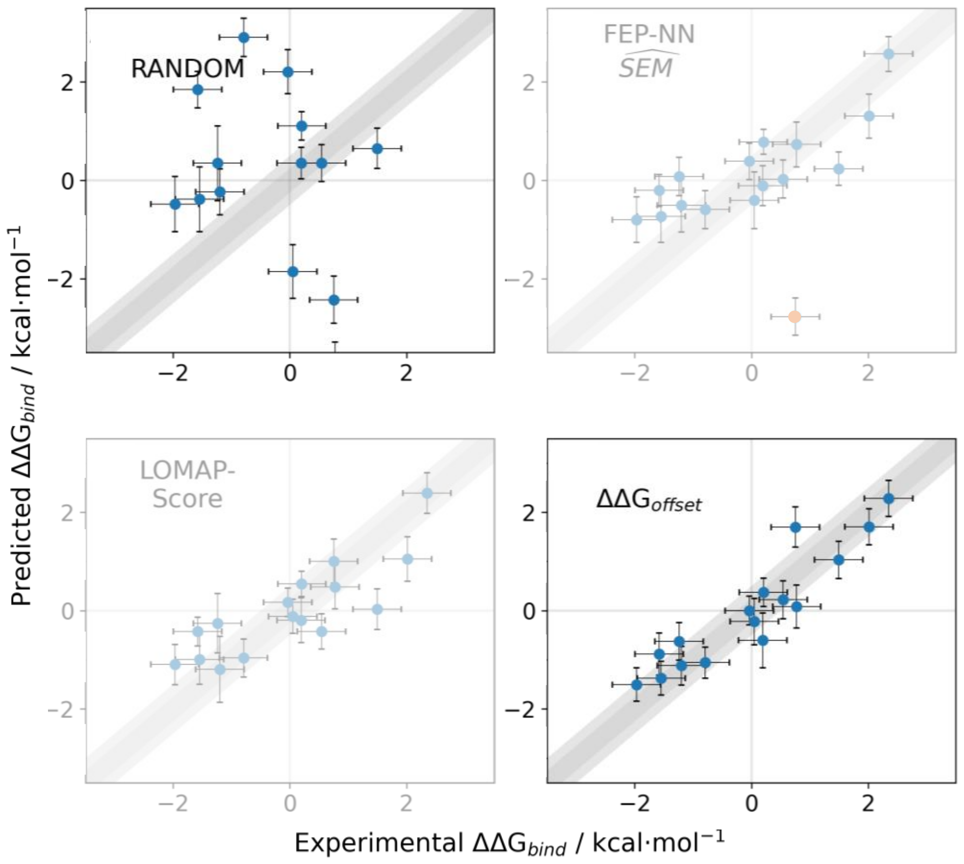
$$\Delta\Delta G_{bind,\ ligand\ 5} = ..$$

THE UNIVERSITY of EDINBURGH cresset EaStCHEM

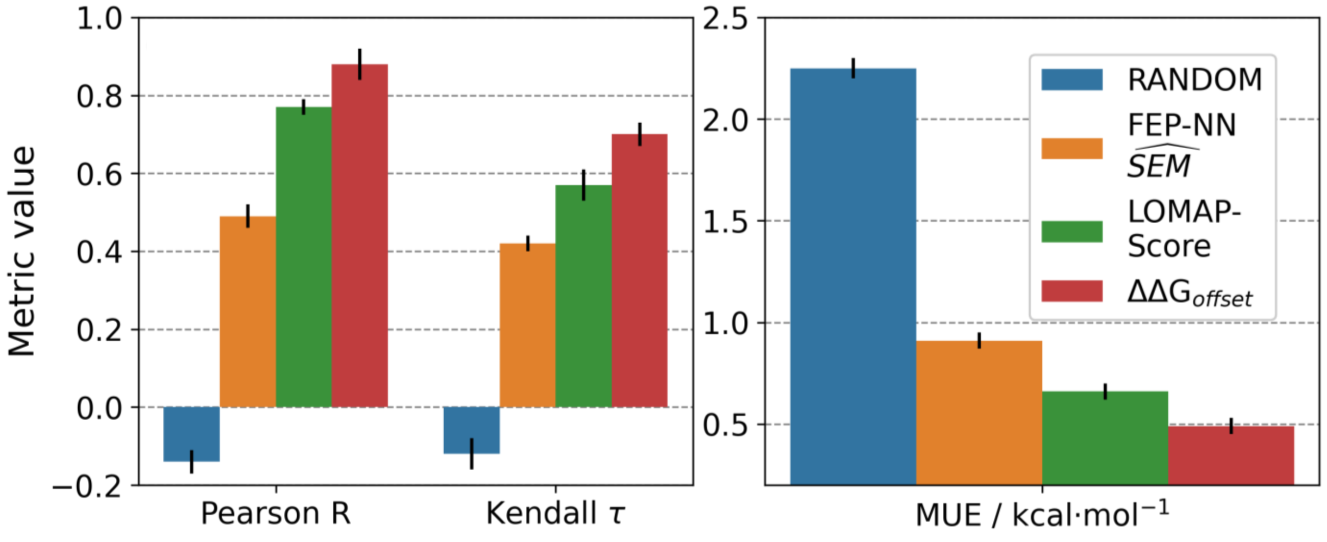
120 FEPs (5x): ~13K GPUh

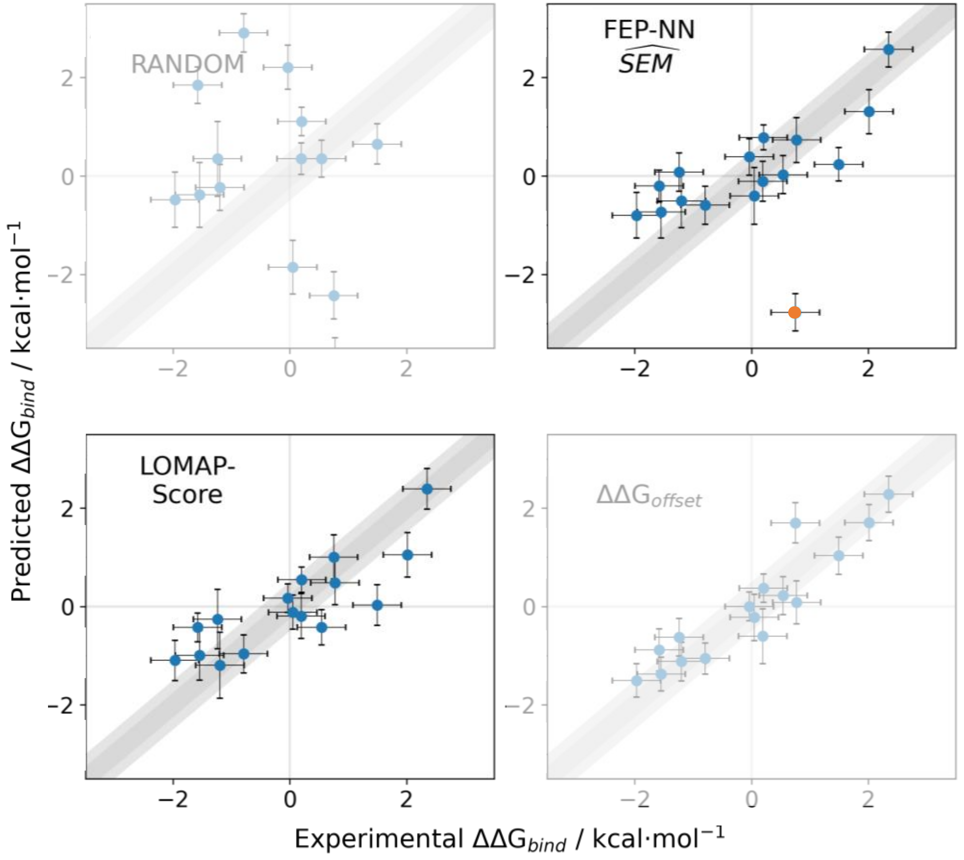
Applying trained model to test set – TYK2 results:



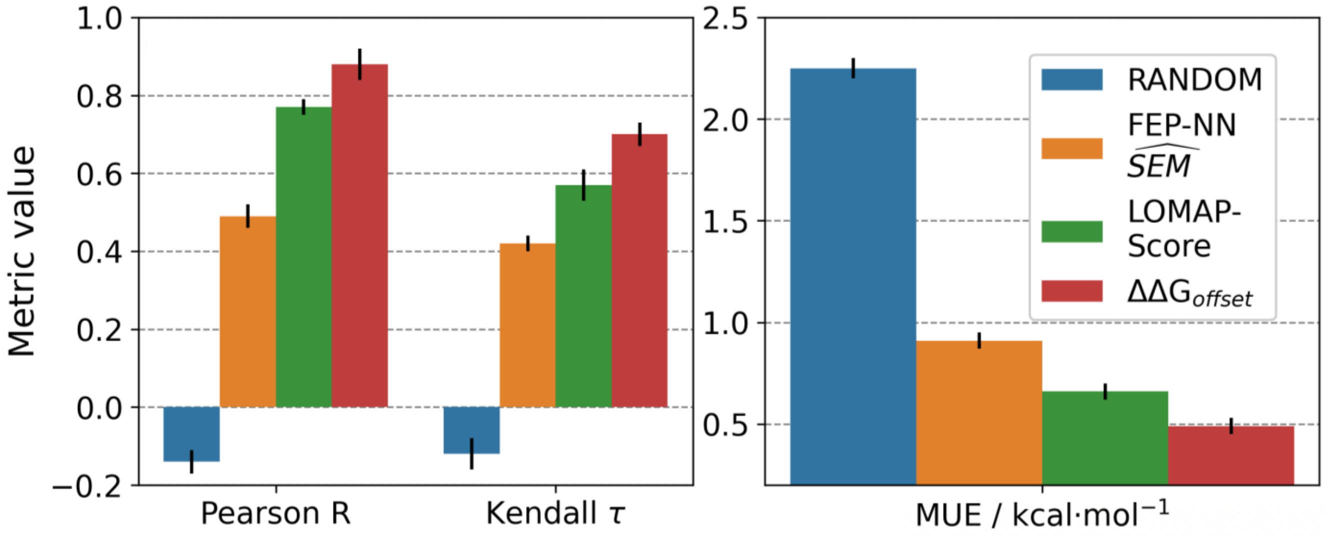


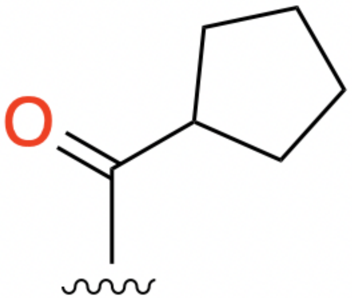
Applying trained model to test set – TYK2 results: FEP-NN graph contains an outlier

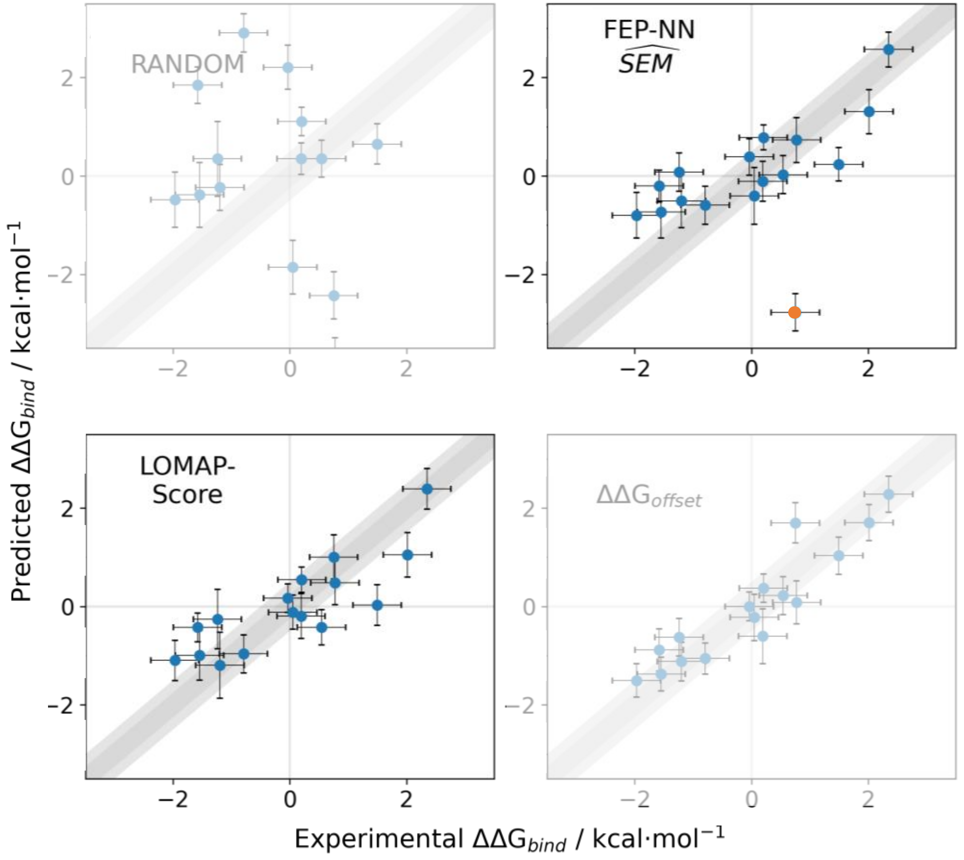




Applying trained model to test set – TYK2 results: FEP-NN graph contains an outlier

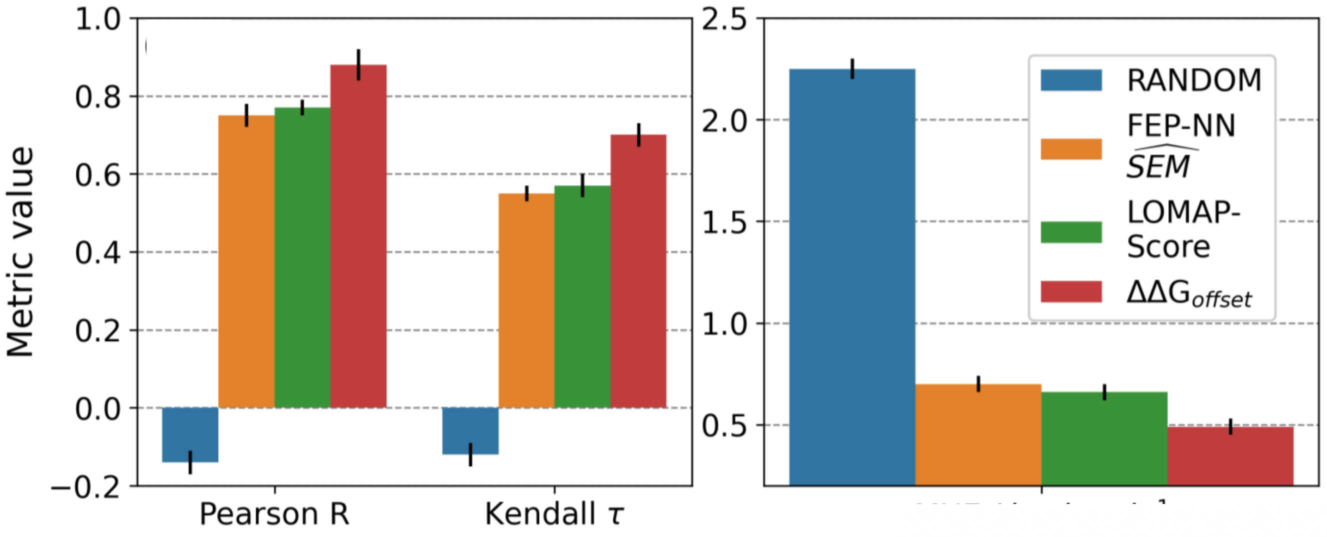


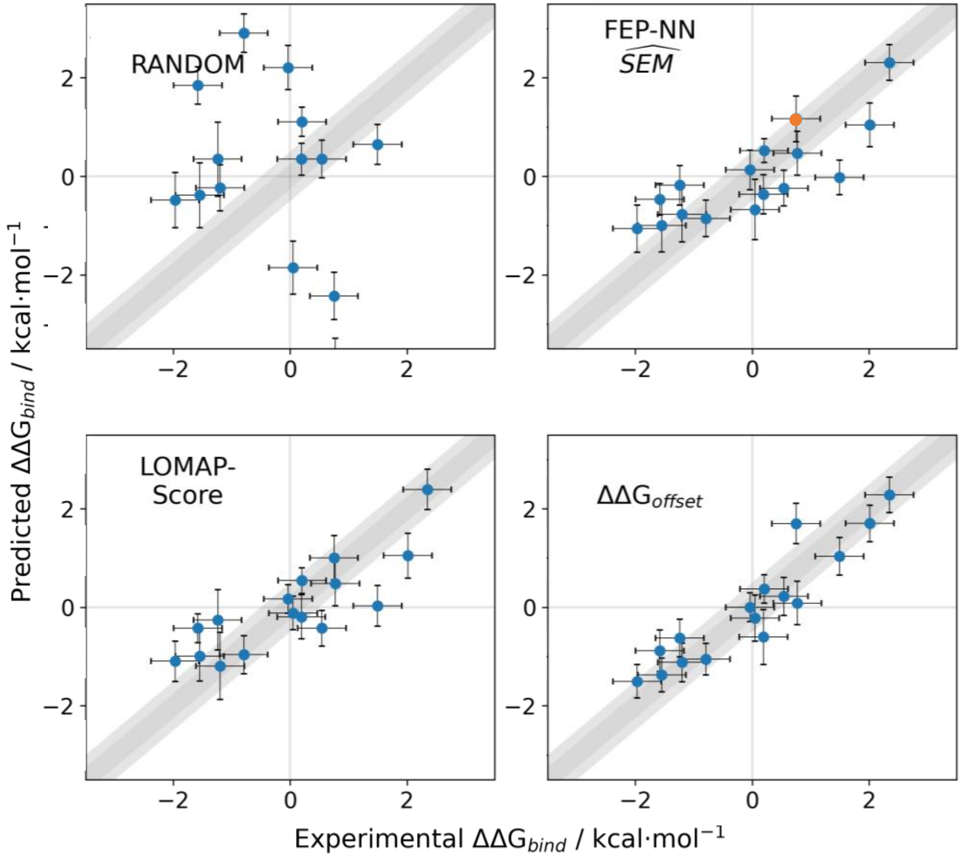


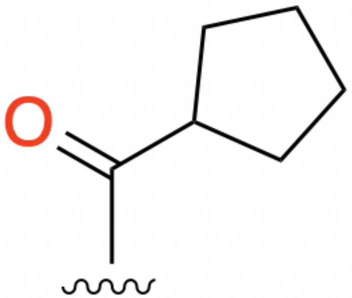


THE UNIVERSITY of EDINBURGH cresset EaStCHEM

Applying trained model to test set – TYK2 results: corrected FEP-NN graph performs competitively with state-of-the-art LOMAP-Score

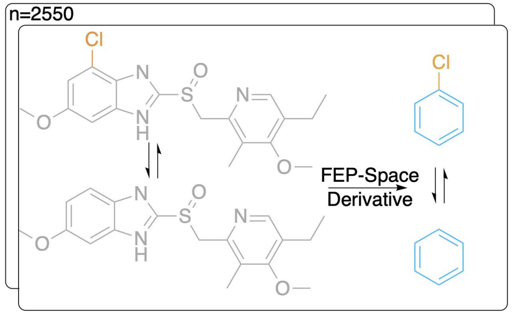






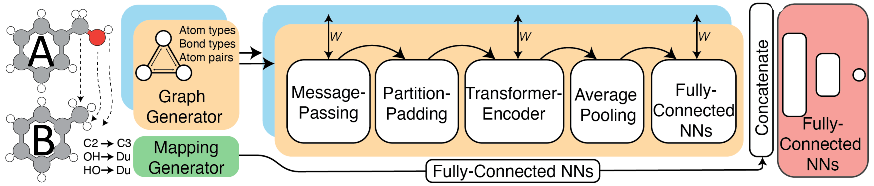
Take-home messages

FEP-Space publicly Available to advance ML for FEP

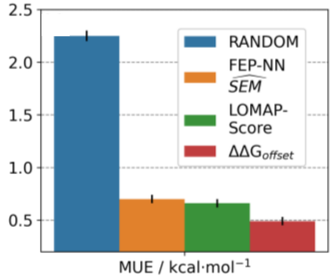


https://github.com/michellab/data_driven_fep_reliabilities

First Siamese-NN-type implementation in FEP



Data-driven network generator is competitive with state-of-the-art generators

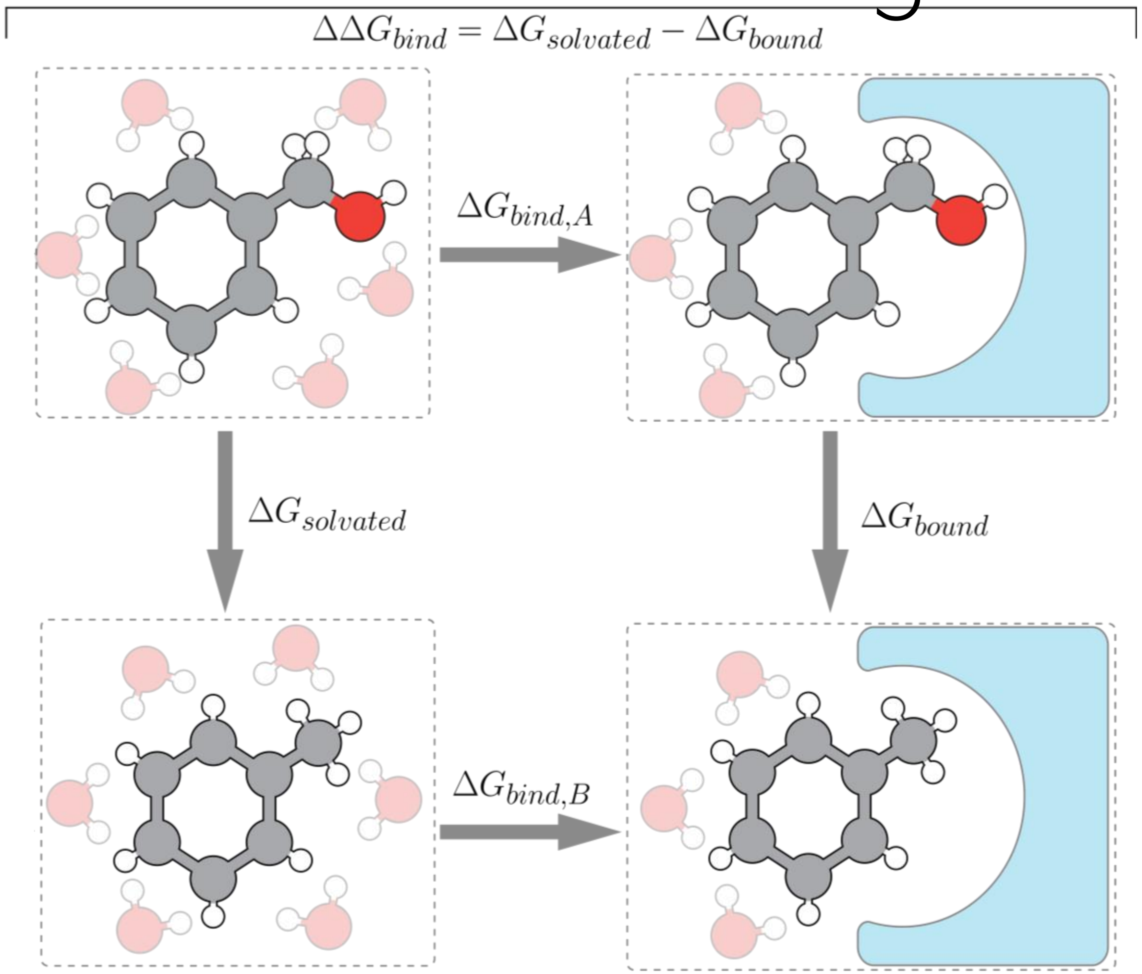


@Jenkescheen

j.scheen@sms.ed.ac.uk

THE UNIVERSITY of EDINBURGH cresset EaStCHEM

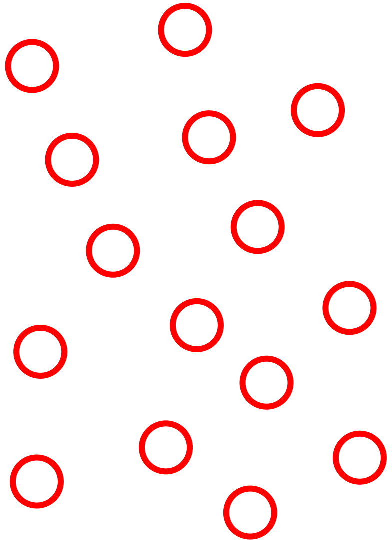
Thermodynamic cycle is the core concept in FEP methodologies



Antonia S. J. S. Mey, Jenke Scheen et al. / https://doi.org/10.33011/livecoms.2.1.18378

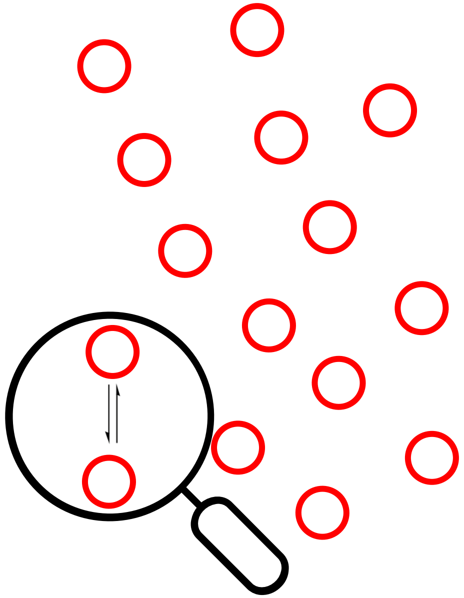
THE UNIVERSITY of EDINBURGH cresset EaStCHEM

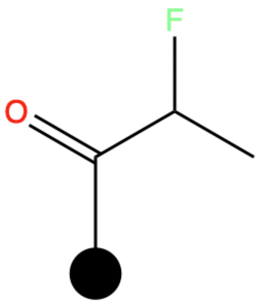
Under the hood: LOMAP scores edges based on expert-knowledge

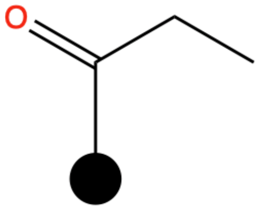


Shuai Liu et al, Lead optimization mapper: automating free energy calculations for lead optimization, J Comput Aided Mol Des 2013 Sep;27(9):755-70. doi: 10.1007/s10822-013-9678-y. Epub 2013 Sep 26.

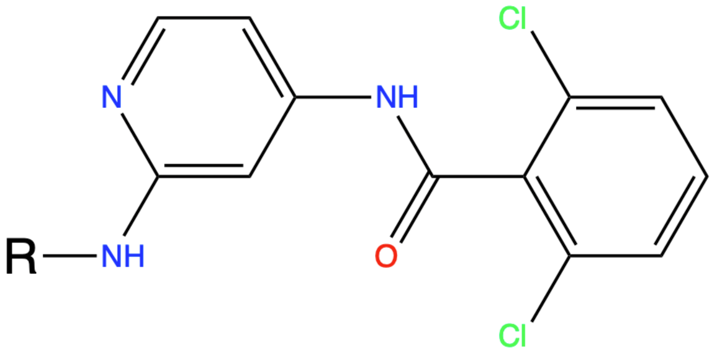
Under the hood: LOMAP scores edges based on expert-knowledge





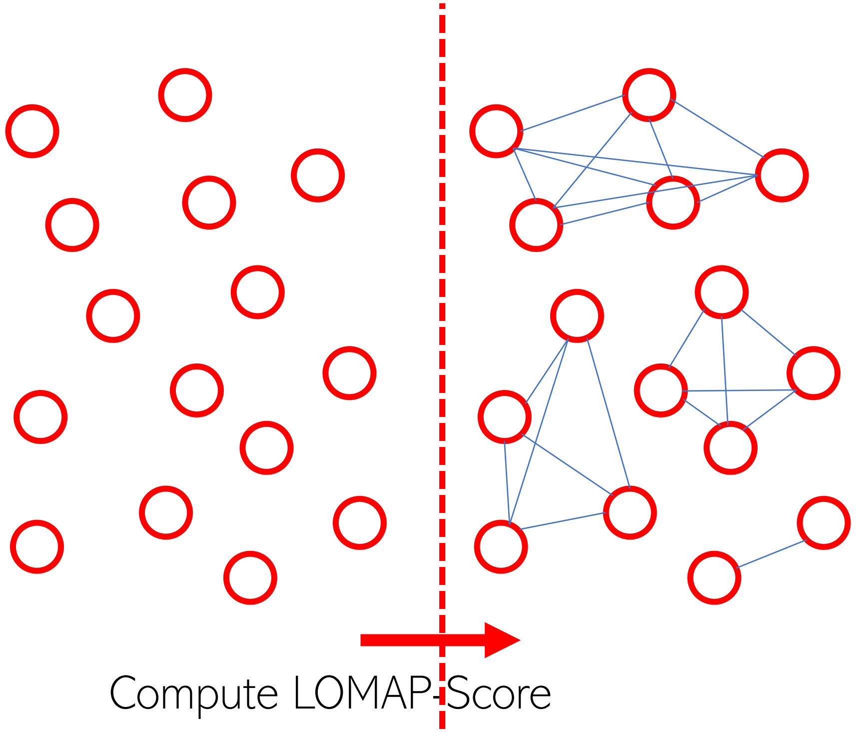


- 4 shared HAs
- 1 added HA



Shuai Liu et al, Lead optimization mapper: automating free energy calculations for lead optimization, J Comput Aided Mol Des 2013 Sep;27(9):755-70. doi: 10.1007/s10822-013-9678-y. Epub 2013 Sep 26.

Under the hood: LOMAP scores edges based on expert-knowledge

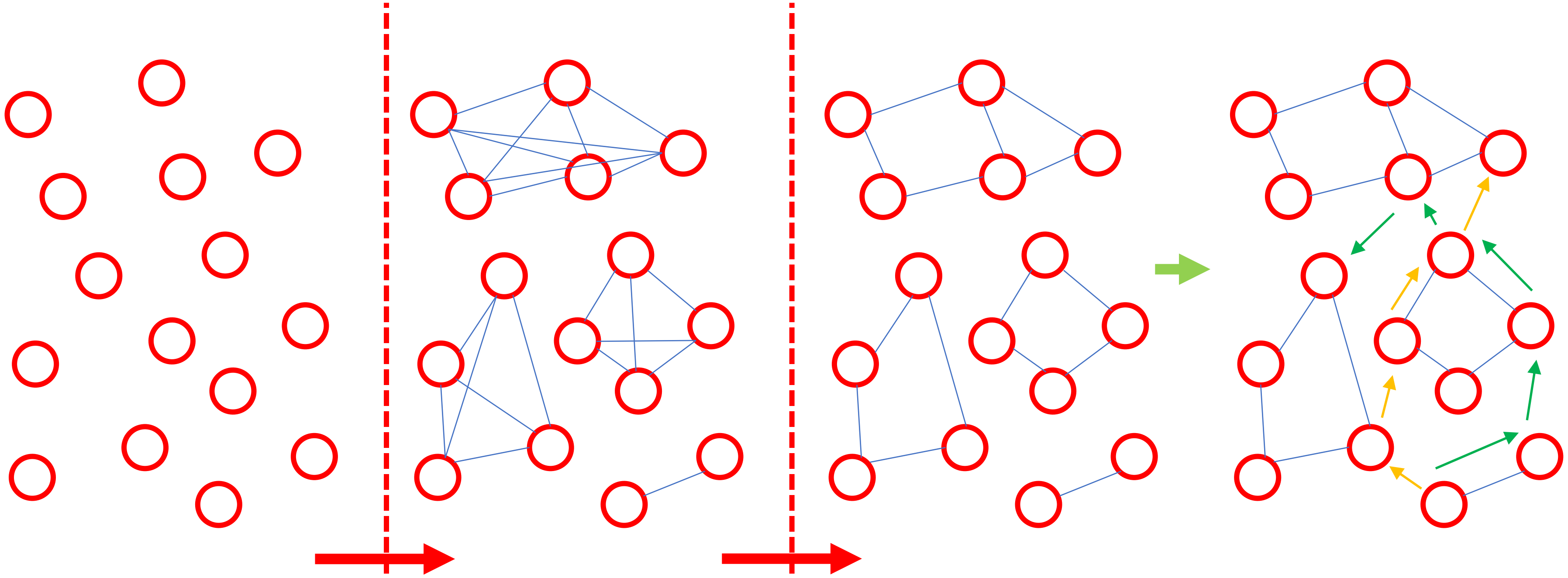


Compute LOMAP Score
(models FEP uncertainty);
keep edges above threshold

THE UNIVERSITY of EDINBURGH
cresset
EaStCHEM

Shuai Liu et al, Lead optimization mapper: automating free energy calculations for lead optimization, J Comput Aided Mol Des 2013 Sep;27(9):755-70. doi: 10.1007/s10822-013-9678-y. Epub 2013 Sep 26.

Under the hood: LOMAP scores edges based on expert-knowledge

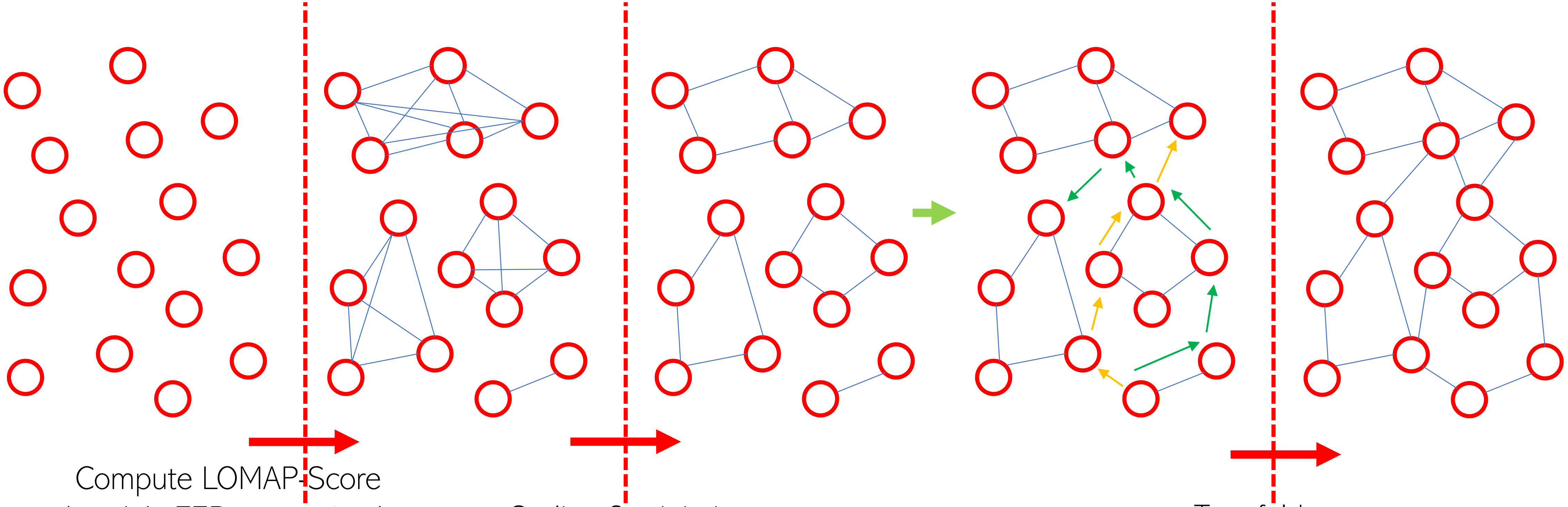


Compute LOMAP-Score (models FEP uncertainty); keep edges above threshold

Cyclize & minimize clusters

Shuai Liu et al, Lead optimization mapper: automating free energy calculations for lead optimization, J Comput Aided Mol Des 2013 Sep;27(9):755-70. doi: 10.1007/s10822-013-9678-y. Epub 2013 Sep 26.

Under the hood: LOMAP scores edges based on expert-knowledge



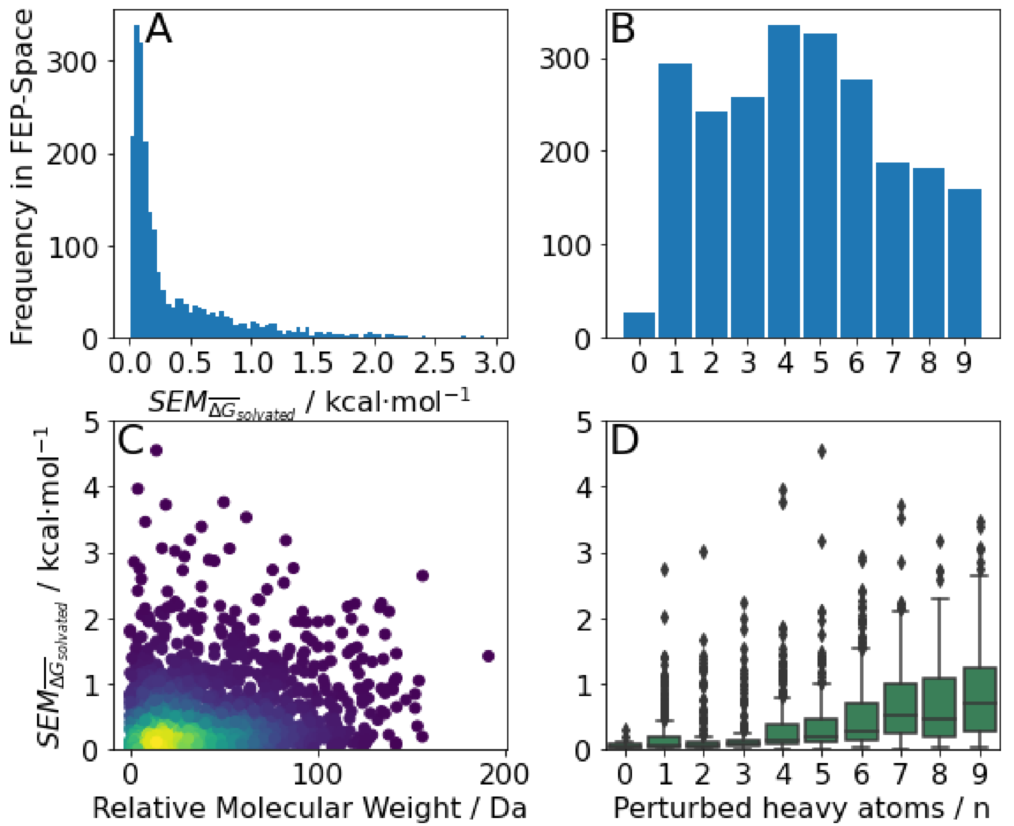
Compute LOMAP-Score (models FEP uncertainty); keep edges above threshold

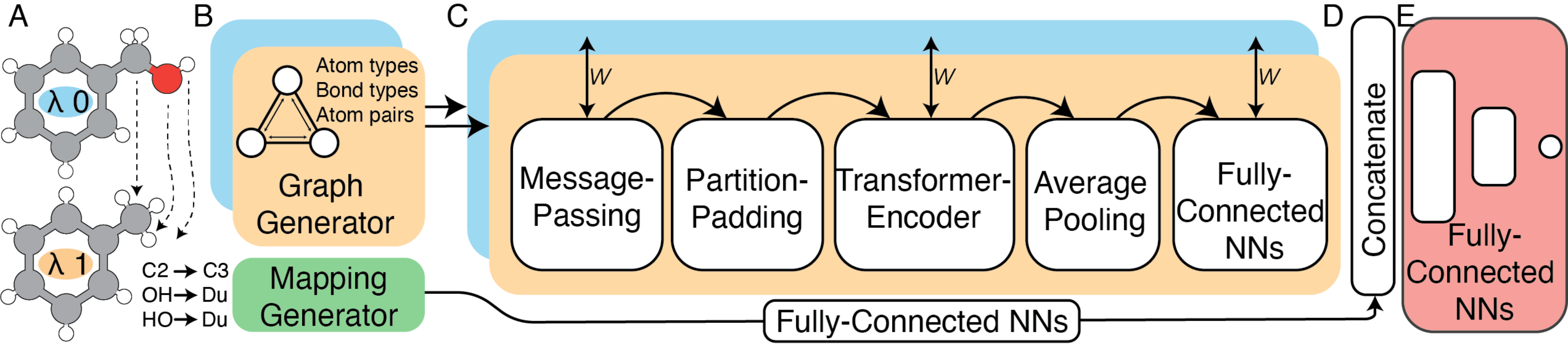
Cyclize & minimize clusters

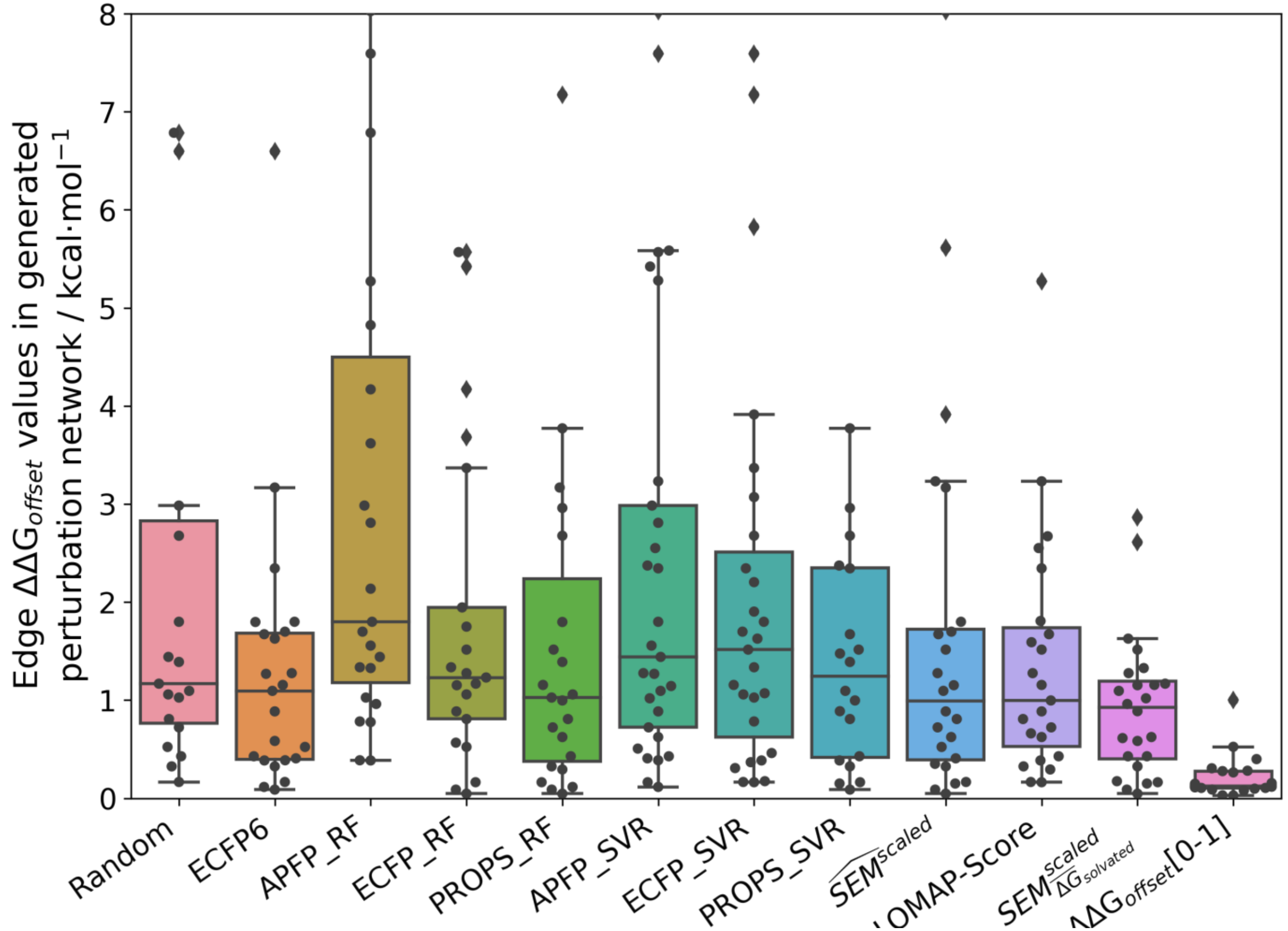
Two-fold non-overlapping *Minimum-Spanning Trees*

THE UNIVERSITY of EDINBURGH cresset EaStCHEM

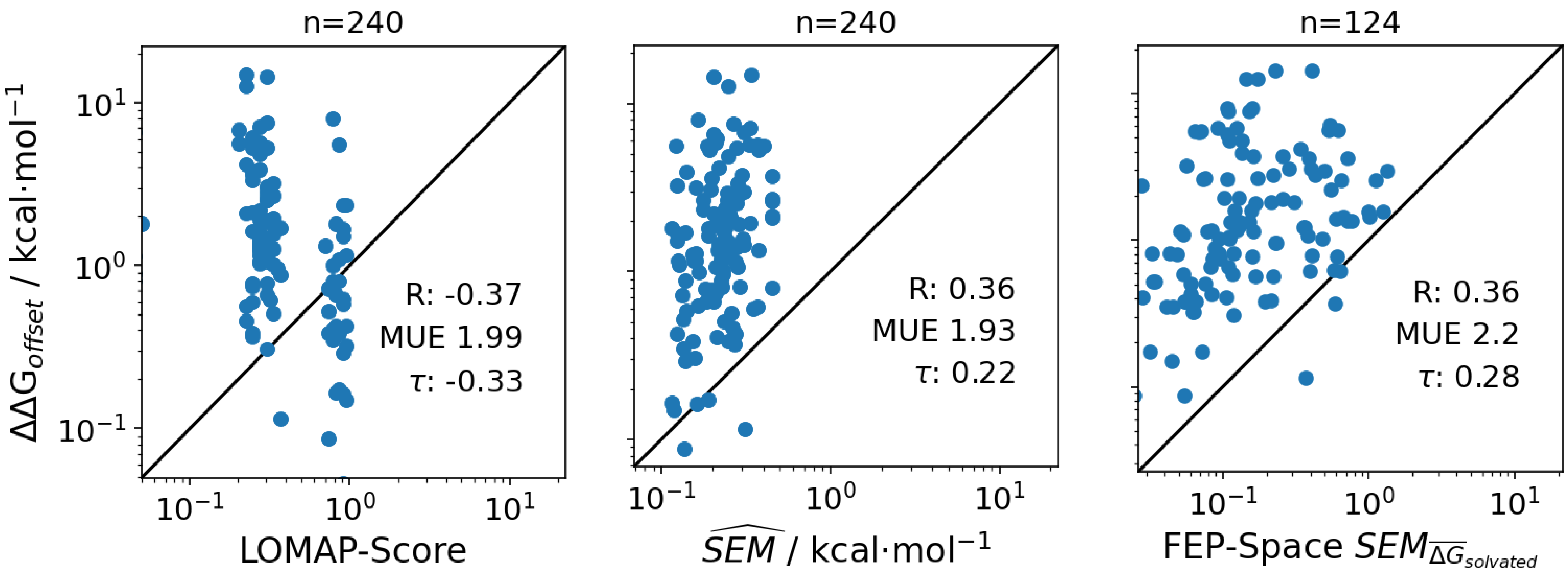
Shuai Liu et al, Lead optimization mapper: automating free energy calculations for lead optimization, J Comput Aided Mol Des 2013 Sep;27(9):755-70. doi: 10.1007/s10822-013-9678-y. Epub 2013 Sep 26.


A
B
C
D
Frequency in FEP-Space
$SEM_{\overline{\Delta G}_{solvated}}$ / kcal·mol^{-1}
Relative Molecular Weight / Da
Perturbed heavy atoms / n

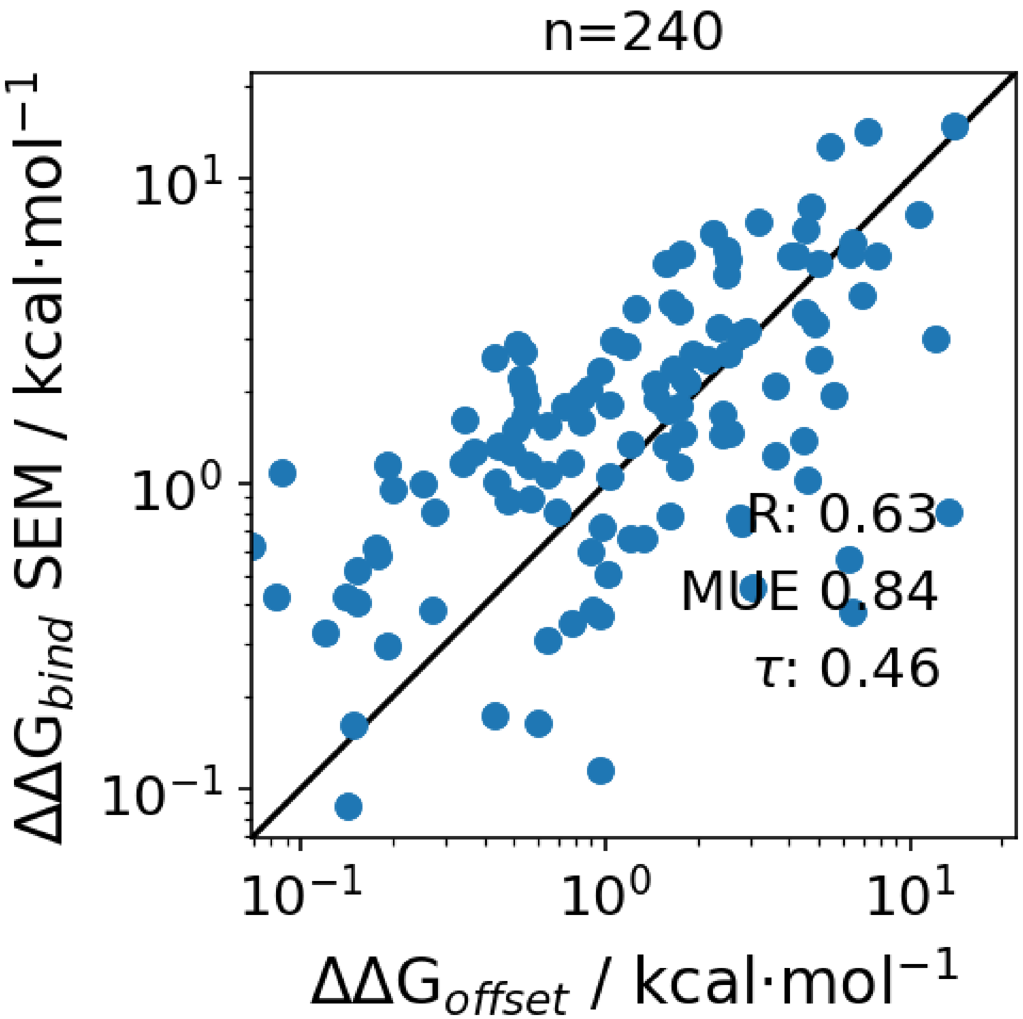

A
B
C
D
E
λ 0
λ 1
C2 → C3
OH → Du
HO → Du
Atom types
Bond types
Atom pairs
Graph Generator
Mapping Generator
W
W
W
Message-Passing
Partition-Padding
Transformer-Encoder
Average Pooling
Fully-Connected NNs
Concatenate
Fully-Connected NNs
Fully-Connected NNs



Edge ΔΔG_offset values in generated perturbation network / kcal·mol⁻¹
0
1
2
3
4
5
6
7
8
Random
ECFP6
APFP_RF
ECFP_RF
PROPS_RF
APFP_SVR
ECFP_SVR
PROPS_SVR
SEM^scaled
OMAP-Score
SEM^scaled ΔG_solvated
ΔG_offset[0-1]



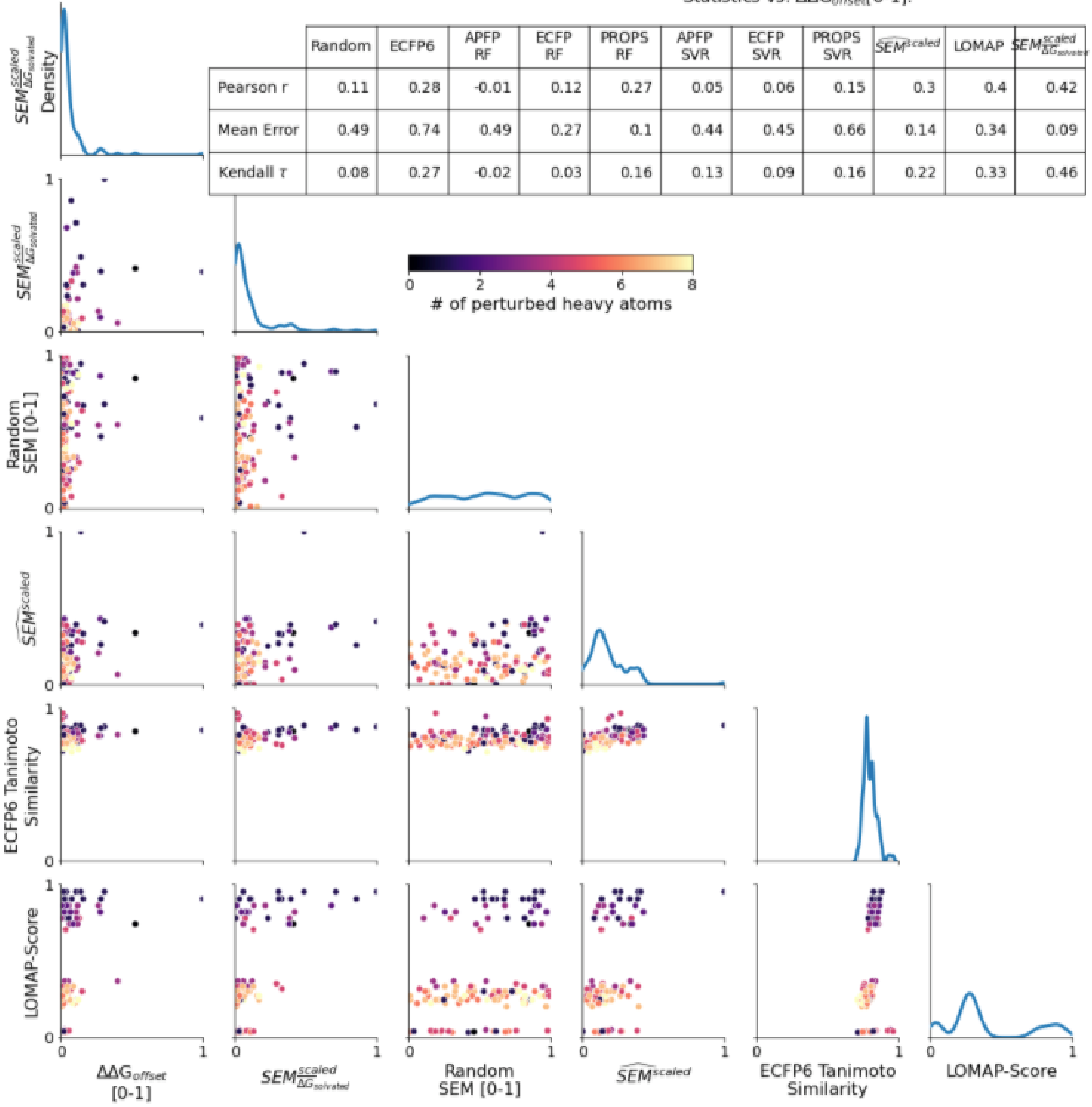
THE UNIVERSITY of EDINBURGH cresset EaStCHEM



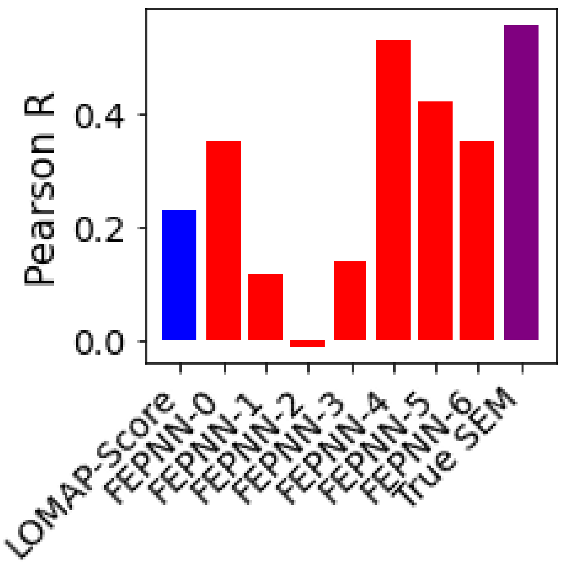
n=240
R: 0.63
MUE 0.84
τ: 0.46
10^{1}
10^{0}
10^{-1}
10^{-1}
10^{0}
10^{1}
$\Delta\Delta G_{bind}$ SEM / kcal·mol^{-1}
$\Delta\Delta G_{offset}$ / kcal·mol^{-1}

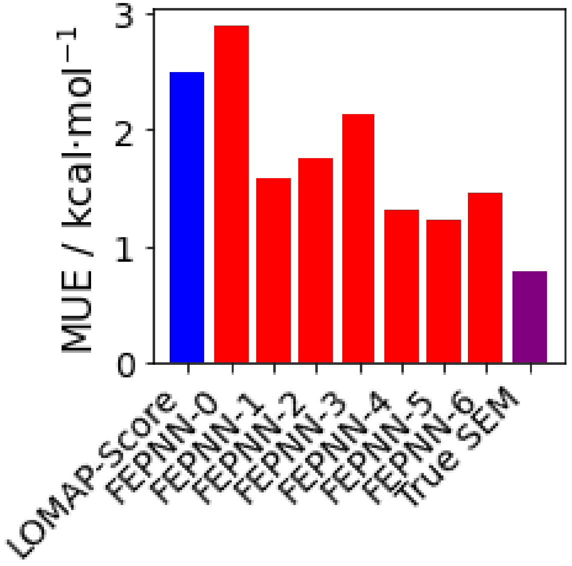
Statistics vs. $\Delta\Delta G_{offset}$[0-1]:

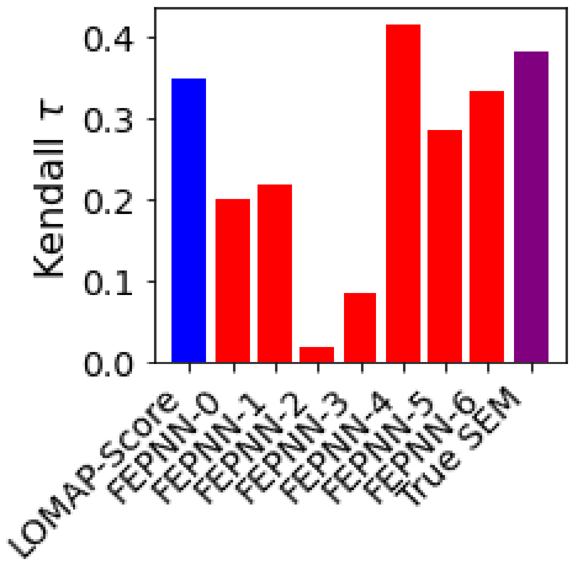
	Random	ECFP6	APFP RF	ECFP RF	PROPS RF	APFP SVR	ECFP SVR	PROPS SVR	$\widehat{SEM}^{scaled}$	LOMAP	$SEM^{scaled}_{\Delta G_{solvated}}$
Pearson r	0.11	0.28	-0.01	0.12	0.27	0.05	0.06	0.15	0.3	0.4	0.42
Mean Error	0.49	0.74	0.49	0.27	0.1	0.44	0.45	0.66	0.14	0.34	0.09
Kendall τ	0.08	0.27	-0.02	0.03	0.16	0.13	0.09	0.16	0.22	0.33	0.46

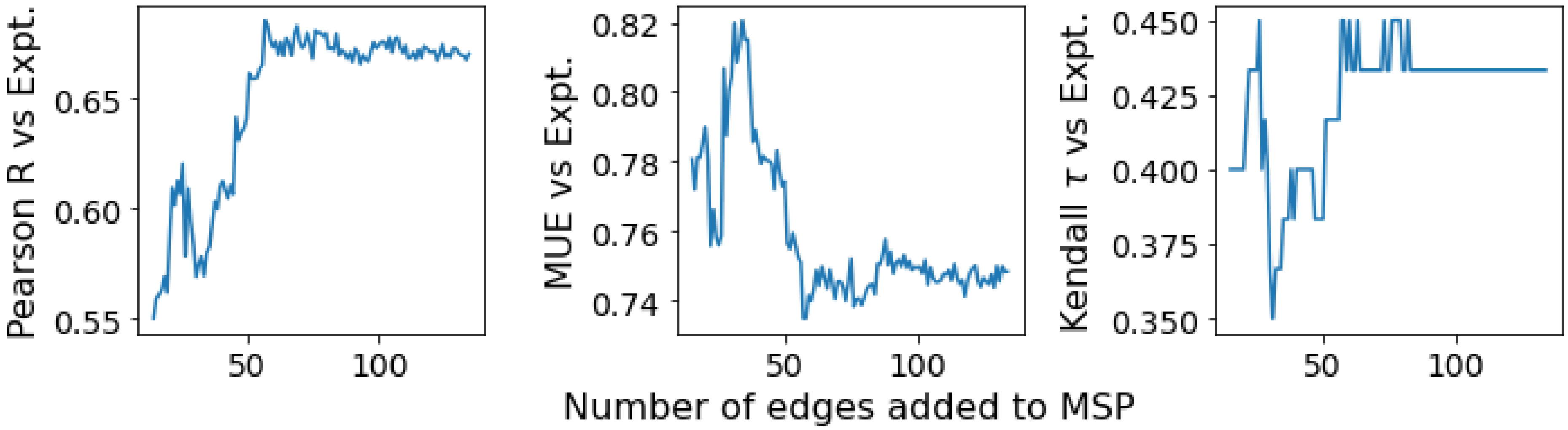


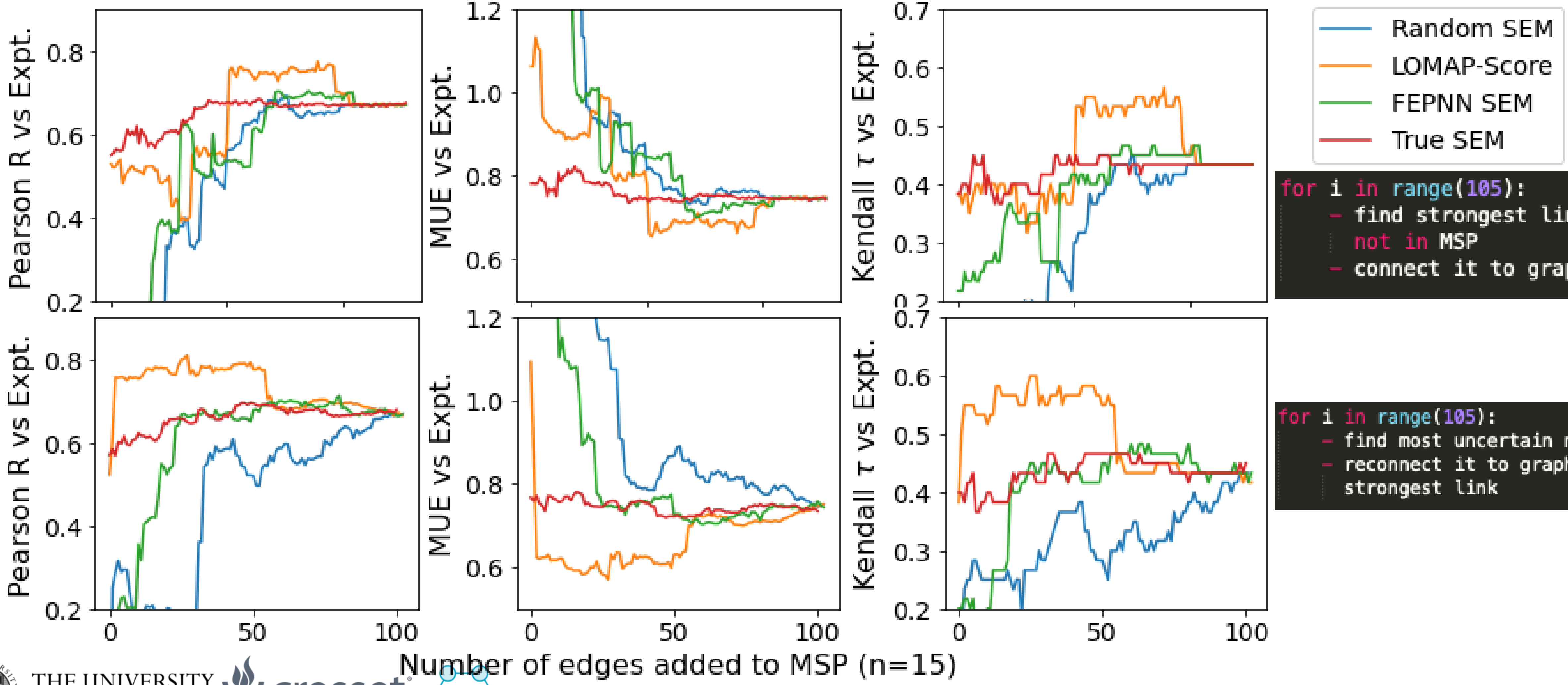
THE UNIVERSITY of EDINBURGH


Pearson R
0.0
0.2
0.4
LOMAP-Score
FEPNN-0
FEPNN-1
FEPNN-2
FEPNN-3
FEPNN-4
FEPNN-5
FEPNN-6
True SEM


MUE / kcal·mol^{-1}
0
1
2
3
LOMAP-Score
FEPNN-0
FEPNN-1
FEPNN-2
FEPNN-3
FEPNN-4
FEPNN-5
FEPNN-6
True SEM

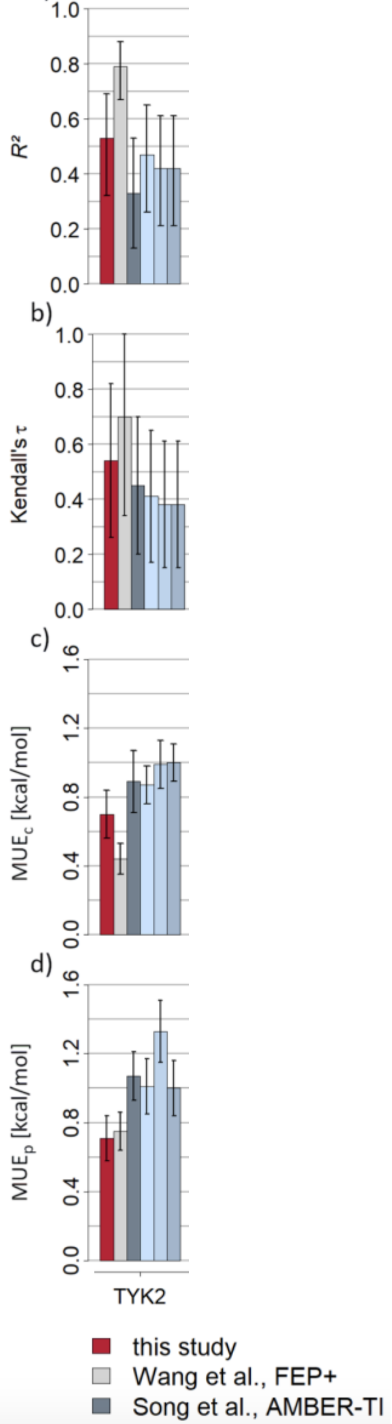

Kendall τ
0.0
0.1
0.2
0.3
0.4
LOMAP-Score
FEPNN-0
FEPNN-1
FEPNN-2
FEPNN-3
FEPNN-4
FEPNN-5
FEPNN-6
True SEM


Pearson R vs Expt.
0.65
0.60
0.55
50
100
MUE vs Expt.
0.82
0.80
0.78
0.76
0.74
50
100
Kendall τ vs Expt.
0.450
0.425
0.400
0.375
0.350
50
100
Number of edges added to MSP

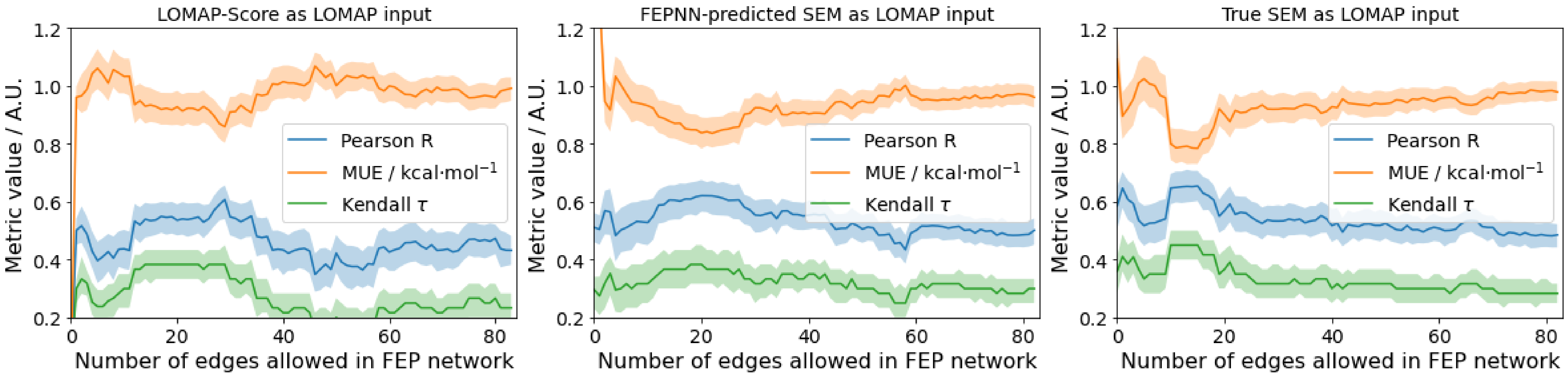



```
for i in range(105):
    - find strongest link
      not in MSP
    - connect it to graph
```

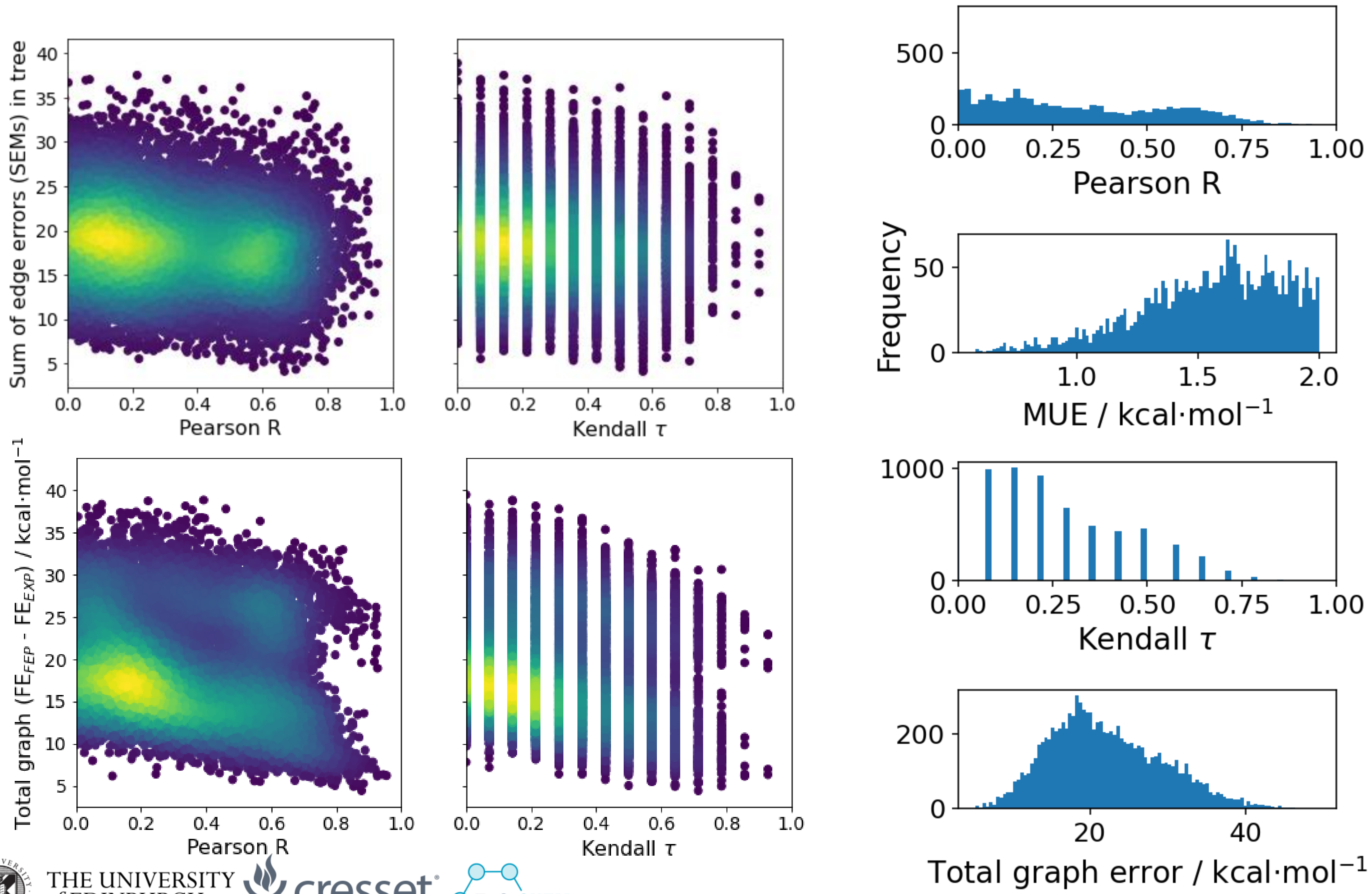
```
for i in range(105):
    - find most uncertain node
    - reconnect it to graph at
      strongest link
```



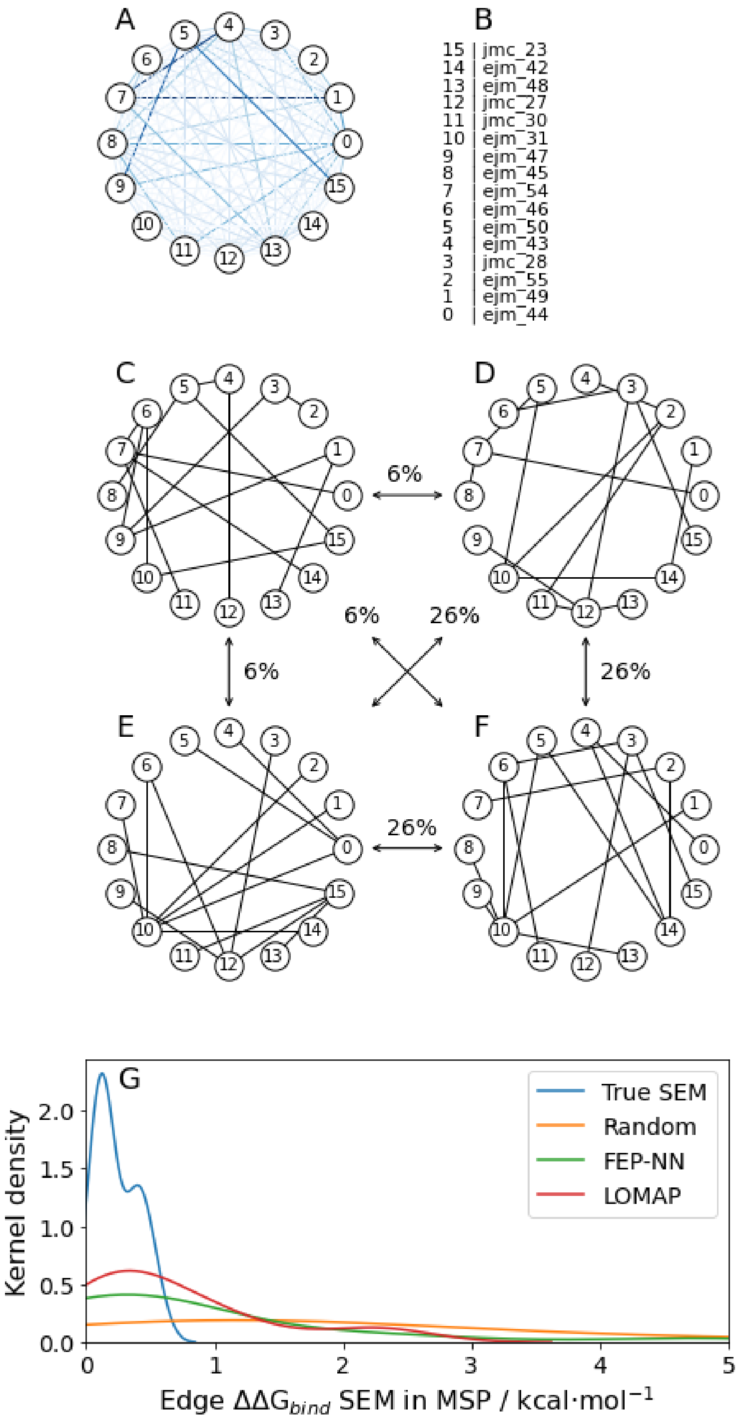
THE UNIVERSITY of EDINBURGH cresset EaStCHEM

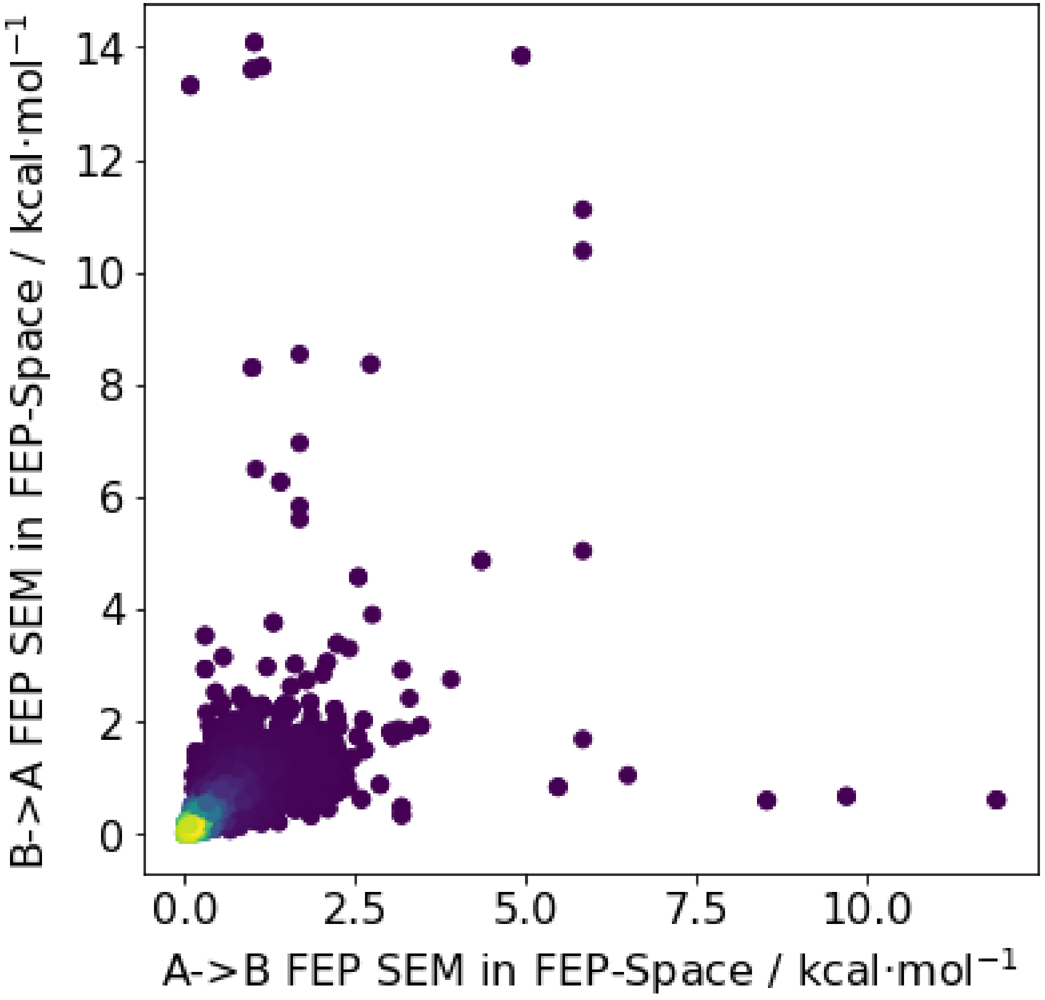


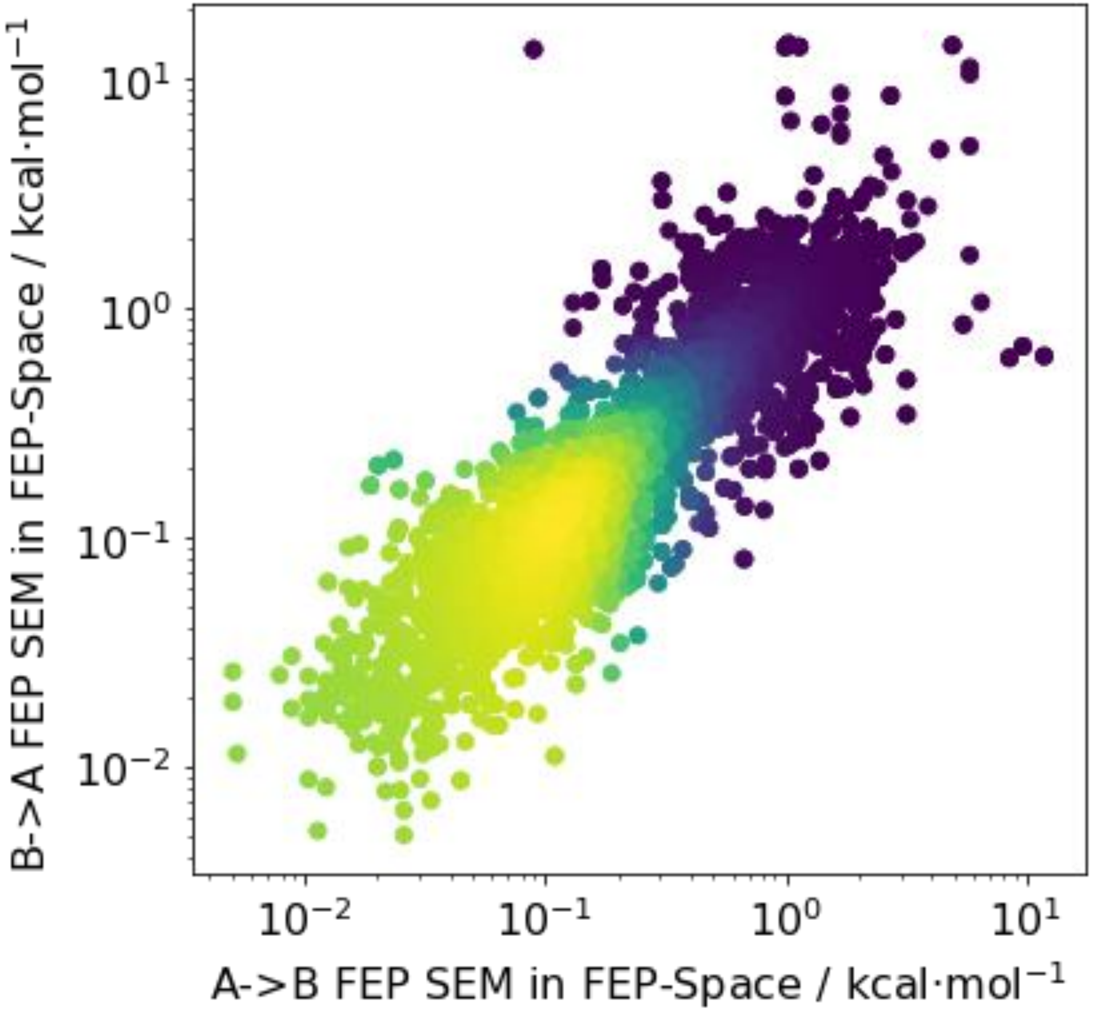
THE UNIVERSITY of EDINBURGH cresset EaStCHEM



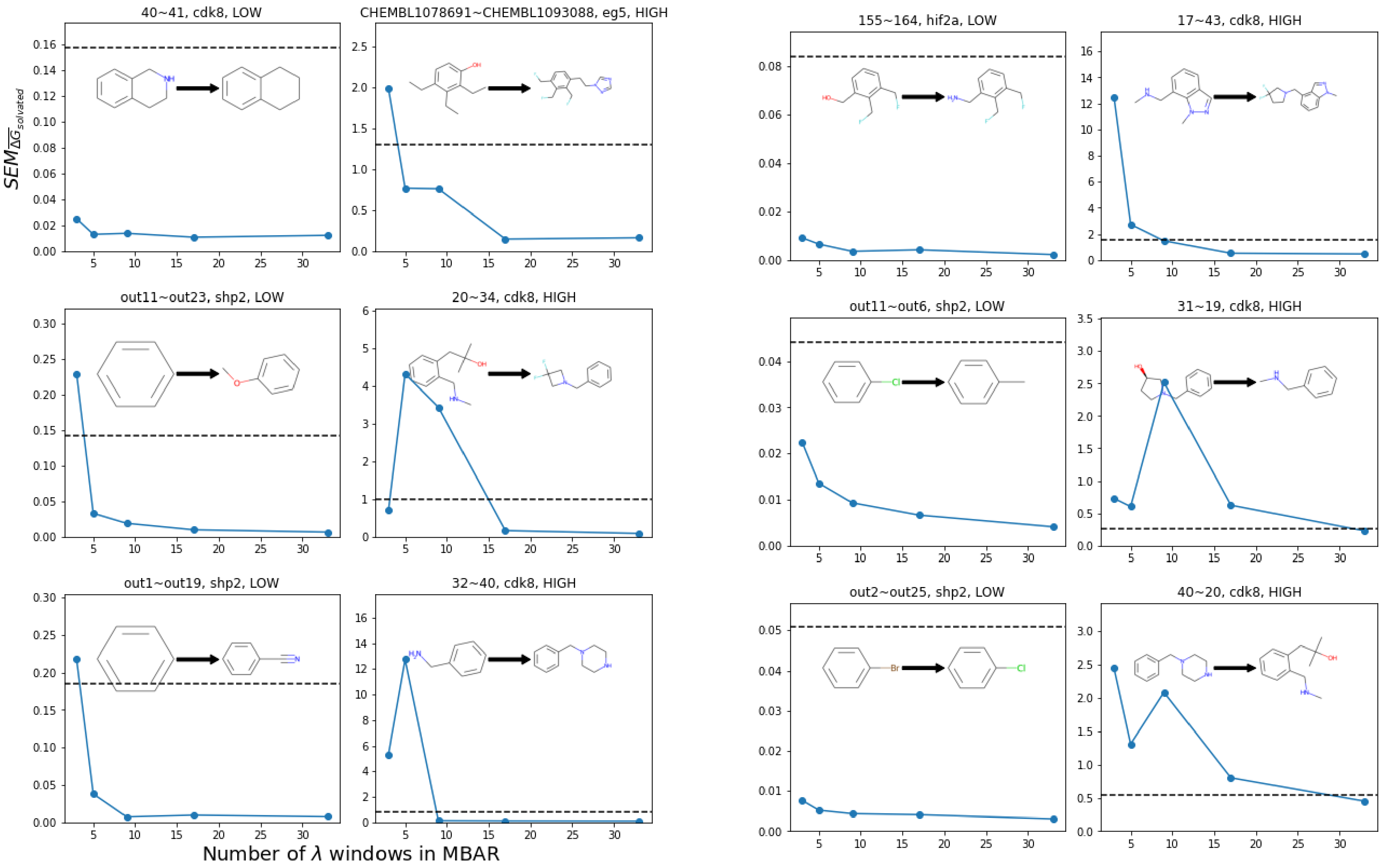
THE UNIVERSITY of EDINBURGH
cresset
EaStCHEM

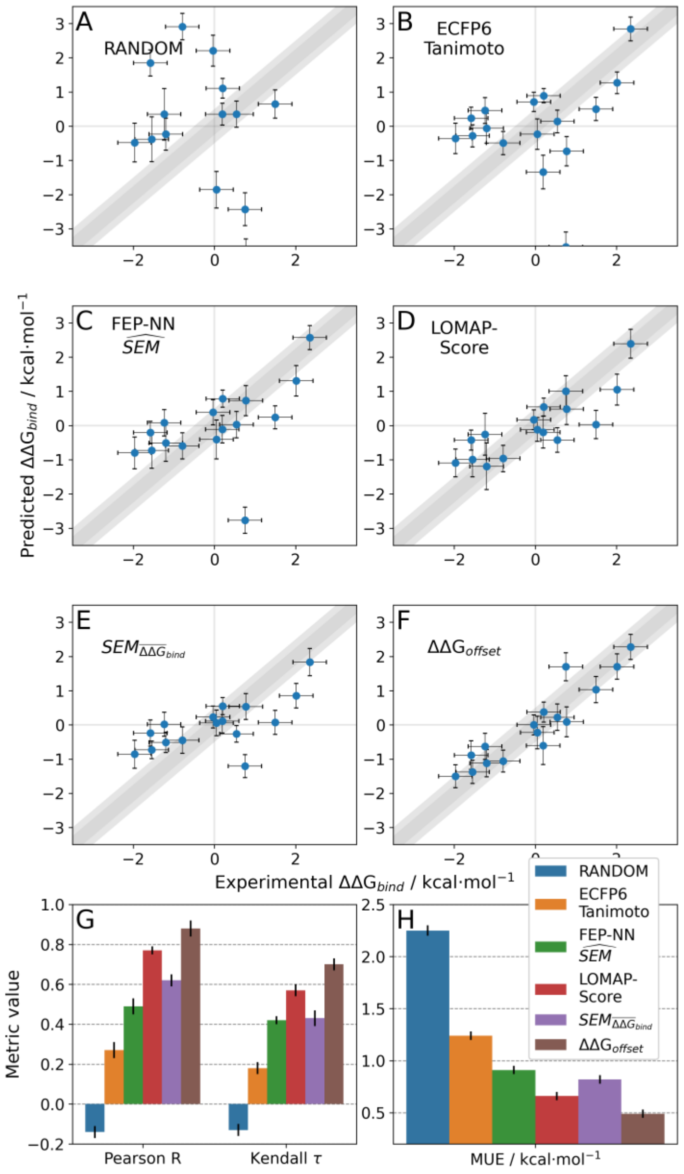

A
B
15 | jmc_23
14 | ejm_42
13 | ejm_48
12 | jmc_27
11 | jmc_30
10 | ejm_31
9 | ejm_47
8 | ejm_45
7 | ejm_54
6 | ejm_46
5 | ejm_50
4 | ejm_43
3 | jmc_28
2 | ejm_55
1 | ejm_49
0 | ejm_44
C
D
E
F
6%
6%
26%
6%
26%
26%
G
True SEM
Random
FEP-NN
LOMAP
Kernel density
Edge ΔΔG$_{bind}$ SEM in MSP / kcal·mol^{-1}

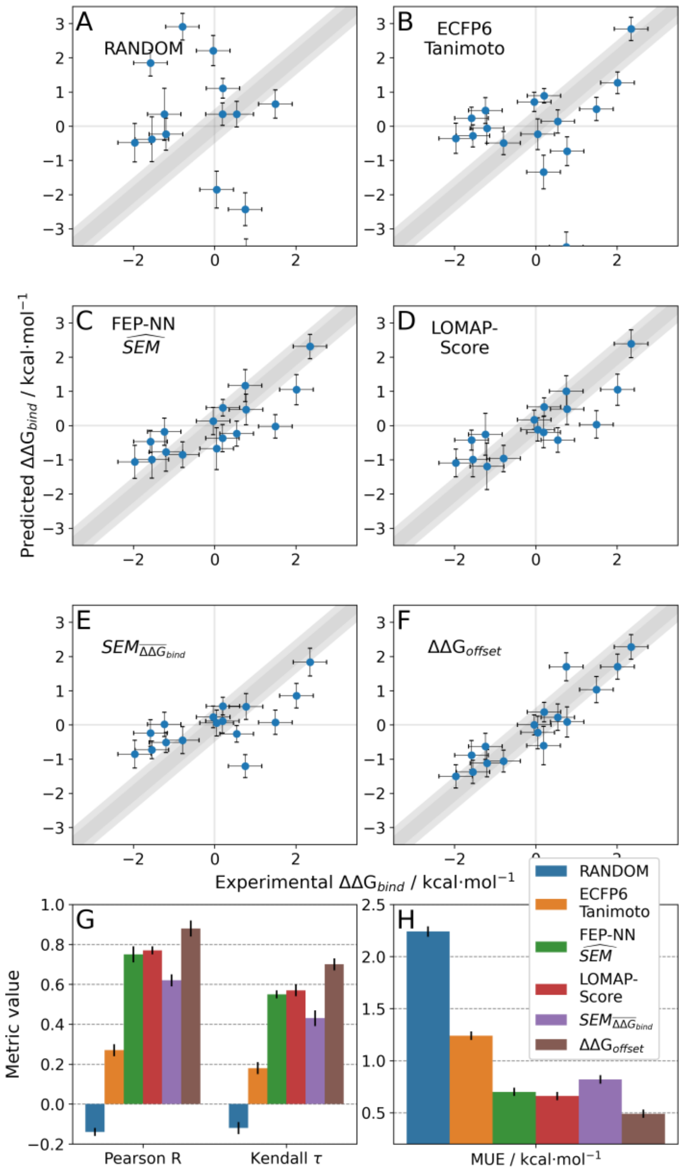




THE UNIVERSITY of EDINBURGH
cresset
EaStCHEM


40~41, cdk8, LOW
CHEMBL1078691~CHEMBL1093088, eg5, HIGH
155~164, hif2a, LOW
17~43, cdk8, HIGH
out11~out23, shp2, LOW
20~34, cdk8, HIGH
out11~out6, shp2, LOW
31~19, cdk8, HIGH
out1~out19, shp2, LOW
32~40, cdk8, HIGH
out2~out25, shp2, LOW
40~20, cdk8, HIGH
$SEM_{\overline{\Delta G}_{solvated}}$
Number of λ windows in MBAR



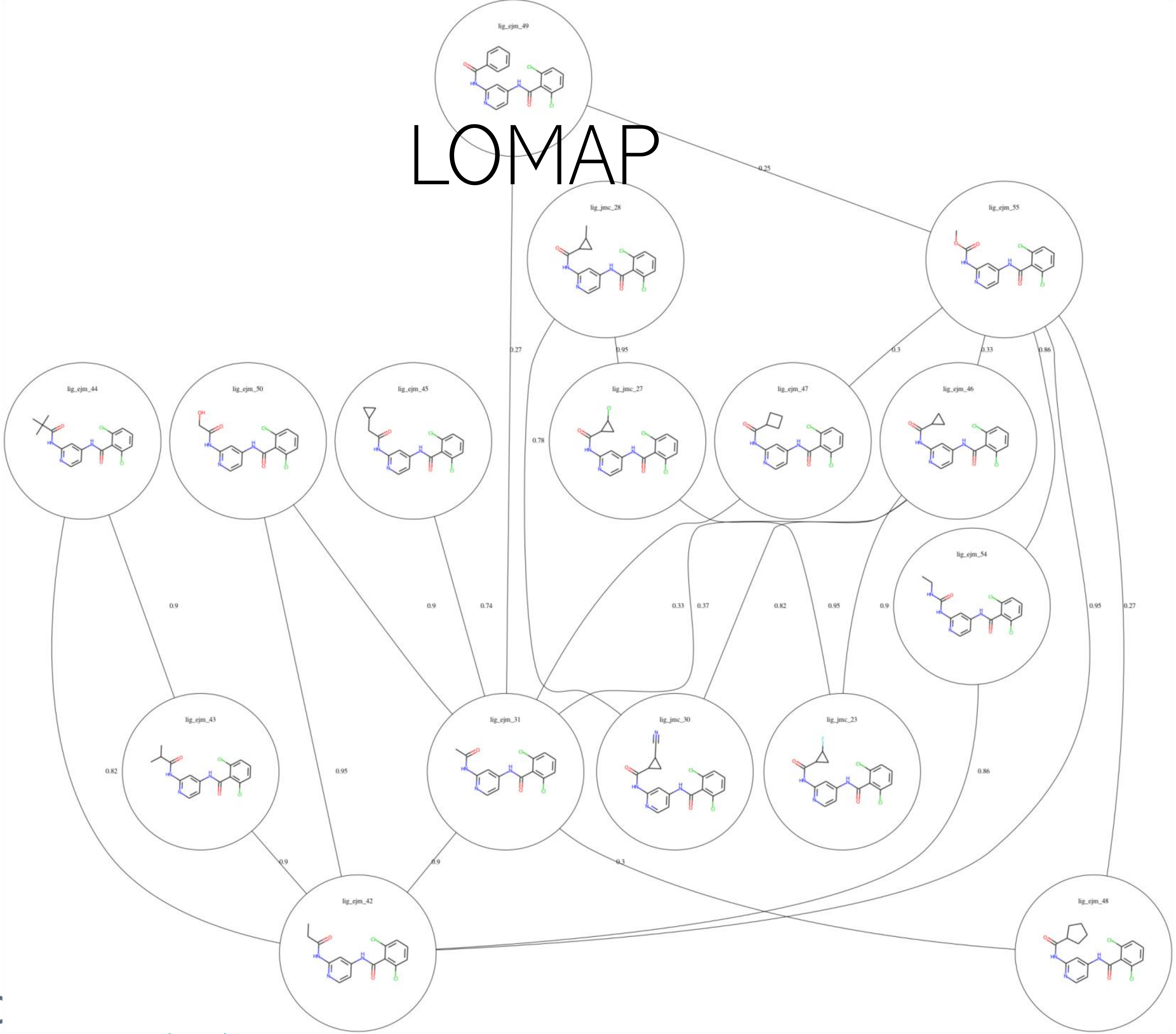


THE UNIVERSITY of EDINBURGH

cresset

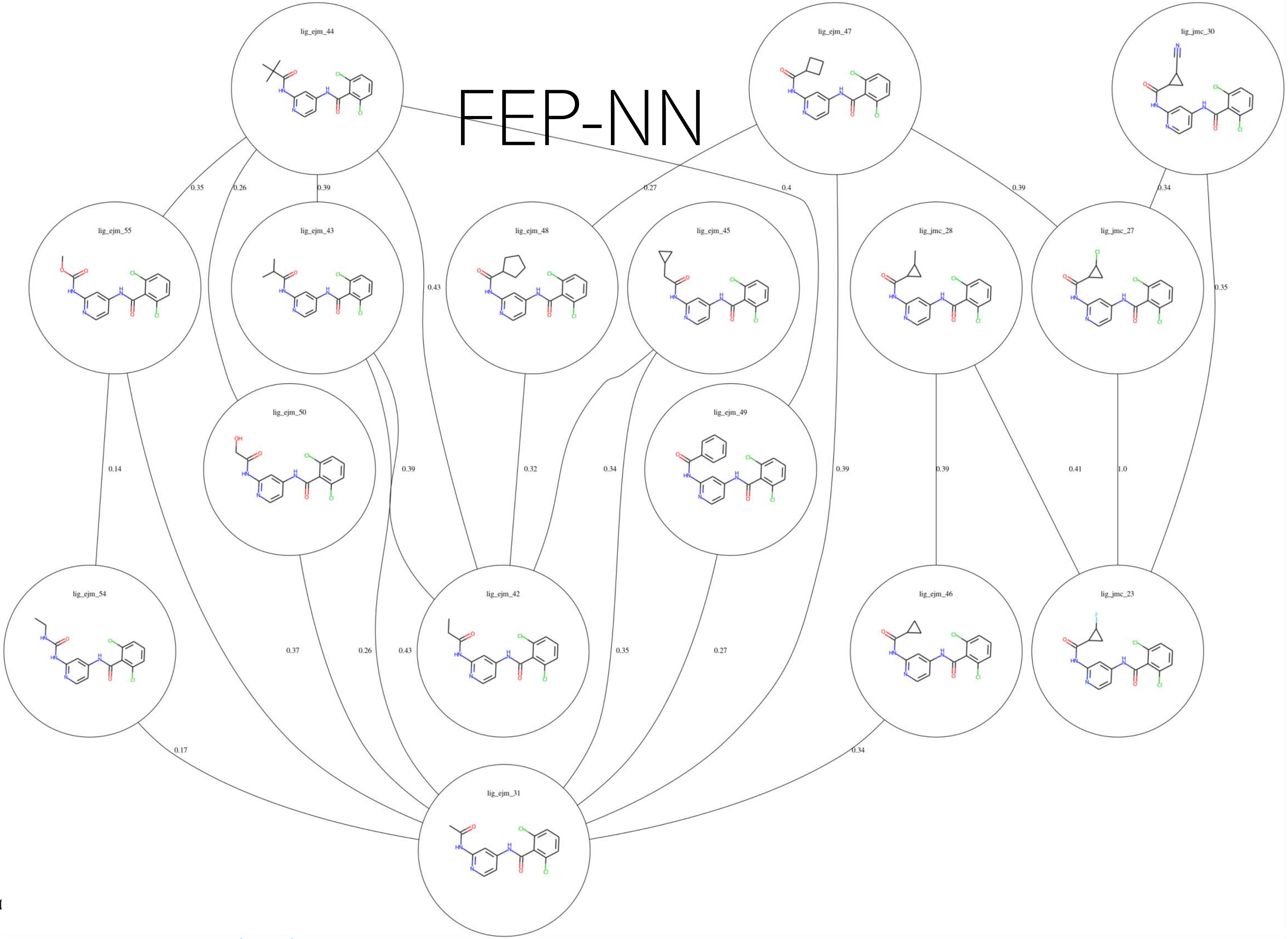
EaStCHEM

LOMAP



THE UNIVERSITY *of* EDINBURGH

FEP-NN


lig_ejm_44
lig_ejm_47
lig_jmc_30
lig_ejm_55
lig_ejm_43
lig_ejm_48
lig_ejm_45
lig_jmc_28
lig_jmc_27
lig_ejm_50
lig_ejm_49
lig_ejm_54
lig_ejm_42
lig_ejm_46
lig_jmc_23
lig_ejm_31
0.35
0.26
0.39
0.43
0.27
0.4
0.39
0.34
0.35
0.14
0.39
0.32
0.34
0.39
0.39
0.41
1.0
0.37
0.26
0.43
0.35
0.27
0.17
0.34