Fragment screening libraries for the identification of protein hot spots and their minimal binding pharmacophores

Rebecca L. Whitehouse[1,2], Wesam S. Alwan[1,3], Olga V. Ilyichova[1,4,5], Ashley J. Taylor[1], Indu R. Chandrashekaran[1,5,6], Biswaranjan Mohanty[1,7], Bradley C. Doak* [1,5,6] and Martin J. Scanlon* [1,5,6].

[1] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia.

[2] Current address: Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, United States of America.

[3] Current address: Agilent Technologies, Mulgrave, VIC 3170, Australia.

[4] Australian Synchrotron, ANSTO, Clayton, VIC 3168, Australia.

[5] ARC Training Centre for Fragment Based Design, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia.

[6] Monash Fragment Platform, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia.

[7] Current address: Sydney Analytical Core Research Facility, The University of Sydney, Sydney, NSW 2006, Australia.

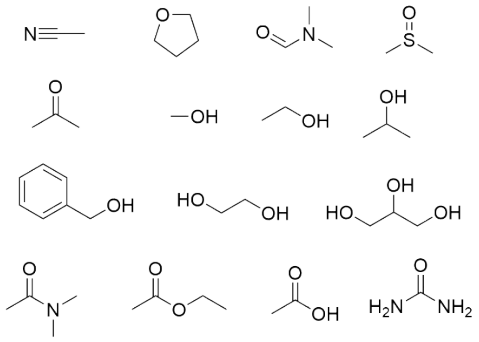
Supporting information

SMILES strings used in FTMAP

Probe	SMILES
Acetaldehyde	CC=O
Acetamide	CC(N)=O
Acetone	CC(C)=O
Acetonitrile	CC#N
Benzaldehyde	O=CC1=CC=CC=C1
Benzene	C1=CC=CC=C1
Cyclohexane	C1CCCCC1
Dimethyl ether	COC
Ethane	CC
Ethanol (EtOH)	CCO
Isobutanol	CC(C)CO
Isopropanol (IPA)	CC(O)C
Methylamine	NC
***N*,*N*-dimethylformamide (DMF)**	O=CN(C)C
Phenol	OC1=CC=CC=C1
Urea	NC(N)=O

SMILES strings and structures of organic solvent screen library

Solvent	SMILES
Acetonitrile	CC#N
Tetrahydrofuran (THF)	C1CCCO1
***N*,*N*-dimethylformamide (DMF)**	O=CN(C)C
Dimethyl sulfoxide (DMSO)	O=S(C)C
Methanol (MeOH)	CO
Ethanol (EtOH)	CCO
Isopropanol (IPA)	CC(O)C
Acetone	CC(C)=O
Benzyl alcohol	OCC1=CC=CC=C1
Ethylene glycol	OCCO
Glycerol	OCC(CO)O
***N*,*N*-dimethylacetamide**	CC(N(C)C)=O
Ethyl acetate (EtOAc)	O=C(OCC)C
Acetic acid	CC(O)=O
Urea	NC(N)=O

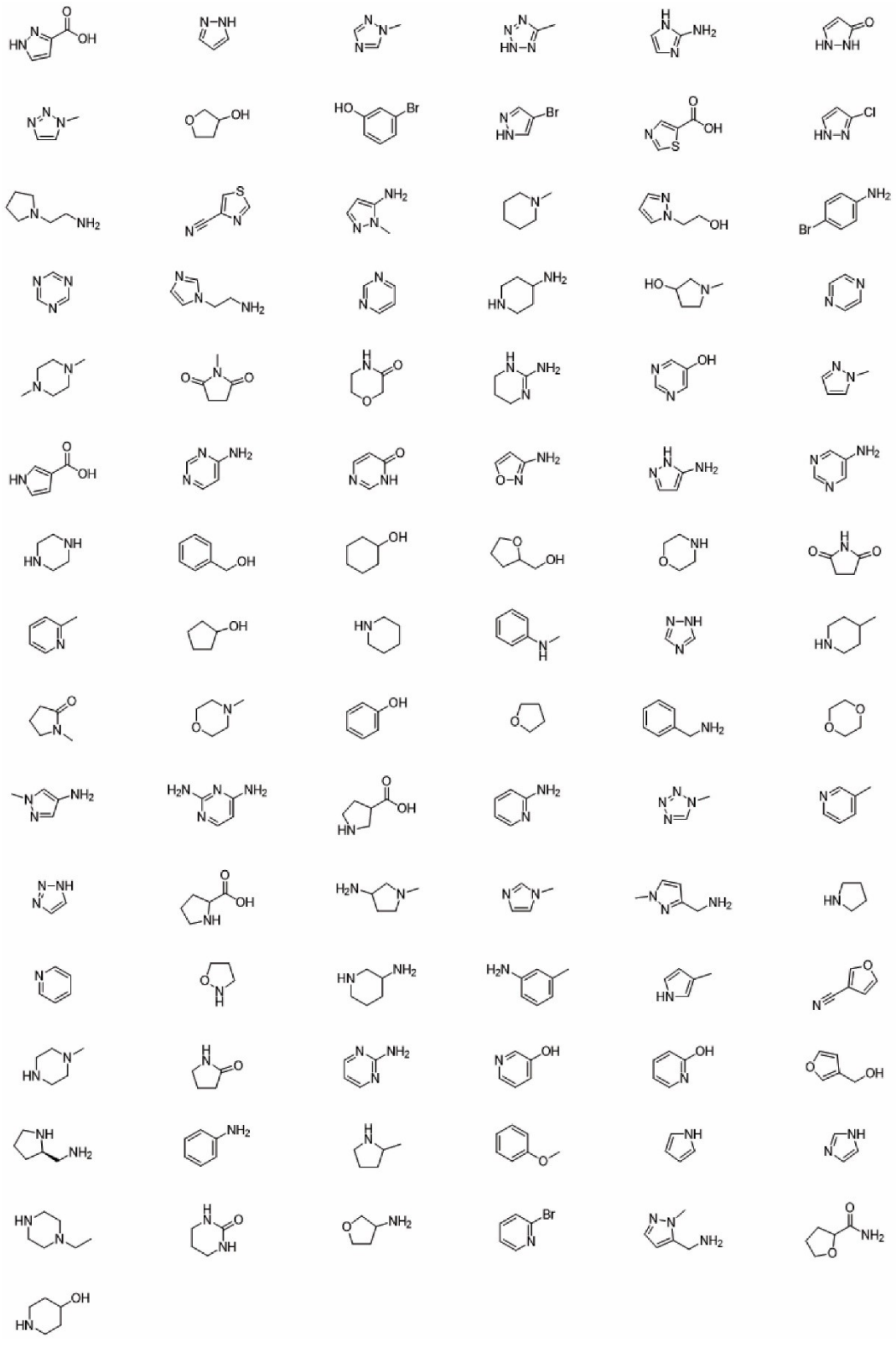


SMILES strings and structures of MicroFrag library

Compound number	SMILES
MF1	OC(=O)C1=NNC=C1
MF2	CN1C=CN=N1
MF3	NCCN1CCCC1
MF4	C1=NC=NC=N1
MF5	CN1CCN(C)CC1
MF6	N1C=CC=N1
MF7	OC1CCOC1
MF8	N#CC1=CSC=N1
MF9	Cl.NCCN1C=CN=C1
MF10	CN1C(=O)CCC1=O
MF11	CN1C=NC=N1
MF12	OC1=CC(Br)=CC=C1
MF13	CN1N=CC=C1N
MF14	C1=CN=CN=C1
MF15	O=C1COCCN1
MF16	CC1=NNN=N1
MF17	BrC1=CNN=C1
MF18	CN1CCCCC1
MF19	NC1CCNCC1
MF20	Cl.NC1=NCCCN1
MF21	NC1=NC=CN1
MF22	OC(=O)C1=CN=CS1
MF23	OCCN1C=CC=N1
MF24	CN1CCC(O)C1
MF25	OC1=CN=CN=C1
MF26	O=C1NNC=C1
MF27	ClC1=NNC=C1
MF28	NC1=CC=C(Br)C=C1
MF29	C1=CN=CC=N1
MF30	CN1C=CC=N1
MF31	CC1=CC=CN=C1
MF32	C1CCNC1
MF33	N#CC1=COC=C1
MF34	CN1N=CC=C1CN
MF35	NC(=O)C1CCCO1

Compound number	SMILES
MF36	CN1C=NN=N1
MF37	CN1C=CC(CN)=N1
MF38	CC1=CNC=C1
MF39	OCC1=COC=C1
MF40	Cl.CC1CCCN1
MF41	Cl.OC(=O)C1CCNC1
MF42	Cl.Cl.CN1CCC(N)C1
MF43	Cl.Cl.NC1CCCNC1
MF44	OC1=CC=CN=C1
MF45	Cl.Cl.NC[C@H]1CCCN1
MF46	NC1=CC=NC(N)=N1
MF47	OC(=O)C1CCCN1
MF48	O=C1CCCN1
MF49	Cl.C1CNOC1
MF50	O=C1NCCCN1
MF51	OC(=O)C1=CNC=C1
MF52	NC1=CC=NC=N1
MF53	O=C1NC=NC=C1
MF54	NC1=NOC=C1
MF55	NC1=CC=NN1
MF56	NC1=CN=CN=C1
MF57	OC1CCNCC1
MF58	N1C=CN=C1
MF59	COC1=CC=CC=C1
MF60	BrC1=NC=CC=C1
MF61	C1CNCCN1
MF62	OCC1=CC=CC=C1
MF63	OC1CCCCC1
MF64	OCC1CCCO1
MF65	C1COCCN1
MF66	O=C1CCC(=O)N1
MF67	N1C=CN=N1
MF68	NC1=NC=CC=N1
MF69	NC1=NC=CC=C1
MF70	OC1=NC=CC=C1

Compound number	SMILES
MF71	CC1=NC=CC=C1
MF72	OC1CCCC1
MF73	C1CCNCC1
MF74	CNC1=CC=CC=C1
MF75	N1C=NC=N1
MF76	CC1CCNCC1
MF77	CN1CCNCC1
MF78	CCN1CCNCC1
MF79	CN1C=CN=C1
MF80	NC1=CC=CC=C1
MF81	CN1CCCC1=O
MF82	CN1CCOCC1
MF83	OC1=CC=CC=C1
MF84	C1CCOC1
MF85	NCC1=CC=CC=C1
MF86	C1COCCO1
MF87	C1=CC=NC=C1
MF88	NC1CCOC1
MF89	CC1=CC=CC(N)=C1
MF90	N1C=CC=C1
MF91	CN1C=C(N)C=N1



MicroFrag library design

All analysis and design was conducted using KNIME Analytic Platform version 4.0.2[1] with nodes developed by KNIME Analytic Platform, RDKit, CDK toolkits, R, Vernalis and Indigo (EPAM Systems). All theoretical topologies were manually created for compounds containing 5-8 heavy atoms and one 5- or 6-membered ring. Topologies were defined as the theoretically connected scaffold graphs of nodes and connecting edges without bond order or atom type.

The PDB analysis was conducted using the dataset of all ligands accessed in November 2018[2, 3], where the following filters were applied: 5 – 8 heavy atoms, SlogP < 2, at least one heteroatom, at least one ring, and a custom list of ≥ 3 undesirable structural alerts[4], ≥1 reactive functional groups or ≥1 PAINS.[5, 6] A score of 1 or 0 was given if the fragments were found in the PDB or not.

The DrugBank analysis was conducted using the dataset accessed in November 2018[7] where the following filters were applied: FDA approved status, oral route of administration, and SlogP between -2 and 6. The remaining 1054 drugs were fragmented into 5250 5 and 8 heavy atom fragments, radicals were converted to hydrogens and aromaticity was resolved using the RDKit normalisation node. Fragments with ≥ 3 undesirable alerts, ≥1 reactive functional groups or ≥1 PAINS were removed as well as any compound that did not contain at least one heteroatom. A normalised frequency of fragment occurrence within the oral drug bank set was then calculated with sd = 1, av = 0.

The MicroFrag library was designed from reagents that were commercially available within the MolPort building blocks reagent list accessed in November 2018. The reagent list was filtered for compounds which contained between 5 – 8 heavy atoms, SlogP < 2, one 5- or 6-membered ring, and removed isotopically enriched compounds.

All commercially available compounds fitting the criteria were analysed by calculation of physicochemical properties (RDKit Descriptors node) number of 2D 2point pharmacophores (from a list of 95 in-house SMARTS substructural 2-point pharmacophore motifs). A scoring function (MicroFrag score) for desired physicochemical properties, presence of Cl, Br or I halogens, complexity and appearance in the PDB dataset or frequency in the DrugBank databases was calculated based on the following formula:

(abs((2.5-$NumHeteroAtoms$))*-2)+(9-$NumHeavyAtoms$)*2+(abs(-0.65-$SlogP$))*-6+($number of 2D 2-point Pharmacophores$*-0.5)+($whether it was found in the PDB or DrugBank$*1)+(normalised frequency of occurrence in drugs*2)+($contains a Cl, Br or I$*1)+($NumAromaticRings$*2)+(abs(45-$TPSA$)*-0.1)+($number of alert FG matches$*-2)

Diversity selection was conducted using an iterative selection process where a balance of normalised MicroFrag score, novel topology coverage, novel 2D 2-point pharmacophore coverage, and sphere exclusion was calculated with Tanimoto similarity of ECFP4[8] and 2D 2-point pharmacophore fingerprints were used to build the library. Diversity selection was calculated based on the following formula:

(-1*$number of novel topology or 2D 2-point pharmacophore$)*(weighting of novel topology/pharmacophore)+(-1*$MicroFrag score$)*(weighting of MicroFrag score)+($Tanimoto similarity to library (ECFP4)$)*(weighting of Tanimoto similarity)+($Tanimoto similarity to library (2D 2-point pharmacophores)$)*(weighting of Tanimoto similarity)

Weighting of variables:

- Novel topology/pharmacophore = 0.5
- MicroFrag score = 1.2
- Tanimoto similarity to library = 1.2

Physicochemical properties, nPMI[9], 2D 2-point pharmacophores, 2D 3-point pharmacophores, SMCM[10], Murcko scaffolds[11], topologies, and Tanimoto similarities of ECFP4, FCFP4, and MACCS fingerprints [8, 12] were used to analyse and compare the resultant commercially available, PDB, DrugBank and MicroFrag compound lists.

Crystallisation and soaking conditions

Organic solvent library

Compound-*Ec*DsbA complexes were prepared by crystal soaking. *Ec*DsbA was crystallised using hanging drop vapour diffusion. 1 µL of 30 mg/mL protein was mixed with an equal volume of crystallisation buffer (11 – 13 % PEG8000, 5 – 6.5 % glycerol, 1 mM $CuCl_2$, 100 mM sodium cacodylate pH 6.1) and equilibrated against 0.5 mL of reservoir buffer at 20 °C. Pre-grown *Ec*DsbA crystals were dehydrated and transferred into 2 µL drops of 50 or 80 % organic solvent diluted in 13 % PEG8000, 6.5 % glycerol, 1 mM $CuCl_2$, 100 mM sodium cacodylate pH 6.1 for 0.5- 5 mins. Crystals were mounted on cryo loops and flash-cooled in liquid nitrogen.

MicroFrag library

Compound-*Ec*DsbA complexes were prepared by crystal soaking. *Ec*DsbA was crystallised using hanging drop vapour diffusion. 1 µL of 30 mg/mL protein was mixed with an equal volume of crystallisation buffer (11 – 13 % PEG8000, 5 – 6.5 % glycerol, 1 mM $CuCl_2$, 100 mM sodium cacodylate pH 6.1) and equilibrated against 0.5 mL of reservoir buffer at 20 °C. Crystals were transferred into 2 µL drops of 13 % PEG8000, 6.5 % glycerol, 1 mM $CuCl_2$, 100 mM sodium cacodylate pH 6.1, 9 % MeOH and the MicroFrag at 1 M and incubated for 0.5-5 minutes. Crystals were mounted on cryo loops and flash-cooled in liquid nitrogen. If the addition of a MicroFrag compound caused cracking or dissolution of the crystal, the final concentration was adjusted to 0.5 M, adjusted to the pH range 4-9, and the soaking experiment was rerun.

Data collection and processing of the dataset

Datasets were collected at the Australian Synchrotron, part of ANSTO, on MX1 and MX2 beamlines and made use of the Australian Cancer Research Foundation (ACRF) detector[13]. The MX1 beamline was equipped with an ADSC Quantum 210r detector and MX2 with an EIGER X 16M pixel detector (Dectris Ltd). Data were processed using the automated data processing pipeline implemented at the beamline, where data were indexed, integrated and scaled with xdsme[14] and AIMLESS[15]. The resulting data collection statistics for each dataset were reviewed and, if necessary, reprocessed with XDS [16] or IMOSFLM [17] and AIMLESS[15, 18]. For all the datasets, we chose a resolution cut-off based on the following criteria in the highest resolution range met: $CC_{1/2}$ was at least 0.6, $\langle I/\sigma(I) \rangle$ was greater than 1.0, and

completeness was greater than 90 %. In cases where data collection statistics did not fulfil our quality requirement or crystals diffracted to lower than 2.5 Å resolution, the crystallographic experiment was repeated.

All datasets were phased by molecular replacement (MR) with Phaser[19] using chain A of apo-*Ec*DsbA structure (PDB:1FVK[20]) as a search model. MR and automated refinement steps were carried out using the automated MR pipeline implemented in Auto-Rickshaw[21, 22]. Briefly, MR was performed using the MOLREP[23] automated refinement protocol. This involved initial refinement in CNS[24] followed by refinement in REFMAC5[25]. In the CNS rigid body refinement, B-factor (BINDIVIDUAL refinement) and positional refinement were undertaken. The resulting model was used in REFMAC5, to perform a maximum likelihood refinement including B-factor refinement.

The final structure was obtained after several rounds of manual refinement using Coot[26] and refinement in phenix.refine[27, 28]. Ligand restraints were generated in phenix.eLBOW[29] or Grade web server[30] and refinement statistics were prepared with “Generate “Table 1” for journal” utility within PHENIX version 1.19.4092[28].

Refinement and analysis statistics

	Urea - 8DG0	DMSO-1 - 8DG1	DMSO-2 - 8DG2
Wavelength	0.9537	0.95372	0.93572
Resolution range	37.54 - 2.5 (2.589 - 2.5)	47.86 - 1.95 (2.02 - 1.95)	37.18 - 1.95 (2.02 - 1.95)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell	116.341 64.324 75.3464 90 125.318 90	117.214 63.205 74.8476 90 125.254 90	117.246 63.095 74.5471 90 125.253 90
Total reflections	31741 (3130)	65422 (6554)	65016 (6416)
Unique reflections	17916 (1565)	32730 (2296)	32516 (2243)
Multiplicity	7.0 (6.8)	7.0 (6.8)	6.9 (6.7)
Completeness (%)	99.87 (99.94)	99.96 (99.88)	99.88 (99.66)
Mean I/sigma(I)	13.6 (2.57)	15.2 (2.0)	9.4 (1.5)
Wilson B-factor	47.12	35.1	33.16
R-merge	0.03076 (0.3038)	0.063 (1.1183)	0.096 (1.243)
R-meas	0.0435 (0.4296)	0.069 (1.304)	0.103 (1.346)
R-pim	0.03076 (0.3038)	0.036 (0.693)	0.039 (0.512)
CC1/2	0.999 (0.805)	0.999 (0.805)	0.998 (0.817)
CC*	1 (0.944)	1 (0.944)	1 (0.948)
Reflections used in refinement	15859 (1564)	32719 (3274)	32480 (3199)
Reflections used for R-free	807 (60)	1739 (167)	1728 (164)
R-work	0.2015 (0.2959)	0.1898 (0.2591)	0.1882 (0.2521)
R-free	0.2569 (0.4225)	0.2211 (0.3010)	0.2213 (0.2730)
CC(work)	0.968 (0.788)	0.959 (0.858)	0.959 (0.887)
CC(free)	0.943 (0.510)	0.958 (0.794)	0.954 (0.811)
Number of non-hydrogen atoms	2819	3041	3026
Macromolecules	2771	2880	2876
Ligands	22	81	63
Solvent	26	80	87
Protein residues	369	375	375
RMS(bonds)	0.008	0.008	0.008
RMS(angles)	0.93	0.90	0.91
Ramachandran favored (%)	96.99	98.65	98.38
Ramachandran	3.01	1.35	1.62

allowed (%)			
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	1.79	1.70	1.02
Clashscore	5.01	3.67	6.52
Average B-factor	55.44	44.15	41.14
Macromolecules	55.55	43.69	40.68
Ligands	55.60	59.75	58.12
Solvent	44.02	44.69	43.76
Number of TLS groups	12	9	14

	MFP10474_E3 - 8CZN	**MFP6710_B5 - 8CZM**	**MFP10475_E11 - 8CXE**
Wavelength	0.95365	0.95365	0.95365
Resolution range	31.59 - 1.7 (1.761 - 1.7)	37.41 - 1.8 (1.864 - 1.8)	49.00 - 1.47 (1.523 - 1.47)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell	118.439 63.19 75.668 90 124.021 90	116.956 64.9 75.009 90 125.676 90	115.825 64.822 75.075 90 125.752 90
Total reflections	101618 (10199)	83850 (8290)	152390 (14589)
Unique reflections	62171 (5112)	45518 (4149)	76352 (3452)
Multiplicity	6.7 (5.9)	7.0 (6.3)	6.9 (6.7)
Completeness (%)	99.88 (99.98)	99.24 (98.71)	99.44 (90.5)
Mean I/sigma(I)	10.6 (1.79)	13.5 (1.40)	15.8 (1.5)
Wilson B-factor	22.88	30.79	18.76
R-merge	0.092 (0.2843)	0.02021 (0.3699)	0.044 (0.809)
R-meas	0.1 (0.4021)	0.02858 (0.5231)	0.048 (0.876)
R-pim	0.038 (0.2843)	0.02021 (0.3699)	0.018 (0.460)
CC1/2	0.999 (0.814)	1 (0.723)	1 (0.799)
CC*	1 (0.947)	1 (0.916)	1 (0.943)
Reflections used in refinement	50994 (5111)	42058 (4148)	76202 (7304)
Reflections used for R-free	784 (84)	962 (109)	960 (80)
R-work	0.1919 (0.2728)	0.1971 (0.2977)	0.1741 (0.2526)
R-free	0.2172 (0.3120)	0.2211 (0.3408)	0.2032 (0.2572)
CC(work)	0.960 (0.842)	0.955 (0.816)	0.966 (0.850)
CC(free)	0.964 (0.807)	0.937 (0.775)	0.957 (0.840)

Number of non-hydrogen atoms	3292	3168	3464
Macromolecules	2887	2873	2942
Ligands	41	56	50
Solvent	364	239	472
Protein residues	374	376	376
RMS(bonds)	0.007	0.008	0.006
RMS(angles)	0.87	0.93	0.80
Ramachandran favored (%)	98.38	98.66	98.66
Ramachandran allowed (%)	1.62	1.34	1.34
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	0.34	0.69	0.33
Clashscore	1.41	2.29	1.55
Average B-factor	27.34	34.19	24.48
Macromolecules	26.29	33.70	22.81
Ligands	33.96	43.18	31.57
Solvent	34.87	38.06	34.12
Number of TLS groups	1	1	11

	MFP10498_F3 - 8CXD	**MFP2018-H8 - 8D12**	**MFP2447_C10 - 8D10**
Wavelength	0.95365	0.95365	0.95365
Resolution range	31.29 - 1.8 (1.864 - 1.8)	30.05 - 1.472 (1.525 - 1.472)	33.04 - 1.47 (1.523 - 1.47)
Space group	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell	117.725 62.577 74.722 90 126.347 90	115.777 64.725 75.052 90 126.208 90	115.56 65.082 74.938 90 126.282 90
Total reflections	81077 (8065)	147200 (14104)	151134 (14467)
Unique reflections	50152 (4031)	73888 (7041)	75786 (3343)
Multiplicity	6.8 (5.4)	7.0 (6.6))	6.9 (6.3)
Completeness (%)	99.67 (99.88)	97.5 (93.58)	99.22 (94.73)
Mean I/sigma(I)	8.6 (1.92)	12.9 (1.54)	13.23 (1.13)
Wilson B-factor	24.42	20.37	22.71
R-merge	0.02865 (0.2312)	0.053 (0.2822)	0.053 (1.057)
R-meas	0.04051 (0.327)	0.057 (0.3992)	0.053 (1.250)

R-pim	0.02865 (0.2312)	0.031 (0.2822)	0.020 (0.485)
CC1/2	0.999 (0.887)	0.999 (0.814)	1 (0.662)
CC*	1 (0.97)	1 (0.947)	1 (0.892)
Reflections used in refinement	40541 (4031)	73575 (7041)	75518 (7189)
Reflections used for R-free	770 (71)	962 (82)	759 (78)
R-work	0.1781 (0.2641)	0.1903 (0.2924)	0.1852 (0.3053)
R-free	0.2012 (0.3139)	0.2080 (0.3058)	0.1999 (0.3098)
CC(work)	0.966 (0.886)	0.959 (0.848)	0.965 (0.763)
CC(free)	0.953 (0.826)	0.958 (0.786)	0.955 (0.702)
Number of non-hydrogen atoms	3223	3310	3344
Macromolecules	2919	2994	2919
Ligands	58	22	33
Solvent	246	294	392
Protein residues	376	376	376
RMS(bonds)	0.007	0.007	0.008
RMS(angles)	0.78	0.97	0.88
Ramachandran favored (%)	98.92	98.92	98.92
Ramachandran allowed (%)	1.08	1.08	1.08
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	0.34	0.65	0.68
Clashscore	1.57	3.07	1.75
Average B-factor	28.34	25.51	29.95
Macromolecules	27.62	24.77	28.76
Ligands	39.58	36.94	41.20
Solvent	34.17	32.23	37.86
Number of TLS groups	8	10	7

Statistical significance of library properties

Statistical significance was calculated using Chi square test.

Heavy atom count

	Expected (total library)	Observed (hits)	Percentage expected (total library)	Percentage observed (hits)	P value
5-8	140	43	11	59	1.22×10^{-51}
Out	1109	30	89	41	

Hydrogen bond acceptors

	Expected (total library)	Observed (hits)	Percentage expected (total library)	Percentage observed (hits)	P value
0-1	151	23	12	32	2.58×10^{-9}
Out	1098	50	88	68	

Hydrogen bond donors

	Expected (total library)	Observed (hits)	Percentage expected (total library)	Percentage observed (hits)	P value
0-1	1041	68	83	93	0.0085
Out	208	5	17	7	

ClogP

	Expected (total library)	Observed (hits)	Percentage expected (total library)	Percentage observed (hits)	P value
-1 – 0.5	184	28	15	38	2.64×10^{-11}
Out	1065	45	85	62	

SMCM

	Expected (total library)	Observed (hits)	Percentage expected (total library)	Percentage observed (hits)	P value
10 - 15	111	29	9	40	2.29×10^{-27}
Out	1138	44	91	60	

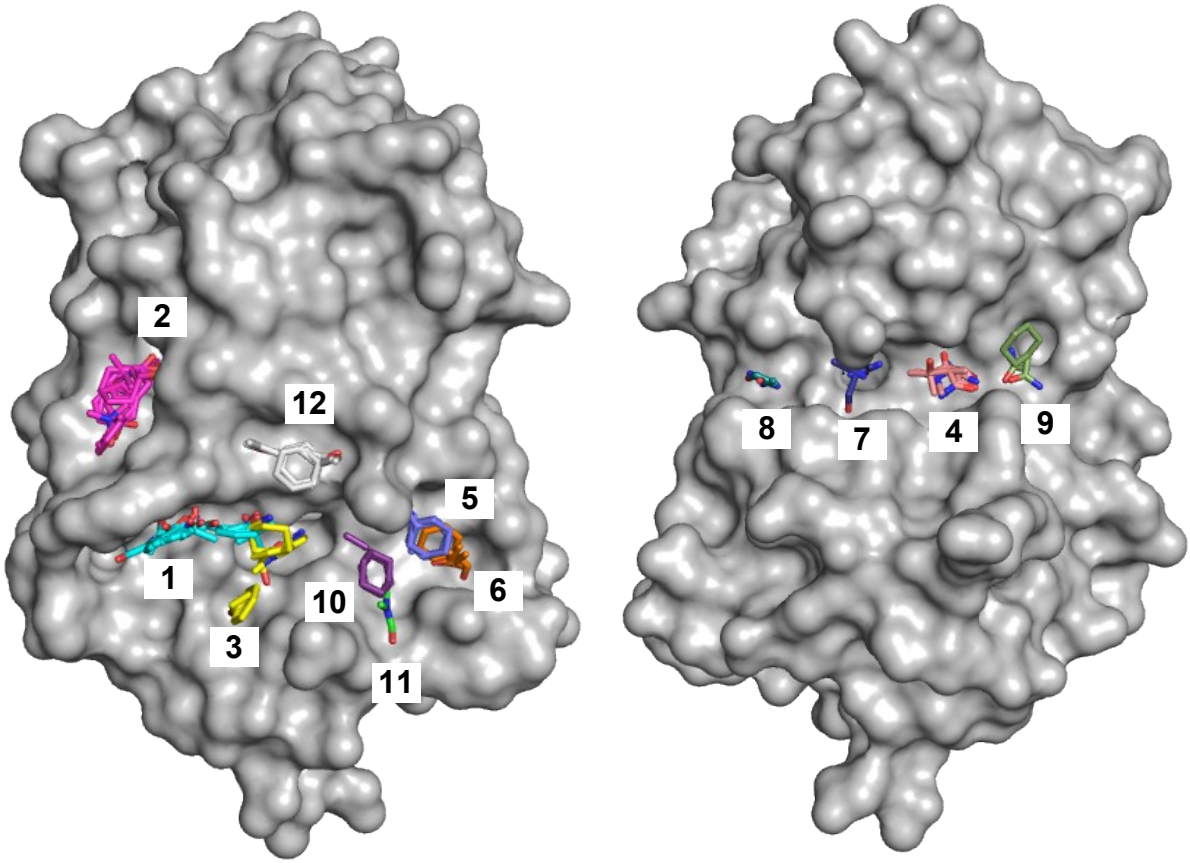
2D 2-point pharmacophores

	Expected (total library)	Observed (hits)	Percentage expected (total library)	Percentage observed (hits)	P value
0-3	311	33	25	45	2.67×10^{-6}
Out	938	40	75	55	

TPSA

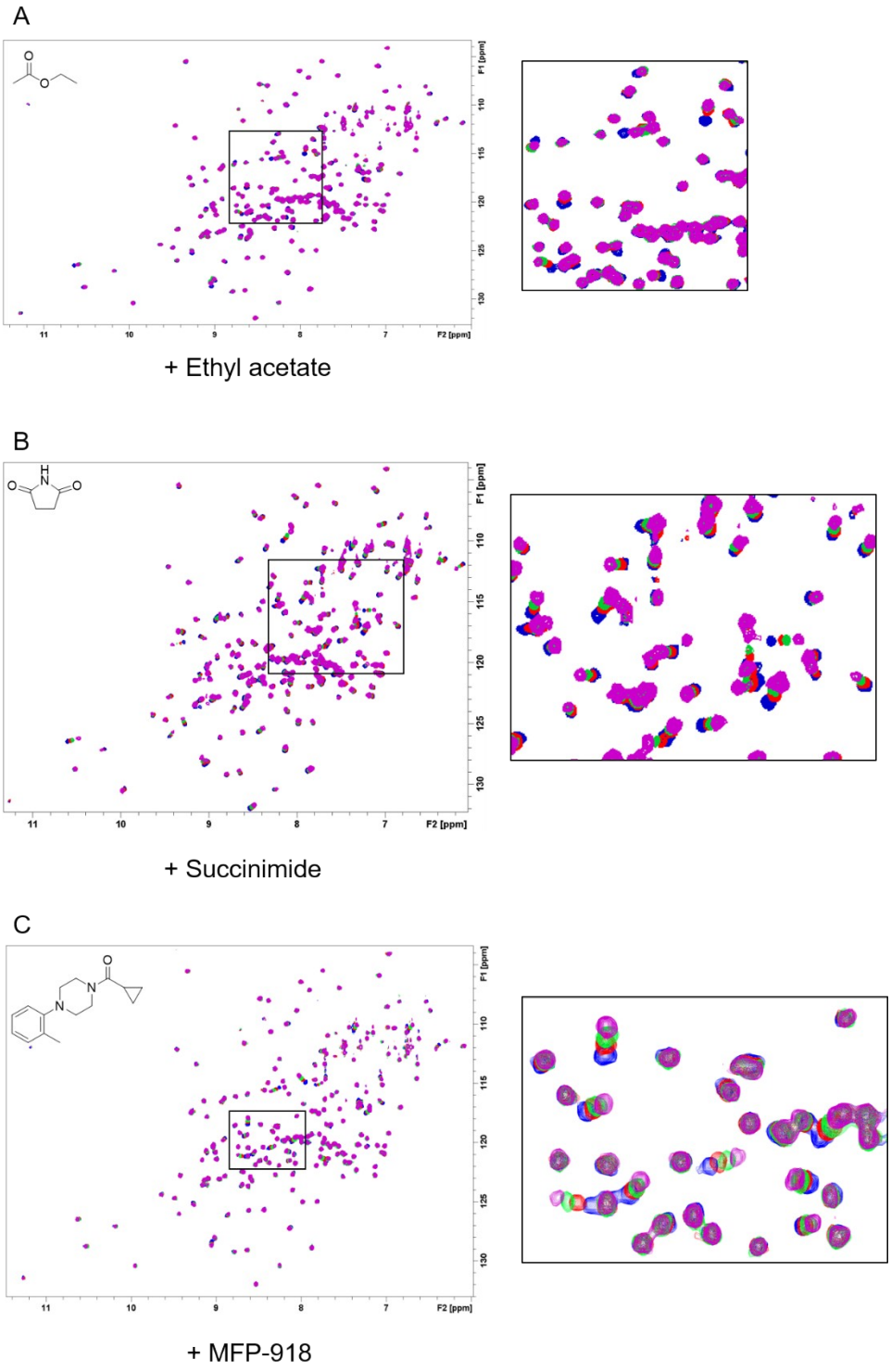
	Expected (total library)	Observed (hits)	Percentage expected (total library)	Percentage observed (hits)	P value
20-40	409	37	33	51	3.33×10^{-4}
Out	840	36	67	49	

Supplementary Figure 1: FTMAP clusters



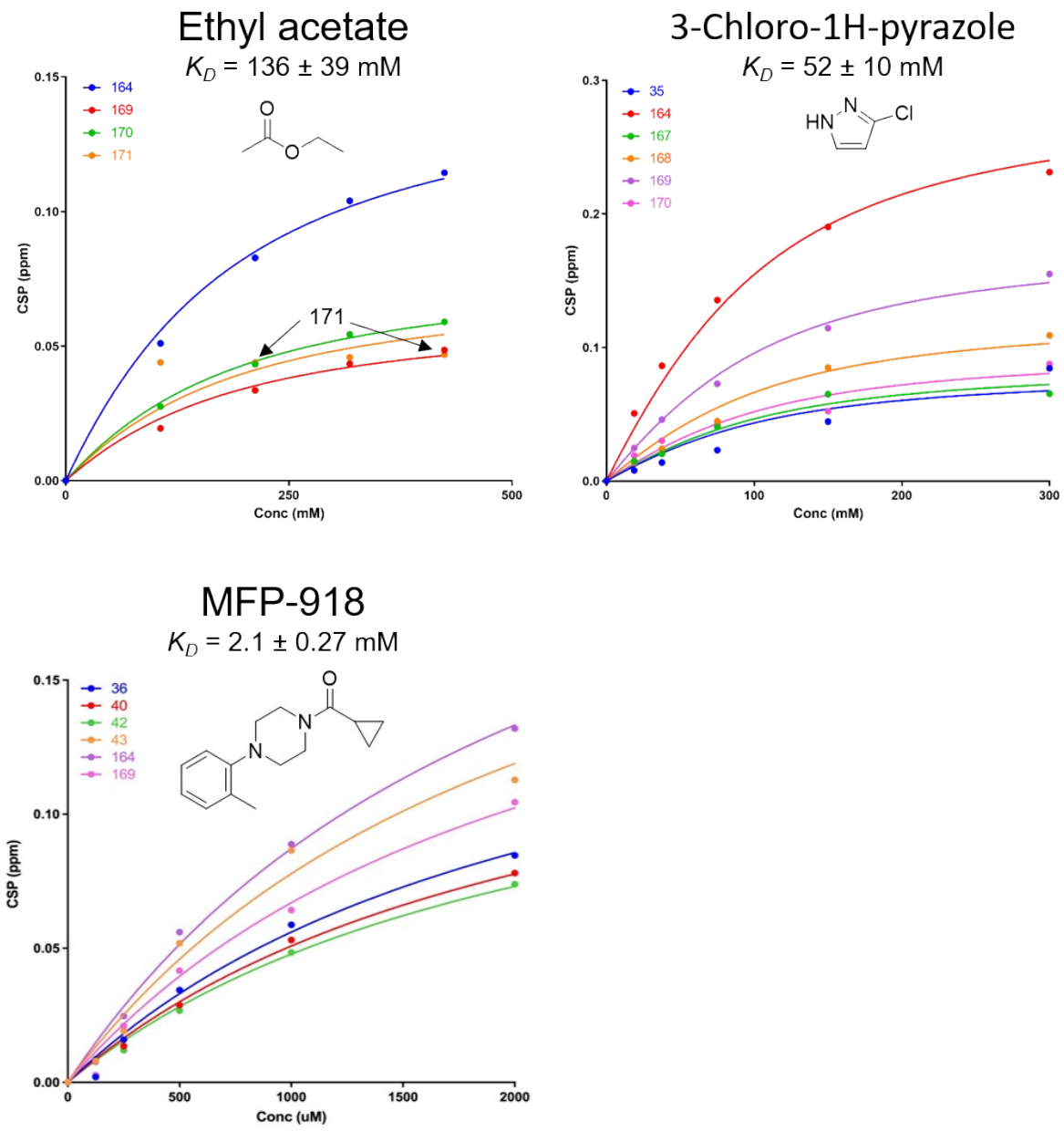
Supplementary Figure 1: Binding hot spots determined by FTMap analysis. Clusters are numbered in order of FTMap ranking. *Ec*DsbA is shown as a grey surface (PDB ID: 1FVK[31]), with each cluster shown as a different colour.

Supplementary Figure 2: HSQC spectral data



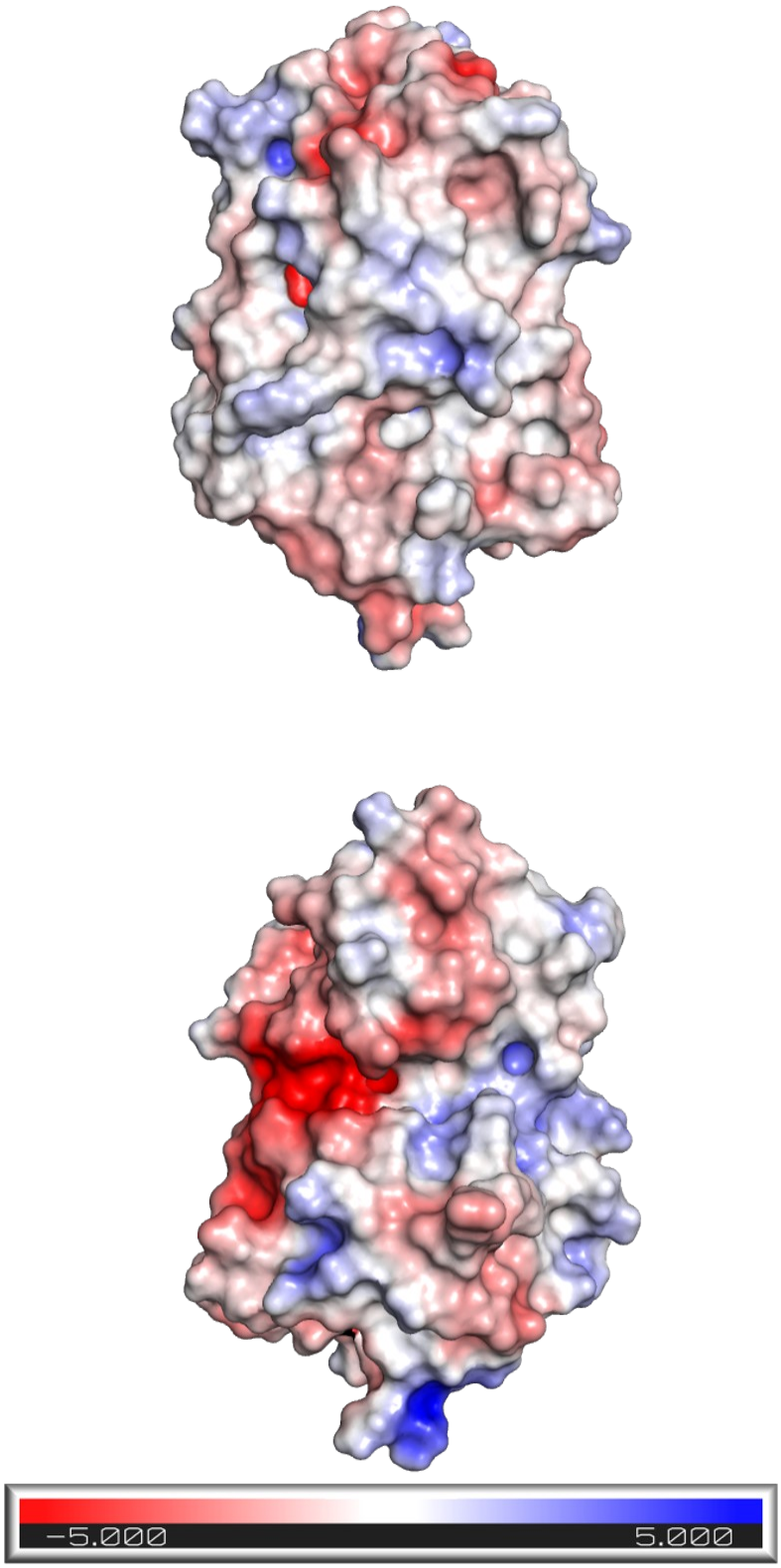
Supplementary Figure 2: Sample ^{15}N-^{1}H HSQC NMR titration spectra. A) ^{15}N-^{1}H HSQC spectra of oxidised ^{15}N *Ec*DsbA (100 µM, blue) in the presence of 2 % (red), 4 % (green) and 5 % (magenta) ethyl acetate. B) ^{15}N-^{1}H HSQC spectra of oxidised ^{15}N *Ec*DsbA (100 µM, blue), in the presence of 0.3 M (red), 0.6 M (green) and 1.5 M (magenta) succinimide and C)) ^{15}N-^{1}H HSQC spectra of oxidised ^{15}N *Ec*DsbA (100 µM, blue), in the presence of 500 µM (red), 1000 µM (green) and 2000 µM (magenta) MFP-918

Supplementary Figure 3: Binding isotherms



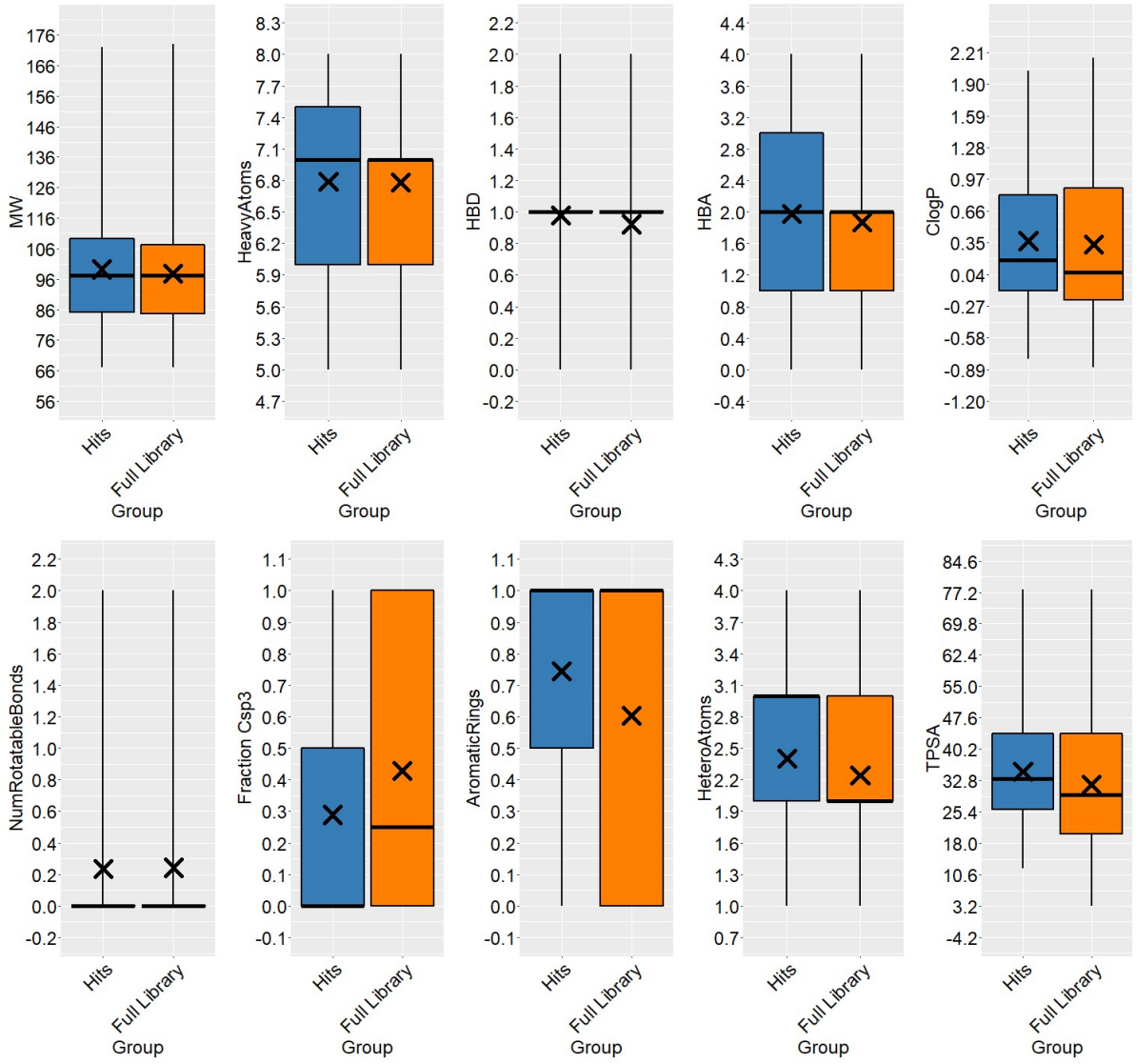
Supplementary Figure 3: Binding affinity estimation based on CSP observed in the ^{15}N-^{1}H HSCQ NMR. Sample affinity data were collected by testing compound titrations against uniformly ^{15}N labelled oxidised EcDsbA by ^{15}N-^{1}H HSCQ NMR. Insets indicate the residues of *Ec*DsbA.

Supplementary Figure 4: Electrostatic potential of *Ec*DsbA



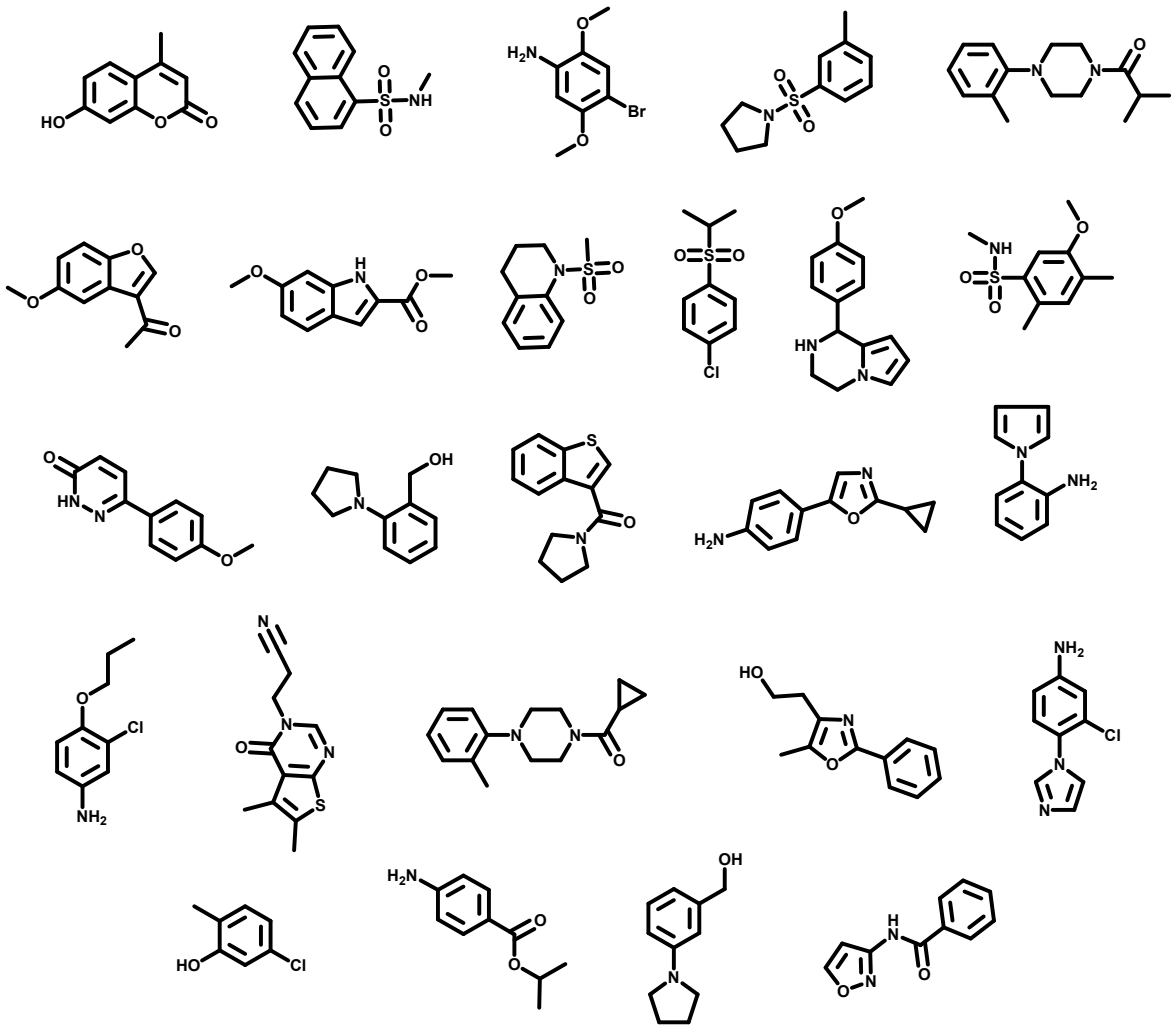
Supplementary Figure 4: Electrostatic potential of *Ec*DsbA. Electrostatic potential calculated in PyMOL and mapped onto the surface of *Ec*DsbA (PDB ID: 1FVK[31]) with a red to blue gradient from -5 to 5. Catalytic face of the protein shown at the top, opposite face (rotated 180°) shown at the bottom.

Supplementary Figure 5: Comparison of MicroFrag hit properties



Supplementary Figure 5: Comparison of properties for MicroFrag hits and the full MicroFrag library. Box plots show bottom whisker = minimum, bottom box = 25th percentile, box bar = median, box top = 75th percentile, top whisker = maximum, cross = average. MW = Molecular weight (Da), HBD = Hydrogen bond donors, HBA = Hydrogen bond acceptors, ClogP = Calculated ocatnol-water partitioning coefficient, TPSA = Total polar surface area (Å^2).

Supplementary Figure 6: Structures of validated hit fragments



Supplementary Figure 6. Validated fragment hits binding to the hydrophobic groove of oxidised *Ec*DsbA. Fragments were considered to be validated hits if they induced chemical shift perturbations ≥ 0.04 ppm in the ^{15}N-^{1}H HSQC NMR spectrum of EcDsbA upon addition of 1 mM compound.

References

1. M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel and B. Wiswedel, Berlin, Heidelberg, 2008.
2. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res*, 2000, **28**, 235-242.
3. P. W. Rose, A. Prlić, A. Altunkaya, C. Bi, A. R. Bradley, C. H. Christie, L. D. Costanzo, J. M. Duarte, S. Dutta, Z. Feng, R. K. Green, D. S. Goodsell, B. Hudson, T. Kalro, R. Lowe, E. Peisach, C. Randle, A. S. Rose, C. Shao, Y.-P. Tao, Y. Valasatava, M. Voigt, J. D. Westbrook, J. Woo, H. Yang, J. Y. Young, C. Zardecki, H. M. Berman and S. K. Burley, *Nucleic Acids Research*, 2017, **45**, D271-D281.
4. I. Sushko, E. Salmina, V. A. Potemkin, G. Poda and I. V. Tetko, *Journal of Chemical Information and Modeling*, 2012, **52**, 2310-2316.
5. J. B. Baell and G. A. Holloway, *Journal of Medicinal Chemistry*, 2010, **53**, 2719-2740.
6. J. B. Baell and J. W. M. Nissink, *ACS Chemical Biology*, 2018, **13**, 36-44.
7. D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox and M. Wilson, *Nucleic Acids Res*, 2018, **46**, D1074-d1082.
8. D. Rogers and M. Hahn, *Journal of Chemical Information and Modeling*, 2010, **50**, 742-754.
9. W. H. B. Sauer and M. K. Schwarz, *Journal of Chemical Information and Computer Sciences*, 2003, **43**, 987-1003.
10. T. K. Allu and T. I. Oprea, *J Chem Inf Model*, 2005, **45**, 1237-1243.
11. G. W. Bemis and M. A. Murcko, *Journal of Medicinal Chemistry*, 1996, **39**, 2887-2893.
12. J. L. Durant, B. A. Leland, D. R. Henry and J. G. Nourse, *Journal of Chemical Information and Computer Sciences*, 2002, **42**, 1273-1280.
13. T. M. McPhillips, S. E. McPhillips, H. J. Chiu, A. E. Cohen, A. M. Deacon, P. J. Ellis, E. Garman, A. Gonzalez, N. K. Sauter, R. P. Phizackerley, S. M. Soltis and P. Kuhn, *J Synchrotron Radiat*, 2002, **9**, 401-406.
14. P. Legrand, *GitHub repository*, 2017, DOI: 10.5281/zenodo.837885.
15. P. R. Evans and G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr*, 2013, **69**, 1204-1214.
16. W. Kabsch, *Acta Crystallogr D Biol Crystallogr*, 2010, **66**, 133-144.
17. T. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell and A. G. Leslie, *Acta Crystallogr D Biol Crystallogr*, 2011, **67**, 271-281.
18. *Acta Crystallogr D Biol Crystallogr*, 1994, **50**, 760-763.
19. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J Appl Crystallogr*, 2007, **40**, 658-674.
20. K. Inaba, S. Murakami, M. Suzuki, A. Nakagawa, E. Yamashita, K. Okada and K. Ito, *Cell*, 2006, **127**, 789-801.
21. S. Panjikar, V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker, *Acta Crystallogr D Biol Crystallogr*, 2005, **61**, 449-457.
22. S. Panjikar, V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker, *Acta crystallographica. Section D, Biological crystallography*, 2009, **65**, 1089-1097.
23. A. Vagin and A. Teplyakov, *Journal of Applied Crystallography*, 1997, **30**, 1022-1025.
24. A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren, *Acta Crystallogr D Biol Crystallogr*, 1998, **54**, 905-921.
25. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr D Biol Crystallogr*, 1997, **53**, 240-255.
26. P. Emsley and K. Cowtan, *Acta Crystallogr D Biol Crystallogr*, 2004, **60**, 2126-2132.

27. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr D Biol Crystallogr*, 2010, **66**, 213-221.
28. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams and P. D. Adams, *Acta Crystallogr D Struct Biol*, 2019, **75**, 861-877.
29. N. W. Moriarty, R. W. Grosse-Kunstleve and P. D. Adams, *Acta Crystallogr. Sect. D. Struct. Biol.*, 2009, **65**, 1074-1080.
30. O. S. Smart, Womack, T. O., Sharff, A., Flensburg, C., Keller, P., and W. Paciorek, Vonrhein, C. and Bricogne, G., Grade, version 1.2.20, https://www.globalphasing.com).
31. L. W. Guddat, J. C. Bardwell, R. Glockshuber, M. Huber-Wunderlich, T. Zander and J. L. Martin, *Protein Sci*, 1997, **6**, 1893-1900.