Supporting Information

Accelerating Lead Identification by High Throughput Virtual Screening: Prospective Case Studies from the Pharmaceutical Industry

Kelly L. Damm-Ganamet[τ,§], Nidhi Arora[τ,§], Stephane Becart[θ], James P. Edwards[τ], Alec D. Lebsack[θ], Heather M. McAllister[τ], Marina I. Nelen[γ], Navin L. Rao[θ], Lori Westover[γ], John J. M. Wiener[θ], Taraneh Mirzadegan[τ,*]

[τ]*Discovery Sciences and* [θ]*Immunology, Janssen Research and Development, 3210 Merryfield Row, San Diego, CA 92121, USA.*

[γ]*Discovery Sciences, Janssen Research and Development, Welsh and McKean Roads, Spring House, PA 19477, USA.*

Corresponding Author
E-mail: TMirzade@its.jnj.com

Table S1: BTK Benchmark Study

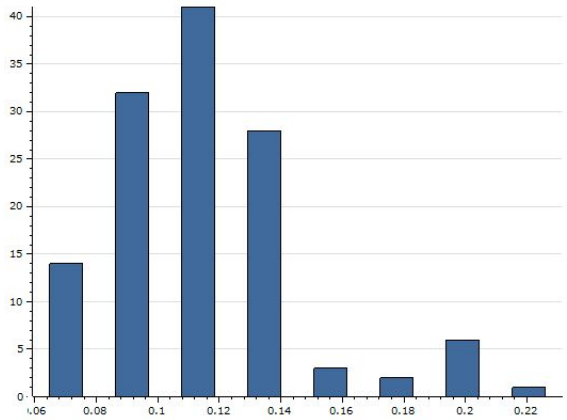
Figure S2: Histograms of the Tanimoto coefficients for five HTVS identified BTK compounds compared to 127 known BTK inhibitors from the literature

Table S3: BTK X-Ray Crystallography Statistics

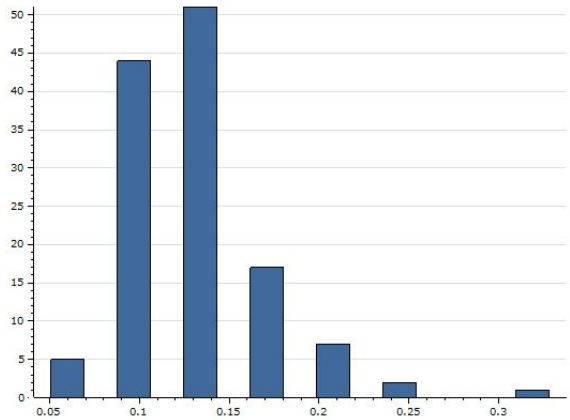
Table S1: BTK Benchmark Study. A benchmark study was completed with the purpose of determining the most optimal software and HTVS parameters for docking into BTK. A drug-like decoy database of 10,993 compounds selected randomly from the 891,982 compound HTVS database along with 16 known actives were docked into both the 3OCS and 3PJ3 structures. The benchmarking screens were carried out using Glide-SP, Glide-XP, FRED with/without constraints, and combinations of FRED/Glide-XP to determine maximum enrichment at sampling of 1% and 2% of database with a reasonably short runtime. The results are shown below.

Docking Protocol	Pocket	Enrichment 1% Database	Enrichment 2% Database	Run Time On 20 cpus (Hours)
Glide SP (No Constraints)	3PJ3	0	9.37	10.43
Glide SP (No Constraints)	3OCS	0	0	17.8
Glide XP (No Constraints)	3PJ3	12.49	9.37	120.37
Glide SP (2HBond constraints)	**3PJ3**	**24.98**	**12.49**	**11.62**
Glide SP (2HBond constraints)	3OCS	6.25	6.25	10.45
Glide XP (1Tautomer) (No Constraints)	3PJ3	6.24	9.37	36.7
Glide XP (1Tautomer) (No Constraints)	**3OCS**	**24.98**	**15.61**	**17.65**
Glide XP(1Tautomer) (2HBond constraints)	3PJ3	12.49	9.37	16.83
Glide XP(1Tautomer) (2HBond constraints)	**3OCS**	**31.23**	**18.73**	**16.95**
FRED (No Constraints)	3PJ3	6.25	6.25	2.36
FRED (No Constraints)	3OCS	0	0	3.44
FRED (1HBond constraints)	3PJ3	6.25	6.25	2.12
FRED (1HBond constraints)	3OCS	0	3.12	1.32
a) FRED (1HBond constraint) b) Top 30% (All tautomers) ->Glide SP (2HBond constraints)	3PJ3	6.21	3.13	1.91
a) FRED (1HBond constraint) b) Top 30% (All tautomers) ->Glide XP (2HBond constraints)	3PJ3	12.4	6.29	16.49
a) FRED (1HBond constraint) b) Top 30% (1 Tautomer) ->Glide XP (2HBond constraints)	**3OCS**	**31.37**	**15.69**	**4.22**

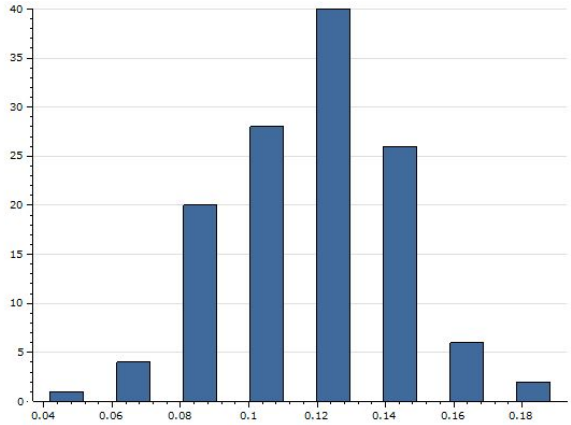
Figure S2: Histograms of the Tanimoto coefficients for five HTVS identified BTK compounds compared to 127 known BTK inhibitors from the literature.



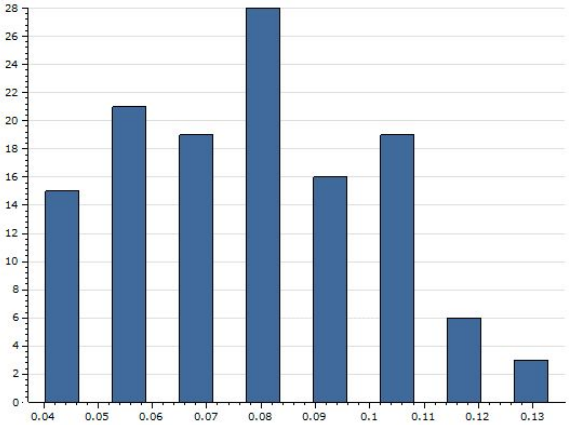
Compound 1



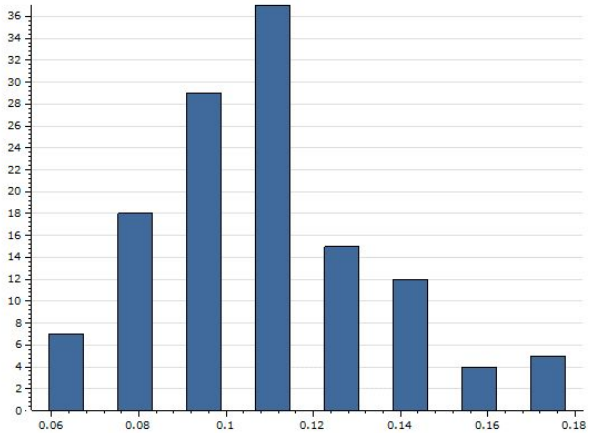
Compound 2



Compound 3



Compound 4



Compound 5

Table S3: BTK X-Ray Crystallography Statistics. Data collection and processing statistics for Compounds 1 and 5 and statistics of the final structure and refinement process.

Compound 1: Data Collection and Processing Statistics

X-ray source	PXI/X06SA (SLS[1])
Wavelength [Å]	0.99987
Detector	PILATUS 6M
Temperature [K]	100
Space group	P 2_1 2_1 2
Cell: a; b; c; [Å]	72.12; 104.37; 38.14
α; β; γ; [°]	90.0; 90.0; 90.0
Resolution [Å]	1.40 (1.65-1.40)
Unique reflections	54763 (21370)
Multiplicity	2.3 (2.4)
Completeness [%]	95.2 (96.9)
R_{sym} [%][3]	5.9 (48.1)
R_{meas} [%][4]	7.5 (61.5)
Mean(I)/sd[5]	10.00 (2.08)

[1] SWISS LIGHT SOURCE (SLS, Villigen, Switzerland)

[2] values in parenthesis refer to the highest resolution bin.

[3] $$Rsym = \frac{\sum_h \sum_i^{n_h} \left| \hat{I}_h - I_{h,i} \right|}{\sum_h \sum_i^{n_h} I_{h,i}} \text{ with } \hat{I}_h = \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$$

where $I_{h,i}$ is the intensity value of the *i*th measurement of *h*

[4] $$Rmeas = \frac{\sum_h \sqrt{\frac{n_h}{n_h - 1}} \sum_i^{n_h} \left| \hat{I}_h - I_{h,i} \right|}{\sum_h \sum_i^{n_h} I_{h,i}} \text{ with } \hat{I}_h = \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$$

where $I_{h,i}$ is the intensity value of the *i*th measurement of *h*

[5] calculated from independent reflections

Compound 5: Data Collection and Processing Statistics

X-ray source	PXI/X06SA (SLS[1])
Wavelength [Å]	0.99987
Detector	PILATUS 6M
Temperature [K]	100
Space group	P 2_1 2_1 2
Cell: a; b; c; [Å]	72.98; 105.55; 38.20
α; β; γ; [°]	90.0; 90.0; 90.0
Resolution [Å]	2.41 (2.60-2.41)
Unique reflections	11994 (2401)
Multiplicity	7.8 (8.1)
Completeness [%]	100.0 (100.0)
R_{sym} [%][3]	6.5 (48.5)
R_{meas} [%][4]	7.4 (55.4)
Mean(I)/sd[5]	20.6 (4.2)

[1] SWISS LIGHT SOURCE (SLS, Villigen, Switzerland)

[2] values in parenthesis refer to the highest resolution bin.

[3] $$Rsym = \frac{\sum_h \sum_i^{n_h} \left| \hat{I}_h - I_{h,i} \right|}{\sum_h \sum_i^{n_h} I_{h,i}} \quad \text{with} \quad \hat{I}_h = \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$$

where $I_{h,i}$ is the intensity value of the ith measurement of h

[4] $$Rmeas = \frac{\sum_h \sqrt{\frac{n_h}{n_h - 1}} \sum_i^{n_h} \left| \hat{I}_h - I_{h,i} \right|}{\sum_h \sum_i^{n_h} I_{h,i}} \quad \text{with} \quad \hat{I}_h = \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$$

where $I_{h,i}$ is the intensity value of the ith measurement of h

[5] calculated from independent reflections

Compound 1: Refinement Statistics

Resolution [Å]	59.33-1.40
Number of reflections (working /test)	49736 / 5022
R_{cryst} [%]	15.6
R_{free}[%][2]	20.5
Total number of atoms:	
Protein	2367
Water	276
Ligand	36
1,2-Ethanediol	24
Deviation from ideal geometry: [3]	
Bond lengths [Å]	0.012
Bond angles [°]	1.43
Bonded B's [Å^2] [4]	3.5
Ramachandran plot: [5]	
Most favoured regions [%]	93.5
Additional allowed regions [%]	6.5
Generously allowed regions [%]	0.0
Disallowed regions [%]	0.0

[1] Values as defined in REFMAC5, without sigma cut-off

[2] Test-set contains 2.4 % of measured reflections

[3] Root mean square deviations from geometric target values

[4] Calculated with MOLEMAN

[5] Calculated with PROCHECK

Compound 5: Refinement Statistics

Resolution [Å]	60.03-2.41
Number of reflections (working /test)	10900 / 1058
R_{cryst} [%]	21.7
R_{free}[%][2]	24.8
Total number of atoms:	
Protein	2021
Water	33
Ligand	29
Sulphate	10
Deviation from ideal geometry: [3]	
Bond lengths [Å]	0.008
Bond angles [°]	1.07
Bonded B's [Å^2] [4]	2.8
Ramachandran plot: [5]	
Most favoured regions [%]	94.0
Additional allowed regions [%]	6.0
Generously allowed regions [%]	0.0
Disallowed regions [%]	0.0

[1] Values as defined in REFMAC5, without sigma cut-off

[2] Test-set contains 2.4 % of measured reflections

[3] Root mean square deviations from geometric target values

[4] Calculated with MOLEMAN

[5] Calculated with PROCHECK