Supporting Information

Fragment Screening at AstraZeneca: Developing the Next Generation Biophysics Fragment Set

Simon C. C. Lucas,*[a] Ulf Börjesson,[c] Mark J. Bostock,[d] John Cuff,[f] Fredrik Edfeldt,[e] Kevin J. Embrey,[d] Per-Olof Eriksson,[e] Andrea Gohlke,[d] Anders Gunnarson,[e] Michael Lainchbury,[a] Alexander G. Milbradt,[d] Rachel Moore,[b] Philip B. Rawlins,[d] Ian Sinclair,[f] Christopher Stubbs,[d] and R. Ian Storer [a]

[a] *Hit Discovery, Discovery Sciences, R&D, AstraZeneca, Cambridge, United Kingdom.*
[b] *Hit Discovery, Discovery Sciences, R&D, AstraZeneca, Alderley Park, United Kingdom.*
[c] *Hit Discovery, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden.*
[d] *Mechanistic and Structural Biology, Discovery Sciences, R&D, AstraZeneca, Cambridge, United Kingdom.*
[e] *Mechanistic and Structural Biology, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden.*
[f] *Compound Synthesis and Management, Discovery Sciences, R&D, AstraZeneca, Alderley Park, United Kingdom.*

Table of Contents

Computational Chemistry

Bemis-Murcko scaffolds[1] were determined using PipelinePilot.[2] Alpha atoms were not included in the generation of assemblies and the stereochemistry of fragments was fixed. cLogP[3] and other molecular descriptors and models used were calculated using in-house software.

FastROCs: The corporate collection of 10^6 available small molecules was processed with Flipper[4] to enumerate undefined stereocenters and double bonds. For molecules containing more than two undefined stereocenters and double bonds, two are selected at random and fully enumerated, while the rest are fixed to a single random stereoisomer. The prepared database of SMILES strings is processed with OMEGA[4, 5] to generate 20 conformers per molecule. The 3D expanded database was searched with the FastROCS server[6] on a local GPU node, sorting hits by the best shape overlay, as quantified by the Tanimoto combo score.

Purity and Solubility Assessment

a) HPLC

All HPLC separations were performed on an Agilent 1200 system utilizing two binary G1312B high-pressure gradient pumps connected, as has been previously described in the literature[1], with the solvent flow from pump 1 delivered through the auto-sampler and that from pump 2 delivered immediately after the separation column. The typical solvent system was a combination of (A) HPLC grade water (Sigma Aldrich, UK) containing 0.1% formic acid (Sigma Aldrich, UK) and (B) HPLC grade acetonitrile (Sigma Aldrich, UK) containing 0.05% formic acid (Sigma Aldrich, UK).

Solvent was delivered from pump 1 at a flow rate of 700 µL per min with a composition at time zero of A=95% and B=5%, developing over a linear gradient such that, after 2 min, B=100%, which was maintained for 0.5 min before being returned to starting conditions.

The solvent composition for pump 2 was delivered in exact opposition to that for pump 1 at all times, starting from A=5% and B=95% at time zero and traversing a mirror-image gradient toward A=100% after 2 min, held for 0.5 min before being returned to starting conditions. The two streams were combined using a T-piece after the UV detector and, by this method, produced a constant 50:50 mix of A and B going forward, and thus minimized any solvent–related drift in signal response.

The separation column used was a 50x2 mm Kinetex C18 3uM column (Phenomenex, Torrence, CA).

b) Peak detection and MS

The HPLC eluent was delivered through an Agilent G1315B diode array detector, scanning from 220 – 300 nm at a rate of 5 Hz, and then split between a single quadrupole MS and a CAD.

The configuration of MS system was a Waters Quattro Micro quadrupole mass detector with integrated ADC, a CTC PAL autosampler (CTC Analytics, Switzerland) and an ESA Corona CAD detector, all operating under MassLynx 4.0.

The corona discharge detector relied on droplet production and transformation into a particle beam, and was particularly sensitive to changes in solvent affecting the size of the droplets, hence the careful control of post UV detector solvent composition. The particle beam was passed through a stream of charged nitrogen molecules where charge was transferred to the analyte molecules contained within the droplets. The charge on these droplets was then detected via an electrometer, with a proportional output acquired at a rate of 5 Hz through the ADC. In this way a chromatogram was produced by plotting electrometer current against time.

The detector had an equivalent response for equal mass-per-volume concentrations of non-volatile analytes.

The CAD detector was complementary to the DAD in that, in the main, it provided confirmation of the sample purity but could also provide information not available from the DAD. This was particularly true when the sample structure did not absorb in the UV detection range, or where the sample was not retained on the chromatographic column but was eluted in the solvent front along with its original solvent (normally dimethyl sulfoxide (DMSO)). In both of these cases, sample purity was not measurable by the DAD and was assessed from the CAD alone.

The two chromatograms produced were integrated within Openlynx (Masslynx) and mass spectra, in both positive and negative ion modes, were produced for each detected peak. The processed data was passed into our in house database where it could be reviewed and annotated as described in the literature.[2]

SPR Clean Screen

A Biacore 4000 (Cytiva) was used for the 'Clean Screen' of the fragment set to remove compounds that bind non-specifically to the biosensor chip surface. Prior to running the fragment screen, a desorb method was run. A Series S NTA biosensor chip (Cat. # BR100034; Cytiva) was docked and running buffer containing 25 mM HEPES, 150 mM NaCl, 1 mM TCEP, 1% DMSO and 0.05% Tween-20 was primed three times. All experiments were conducted at 25 °C using 1 Hz data collection.

NTA capture-coupling[7] was used across all four spots of all flow cells. 6His-streptavidin was immobilised on spot 1 (streptavidin surface). A well-behaved protein was immobilised on spot 5. Bare Ni-NTA surfaces were left on spots 2 and 4. Briefly, a 1 min injection of 500 mM EDTA was followed by a 1 min injection of 500 μM Nickel(II) chloride hexahydrate (Merck). The surface was activated with a 5 min injection of 200 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC; ThermoFisher Scientific/50 mM N-hydroxysuccinimide (NHS; Merck), 25 μg/mL his-streptavidin, 20 μg/mL protein or buffer was injected for 5 min and followed by inactivation with a 5 min injection of 1 M ethanolamine-HCl, pH 8.5 (Cytiva). Unbound streptavidin sites on spot 1 were blocked with a 1 min injection of 50 μM EZ-Link™ Amine-PEG2-Biotin (ThermoFisher Scientific).

4060 fragments were screened at 1 mM by diluting 1 μL of 100 mM compound stock (in 100% DMSO (*v/v*)) in 100 μL of buffer containing 25 mM HEPES, 150 mM NaCl, 1 mM TCEP and 0.05% Tween-20. Compounds that visibly precipitated were manually removed from the microplates prior to testing and flagged for removal from the set. Each compound was injected for 30 seconds with a 12 second dissociation. The 'early stability' report point was exported for analysis from spots 1 (streptavidin), 2 (Ni-NTA) and 3 (untreated NTA surface). Fragments were flagged for removal if they were above 3x the robust standard deviation of the median signal obtained from the early stability report point (baseline subtracted but not reference subtracted) in at least one out of the three spots. Sensorgrams were checked for false removal due to drift caused by a previous compound. In addition, sensorgrams on spot 5 (protein) were checked for any bad behaviour not identified from the other spots. Data was analysed in the Biacore 4000 Evaluation software (Cytiva) and exported to excel for statistical analysis.

NMR QC

Compounds were obtained as 2 μl aliquots of 100 mM DMSO solutions. 2 μl compound was dissolved in 200 μl sample buffer (50 mM Tris-d_{11} pH 7.4, 150 mM NaCl, 1 mM TCEP, 10 μM DSS, 20% D_2O) using a TECAN Evo 100 liquid handling robot and transferred to 3 mm NMR tubes in Bruker SampleJet racks. All NMR spectra were acquired at 298 K on a Bruker Avance Neo 600 MHz spectrometer equipped

with a 5 mm QCI-F probe. 1D double spin echo spectra were acquired with excitation sculpting[8, 9] (Bruker zgesgppr with extended phase cycling) and off-resonance presaturation of the DMSO signal. WaterLOGSY experiments used a modified version of the ephogsy sequence[10-12] with excitation sculpting and mixing time set to 1.8 s. 1D ^{19}F spectra were collected with inverse gated ^{1}H decoupling during acquisition (Bruker zgig30 sequence), and a 1D spin echo sequence with adiabatic refocusing pulses (Bruker zgseiggpad sequence). Data was acquired and processed using Bruker Topspin 4.0.4 with automation handled using the SampleJet robotics system and IconNMR software (Bruker). A subset of compounds was put through a second round of NMR QC. For this, 2 µL 100 mM solution in DMSO was dried by evaporation under vacuum, redissolved in d_6-DMSO and then prepared as a 1 mM solution by addition of deuterated phosphate buffer (50 mM phosphate pH 7.4, 150 mM NaCl, 10 µM DSS). 1D NOESY spectra with presaturation (Bruker noesygppr1d.2) were acquired. This limited the suppression of compound signals in the water region.

TCEP AMI-MS Redox Assay

50 nL of 10 mM compound was acoustically dispensed using Labcyte Echo 555 liquid handling units into each well of a 384-well Labcyte P-05525 PP clear plate for a final assay concentration of 10 µM. Positive and neutral controls (50 nL of either 10 mM NCS-663284 and 50 nL of 100 % DMSO, respectively) were present on each plate. 50 µL of 0.3 mM Tris(2-carboxyethyl)phosphine hydrochloride solution (TCEP) diluted in 20 mM Tris hydrochloride buffer solution pH 7.5 was dispensed into plates using a Thermo Multidrop Combi with standard cassette. The reaction was then incubated for 24 hrs prior to data acquisition on using Acoutstic Mist Ionisation-MS (AMI-MS). The AMI-MS platform combines a modified Labcyte Echo 555 liquid handling unit with a Waters Xevo G2-XS Q Tof mass spectrometer. TCEP and TCEPO ions were measured using a desolvation temperature of 300 °C, cone gas flow between 30-50 L/h and charging voltage of -2 kV. The resulting AMI-MS raw data was processed through MSExport in Masslynx (Waters, Wilsmslow, UK) generating a single XY text file of the spectra for each well. These text files were queried in MS Parser (an in house Java based program) to produce an ion area for each of the target ions. The ion area of *m/z* peaks representing TCEP and TCEPO (249 *m/z* and 265 *m/z* respectively) were imported into Genedata Screener (Genedata, Basel, CH) where TCEP oxidation was calculated using the equation (TCEPO / (TCEP + TCEPO). Data was then normalised to positive controls (10 µM NSC-663284) minus neutral (DMSO) controls and presented as % TCEP oxidation.

Case Studies

Four case studies, three screening the full biophysics set in SPR and one screening layer 1 *via* NMR are included to exemplify the use of the revamped AstraZeneca biophysics fragment set.

Target 1 Experimental

SPR experiments were performed on a Biacore 8K+ instrument (Cytiva) at 20 °C using 1 Hz data collection. A Series S streptavidin (SA) sensor chip (Cytiva) was docked into the system in buffer consisting of 10 mM HEPES (pH 7.4), 150 mM NaCl, 1 mM TCEP (tris(2-carboxyethyl)phosphine), 0.05% (*v/v*) Tween-20 (polyoxyethylene(20)sorbitan monolaurate) and 1% (*v/v*) DMSO. The Sensor Chip surface was conditioned with 3 x 60 s injections of 50 mM NaOH/1 M NaCl. Biotinylated Peptide was immobilized at ~1000 RU. Reference surfaces were prepared with the same method, in the absence of peptide. Compounds were prepared in assay buffer in a 384-well polypropylene microplate (Greiner) by diluting 1 µL of compound at 100 mM in 100% DMSO (*v/v*) in 100 µL of running buffer containing 50 nM of Target 1. Samples were injected for 60 s followed by 30 s dissociation at 30 µL/min

before 30 s surface regeneration with 0.5% SDS, 20 mM TCEP. Analysis was made in Insight software (Cytiva) using a hit cut-off set to >10% reduction in protein binding versus blank. For determination of fragment affinities of hits, a 7-point concentration response was used from 1 mM top concentration with 2-fold dilutions. Affinity parameters were determined by global fitting to a 1:1 binding model using Matlab (Mathworks).

Target 2 Experimental

SPR experiments were performed on a Biacore 8K+ instrument (Cytiva) at 25 °C using 1 Hz data collection. Prior to immobilising protein, a desorb method was run. A Series S streptavidin (SA) sensor chip (Cytiva) was docked into the system in buffer consisting of 20 mM Tris (pH 7.5), 150 mM NaCl, 50 μM EDTA, 1 mM TCEP (tris(2-carboxyethyl)phosphine), 0.05% (*v/v*) Tween-20 (polyoxyethylene(20)sorbitan monolaurate) and 1% (*v/v*) DMSO, and primed three times. The Sensor Chip surface was conditioned with 3 x 60 s injections of 50 mM NaOH/1 M NaCl at 30 μL/min. Biotinylated-Target 2; 25 μg/mL was immobilized for 10 min at a flow rate of 5 μL/min. Unbound streptavidin sites were blocked with 50 μM EZ-Link™ Amine-PEG2-Biotin (ThermoFisher). Reference surfaces were prepared with the same method, in the absence of protein. Compounds were prepared in assay buffer in a 384-well polypropylene microplate (Greiner) by diluting 0.3 μL of compound in 100% DMSO (*v/v*) in 100 μL of running buffer containing 0.7% DMSO to make 1% (*v/v*) DMSO final concentration. Samples (analyte) was injected for 30 s followed by 12 s dissociation at 30 μL/min. Prior to analysis, solvent calibration and reference subtraction were made to eliminate bulk refractive index changes, injection noise, and data drift. The Insight software (Cytiva) was used for analysis of the fragment screen and report points were exported to excel for further analysis. For determination of fragment affinities, a seven point concentration response was tested from 1 mM top concentration with 2-fold dilutions. Affinity parameters were determined by global fitting to a 1:1 binding model within the Insight Software (Cytiva).

Target 3 Experimental

SPR experiments were performed on a Biacore 8K+ instrument (Cytiva) at 25 °C using 1 Hz data collection. Prior to immobilising protein, a desorb method was run. A Series S Streptavidin (SA) sensor chip (Cytiva) was docked into the system in buffer consisting of 10 mM Tris (pH 7.5), 100 mM NaCl, 50 μM EDTA, 0.1 mM TCEP (tris(2-carboxyethyl)phosphine), 0.05% (*v/v*) Tween-20 (polyoxyethylene(20)sorbitan monolaurate) and 1% (*v/v*) DMSO, and primed three times. The Sensor Chip surface was conditioned with 3 x 60 s injections of 50 mM NaOH/1 M NaCl at 30 μL/min. Biotinylated protein Target 3 was then injected for 5 min at a flow rate of 10 μL/min. Unbound biotin binding sites were blocked by injecting 100 μM EZ-Link™ Amine-PEG2-Biotin (ThermoFisher) for 1 min at 30 μL/min. Reference surfaces were prepared with the same method, omitting the protein injection. Compounds were prepared in assay buffer in a 384-well polypropylene microplate (Greiner) by diluting 1 μL of 30 mM compound in 100% DMSO (*v/v*) with 100 μL of running buffer containing 0% (*v/v*) DMSO to give 1% (*v/v*) DMSO final concentration. Samples were injected for 30 s followed by 15 s dissociation at 30 μL/min. 10-point 2-fold serial dilution titrations of a control compound (K_d = 400 nM, 1 μM top concentration) were performed at the start and end of the experiment, with additional single point control injections (1 μM) every 16 sample cycles to monitor surface activity. Prior to analysis, solvent calibration and reference subtraction were made to eliminate bulk refractive index changes, injection noise, and data drift. The Insight software (Cytiva) was used for analysis of the fragment screen and report points were exported to Excel for further analysis. For determination of

fragment affinities, a seven point concentration response was tested from 1 mM top concentration with 2-fold dilutions. Affinity parameters were determined by global fitting to a 1:1 binding model within the Insight Software (Cytiva).

The primary screen was analysed by considering the blank subtracted binding early (median response 6 s after injection start, 5 s window; to identify significant binding responses) and stability early (median response 5 s after injection end, 5 s window; to identify residual binding) report points. For each report point, the robust standard deviation (RSD) and corresponding robust Z-scores (RZ) were calculated, as below.

$$\mathrm{RSD(ReportPoint)} = 1.4826 \times \mathrm{median}(|\mathrm{ReportPoint}_i - \mathrm{median(ReportPoint)}|)$$

$$\mathrm{RZ}_i = \frac{\mathrm{ReportPoint}_i - \mathrm{median(ReportPoint)}}{\mathrm{RSD(ReportPoint)}}$$

Fragments were then scored according to the scheme below, and hits defined as compounds with a score = 2. The sensorgrams for all hits were manually checked for bad behaviours.

$$\mathrm{Combined\ Score} = \mathrm{Score(Binding\ Early)} \times (2 + \mathrm{Score(Stability\ Late)})$$

where

If RZ(Binding Early)	> 3	Score(Binding Early)	= 1
	≤ 3		= 0
If Absolute Value (RZ(Stability Late))	> 5	Score(Stability Late)	= -1
	≤ 5		= 0

Target 4 Experimental

Target 4, a protein-protein complex, was screened using NMR spectroscopy against the 1,152 compounds in layer 1. The screen was prepared in pools of four compounds, resulting in 288 pools. 1 µl of each compound at 100 mM in DMSO was mixed with sample buffer (50 mM d-TRIS pH 7.4, 100 mM NaCl, 5 mM $MgCl_2$, 50 µM GMPPNP, 1 mM d-TCEP, 10 µM DSS, 10% D_2O) giving a final compound concentration of 500 µM. For each pool two samples were prepared, with and without 10 µM protein. Samples were transferred using a TECAN Genesis RSP 100 liquid handling robot to 3 mm NMR tubes in Bruker SampleJet racks. NMR spectra were acquired at 298 K on a Bruker Avance Neo 600 MHz spectrometer (^{1}H frequency, 600 MHz) equipped with a 5 mm TXI cryoprobe. Data was acquired and processed using Bruker Topspin with automation handled using the SampleJet robotics system and IconNMR software (Bruker).

1D spin echo spectra were acquired with excitation sculpting[8, 9] (Bruker zgesgppr) and off-resonance presaturation of the DMSO signal. Compound-protein interactions were assessed using CPMG and WaterLOGSY experiments[8, 11, 12] using Bruker cpmgesgppr and ephogsygpno sequences. The CPMG sequence used a 400 ms delay with water suppression using excitation sculpting and off-resonance presaturation at the DMSO position. The WaterLOGSY used a 2 s mixing period and excitation sculpting for water suppression. Both experiments were recorded with 128 scans and a two second interscan

delay. Subsequently 100 µM of a competitor compound, previously identified in a high-throughput screen, was added to the protein-containing wells and the experiments re-run.

Data was analysed in Bruker Topspin. Custom written perl scripts were used to import reference spectra for each compound and the 1D spin echo spectrum was compared to these to assess compound integrity in the mixtures. CPMG spectra +/- protein were compared and the % drop in signal intensity was recorded. Separately CPMG spectra without protein and containing protein and competitor were compared in a similar manner. From this data, the % recovery in signal intensity on addition of competitor was calculated. A cut-off of 80% residual signal intensity on addition of protein was used to discriminate compounds showing binding. Based on the % recovery in signal intensity on addition of competitor a provisional assessment of the mechanism of action could be assigned, with some compounds showing clear signs of competitive inhibition. For the WaterLOGSY spectra, the sign change of signals on addition of protein was used to assess protein-compound interaction. For the compound-only spectra, positive signals indicated the potential presence of soluble aggregates and these compounds were excluded from further consideration. In comparison with the reference compound spectra, an initial assignment of the active compounds in the pools could be determined.

Subsequently a deconvolution screen was run for 48 compounds considered active in the primary screen, using the same conditions as described above. 11 out of 48 compounds were removed at this stage. Further deconvolution was carried out against the individual components of the protein complex. Consequently, five compounds showing unambiguous binding only in the presence of the protein-protein complex were identified. A further nine compounds showing clear binding in the presence of the complex, but also a possible interaction with one of the individual binding partners were identified. Other compounds were excluded on the basis of clear interaction with one of the constitutents of the protein-protein complex, or ambiguous interaction with the complex.

References

1. G. W. Bemis and M. A. Murcko, *J. Med. Chem.*, 1996, **39**, 2887-2893.
2. BIOVIA, Dassault Systèmes, PipelinePilot, 20.1.0, San Diego: Dassault Systèmes, 2019
3. ClogP 4.3.1, BioByte Corp, Covina
4. Omega Toolkit. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.
5. P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson and M. T. Stahl, *J. Chem. Inf. Model.*, 2010, **50**, 572-584.
6. FastROCS Toolkit. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com (accessed March 31, 2022)
7. A. J. Kimple, R. E. Muller, D. P. Siderovski and F. S. Willard, *Methods Mol. Biol.*, 2010, **627**, 91-100.
8. T. L. Hwang and A. J. Shaka, *J. Magn. Reson.*, 1995, **112**, 275-279.
9. K. Stott, J. Stonehouse, J. Keeler, T.-L. Hwang and A. J. Shaka, *J. Am. Chem. Soc.*, 1995, **117**, 4199-4200.
10. C. Dalvit, P. Pevarello, M. Tatò, M. Veronesi, A. Vulpetti and M. Sundström, *J. Biomol. NMR*, 2000, **18**, 65-68.
11. C. Dalvit, *J. Biomol. NMR*, 1998, **11**, 437-444.
12. C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi and B. Stockman, *J. Biomol. NMR*, 2001, **21**, 349-359.