Supporting Information

Co-folding, the Future of Docking – Prediction of Allosteric and Orthosteric Ligands

Eva Nittinger[1,*], Özge Yoluk[3], Alessandro Tibo[2], Gustav Olanders[1] and Christian Tyrchan[1]

[1]Medicinal Chemistry, Research and Early Development, Respiratory and Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, 43183, Gothenburg, Sweden

[2]Molecular AI, Discovery Sciences, R&D, AstraZeneca, 43183, Gothenburg, Sweden

[3]Hit Discovery, Discovery Sciences, R&D, AstraZeneca, 43183, Gothenburg, Sweden

Corresponding author: Eva Nittinger, eva.nittinger@astrazeneca.com

S1. Dataset curation

The dataset for this study based on that published by Xie et al.[22]. Several adaptions have been conducted in order to enhance the quality of the structures used, containing the removal of structures with

a) close analogues binding to both orthosteric and allosteric binding site
b) with different UniProt Ids
c) where ligands bind to different proteins
d) with covalent ligands or peptides

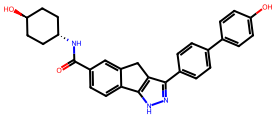
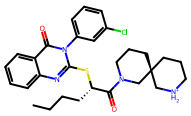
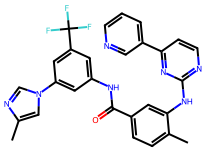
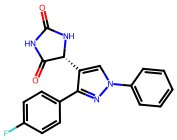
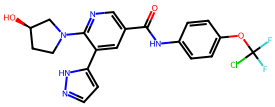
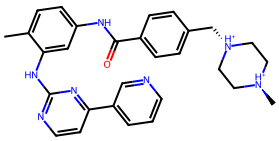
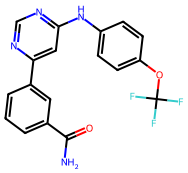
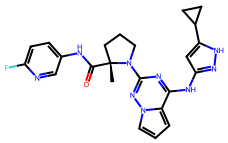
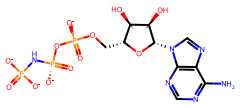
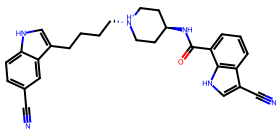
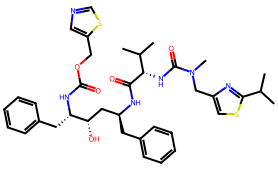
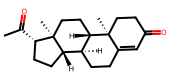
To further improve the quality of selected structures, we replaced entries with worse than 2.5Å resolution. If no structures with better resolution were available, the structures were retained. Last, endogenous ligands in the orthosteric site were – if possible – substituted with more drug-like ligands, in some cases we kept both the natural ligand and a more drug-like ligand.

The final dataset can be seen in Table S 1 and Table S 2.

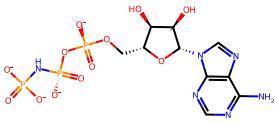
Table S 1. Curated data set with structural substitutions and information about ligands, uniport ids and exclusion criteria. 1: Xiu et al., 2: Olanders et al.

Source	Protein Name	OS replacement	OS Source PDBid	AS replacement	AS Source PDBid	UniProt IDs	Gene Name	Reasons for changes/exclusion
1	Serine/threonine protein kinase Chk1	2e9n	2E9N	3f9n	3JVR	O14757	CHEK1	change to more drug like ligand 3f9n
1	Tyrosine-protein kinase ABL1	5mo4	3PYY	3pyy, 5mo4	3PYY	P00519	ABL1	Remove 3pyy from OL due to structural similarities to 5mo4 AL
1	Tyrosine protein kinase ABL1	3k5v	3K5V	3k5v	3K5V	P00520	Abl1	
1	Insulin-like growth factor 1 receptor	1jqh, 3i81	1JQH	3lw0	3LW0	P08069	IGF1R	3i81 more drug like ligand
1	Cytochrome P450 3A4	3nux	1W0F	5a1r	1W0F	P08684	CP3A4	better resolution
1	Myosin II heavy chain	1mmn	2JHR	1yv3	2JHR	P08799	mhcA	better resolution, more drug like
1	Androgen receptor	2ax9, 2pio	2PIO	2yhd, 2ylo	2YHD	P10275	AR	more drug like, additional allosteric pocket
1	Tyrosine-protein phosphatase nonreceptor type 1	1bzc	1BZC	1t49	1T49	P18031	PTPN1	
1	RAC-alpha serine/threonine-protein kinase	3mvh	4EKK	4ejn	3O96	P31749	AKT1	4ejn better resolution, 3mvh more drug like
1	Glucokinase	3fgu	3FGU	1v4s	1V4S	P35557	GCK	
1	Mitogen-activated protein kinase 8	4izy	1UKI	3o2m	3O2M	P45983	MAPK8	4izy better resolution; 3o2m kept - no better resolution examples found
2	Nuclear receptor ROR-gamma	5aph	5APH	4ypq	4ypq	P51449	RORC	
1	Kinesin-like protein KIF11	3zcw	3ZCW	2gm1, 3wpn	3ZCW	P52732	KIF11	Two allosteric pockets, 3wpn kept - no better resolution examples found
1	Beta-lactamase TEM	1axb	1AXB	1pzp	1PZO	P62593	bla	replace covalent ligand 1PZO
1	Casein kinase 2a	3pe2	3H30	6fvf	3H30	P68400	CSNK2A1	original binding both pockets
1	Focal adhesion kinase 1	2ijm, 3bz3	2IJM	4ebv	4EBW	Q05397	PTK2	resolution, more drug like
1	Mitogen-activated protein kinase 14	3kf7	3KF7	3new	3NEW	Q16539	MAPK14	3new kept - no better resolution examples found

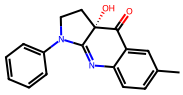
Table S 2. Orthosteric and allosteric ligand sets with their UniProt identifier and respective PDBids (molecule label: <UniProtID>_<Protein Name>_<PDBid>_<allosteric or orthosteric>_ligand).

1.	 O14757_CHEK1_2E9N_orthosteric	 O14757_CHEK1_3F9N_allosteric	
2.	 P00519_ABL1_5MO4_orthosteric	 P00519_ABL1_3PYY_allosteric	 P00519_ABL1_5MO4_allosteric
3.	 P00520_Abl1_3K5V_orthosteric	 P00520_Abl1_3K5V_allosteric	
4.	 P08069_IGF1R_3I81_orthosteric	 P08069_IGF1R_1JQH_orthosteric	 P08069_IGF1R_3LW0_allosteric
5.	 P08684_CP3A4_3NXU_orthosteric	 P08684_CP3A4_5A1R_allosteric	

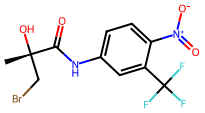
6.



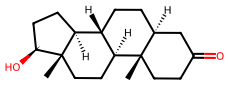
P08799_mhcA_1MMN_orthosteric



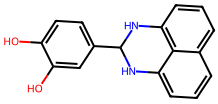
P08799_mhcA_1YV3_allosteric



P10275_AR_2AX9_orthosteric

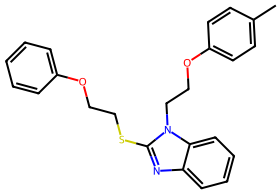


P10275_AR_2PIO_orthosteric



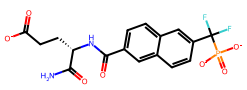
P10275_AR_2YHD_allosteric

7.

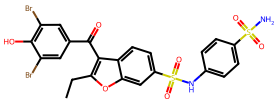


P10275_AR_2YLO_allosteric

8.



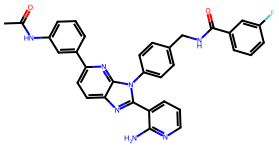
P18031_PTPN1_1BZC_orthosteric



P18031_PTPN1_1T49_allosteric

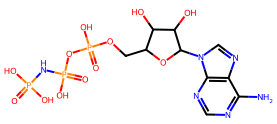
9.

P31749_AKT1_3MVH_orthosteric



P31749_AKT1_4EJN_allosteric

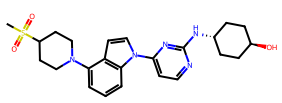
10.



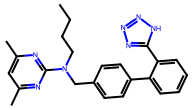
P35557_GCK_3FGU_orthosteric

P35557_GCK_1V4S_allosteric

11.

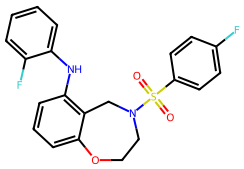


P45983_MAPK8_4IZY_orthosteric

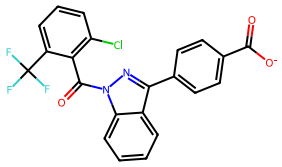


P45983_MAPK8_3O2M_allosteric

12.

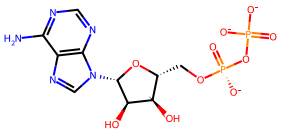


P51449_RORC_5APH_orthosteric

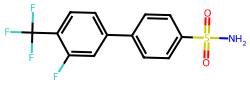


P51449_RORC_4YPQ_allosteric

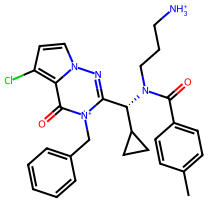
13.



P52732_KIF11_3ZCW_orthosteric

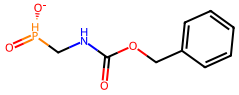


P52732_KIF11_3WPN_allosteric

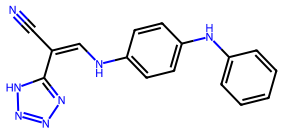


P52732_KIF11_2GM1_allosteric

14.

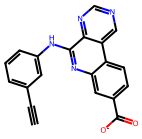


P62593_bla_1AXB_orthosteric

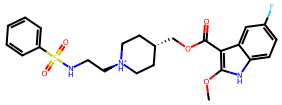


P62593_bla_1PZP_allosteric

15.

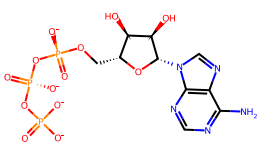


P68400_CSNK2A1_3PE2_orthosteric

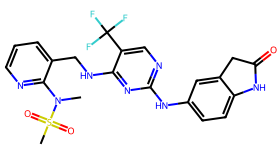


P68400_CSNK2A1_6FVF_allosteric

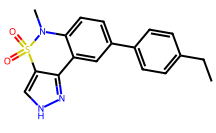
16.



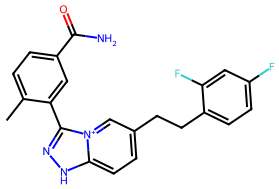
Q05397_PTK2_2IJM_orthosteric

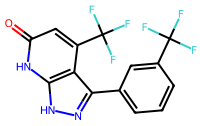


Q05397_PTK2_3BZ3_orthosteric



Q05397_PTK2_4EBV_allosteric





17. Q16539_MAPK14_3KF7_orthosteric Q16539_MAPK14_3NEW_allosteric

Table S 3. Ligand identifier for the allosteric and orthosteric X-ray structures. * = ligands bind to the same binding site and are labeled S1/S2 to facilitate differentiation in plots and evaluation.

UniProt IDs	Gene Name	OS PDBid	OS Lig Tag	Lig id	AS PDBid	As Lig Tag	Lig id
O14757	CHEK1	2e9n		76A	3f9n		38M
P00519	ABL1	5mo4		NIL	3pyy, 5mo4	S1, S2*	3YY, AY7
P00520	Abl1	3k5v		STI	3k5v		STJ
P08069	IGF1R	1jqh, 3i81	S1, S2*	ANP, EBI	3lw0		CCX
P08684	CP3A4	3nxu		RIT	5a1r		STR
P08799	mhcA	1mmn		ANP	1yv3		BIT
P10275	AR	2ax9, 2pio	S1, S2*	BHM, DHT	2yhd, 2ylo	S1, S2	AV6, YLO
P18031	PTPN1	1bzc		TPI	1t49		892
P31749	AKT1	3mvh		WFE	4ejn		0R4
P35557	GCK	3fgu		ANP	1v4s		MRK
P45983	MAPK8	4izy		1J2	3o2m		46A
P51449	RORC	5aph		VYI	4ypq		4F1
P52732	KIF11	3zcw		ADP	2gm1, 3wpn	S1, S2	2AZ, B4S
P62593	bla	1axb		FOS	1pzp		FTA
P68400	CSNK2A1	3pe2		E1B	6fvf		503
Q05397	PTK2	2ijm, 3bz3	S1, S2*	ATP, YAM	4ebv		0O7
Q16539	MAPK14	3kf7		L9G	3new		3NE

Table S 4. Information on amino acids lengths. Amino acids crystallized vs full length and manually truncated versions used for co-folding predictions.

UniProt IDs	Gene Name	OS PDBid	AA_{PDB} length	AS PDBid	AA_{PDB} length	Full length	Truncated
O14757	CHEK1	2e9n	269	3f9n	253	476	
P00519	ABL1	5mo4	429	3pyy, 5mo4	265, 429	1130	1-550
P00520	Abl1	3k5v	286	3k5v	286	1123	1-550
P08069	IGF1R	1jqh, 3i81	291, 296	3lw0	292	1367	950-1367
P08684	CP3A4	3nxu	457	5a1r	462	503	
P08799	mhcA	1mmn	737	1yv3	703	2116	1-800
P10275	AR	2ax9, 2pio	238, 247	2yhd, 2ylo	249, 250	920	650-920
P18031	PTPN1	1bzc	297	1t49	282	435	
P31749	AKT1	3mvh	301	4ejn	376	480	
P35557	GCK	3fgu	447	1v4s	448	465	
P45983	MAPK8	4izy	322	3o2m	358	427	
P51449	RORC	5aph	245	4ypq	243	518	
P52732	KIF11	3zcw	321	2gm1, 3wpn	328, 309	1056	1-400
P62593	bla	1axb	263	1pzp	263	286	
P68400	CSNK2A1	3pe2	327	6fvf	327	391	
Q05397	PTK2	2ijm, 3bz3	260, 259	4ebv	262	1052	400-700
Q16539	MAPK14	3kf7	337	3new	334	360	

Table S 5. Average runtimes for all co-folding methods on A100 80GB.

	Average runtime (sec)	Standard deviation (sec)
Boltz-1	145.95	118.88
Boltz-1 dual	173.68	203.92
Boltz-1 (with msa)	188.04	140.81
Boltz-1 dual (with msa)	194.3	146.71
NeuralPLexer (fast)	29.44	1.65
NeuralPLexer (pdb)	68.97	31.09
RoseTTAFold All-Atom	1831.44	2534.08

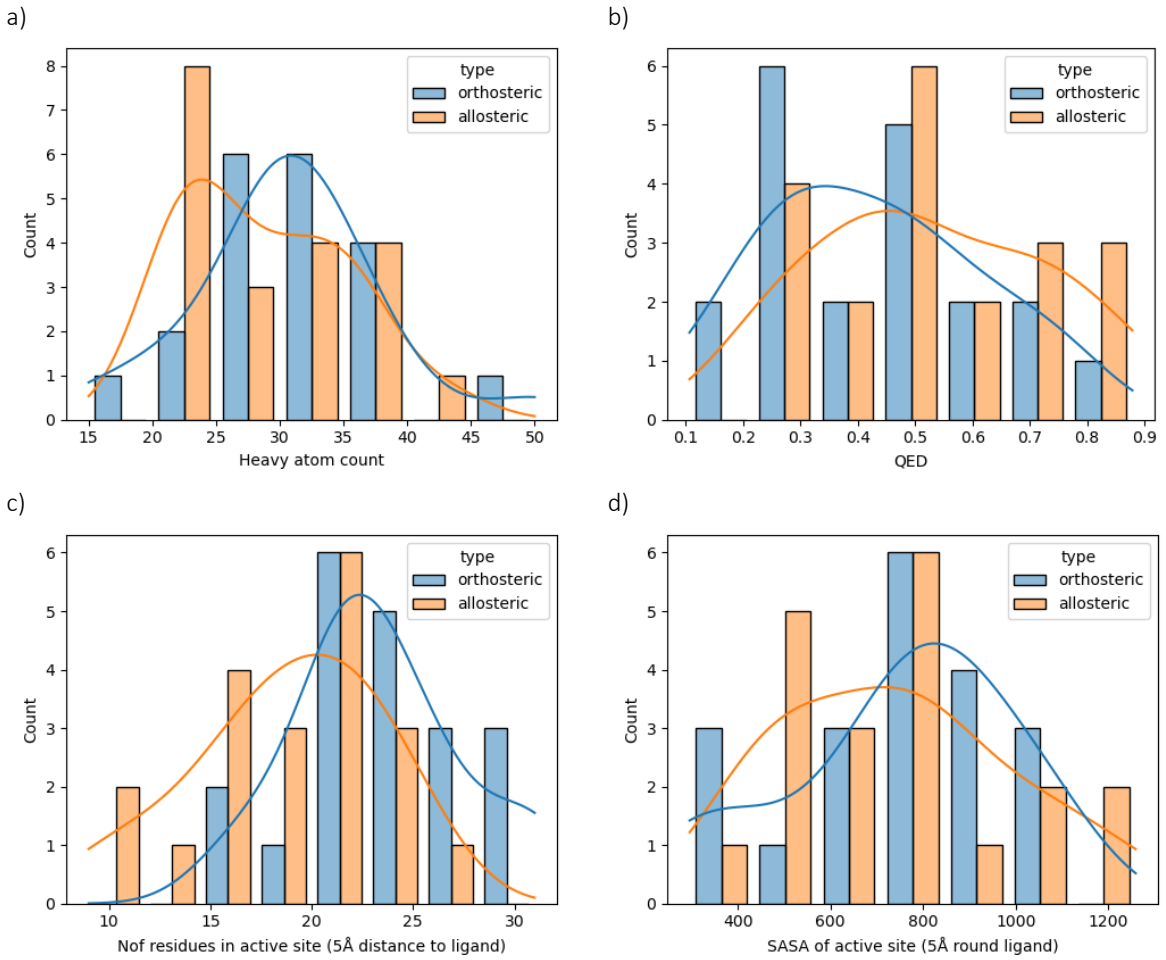


Figure S 1. Comparisons between orthosteric (blue) and allosteric (orange) ligand properties (a, b) and binding sites (c, d). The binding site is defined by a radius of 5Å around any ligand atom based on the X-ray structures; the solvent accessible surface area (SASA) was calculated for the empty pocket.

S2. PoseBusters Setting and Results

Table S 6. PoseBusters settings for default and stricter evaluation of physical-chemical validity of ligand predictions.

	default	strict
threshold_bad_bond_length	0.25	0.15
threshold_bad_angle	0.25	0.15
threshold_clash	0.25	0.15
threshold_flatness (ring)	0.25	0.15
threshold_flatness (double bond)	0.25	0.15
clash_cutoff (minimum distance)	0.75	0.85
clash_cutoff (volume overlap)	0.075	0.050

Table S 7. PoseBusters results (with strict settings) for all 40 predicted ligands by Boltz-1 (five predictions per ligand), NeuralPLexer (16 predictions per ligand) and RoseTTAFold All-Atom (one prediction per ligand).

program	success (%)	failed sanitization (%)	nonconnected (%)	wrong_bond_length (%)	wrong_bond_angles (%)	internal_steric_clash (%)	aromatic_ring_nonflat (%)	double_bond_nonflat (%)	internal_strain (%)	protein_clash (%)	volume overlap (%)	any_issue (%)
Boltz-1	200 (100)	5 (2.5)	0 (0)	4 (2)	0 (0)	19 (9.5)	0 (0)	0 (0)	6 (3)	168 (84)	34 (17)	65 (95.9)
NeuralPLexer	640 (100)	22 (3.4)	2 (0.3)	262 (40.9)	153 (23.9)	103 (16.1)	4 (0.6)	5 (0.8)	29 (4.5)	612 (95.6)	481 (75.2)	546 (86.5)
RosettaFold	37 (92.5)	2 (5)	1 (2.5)	10 (25)	16 (40)	13 (32.5)	0 (0)	1 (2.5)	1 (2.5)	37 (100)	26 (65)	37 (100)