Supporting Information

FEP-Augmentation as a Means to Solve Data Paucity Problems for Machine Learning in Chemical Biology

Pieter B Burger[†], Xiaohu Hu[‡], Ilya Balabin[†], Morné Muller[†], Megan Stanley[#], Fourie Joubert[ł] and Thomas M Kaiser[†*]*

[†]Avicenna Biosciences Inc. 101 W. Chapel Hill Street, Suite 210, Durham, NC, 27001. [‡]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States. [#]Microsoft Research AI4Science, Cambridge CB1 2FB, U.K. [ł]Centre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0001, South Africa

Corresponding Author

For FEP Inquiries: Pieter B Burger, Avicenna Biosciences Inc. 101 W. Chapel Hill Street, Suite 210, Durham, NC, 27001. pburger@avicenna-bio.com

For Machine Learning and Chemical Biology Inquiries: Thomas M Kaiser, Avicenna Biosciences Inc. 101 W. Chapel Hill Street, Suite 210, Durham, NC, 27001. tkaiser@avicenna-bio.com

Additional structure description: The selected structures PDBid's 3EN4 and 3EN5 were prepared for free energy calculations. Firstly, we constructed the missing activation loops for both 3EN4 and 3EN5. For structure 3EN4, residues 411 through 419 were missing, and for structure 3EN5, residues 408 through 424 were missing. To provide context of the proximity of the missing loops to the ligands, we found that the origins of the unresolved loops (closest point to the ligand) were ~17 and ~7Å away from the nearest atom of the co-resolved ligand for the PDB structures 3EN4 and 3EN5, respectively. Next the protein was prepared as described in the methods section.

Sampling of the dual binding mode: No conformational interconversion took place during the ABFE calculation, and indeed, none was observed. If conformer interconversion were to happen during the calculation window, it would provide enough sampling of the energy landscape and a single FEP calculation would suffice. We also looked at the conformational dynamics in RBFE calculations with various times scales during the initial FEP+ runs (5, 10, 20 and 100 ns time windows) for a limited set of compounds and no compounds showed conformational interconversions during these longer FEP runs. Finally, we conducted 500ns molecular dynamics (MD) simulations as an added control on the S1 and PP494 protein-ligand complexes in both the 3EN4 and 3EN5 structures. Throughout these simulations, we did not observe any conformational interconversion indicating that the ring conformation remained

stable during the simulation period. This suggests that sufficient sampling of the ring conformation would most likely not be achieved through a single calculation.

Protein relaxation prior to FEP calculations: As the last step before initiating FEP calculations, a relaxation and evaluation process was carried out on all six protein-reference ligand complexes used in this study. This process involved subjecting the complexes to a 10ns molecular dynamics (MD) simulation to assess their stability and ensure the utilization of a locally minimized system that reflected the global minimum. The primary objective was to observe and monitor the complexes for any unexpected ligand movement or protein structural changes covering the standard FEP calculation simulation time window (5 ns) + an additional 5 ns window. These steps ensured the appropriate positioning of the ligands and provided a reliable starting configuration for the subsequent MD simulation. If no issues or anomalies were observed during the MD simulations, a protein-ligand complex was selected and captured after 1.2ns of simulation time. This specific cutoff of 1.2ns was chosen based on the observation that the protein reached an equilibrated state and exhibited consistent behavior throughout the remaining simulation time. While it is possible to use different cutoffs, maintaining consistency with the same cutoff was preferred for the sake of comparability and ensuring that well-behaved simulations were treated uniformly. The captured complex was then subjected to a minimization and the optimized protein-ligand complex was used for further studies. The MD report are provided as pdf's.

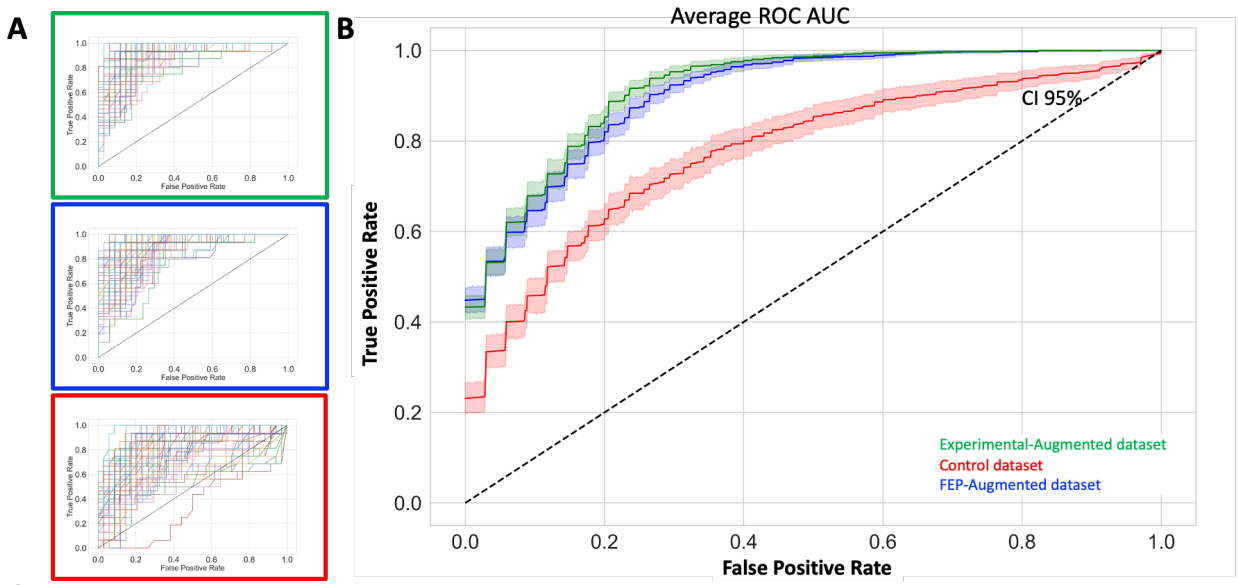
Assignment of Scaffold 1 to the inhibitors classes Type 1 and Type 1 ½ : In this study we assigned the Scaffold 1 inhibitors into the two classes of inhibitors as if we were performing a real-world lead optimization project, starting from a limited SAR dataset and a handful of crystal structures. We thus only used the information that 2 inhibitor types were present and that the

difference between these classes laid in the structural change where a hydroxyl was change to a methoxy. Therefor the classification we used here are not comprehensive in assigning all compounds to the correct type of inhibitor but give us a proxy of what to expect in real world application.

Table S1:

The statistical evaluation of machine learning algorithms produced under three distinct experimental conditions, each assessed at three different decision thresholds: 31nM, 50nM, and 239nM.

Decision Value		ROC AUC	Sensitivity	Precision	Enrichment Factor at 10%	Enrichment Factor at 20%
Condition	**DV**	**Mean ± Std Dev**	**Mean ± Std Dev**	**Mean ± Std Dev**	**Mean ± Std Dev**	**Mean ± Std Dev**
Experimental-Augmented dataset	31	0.898 ± 0.045	0.555 ± 0.137	0.808 ± 0.160	3.80 ± 0.69	2.98 ± 0.51
11 Compound dataset	31	0.779 ± 0.097	0.258 ± 0.247	0.699 ± 0.277	2.87 ± 1.03	2.35 ± 0.63
FEP- Augmented dataset	31	0.868 ± 0.054	0.568 ± 0.140	0.655 ± 0.144	3.57 ± 0.73	2.73 ± 0.56
Experimental-Augmented dataset	50	0.921 ± 0.034	0.656 ± 0.137	0.764 ± 0.116	3.18 ± 0.48	2.73 ± 0.43
11 Compound dataset	50	0.794 ± 0.100	0.307 ± 0.251	0.768 ± 0.203	2.57 ± 0.78	2.26 ± 0.46
FEP-Augmented dataset	50	0.910 ± 0.045	0.737 ± 0.144	0.709 ± 0.097	3.14 ± 0.49	2.64 ± 0.42
Experimental-Augmented dataset	239	0.909 ± 0.036	0.819 ± 0.081	0.844 ± 0.072	2.02 ± 0.18	1.99 ± 0.19
11 Compound dataset	239	0.806 ± 0.062	0.857 ± 0.173	0.624 ± 0.130	1.92 ± 0.29	1.82 ± 0.25
FEP-Augmented dataset	239	0.898 ± 0.042	0.833 ± 0.072	0.770 ± 0.066	2.02 ± 0.18	1.95 ± 0.21



C

Decision Value		ROC AUC	Sensitivity	Precision	Enrichment Factor at 10%	Enrichment Factor at 20%
Dataset	**DV**	**Mean ± Std Dev**	**Mean ± Std Dev**	**Mean ± Std Dev**	**Mean ± Std Dev**	**Mean ± Std Dev**
Experimental-Augmented	50	0.916 ± 0.038	0.626 ± 0.133	0.764 ± 0.107	3.15 ± 0.25	2.64 ± 0.32
Control	50	0.777 ± 0.113	0.289 ± 0.246	0.719 ± 0.244	2.54 ± 0.78	2.24 ± 0.47
FEP-Augmented	50	0.905 ± 0.044	0.731 ± 0.160	0.698 ± 0.111	3.16 ± 0.27	2.62 ± 0.42

Figure S1: A) This figure illustrates the results when the in vitro IC_{50} values are used when classifying subset E instead of the FEP derived IC_{50} values. The ROC curves corresponding to all 100 iterations across the three distinct experimental scenarios using a decision value (DV) of 50nM are displayed. The curve representing the Experimental-Augmented dataset is color-coded in green, the Control dataset is depicted in red, and the FEP-Augmented dataset is illustrated in blue. B) The mean ROC curve, aggregated from all 100 iterations, was graphed for each experimental condition, and it is accompanied by a 95% confidence interval (shaded areas). C) A table is presented that compares the average statistical performance of the three machine learning experiments, each evaluated over 100 iterations and at a DV of 50 nM.