Supporting Information

Confidence in Inactive and Active Predictions from Structural Alerts

Andrew J. Wedlake,† Timothy E. H. Allen,†‡ Jonathan M. Goodman,† Steve Gutsell, § Predrag Kukic, § and Paul J. Russell§*

†Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

‡MRC Toxicology Unit, University of Cambridge, Gleeson Building, Tennis Court Road, Cambridge CB2 1QR, United Kingdom

§Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire, MK44 1LQ, United Kingdom

Corresponding author email address: jmg11@cam.ac.uk

Contents:

Table S2: The mean inhibition for each compound against each biological target, as measured in the SafetyScreen44™ panel.

Table S3: Statistical performance metrics for the structural alert models against each biological target.

A. Constructing Structural Alerts

Structural alerts were generated using an automated workflow in KNIME (version 3.5.3). Maximum common substructures occurring in at least two active chemicals were generated by the MoSS node with ring mining (rings were treated as indivisible units). Substructures must occur in no greater than 5% of the inactive chemicals. The occurrence of each substructure in the active and inactive chemicals was analysed statistically to identify the best performing substructure. A hypothesis test was done for each substructure, comparing the probability of the occurrence of active and inactive being given by two different binomial distributions - one distribution bias towards active chemicals (probability of giving active = 0.95) and the other a random distribution (probability of giving active = 0.5). Bayes factor was calculated as the probability given by the bias distribution divided by the probability given by the random distribution. Setting the hypothesis test up in this way allows the user to adjust the relative weighting of active and inactive chemicals - the larger the bias towards active of the bias distribution, the more active chemicals required for each inactive chemical containing the substructure. In this work, the bias was fixed as the probability of giving active of 0.95.

$$Bayes\ factor = \frac{0.95^{actives}(1-0.95)^{inactives}}{0.5^{actives}0.5^{inactives}}$$

At each iteration, the maximal common substructure with the largest Bayes factor is chosen to be a structural alert. All active chemicals containing the substructure are removed from the training set and the iterative process is repeated until no remaining substructure has a Bayes factor greater than or equal to the value of Bayes factor given by two actives and one inactive. The workflow outputs a list of structural alerts.

B. Reformatting substructures from SMILES to SMARTS

The structural alert substructures were written by the MoSS node in a SMILES format. Whilst this was not a problem within the Indigo nodes in KNIME, some SMILES substructures with multiple aromatic nitrogen atoms in five-member rings will not be read by the RDKit[27] nodes in KNIME, or equivalent in Python (it is ambiguous which nitrogen atom should have bonds coming out of the ring). Furthermore, writing substructures in SMARTS formats allows more advanced operations for substructure searching. In theory, SMILES strings should be valid SMARTS strings. However, there is one key difference regarding terminal atoms within a substructure. Using carbon atoms as an example, in SMILES format, a terminal "C" could be any carbon - aliphatic or aromatic. In SMARTS format, a terminal "C" can only be an aliphatic carbon. In a SMARTS string, "[C,c]" should be used to refer to aliphatic or aromatic carbons. Hence, if a SMILES string is used as a SMARTS strings with no changes to the string, chemicals with aromatic groups at the edges of the substructure will not be identified by the substructure. This small difference can change the intended results of a structural alert or a substructure search.

Within a SMILES string with bonds explicitly written, terminal carbons intended to refer to "aliphatic or aromatic" carbons were identified as:

1. "-C)" anywhere in the string
2. "C-" at the start of the string
3. "-C" at the end of the string

The same rules apply for identifying terminal nitrogen atoms intended to refer to "aliphatic or aromatic" nitrogen atoms.

A KNIME workflow (version 3.5.3) including Javascript snippets has been written to reformat SMILES substructures as SMARTS substructures. Each character in the SMILES string was split into an individual cell in an array, apart from "Cl" (chlorine atom) or "Br" (bromine atom) for which both characters occupy one cell. The sequence of cells was checked for cases where consecutive cells meet any of the three rules for terminal carbons. Where terminal carbons were identified, the cell containing "C" was replaced by "[C,c]". Similarly, terminal nitrogen atoms were identified and replaced with "[N,n]". The entire series of cells was then concatenated sequentially to give the new SMARTS string. An example of this procedure is shown in Figure S1.

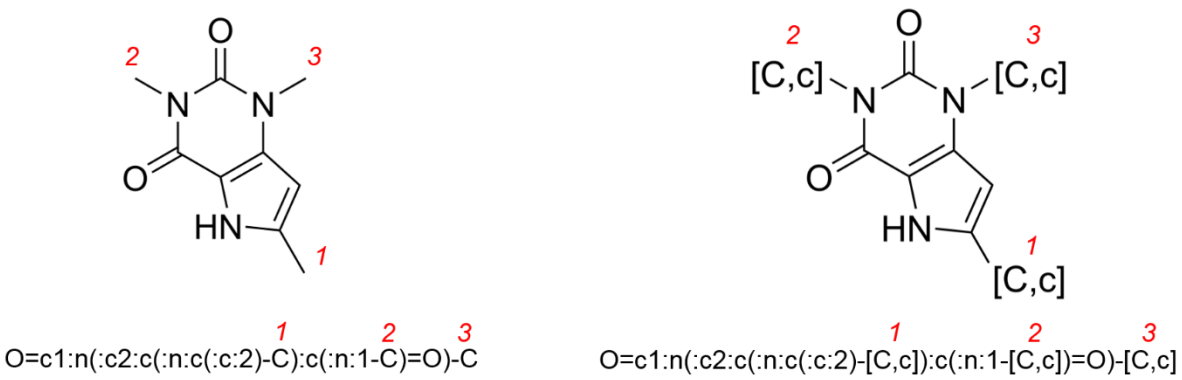

Figure S1: An example of a substructure with three terminal carbons, labelled "1", "2", and "3". On the left is the substructure when written in SMILES format. On the right is the substructure when written in SMARTS format.

C. Equations for model validation statistics

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$ACC = \frac{TN}{TN + FN}$$

$$F1 = 2 \times \frac{SE \times PPV}{SE + PPV}$$

Protein Name	Gene	Uniprot ID	Effects	Reference Compound	IC50 (M)	Ki (M)	Hill Coefficient
G protein coupled receptors							
5-hydroxytryptamine receptor 2A	**HTR2A**	P28223	agonist	(±)DOI	1.30E-10	9.40E-11	0.8
Adenosine receptor A2a	**ADORA2A**	P29274	agonist	NECA	5.30E-08	4.40E-08	0.7
Alpha-2A adrenergic receptor	**ADRA2A**	P08913	antagonist	yohimbine	7.40E-09	3.30E-09	0.9
Beta-1 adrenergic receptor	**ADRB1**	P08588	agonist	atenolol	2.40E-07	1.30E-07	1.3
Beta-2 adrenergic receptor	**ADRB2**	P07550	antagonist	ICI 118551	5.90E-10	2.00E-10	0.8
D(1A) dopamine receptor	**DRD1**	P21728	antagonist	SCH 23390	4.30E-10	1.70E-10	0.8
D(2) dopamine receptor	**DRD2**	P14416	agonist	7-OH-DPAT	2.10E-09	8.40E-10	0.9
Delta-type opioid receptor	**OPRD1**	P41143	agonist	DPDPE	3.30E-09	1.80E-09	1.1
Endothelin-1 receptor	**EDNRA**	P25101	agonist	endothelin-1	5.00E-11	2.50E-11	1.0
Histamine H1 receptor	**HRH1**	P35367	antagonist	pyrilamine	2.00E-09	1.30E-09	1.3
Muscarinic acetylcholine receptor M1	**CHRM1**	P11229	antagonist	pirenzepine	3.00E-08	2.60E-08	1.0
Muscarinic acetylcholine receptor M2	**CHRM2**	P08172	antagonist	methoctramine	4.90E-08	3.40E-08	0.8
Muscarinic acetylcholine receptor M3	**CHRM3**	P20309	antagonist	4-DAMP	1.40E-09	1.00E-09	0.8
Mu-type opioid receptor	**OPRM1**	P35372	agonist	DAMGO	9.30E-10	3.80E-10	1.1
Vasopressin V1a receptor	**AVPR1A**	P37288	agonist	[d(CH2)51,Tyr(Me)2]-AVP	1.20E-09	7.80E-10	1.1
Ion channels							
Potassium voltage-gated channel subfamily H member 2	**KCNH2**	Q12809	antagonist	terfenadine	5.80E-08	4.00E-08	1.0
5-hydroxytryptamine receptor 3A	**HTR3A**	P46098	antagonist	MDL 72222	1.20E-08	8.40E-09	1.0
Enzymes							
Acetylcholinesterase	**ACHE**	P22303		galanthamine	7.00E-07		0.8
Tyrosine-protein kinase Lck	**LCK**	P06239		staurosporine	5.50E-08		2.0
Transporters							
Sodium-dependent noradrenaline transporter	**SLC6A2**	P23975	antagonist	protriptyline	3.10E-09	2.30E-09	1.5
Sodium-dependent dopamine transporter	**SLC6A3**	Q01959	antagonist	BTCP	1.30E-08	7.00E-09	1.0
Sodium-dependent serotonin transporter	**SLC6A4**	P31645	antagonist	imipramine	2.00E-09	9.10E-10	1.0
Nuclear receptors							
Androgen receptor	**AR**	P10275	agonist	testosterone	2.90E-09	1.30E-09	1.0
Glucocorticoid receptor	**NR3C1**	P04150	agonist	dexamethasone	4.00E-09	2.00E-09	1.0

Table S1: The reference compounds used for each protein target in the SafetyScreen44^{TM} panel and the protein.

Protein Name	Gene	Uniprot ID	Readout	Measurement	TMPH hydrochloride	Melamine	Ranolazine dihydrochloride	Olanzapine	Azatadine maleate	Penfluridol	Lofepramine	Trequinsin hydrochloride	4-Chloro-DL-phenylalanine	Cyamemazine	ATC 0175	Gaboxadol hydrochloride
5-hydroxytryptamine receptor 2A	**HTR2A**	P28223	Specific binding	Mean %Inhibition	13.6442	6.27278	31.9613	98.7778	94.5817	98.5294	65.9742	0.21569	4.66013	98.7778	97.2876	4.86533
Adenosine receptor A2a	**ADORA2A**	P29274	Specific binding	Mean %Inhibition	-3.12E+00	-4.89E+00	-1.27E+01	-3.24E+00	-3.07E+00	1.99629	-8.39E+00	17.3107	-1.16E+01	-4.61E+00	21.6099	-5.67E+00
Alpha-2A adrenergic receptor	**ADRA2A**	P08913	Specific binding	Mean %Inhibition	16.7198	-6.80E+00	58.1651	84.5289	37.5339	78.1698	-9.76E-01	32.8034	-2.79E+00	51.6479	98.4878	-2.34E+00
Beta-1 adrenergic receptor	**ADRB1**	P08588	Specific binding	Mean %Inhibition	0	5.28109	62.6065	-7.46E+00	-7.24E-02	98.9375	13.2241	-2.15E+00	-7.24E-02	13.2094	93.673	3.53492
Beta-2 adrenergic receptor	**ADRB2**	P07550	Specific binding	Mean %Inhibition	18.6773	4.23328	87.5182	18.2095	-4.26E+00	84.7138	1.77549	-1.52E+00	-1.10E+00	36.2363	87.1498	2.70712
D(1A) dopamine receptor	**DRD1**	P21728	Specific binding	Mean %Inhibition	2.57349	9.19529	16.797	100.661	82.1068	98.4455	82.8511	10.2034	-3.19E+00	98.1809	98.7432	7.20578
D(2) dopamine receptor	**DRD2**	P14416	Specific binding	Mean %Inhibition	5.76789	-2.06E+00	14.0607	100.196	97.5031	99.9021	14.5557	18.3843	4.9694	99.4125	100.392	3.31249
Delta-type opioid receptor	**OPRD1**	P41143	Specific binding	Mean %Inhibition	6.62137	4.36605	7.09845	6.8583	-6.15E+00	99.1931	-3.39E+00	13.7166	3.53001	7.00958	86.7877	4.88651
Endothelin-1 receptor	**EDNRA**	P25101	Specific binding	Mean %Inhibition	11.5344	-2.04E+00	8.5595	3.58632	3.30831	29.886	3.79231	2.64109	3.86433	6.97804	10.2029	8.45511
Histamine H1 receptor	**HRH1**	P35367	Specific binding	Mean %Inhibition	43.4047	14.5956	50.0341	100.633	100.058	96.6033	74.3172	2.18768	-7.48E+00	98.4456	76.7991	13.859
Muscarinic acetylcholine receptor M1	**CHRM1**	P11229	Specific binding	Mean %Inhibition	30.9786	7.19201	14.9565	99.8169	99.3593	93.1655	83.6496	5.23265	-1.15E+01	99.6949	86.1785	6.4247
Muscarinic acetylcholine receptor M2	**CHRM2**	P08172	Specific binding	Mean %Inhibition	59.5692	13.0311	8.05178	99.793	100.673	83.0228	44.2585	21.4286	-4.04E+00	98.8354	91.6925	7.18161
Muscarinic acetylcholine receptor M3	**CHRM3**	P20309	Specific binding	Mean %Inhibition	-8.28E+00	-8.28E+00	-9.85E+00	98.8672	103.32	98.7891	78.6723	-1.12E+01	-1.23E+01	96.3672	73.5547	-4.76E+00
Mu-type opioid receptor	**OPRM1**	P35372	Specific binding	Mean %Inhibition	6.46782	-1.35E+00	8.9168	6.28684	10.4126	97.8389	17.1271	12.3772	-3.44E+00	46.2672	98.5265	-3.14E+00
Vasopressin V1a receptor	**AVPR1A**	P37288	Specific binding	Mean %Inhibition	-6.61E+00	1.36437	1.57838	-9.38E+00	5.099	52.2918	-1.14E+01	61.4049	2.33252	0.54245	12.422	-4.01E-01
Ion channels																
Potassium voltage-gated channel subfamily H member 2	**KCNH2**	Q12809	Specific binding	Mean %Inhibition	22.2898	2.50447	14.3113	26.2217	11.5609	99.9635	2.23081	35.4486	4.52225	69.7666	98.5412	1.25224
5-hydroxytryptamine receptor 3A	**HTR3A**	P46098	Specific binding	Mean %Inhibition	9.51821	4.46533	-5.29E-01	96.0922	22.3575	17.4888	14.0755	1.28123	0.06406	63.9334	2.56246	3.23149
Enzymes																
Acetylcholinesterase	**ACHE**	P22303	Enzymatic activity	Mean %Inhibition	2.13314	-2.06E+00	1.91247	2.78746	6.66999	4.28074	-1.55E+00	5.1767	9.35789	20.7068	68.4918	-9.56E-01
Tyrosine-protein kinase Lck	**LCK**	P06239	Enzymatic activity	Mean %Inhibition	10.083	15.1941	-4.29E+00	-2.48E+00	11.3823	86.5471	11.8547	13.5889	2.79381	32.343	88.4993	7.27394
Transporters																
Sodium-dependent noradrenaline transporter	**SLC6A2**	P23975	Specific binding	Mean %Inhibition	33.5839	-1.49E+00	-3.70E+00	10.3412	51.2484	97.6904	87.0273	8.01082	-2.52E+00	62.526	80.9613	-3.54E+00
Sodium-dependent dopamine transporter	**SLC6A3**	Q01959	Specific binding	Mean %Inhibition	-6.55E+00	-8.55E-01	1.91415	-6.83E+00	9.12363	98.0832	22.2005	5.9764	3.69967	23.2025	97.2294	5.56415
Sodium-dependent serotonin transporter	**SLC6A4**	P31645	Specific binding	Mean %Inhibition	65.5863	-1.77E-02	15.5331	60.5521	13.0971	97.8269	68.0429	-6.56E+00	-6.95E+00	62.5489	-3.35E+01	1.8243
Nuclear receptors																
Androgen receptor	**AR**	P10275	Specific binding	Mean %Inhibition	-1.10E+01	-2.70E+01	-2.76E+01	-2.42E+01	-1.66E+01	-2.33E+01	-1.21E+01	-1.86E+01	-2.88E+01	-8.94E+00	-1.44E+01	-1.44E+01
Glucocorticoid receptor	**NR3C1**	P04150	Specific binding	Mean %Inhibition	4.77164	1.52243	4.3425	-1.59E+00	3.77299	28.9356	9.69097	11.8058	-8.77E+00	6.235	8.32145	0.27588

Mequitazine	Oleanolic acid	SR144528	Caffeine	Carbamazepine	Coumarin	DL sulphoraphane	Doxorubicin	Ethinylestradiol	Niacinamide	Phenoxyethanol	Pioglitazone hydrochloride	Rosiglitazone	Triclosan	Troglitazone
95.7843	-3.97E+01	48.1895	-1.17E+01	-7.49E-01	-6.58E+00	7.38021	-1.63E+01	18.7322	0.14974	-3.88E+00	11.3448	12.6141	25.7487	37.1863
-4.93E+00	-1.22E+01	-1.81E+00	61.6976	-4.39E+00	-4.26E+00	-1.26E+01	13.7328	35.8704	-1.10E+01	3.45656	-1.02E+01	2.90926	21.214	8.60198
76.929	-3.18E+00	-3.64E+00	0.23847	5.00447	1.87668	5.4513	10.0954	7.14924	-5.09E+00	0.55644	8.0429	31.3673	22.1626	30.7417
-1.67E+00	0.79691	45.1823	4.21799	-4.50E+00	1.82388	-1.82E+00	9.80062	-2.39E+00	2.67883	3.52681	0.56996	-7.24E+00	1.65289	1.70989
21.8027	3.16687	-5.24E+00	-6.76E+00	3.92484	-1.17E+00	-2.17E+00	1.32048	1.92067	-1.59E+00	6.54958	3.08977	3.75783	-1.92E+00	-2.76E+00
87.0349	-4.71E+00	21.0518	3.57496	-5.21E-01	-1.18E+00	4.5766	13.9539	19.1011	-2.82E+00	8.27018	3.97369	1.01398	14.6067	48.9175
98.9229	4.43084	32.2399	5.41298	-4.49E+00	-7.44E+00	4.823	10.8475	-1.01E+01	5.34569	10.33	0.9741	-3.30E+00	8.81445	5.72583
29.3495	-9.78E+00	12.6072	-2.75E+00	-8.80E+00	-1.05E+01	1.08351	8.42684	8.0074	-1.49E+01	-4.12E-01	-3.04E+00	24.6564	50.7135	10.2801
6.92244	1.97387	5.53239	-1.73E+01	-4.41E+00	6.46677	-9.02E+00	-1.62E+01	-1.09E+01	-1.25E+01	-2.16E+01	-1.18E+01	-2.43E+01	-6.82E+00	-8.88E+00
96.4882	0.46056	8.4053	4.67626	-5.99E-01	-7.40E+00	-3.27E-01	5.17086	1.30648	-8.27E+00	5.17086	3.4295	-1.52E+00	3.64725	-2.23E+00
99.8169	-1.14E+01	-1.72E+01	5.53731	0.16936	9.23658	7.90776	39.587	18.0433	7.51694	3.82919	5.6149	13.0928	-1.26E+01	15.4377
100.44	-8.02E+00	-1.00E+01	-1.73E+00	-1.23E+01	-6.65E+00	0.651	32.7755	-8.50E-01	-6.65E+00	7.47259	1.66611	1.31303	28.1033	27.9709
100.273	-1.21E+01	5.27344	21.6869	-9.63E+00	-1.18E+01	-3.10E+00	7.48401	-1.92E+00	-3.45E+00	3.53877	0.83995	7.9664	0.24936	11.6674
66.8959	-6.58E+00	23.1827	16.4379	0.88833	-8.84E+00	-2.07E+00	13.1864	34.2217	3.17259	12.1026	-1.65E+00	10.6176	48.2657	27.1151
21.1554	7.54001	5.26173	-4.23E+00	5.89187	-1.14E+01	1.59427	-1.27E-01	1.45564	-6.54E+00	1.18569	-3.63E+00	1.03974	12.3152	-8.11E+00
92.2684	-1.09E+00	55.1787	-6.07E-01	-1.00E+01	-1.31E+01	-7.65E+00	-1.52E-02	1.72456	-5.76E+00	-1.61E+00	-5.56E+00	0.51008	7.45689	18.5815
90.7751	-5.89E+00	-3.33E+00	2.04634	-3.19E+00	-3.33E+00	-9.11E+00	3.18989	-8.84E+00	-5.72E-01	9.44929	7.36382	2.65636	-3.06E+00	4.06859
8.16327	4.57939	11.1498	-4.03E-01	35.4354	19.9199	22.4224	20.1613	9.90991	16.9169	-4.03E-01	3.4034	31.9319	6.40641	15.9159
51.3181	6.60554	16.2863	-9.65E-02	-6.11E-01	-1.02E+01	-8.07E-01	51.7756	-5.74E+00	-1.08E+01	0.75278	5.52073	-1.83E-01	-6.19E+00	-6.10E+00
99.563	-1.07E+01	23.1169	0.41218	-1.99E+01	-2.50E+01	-9.10E+00	0.69753	93.4805	-1.36E+01	0.50729	-5.88E+00	-1.24E+01	95.3786	44.0479
71.0639	-1.01E+01	37.8672	-3.92E+00	-9.64E-01	-1.40E-01	-8.32E+00	-2.93E-01	92.957	-5.50E+00	4.54927	-1.38E+00	-5.22E+00	78.7827	41.9561
48.5709	-8.87E+00	-1.85E+01	-3.40E+00	-1.19E+00	-4.75E+00	2.69133	0.59685	97.2533	-1.27E+00	-2.31E+00	1.00786	-3.56E+00	8.29549	3.73242
-1.04E+01	-2.69E+01	-1.79E+01	-4.39E-01	7.33395	9.25008	1.54943	0.10976	100.815	2.09173	5.56136	-1.08E+01	10.5154	79.6132	64.7454
17.8565	4.16942	9.96975	1.31955	2.47584	6.44777	-7.08E+00	-9.54E+00	94.0156	-1.01E+01	-2.08E+00	-6.75E-01	1.15186	41.0301	82.6294

Table S2: The mean inhibition for each compound against each biological target, as measured in the SafetyScreen44™ panel.

Target	TP	FP	FN	TN	SE	SP	ACC	MCC	PPV	NPV	F1
Acetylcholinesterase	1842	149	162	1318	0.919	0.898	0.910	0.817	0.925	0.891	0.922
Adenosine A2a receptor	2879	121	78	1461	0.974	0.924	0.956	0.903	0.960	0.949	0.967
Alpha-2a adrenergic receptor	596	28	53	741	0.918	0.964	0.943	0.885	0.955	0.933	0.936
Androgen receptor	1499	85	490	5384	0.754	0.984	0.923	0.798	0.946	0.917	0.839
Beta-1 adrenergic receptor	930	67	30	739	0.969	0.917	0.945	0.890	0.933	0.961	0.950
Beta-2 adrenergic receptor	1253	91	209	1410	0.857	0.939	0.899	0.800	0.932	0.871	0.893
Delta opioid receptor	2187	117	53	787	0.976	0.871	0.946	0.866	0.949	0.937	0.963
Dopamine D1 receptor	881	56	148	1428	0.856	0.962	0.919	0.832	0.940	0.906	0.896
Dopamine D2 receptor	4197	136	65	719	0.985	0.841	0.961	0.855	0.969	0.917	0.977
Dopamine transporter	1748	119	122	1326	0.935	0.918	0.927	0.852	0.936	0.916	0.936
Endothelin receptor ET-A	941	59	21	800	0.978	0.931	0.956	0.912	0.941	0.974	0.959
Glucocorticoid receptor	1816	95	458	5148	0.799	0.982	0.926	0.823	0.950	0.918	0.868
HERG	3098	238	519	2183	0.857	0.902	0.875	0.747	0.929	0.808	0.891
Histamine H1 receptor	915	60	48	762	0.950	0.927	0.939	0.878	0.938	0.941	0.944
Mu opioid receptor	2576	117	85	1606	0.968	0.932	0.954	0.903	0.957	0.950	0.962
Muscarinic acetylcholine receptor M1	1403	93	81	844	0.945	0.901	0.928	0.848	0.938	0.912	0.942
Muscarinic acetylcholine receptor M2	1154	82	71	1424	0.942	0.946	0.944	0.887	0.934	0.953	0.938
Muscarinic acetylcholine receptor M3	1132	71	58	768	0.951	0.915	0.936	0.869	0.941	0.930	0.946
Norepinephrine transporter	2099	113	124	1336	0.944	0.922	0.935	0.865	0.949	0.915	0.947
Serotonin 2a (5-HT2a) receptor	2754	91	43	680	0.985	0.882	0.962	0.887	0.968	0.941	0.976
Serotonin 3a (5-HT3a) receptor	321	20	27	758	0.922	0.974	0.958	0.902	0.941	0.966	0.932
Serotonin transporter	3002	95	55	757	0.982	0.888	0.962	0.886	0.969	0.932	0.976
Tyrosine-protein kinase LCK	1246	50	37	355	0.971	0.877	0.948	0.857	0.961	0.906	0.966
Vasopressin V1a receptor	446	21	10	766	0.978	0.973	0.975	0.947	0.955	0.987	0.966
Average					0.930	0.924	0.939	0.863	0.947	0.926	0.937

Table S3a: Statistical performance metrics for the structural alert models against each biological targets training data.

Target	TP	FP	FN	TN	SE	SP	ACC	MCC	PPV	NPV	F1
Acetylcholinesterase	517	61	93	435	0.848	0.877	0.861	0.721	0.894	0.824	0.870
Adenosine A2a receptor	941	49	45	452	0.954	0.902	0.937	0.858	0.951	0.909	0.952
Alpha-2a adrenergic receptor	157	13	39	232	0.801	0.947	0.882	0.764	0.924	0.856	0.858
Androgen receptor	436	42	212	1773	0.673	0.977	0.897	0.723	0.912	0.893	0.774
Beta-1 adrenergic receptor	273	28	28	248	0.907	0.899	0.903	0.806	0.907	0.899	0.907
Beta-2 adrenergic receptor	358	70	125	443	0.741	0.864	0.804	0.611	0.836	0.780	0.786
Delta opioid receptor	733	53	33	262	0.957	0.832	0.920	0.805	0.933	0.888	0.945
Dopamine D1 receptor	247	26	76	481	0.765	0.949	0.877	0.740	0.905	0.864	0.829
Dopamine D2 receptor	1382	59	52	223	0.964	0.791	0.935	0.762	0.959	0.811	0.961
Dopamine transporter	565	47	76	425	0.881	0.900	0.889	0.777	0.923	0.848	0.902
Endothelin receptor ET-A	304	24	19	269	0.941	0.918	0.930	0.860	0.927	0.934	0.934
Glucocorticoid receptor	539	59	205	1671	0.724	0.966	0.893	0.739	0.901	0.891	0.803
HERG	882	145	396	680	0.690	0.824	0.743	0.502	0.859	0.632	0.765
Histamine H1 receptor	284	25	29	259	0.907	0.912	0.910	0.819	0.919	0.899	0.913
Mu opioid receptor	889	42	61	543	0.936	0.928	0.933	0.859	0.955	0.899	0.945
Muscarinic acetylcholine receptor M1	481	41	50	264	0.906	0.866	0.891	0.767	0.921	0.841	0.914
Muscarinic acetylcholine receptor M2	376	31	35	496	0.915	0.941	0.930	0.857	0.924	0.934	0.919
Muscarinic acetylcholine receptor M3	317	23	30	252	0.914	0.916	0.915	0.828	0.932	0.894	0.923
Norepinephrine transporter	624	38	65	454	0.906	0.923	0.913	0.823	0.943	0.875	0.924
Serotonin 2a (5-HT2a) receptor	932	34	28	229	0.971	0.871	0.949	0.849	0.965	0.891	0.968
Serotonin 3a (5-HT3a) receptor	91	10	13	267	0.875	0.964	0.940	0.847	0.901	0.954	0.888
Serotonin transporter	945	32	41	251	0.958	0.887	0.942	0.836	0.967	0.860	0.963
Tyrosine-protein kinase LCK	416	21	33	98	0.927	0.824	0.905	0.725	0.952	0.748	0.939
Vasopressin V1a receptor	150	6	13	266	0.920	0.978	0.956	0.907	0.962	0.953	0.940
Average					0.874	0.902	0.902	0.783	0.924	0.866	0.897

Table S3b: Statistical performance metrics for the structural alert models against each biological targets test data.

Target	TP	FP	FN	TN	SE	SP	ACC	MCC	PPV	NPV	F1
Acetylcholinesterase	157	24	131	147	0.545	0.860	0.662	0.400	0.867	0.529	0.670
Adenosine A2a receptor	234	30	19	27	0.925	0.474	0.842	0.434	0.886	0.587	0.905
Alpha-2a adrenergic receptor	13	4	7	19	0.650	0.826	0.744	0.486	0.765	0.731	0.703
Androgen receptor	93	10	56	25	0.624	0.714	0.641	0.268	0.903	0.309	0.738
Beta-1 adrenergic receptor	3	13	3	37	0.500	0.740	0.714	0.164	0.188	0.925	0.273
Beta-2 adrenergic receptor	36	22	18	37	0.667	0.627	0.646	0.294	0.621	0.673	0.643
Delta opioid receptor	382	55	50	169	0.884	0.754	0.840	0.642	0.874	0.772	0.879
Dopamine D1 receptor	81	20	8	30	0.910	0.600	0.799	0.549	0.802	0.789	0.853
Dopamine D2 receptor	459	40	43	18	0.914	0.310	0.852	0.220	0.920	0.295	0.917
Dopamine transporter	105	35	28	16	0.789	0.314	0.658	0.108	0.750	0.364	0.769
Endothelin receptor ET-A	16	3	0	27	1.000	0.900	0.935	0.871	0.842	1.000	0.914
Glucocorticoid receptor	113	2	19	38	0.856	0.950	0.878	0.723	0.983	0.667	0.915
HERG	182	210	233	437	0.439	0.675	0.583	0.115	0.464	0.652	0.451
Histamine H1 receptor	71	12	17	15	0.807	0.556	0.748	0.343	0.855	0.469	0.830
Mu opioid receptor	853	70	17	45	0.980	0.391	0.912	0.492	0.924	0.726	0.951
Muscarinic acetylcholine receptor M1	75	19	16	36	0.824	0.655	0.760	0.484	0.798	0.692	0.811
Muscarinic acetylcholine receptor M2	35	15	27	43	0.565	0.741	0.650	0.310	0.700	0.614	0.625
Muscarinic acetylcholine receptor M3	70	6	46	54	0.603	0.900	0.705	0.482	0.921	0.540	0.729
Norepinephrine transporter	100	12	12	26	0.893	0.684	0.840	0.577	0.893	0.684	0.893
Serotonin 2a (5-HT2a) receptor	307	63	18	74	0.945	0.540	0.825	0.554	0.830	0.804	0.883
Serotonin 3a (5-HT3a) receptor	59	10	188	54	0.239	0.844	0.363	0.080	0.855	0.223	0.373
Serotonin transporter	150	34	40	33	0.789	0.493	0.712	0.275	0.815	0.452	0.802
Tyrosine-protein kinase LCK	76	102	8	65	0.905	0.389	0.562	0.305	0.427	0.890	0.580
Vasopressin V1a receptor	61	9	2	27	0.968	0.750	0.889	0.759	0.871	0.931	0.917
Average					0.759	0.654	0.740	0.414	0.781	0.638	0.751

Table S3c: Statistical performance metrics for the structural alert models against each biological targets validation data.