Journal of Computer-Aided Molecular Design manuscript No.
(will be inserted by the editor)

Supporting Information for "Lead Optimization Mapper: Automating free energy calculations for lead optimization"

Shuai Liu, Yujie Wu, Teng Lin, Robert Abel, Jonathan P. Redmann, Christopher M. Summa, and David L. Mobley

Received: date / Accepted: date

Abstract This provides an overview of the Supporting Information supplementary files for the associated paper, and details of construction of the trypsin dataset.

Keywords binding free energy · alchemical · planning · molecular dynamics · molecular simulations · lead optimization

1 Supporting Files

The work here examined four compound sets, FXa, SAMPL3, trypsin, and our 504 molecule "fragment set". Each dataset has a corresponding folder in the .tar.gz file distributed with this Supporting Information. In each folder there are the subfolders: `mol2_file`, `mol_pdf_file`, `graph_svg_file` and `graph_dot_file`, which are described below. In addition, the we provide a "graph.pdf" file in the FXa and SAMPL3 folders which contains the final (automatically generated) graph of planned calculations. (This is not provided for the trypsin and fragment datasets as the graph has become prohibitively large). In these figures, structures inside circles are molecules in each dataset, and molecules titles are on the bottom of the circles. Edges represent the planned calculations, the darker the edge is, the higher the similarity is and the more favorable the calculation is. Each edge is labeled with the trimmed maximum common substructure identified for that planned calculation. Titles on top of these substructures correspond to titles of their parents' structures.

Inside the subfolders, the `graph_svg_file` contains:

- An svg file displaying the whole graph (as provided in pdf format as "graph.pdf", discussed just prior); this is named FXa.svg, sampl3.svg, and trypsin.svg, respectively
- Individual svg files showing the 2D structure of each molecule and/or the trimmed common substructure, as indicated by their file names

The subfolder `graph_dot_file` contains one graph .dot file showing the full map of planned relative free energy calculations. And the subfolder `mol_pdf_file` contains individual 2D structures of molecules as images. The subfolder `mol2_file` contains the mol2 files of molecules in each set.

2 Trypsin dataset

The trypsin dataset used here consists in part of the Maybridge fragment library screened for binding for trypsin [7] and used in the SAMPL3 challenge[8]. This was supplemented with a set of known trypsin inhibitors pulled from

S. Liu · D. Mobley
Department of Pharmaceutical Sciences
and Department of Chemistry
147 Bison Modular
University of California, Irvine
Irvine, CA 92697
Tel.: 949-824-6383
Fax: 949-824-2949
E-mail: dmobley@mobleylab.org

D. Mobley also
University of New Orleans
2000 Lakeshore Drive
New Orleans, LA 70148

Y. Wu, T. Lin and R. Abel
Schrödinger
120 West 45th Street
17th Floor, Tower 45
New York, NY 10036-4041

J. Redmann and C. Summa
Department of Computer Science
University of New Orleans
2000 Lakeshore Drive
New Orleans, LA 70148

BindingDB[6] and from the literature. Some of these are very like benzamidine, the "classic" trypsin inhibitor, while others are substantially different and/or bulkier (though many if not all include benzamidine as ca component). Files for all of the compounds, including the Maybridge library, are provided in the supporting information. Specifically, from BindingDB, we added *p*-*n*-butylbenzamidine, *p*-*n*-ethyl-benzamidine, *p*-isopropyl-benzamidine, *p*-*n*-pentyl-benzamidine, *p*-*n*-propyl-benzamidine, benzamidine, and *p*-methylbenzamidine from ref.[9], benziothiazole analogs 1, 2, 3, 4 and phenylglycine derivatives 1, 5, 6, 11 and 16 from ref. [1], and APC-1-762, CRA-8249, CRA-15566, APC-1144, CRA-16847, CRA-18305, and CRA 19858 from ref. [3], as well as APC-7377 and APC-7528 from ref. [4], ZZ-BABCH and CHEBI254833 from ref. [2] and pyridine template III from ref.[10]. We also added benzylammonium, 2-phenylethylammonium, and 3-phenylpropylammonium from ref. [5]. This consisted a total of 33 knowns added to the set, and included a variety of different net charges to provide a reference for most of the different net charges in the Maybridge library.

3 Randomly selected compounds from the fragment set

To illustrate the necessity of having an automated algorithm for planning calculations spanning large sets, we took our 504 molecule fragment set and randomly selected 10 pairs of compounds using the Python random module. These are displayed in Figure 1. As discussed in the main text, it is fairly clear that randomly picking compound pairs for relative free energy calculations typically results in pairs which most practitioners would disfavor.

References

1. Groebke Zbinden, K., Banner, D.W., Hilpert, K., Himber, J., Lavé, T., Riederer, M.A., Stahl, M., Tschopp, T.B., Obst-Sander, U.: Dose-dependent antithrombotic activity of an orally active tissue factor/factor VIIa inhibitor without concomitant enhancement of bleeding propensity. Bioorganic & Medicinal Chemistry **14**(15), 5357–5369 (2006)
2. Guilford, W.J., Shaw, K.J., Dallas, J.L., Koovakkat, S., Lee, W., Liang, A., Light, D.R., McCarrick, M.A., Whitlow, M., Ye, B., Morrissey, M.M.: Synthesis, Characterization, and StructureActivity Relationships of Amidine-Substituted (Bis)benzylidene-Cycloketone Olefin Isomers as Potent and Selective Factor Xa Inhibitors1,2. J Med Chem **42**(26), 5415–5425 (1999)
3. Katz, B.A., Elrod, K., Verner, E., Mackman, R.L., Luong, C., Shrader, W.D., Sendzik, M., Spencer, J.R., Sprengeler, P.A., Kolesnikov, A.: Elaborate Manifold of Short Hydrogen Bond Arrays Mediating Binding of Active Site-directed Serine Protease Inhibitors. J. Mol. Biol. **329**(1), 93–120 (2003)
4. Katz, B.A., Mackman, R., Luong, C., Radika, K., Martelli, A., Sprengeler, P.A., Wang, J., Chan, H., Wong, L.: Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. Chemistry & biology **7**(4), 299–312 (2000)
5. Leiros, H., Brandsdal, B.O., Andersen, O., Os, V., Leiros, I., Helland, R., Otlewski, J., Willassen, N., Smalås, A.: Trypsin specificity as elucidated by LIE calculations, Xray structures, and association constant measurements. Protein Science **13**(4), 1056–1070 (2004)
6. Liu, T., Lin, Y., Wen, X., Jorissen, R.N., Gilson, M.K.: BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic acids research **35**(suppl 1), D198–D201 (2007)
7. Newman, J., Dolezal, O., Fazio, V., Caradoc-Davies, T., Peat, T.S.: The DINGO dataset: a comprehensive set of data for the SAMPL challenge. J Comput Aided Mol Des (2011)
8. Skillman, A.G.: SAMPL3: blinded prediction of host-guest binding affinities, hydration free energies, and trypsin inhibitors. J Comput Aided Mol Des (2012)
9. Talhout, R., Villa, A., Mark, A.E., Engberts, J.B.F.N.: Understanding Binding Affinity: A Combined Isothermal Titration Calorimetry/Molecular Dynamics Study of the Binding of a Series of Hydrophobically Modified Benzamidium Chloride Inhibitors to Trypsin. J Am Chem Soc **125**, 10,570–10,579 (2003)
10. Whitlow, M., Arnaiz, D.O., Buckman, B.O., Davey, D.D., Griedel, B., Guilford, W.J., Koovakkat, S.K., Liang, A., Mohan, R., Phillips, G.B., Seto, M., Shaw, K.J., Xu, W., Zhao, Z., Light, D.R., Morrissey, M.M.: Crystallographic analysis of potent and selective factor Xa inhibitors complexed to bovine trypsin. Acta Crystallographica Section D: Biological Crystallography **55**(8), 1395–1404 (1999)

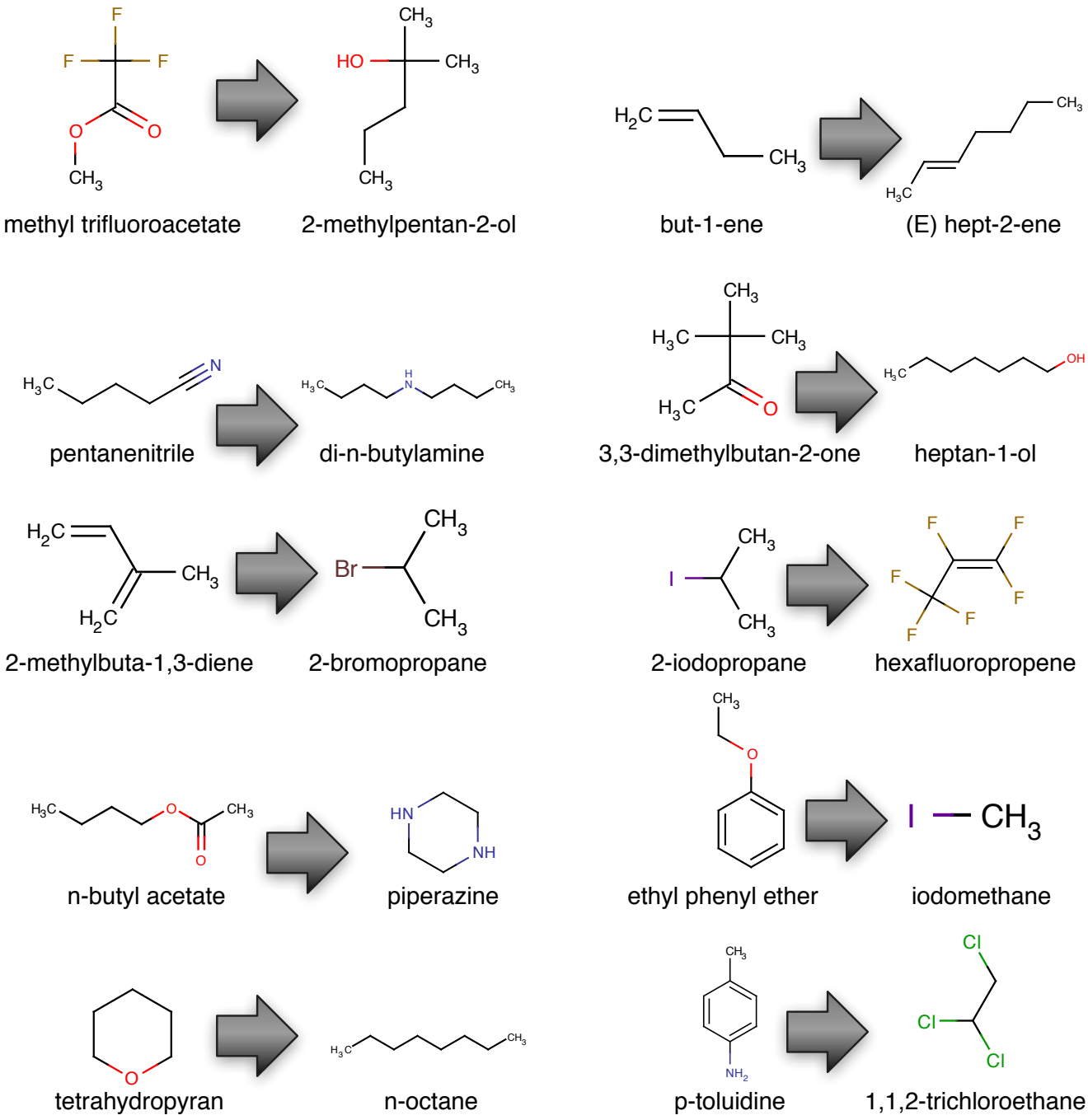

Fig. 1 Randomly selecting compounds from the fragment set for relative free energy calculations results in pairs which typically would be disfavored by practitioners of free energy calculations, as discussed in the main text of the paper.