

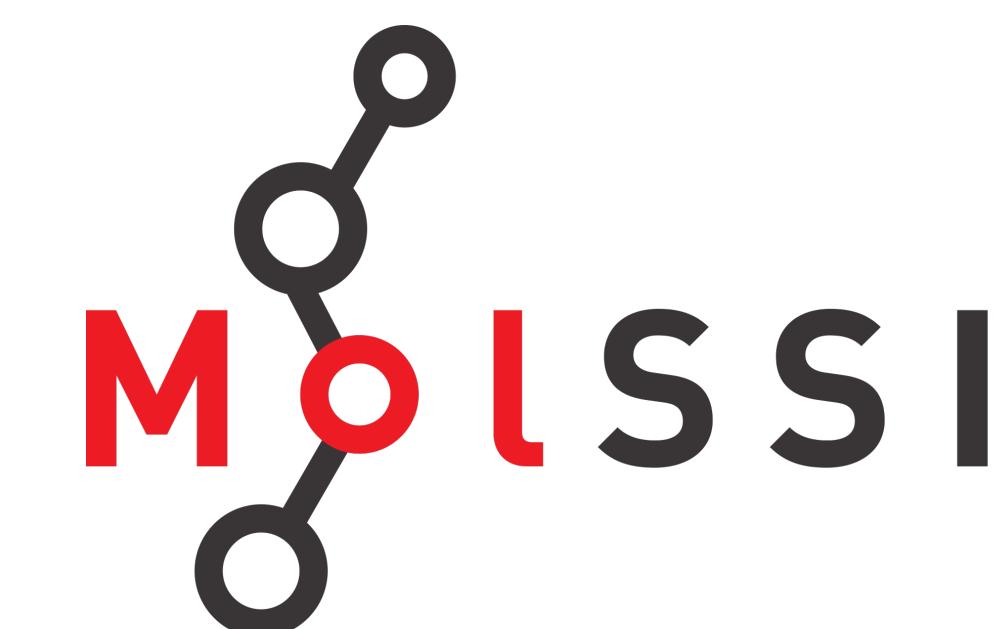
To Design Scalable Free Energy Perturbation Networks, Optimal Is Not Enough

Mary Pitman[†], David Hahn[‡], Gary Tresadern[‡], David Mobley[†]

[†]Department of Pharmaceutical Sciences, University of California, Irvine

[‡]Computational Chemistry, Janssen Research & Development, Belgium

Software Scientist Mentors: Samuel Ellis, Levi Naden



Introduction

Alchemical simulations can measure the change in Gibbs free energy, ΔG , using non-physical intermediates between end states. For ligand binding, ΔG is related to affinity. But, large transformations may be difficult and converge slowly.

To find the set of transformations that minimize errors, we developed the High Information Mapper, HiMap (Figure 1A)[1]. In HiMap a design matrix, \mathbf{A} , maps the true ΔG 's onto our observations. Each observation or calculated value has an unknown error, e_G . The map \mathbf{A} is a graph that defines what ligands (nodes) that transformations (edges) connect (Figure 1B). The weighted linear regression solution to minimize e_G is

$$\hat{\mathbf{G}} = (\mathbf{A}^T \mathbf{W}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}^{-1} \mathbf{G}^{\text{obs}}. \quad (1)$$

From eq 1 we decrease prediction error using the covariance matrix of the parameter errors, \mathbf{C} :

$$\mathbf{C} = (\mathbf{A}^T \mathbf{W}^{-1} \mathbf{A})^{-1}. \quad (2)$$

For the optimization, we hold the edge count fixed and sample connected graphs. After each iteration, if numeric criteria of \mathbf{C} decrease, we accept the design $\mathcal{G}_i(\mathbf{n}, \mathbf{k})$. The criteria we use to find an optimal graph, \mathcal{G}^* , are

$$\mathcal{G}^* = \begin{cases} \min(\det \mathbf{C}) = \min \prod_{i=1}^n \lambda_i(\mathbf{C}) & \text{if D-optimal} \\ \min(\text{Tr } \mathbf{C}) = \min \sum_{i=1}^n \lambda_i(\mathbf{C}) & \text{if A-optimal} \end{cases} \quad (3)$$

where $\lambda_i(\mathbf{C})$ are the eigenvalues of \mathbf{C} with multiplicity.

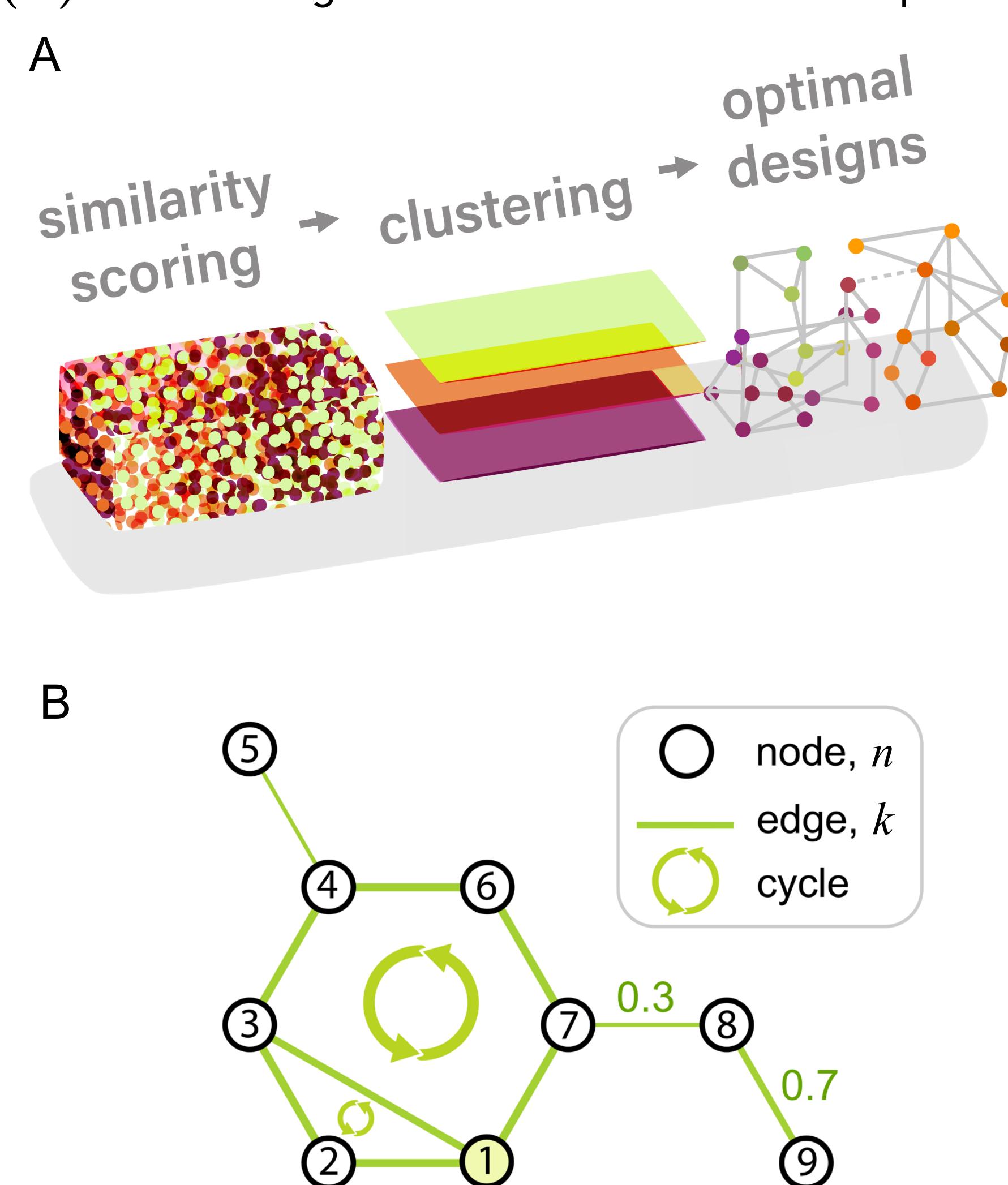


Figure 1: (A) The HiMap algorithm. Ligands are scored by a similarity metric. Then, we cluster ligands with DBSCAN. Finally, we optimize designs. (B) An example design where nodes or ligands, n , are black circles, numbered. The number of edges or perturbations, k , are shown as green lines. Numbered edges depict edge weights. Circular arrows show cycles.

Software Advances

With guidance from MolSSI Software Mentors, we:

- did software profiling and parallelization (Figure 2)
- wrote a vectorized Fedorov Exchange algorithm
- implemented testing with pytest
- used Git version control and CI

With MolSSI support we:

- published our findings in JCIM
- released HiMap with version 2 in development

Performance Improvements

We added features such as deterministic random seed design [2] to decrease the time for minimally connected seed design search. We also polled experts and users who responded that the time for optimization for designs with $n \approx 100$ should be under an hour. Due to the combinatorial expansion of the search, we worked to improve the speed of optimization (Figure 2).

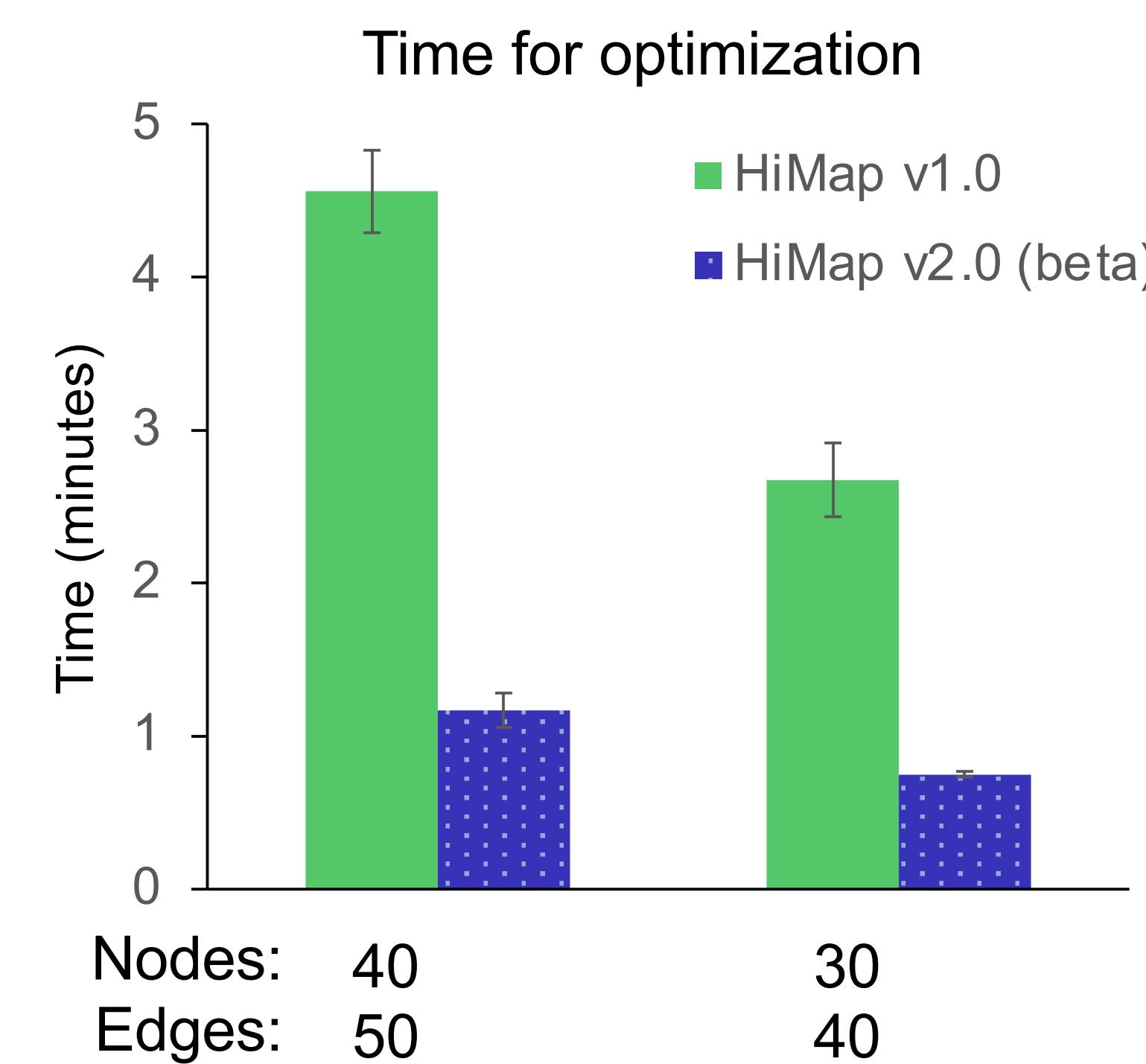


Figure 2: With vectorized and parallelized Fedorov Exchange for optimization, we improved the speed for graph optimization. The relative improvement increases as designs grow in size. Error bars are the standard error of 5 replicates.

Example Outputs

Example HiMap optimization of β -secretase enzyme 1, BACE1, perturbation maps (Figure 3). Designs are output for visual inspection and are returned in python API for automated alchemical simulation setup.

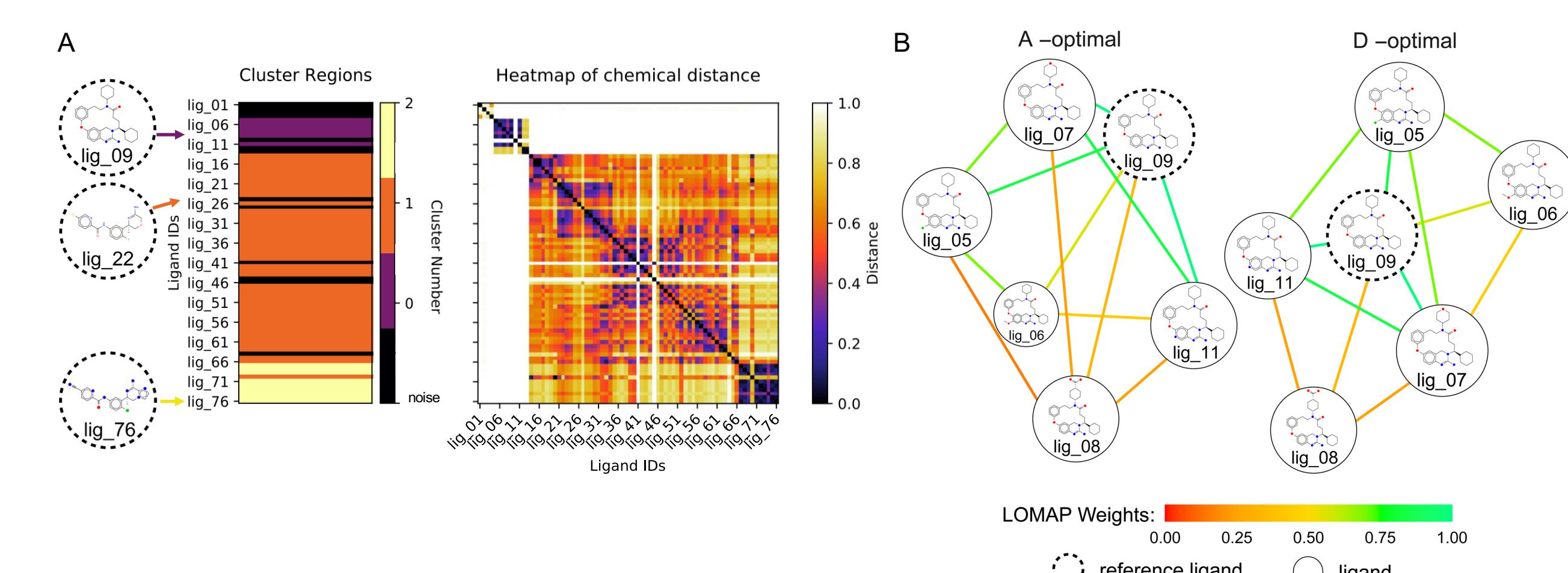


Figure 3: (A) Cluster regions show 3 colored clusters and noise points (black). The most similar ligand in a cluster is shown on the left. The cluster region plot is vertically aligned with the heatmap of chemical distance. Here a value of 0.0 (black) means ligands are identical; a value of 1.0 (white) means ligands are dissimilar. (B) Example perturbation maps for BACE cluster. Low LOMAP weight means the ligands are dissimilar.

References

- (1) Pitman, M.; Mobley, D. HiMap V1.0.0. 2022; <https://github.com/MobleyLab/HiMap>.
- (2) Pitman, M; Hahn, C; Tresadern, G; Mobley, D., J. Chem. Inf. Model. 2023, 63, 6, 1776–1793.

Acknowledgements

Software Fellow Mary Pitman was supported by a fellowship from the Molecular Science Software Institute under NSF grant CHE-2136142 and Janssen Pharmaceuticals.

