Supplemental Information for

CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma

Heather A. Carlson*[1], Richard D. Smith[1], Kelly L. Damm-Ganamet[1], Jeanne A. Stuckey[2], Aqeel Ahmed[1], Maire A. Convery[3], Donald O. Somers[3], Michael Kranz[3,4], Patricia A. Elkins[5], Guanglei Cui[5], Catherine E. Peishoff[5], Millard H. Lambert[5], James B. Dunbar, Jr.[1]

[1] Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church St., Ann Arbor, Michigan 48109-1065, United States

[2] Center for Structural Biology, University of Michigan, 3358E Life Sciences Institute, 210 Washtenaw Ave., Ann Arbor, Michigan 48109-2216, United States

[3] Computational and Structural Sciences, GlaxoSmithKline Research & Development, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, United Kingdom

[4]Deceased, April 22, 2014

[5] Computational and Structural Sciences, GlaxoSmithKline Research & Development, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426, United States

* Corresponding author email: carlsonh@umich.edu

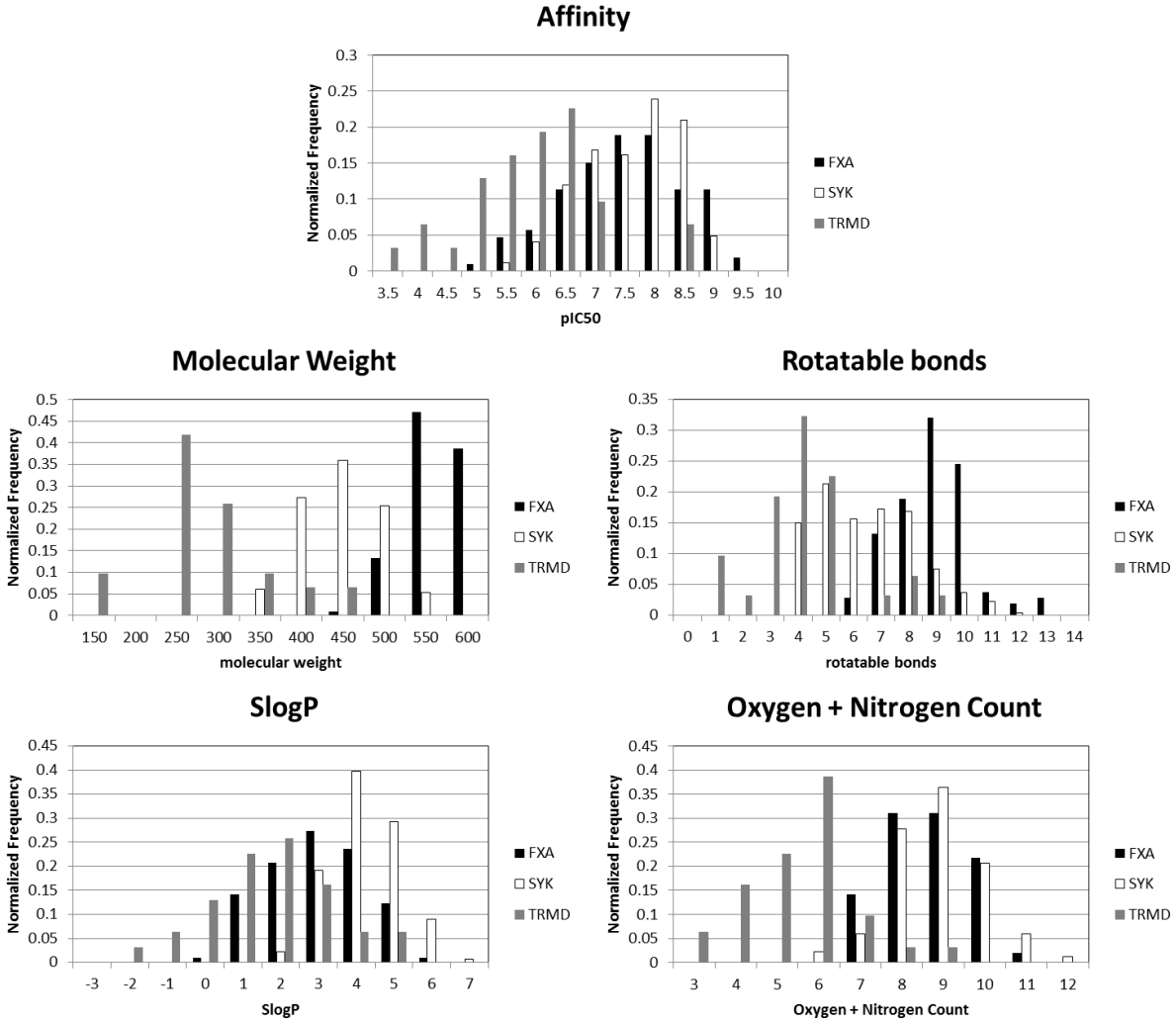

Figure S1. The distributions of affinity and key physical properties for the unique ligands used in Phase 2 of the 2014 CSAR exercise. Data for TrmD, SYK, and FXa are given separately.

Evaluations of scoring docked poses – Phase 1

The initial data given to participants in Phase 1 is in the supplemental file Phase1-start.zip. Answers are given in Table S1.

Table S1. Each protein-ligand complex from Phase 1 had 200 ligand poses to evaluate. Only one near-native pose was included in each set, and their ID numbers in the lists of poses are given below.

TrmD Complexes	ID# of near-native pose in Phase 1
gtc000445	126
gtc000446	76
gtc000447	18
gtc000448	3
gtc000451	5
gtc000452	79
gtc000453	32
gtc000456	185
gtc000457	88
gtc000458	116
gtc000459	195
gtc000460	27
gtc000464	118
gtc000465	154

SYK Complexes	ID# of near-native pose in Phase 1
gtc000224	116
gtc000225	51
gtc000233	71
gtc000249	48
gtc000250	141

FXa Complexes	ID# of near-native pose
gtc000101	146
gtc000398	190
gtc000401	37

Evaluations of submitted scores/ranks – Phase 2

The initial data given to participants in Phase 2 is in the file Phase2-start.zip and the "answers" are in the experimental data below. Participants submitted predicted poses and scores/ranks for each ligand. Each method is evaluated separately. Docked poses were evaluated using symmetry-corrected RMSD to the coordinates in the crystal structure. The data is provided on the pages 3, 5, and 7 of the supplemental file Phase2-Eval-Metrics.xlsx (data on protein-ligand contacts is also provided). Several metrics were calculated to evaluate the ranking of the inhibitors in Phase 2: R^2, Pearson R, Spearman ρ, and Kendall τ (each calculated using JMP[1]). Standard deviations (σ) and 95%-confidence intervals (95% CI) are also given for all metrics. Detailed analysis is provided in pages 4, 6, and 8-11 of the supplemental file Phase2-Eval-Metrics.xlsx. The results for docking and scoring together are summarized in the first two pages of Phase2-Eval-Metrics.xlsx. Histograms in Figure S2 supplement Figure 4 in the paper.

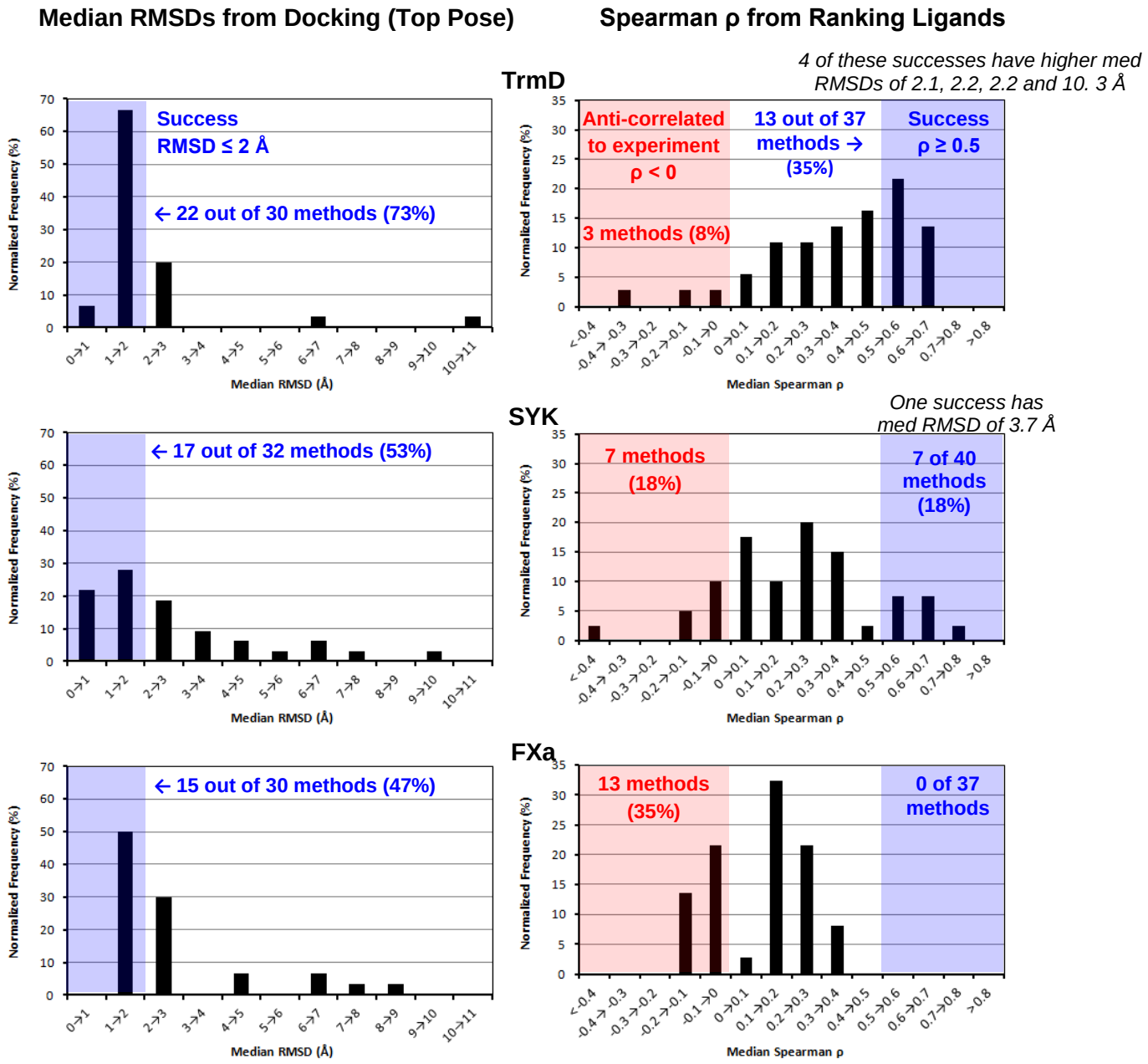

Figure S2. Docking was evaluated by the RMSD between the submitted docked poses and the crystal structures from GSK that were saved as new data for Phase 2. The scores/ranks were evaluated using ρ calculated by comparing scores for each of the ligands to the affinity data measured at GSK. Our criteria for "good performance" for this study are median RMSD ≤ 2 Å on the left and ≥ 0.5 for ρ on the right, highlighted in blue above. The red regions in the ρ distributions show where there were negative correlations to experiment.

Experimental Values

The experimental affinity data from GSK are given in the supplemental file Phase2-Expt-Affinities.xlsx. Table S2 below gives the PDB codes for each crystal structure.

Table S2. PDB ID codes for the complexes from GSK that were used in Phase 1 (HiQ structures) and Phase 2.

TrmD crystal structures

Ligand	Phase 1	Phase 2
gtc000444		4YPW
gtc000445	4YPX	
gtc000446	4YPY	
gtc000447	4YPZ	
gtc000448	4YQ0	
gtc000449		4YQ1
gtc000450		4YQ2
gtc000451	4YQ3	
gtc000452	4YQ4	
gtc000453	4YQ5	
gtc000454		4YQ6
gtc000455		4YQ7
gtc000456	4YQ8	
gtc000457	4YQ9	
gtc000458	4YQA	
gtc000459	4YQB	
gtc000460	4YQC	
gtc000461		4YQD
gtc000462		4YQG
gtc000463		4YQI
gtc000464	4YQJ	
gtc000465	4YQK	
gtc000466		4YQL
gtc000467		5D9F
gtc000468		4YQN
gtc000469		4YQO
gtc000470		4YQP
gtc000471		4YQQ
gtc000472		4YQR
gtc000473		4YQS
gtc000474		4YQT

SYK crystal structures

Ligand	Phase 1	Phase 2
gtc000222		4YJO
gtc000223		4YJP
gtc000224	4YJQ	
gtc000225	4YJR	
gtc000226		4YJS
gtc000233	4YJT	
gtc000249	4YJU	
gtc000250	4YJV	

FXa crystal structures

Ligand	Phase 1	Phase 2
gtc000006		4ZH8
gtc000101	4Y6D	
gtc000102		4ZHA
gtc000398	4Y71	
gtc000401	4Y76	
gtc000406		4Y79
gtc000422		4Y7A
gtc000441		4Y7B

Experimental Methods

Factor-Xa

Crystallography details

Purified des-Gla factor-Xaβ was purchased from Enzyme Research Laboratories (South Bend, IN). Note that FXa is converted autocatalytically to FXaβ by the loss of a 17-residue glycopeptide from the C-terminus of the heavy chain. Purification and crystallization is as described previously[2]. Refinement was done with refmac[3] in all cases. The starting model was 1ezq (with ligand removed) in all cases.

Assay Details

In vitro assays for inhibition of Factor Xa (1)

The chromogenic assay used for set 1 was described in patent WO2003043981 as follows. Compounds were tested for their Factor Xa inhibitory activity as determined *in vitro* by their ability to inhibit human Factor Xa in a chromogenic assay, using N-α-benzyloxycarbonyl-D-Arg-Gly-Arg-p-nitroanilide as the chromogenic substrate. Compounds were diluted from a 10 mM stock solution in dimethylsulfoxide at appropriate concentrations. The assay was performed at room temperature using buffer consisting of: 50 mM Tris-HCl, 150 mM NaCl, 5 mM $CaCl_2$, pH 7.4 containing human Factor Xa (final concentration. of 0.0015 $U.mL^{-1}$). Compound and enzyme were preincubated for 15min prior to addition of the substrate (final conc. of 200 μM). The reaction was stopped after 30 min with the addition of soybean trypsin inhibitor or H-D-PHE-PRO-ARG-Chloromethylketone. BioTek EL340 or Tecan SpectraFluor Plus plate readers were used to monitor the absorbance at 405nm. To obtain IC_{50} values, the data were analyzed using ActivityBase® and XLfit®.

Compounds used in set2 and set3 were tested for their Factor Xa inhibitory activity as determined *in vitro* by their ability to inhibit human Factor Xa in a fluorogenic assay, using Rhodamine 110, bis-CBZ-glycylglycyl-L-arginine amide as the fluorogenic substrate. Compounds were diluted from a 10 mM stock solution in dimethylsulfoxide at appropriate concentrations. The assay was performed at room temperature using buffer consisting of: 50 mM Tris-HCl, 150 mM NaCl, 5 mM $CaCl_2$, pH 7.4 containing human Factor Xa (final concentration

of 0.0003U.mL-1). Compound and enzyme were preincubated for 15 min prior to addition of the substrate (final concentration of 10 μM). The reaction was stopped after 3 hrs with the addition of H-D-Phe-Pro-Arg-Chloromethylketone. An LJL-Analyst fluorimeter was used to monitor fluorescence with 485 nm excitation/535 nm emission. To obtain IC_{50} values, the data were analyzed using ActivityBase® and XLfit®. The compound affinities for set2 and set3 were both obtained with this assay but were processed somewhat differently.

The assay results reported with this data submission are raw, unfiltered data downloaded directly from the corporate assay results database, with no further editing or exclusion of replicates. This data should have an accuracy comparable to that typically obtained in the pharmaceutical industry at the time of the experiments, 2000-2006. Users should recognize that data of this sort is subject to many possible errors, ranging from small pipette errors, interference with fluorescence, or partial insolubility or degradation of compound, up to rare cases where an entirely different compound might have been dispensed.

The current, common practice at GSK is to run several "standard" compounds through the assay each week to monitor the consistency of the assay. As a result, the standards get re-tested many times. For FXa, the data from the chromogenic assay is very consistent for all standard compounds. In the flint1 data, the standard deviations for compounds 1 and 6 are inflated from a few "dead runs", which can happen for a variety of reasons. The flint2 assay is less accurate than the flint1 or the chromogenic assays, but the exact accuracy cannot be well established as no suitable standards were tested more than 15 times (the other assays had standards tested over 100 times). Flint2 was handled differently because it was never the primary FXa assay during the project. It was an alternative complement to flint1 that was used less frequently.

FXa Standards	Chromogenic Assay (set 1)			Flint 1 Assay (set 2)			Flint 2 Assay (set 3)		
	Ave pIC_{50}	Std Dev	Replicates	Ave pIC_{50}	Std Dev	Replicates	Ave pIC_{50}	Std Dev	Replicates
1	7.39	0.16	71	8.01	0.76	126	7.24	1.50	15
2	6.79	0.19	109	7.74	0.12	3			
3	7.43	0.09	21	8.18	0.17	26			
4	7.65	0.10	17	8.30	0.27	19	8.12	0.17	4
5				6.57	0.20	7	6.56	0.20	10
6				7.26	1.09	7	7.02	1.38	10
Ave Std Dev		0.1			0.4			0.8	

Syk

Crystallography details

SYK protein was purified, crystallized and data collected and processed as described by Liddle *et al.*[4] The gtc000224, gtc000225, gtc000233,gtc000249 and gtc000250 complexed structures were further fit and refined with Coot[5] and Buster[6], respectively. The coordinates and restraints for the ligands were produced from smile strings in Grade[6] with the mogul+qm option. Validation was done with MolProbity[7].

Assay details

Homogeneous Time Resolved Fluorescence (HTRF) was used to monitor catalytic activity of Syk Kinase. Syk kinase phosphorylates the biotinylated-peptide. The phosphorylated peptide binds a specific antiphosphotyrosine antibody labelled with europium (Eu-Ab). Streptavidin Allophycocyanin(Stre-APC) binds the biotin in the peptide resulting in FRET between the two fluorophores when in close proximity upon excitation at 337nm. Emission fluorescence by APC at 665nm correlates with the amount of phosphorylated product formed. For SYK's standards, the compounds have been tested over 1000 times.

SYK Standards	Ave pIC_{50}	Std Dev	Replicates
1	5.66	0.36	1738
2	6.18	0.34	1693
3	6.28	0.24	1710
4	7.24	0.32	1304
5	6.29	0.27	1259
Ave Std Dev		0.3	

Potential for False Positives: Compound fluorescence can add to the signal in either channel or both Potential for False Negatives: Compound absorbance can compete with Eu-Ab molecule for excitation energy, reducing fluorescence in both channels A and B. Absorption of emitted photons by unknowns can reduce detected fluorescence in either channel or both.

Target: Homo sapiens; Accession No: NM_003177.3; 6His-tagged; DMSO (Maximum 3% final concentration in assay).

Starting concentration of test compound: Variable, typically 2.5E-5M; a 4-fold dilution factor was used.

Points per curve: 11;

Data analysis: Process through the XC50v2 curve fit algorithm using the 4 parameter logistic equation in Activity Base; plate format: 384.

- assay buffer: 40 mM HEPES, pH 7.4, 0.01% BSA;
- substrate reagent: 1 μM biotinylated src peptide [0.5 μM IN ASSAY], 20 mM $MgCl_2$ [10 mM IN ASSAY], 60 μM ATP [30 μM in assay], in Assay Buffer
- enzyme reagent: Syk in Assay Buffer. At concentration to ensure assay displays linearity proportional to time and enzyme concentration. Typically 4 nM in reagent [typically 2 nm in assay].
- Stop/read buffer: 40 mM HEPES, 150 mM NaCl, 0.03% w/v BSA, 60 mM EDTA
- Read reagent: Streptavidin-APC (based on Streptavidin concentration) typically 150 nM [final 50 nM], Eu-Ab typically 1.5 nM [final 0.5 nM], in Read Buffer

The following steps were performed

Blanks

1 Add 3 μL Assay Buffer (minus enzyme reagent) to 'blank' wells

Pre-Activation of Syk enzyme

1 Incubate for 30 minutes at 20 °C in the presence of 16.6 mM $MgCl_2$ and 8.3 mM ATP

Addition of Assay Components

1 Add 3 μL Substrate Reagent to test and control wells

2 Add 3 μL Enzyme Reagent to test and high control wells

3 Incubate at room temp for 30 minutes.

Add Stop / Read reagent

1 Add 3 μL of Stop / Read reagent to all wells. Leave for at least 45 min at room temp and read on suitable HTRF reader

TrmD

Crystallography details

H. infl. TrmD was expressed and purified from *E. coli*[8]. Protein at 12 mg/mL in 100 mM HEPES pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, and 2 mM DTT was crystallized at room temperature via vapor diffusion in sitting drops. 4 µL of 100 mM S-adenosyl methionine (SAM) in H_2O was added to a 100 µL protein drop and allowed to incubate on ice for 1 hour. The solution was then plated at a 1:1 ratio with a well solution consisting of 20% PEG 3350 and 0.2 M potassium citrate tribasic monohydrate. Rounds of seeding were necessary to obtain larger single crystals. Compound soaks were done in the crystallization plate which originally grew crystals. Compound stocks in DMSO were added up to a final drop concentration of 4.8% DMSO and lasted for 4-6 hours. Duration of soaks was dependent upon crystal appearance, and crystals were harvested if they appeared to start decaying. The crystals were then briefly (<20 seconds) transferred to a cryoprotective solution consisting of 20% glycerol/80% well solution before flash freezing in liquid nitrogen.

The data were collected at Beamline 21-ID-G, 21-ID-D, and 21-ID-F at LS-CAT, APS, Argonne National Labs and at Beamline 8.3.1 at the ALS, Berkeley Labs. The data were processed with HKL2000[9] or Mosflm. The structures were determined by molecular replacement using Phaser[10] (CCP4)[11] using an in-house structure as a model and refined with Phenix[12]. Coot[5] was used for model building and validation was done with MolProbity[7].

Assay Details

Haemophilus influenzae TrmD inhibition was evaluated in assays containing 4-10 nM H. influenza TrmD, 0.11 µM [3H]SAM, 1 µM tRNALeu 50 mM HEPES pH 7.5, 7 $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 1 mM CHAPS and test compound. Assays were incubated at ambient temperature for 60 minutes, after which PEI-coated PS SPA beads (polyethyleneimine polystyrene scintillation proximity assay beads, eg, GE Healthcare Bio-Sciences RPNQ0098) were added at 5 mg/ml final concentration to quench the reaction. The quenched assay plates were then imaged on a Perkin-Elmer Viewlux to quantify turnover. Dose response values were performed using an 11-point serial dilution of test compound, the data fit in ActivityBase using a

four parameter logistic fit, and potency values expressed as pXC50. Essentially, TrmD incorporates tritiated methyl groups into substrate tRNA, which is subsequently captured by the PEI-coated SPA beads, allowing for quantitation of the incorporated tritium. TrmD inhibitors will decrease the intensity of the observed signal. Several virtually identical assays were available, differing only in plate type (384 vs. 1536), enzyme concentration (4 nM or 10 nM TrmD), or assay volume (4 μL, 5 μL, or 10 μL). These changes do not affect the pharmacology.

For TrmD, the standard compound run in the GSK affinity assays has an average pIC_{50} of 6.8 with a standard deviation of 0.2, based on 124 replicates.

References

(1) *JMP Pro 10*; SAS Institute, Inc.: Cary, N.C., accessed: February 23, 2016.

(2) Chan, C.; Borthwick, A. D.; Brown, D.; Burns-Kurtis, C. L.; Campbell, M.; Chaudry, L.; Chung, C.; Convery, M. A.; Hamblin, J. N.; Johnstone, L.; Kelly, H. A.; Kleanthous, S.; Patikis, A.; Patel, C.; Pateman, A. J.; Senger, S.; Shah, G. P.; Toomey, J. R.; Watson, N. S.; Weston, H. E.; Whitworth, C.; Young, R. J.; Zhou, P. Factor Xa Inhibitors: S1 Binding Interactions of a Series of N-{(3S)-1-[(1S)-1-Methyl-2-Morpholin-4-Yl-2-Oxoethyl]-2-Oxopyrrolidin-3-Yl}sulfonamides. *J. Med. Chem.* **2007**, *50*, 1546–1557. DOI: 10.1021/jm060870c

(3) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367. DOI: 10.1107/S0907444911001314

(4) Liddle, J.; Atkinson, F. L.; Barker, M. D.; Carter, P. S.; Curtis, N. R.; Davis, R. P.; Douault, C.; Dickson, M. C.; Elwes, D.; Garton, N. S.; Gray, M.; Hayhow, T. G.; Hobbs, C. I.; Jones, E.; Leach, S.; Leavens, K.; Lewis, H. D.; McCleary, S.; Neu, M.; Patel, V. K.; Preston, A. G. S.; Ramirez-Molina, C.; Shipley, T. J.; Skone, P. A.; Smithers, N.; Somers, D. O.; Walker, A. L.; Watson, R. J.; Weingarten, G. G. Discovery of GSK143, a Highly Potent, Selective and Orally Efficacious Spleen Tyrosine Kinase Inhibitor. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 6188–6194. DOI: 10.1016/j.bmcl.2011.07.082

(5) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501. DOI: 10.1107/S0907444910007493

(6) Roversi, P.; Sharff, A.; Smart, O. S.; Vonrhein, C.; Womak, T. O. *BUSTER v. 2.11.1*; Cambridge, United Kingdom: Global Phasing Ltd., 2011, accessed: February 23, 2016.

(7) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *MolProbity* : All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, *66*, 12–21. DOI: 10.1107/S0907444909042073

(8) Kim, H.-W.; Ahn, H. J.; Yoon, H.-J.; Kim, H.-W.; Baek, S.-H.; Suh, S. W. Crystallization and Preliminary X-Ray Crystallographic Analysis of tRNA(m1G37)methyltransferase from Haemophilus Influenzae. *Acta Crystallogr. Sect. D* **2003**, *59*, 183–184. DOI: 10.1107/S0907444902019716
(9) Otwinowski, Z.; Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology*; Charles W. Carter, J., Ed.; Macromolecular Crystallography Part A; Academic Press, 1997; Vol. 276, pp 307–326.
(10) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674. DOI: 10.1107/S0021889807021206
(11) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the *CCP* 4 Suite and Current Developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242. DOI: 10.1107/S0907444910045749
(12) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *PHENIX* : A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221. DOI: 10.1107/S0907444909052925