

LSA: a local-weighted structural alignment tool for pharmaceutical virtual screening

Xiuming Li^{§a}, Xin Yan^{§*a}, Yuedong Yang^c, Qiong Gu^a, Huihao Zhou^a, Yunfei Du^c, Yutong Lu^c, Jielou Liao^d,
and Jun Xu^{*a,b}

^a Research Center for Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-Sen University, 132 East Circle at University City,
Guangzhou 510006, China

^b School of Computer Science & Technology, Wuyi University, 99 Yingbin Road, Jiangmen 529020, China

^c National Supercomputer Center in Guangzhou & School of data and computer science, Sun Yat-Sen University, 132 East Circle at University
City, Guangzhou 510006, China

^d Department of Chemical Physics, University of Science and Technology of China, Jinzhai Road. 96, Hefei 230026, China

§ Equal contributors. *To whom correspondence should be addressed.

Supplementary Materials

Table S1. AUC values and EF values of 102 compound libraries in DUD-E Collection by LSA and WEGA

Target	LSA				WEGA			
	AUC	EF ^{1%}	EF ^{5%}	EF ^{10%}	AUC	EF ^{1%}	EF ^{5%}	EF ^{10%}
AA2AR	0.79	31.82	8.55	4.98	0.65	9.98	4.32	2.95
ABL1	0.85	13.20	10.78	6.32	0.74	14.85	5.83	3.63
ACE	0.98	21.91	18.92	9.62	0.76	11.14	5.34	3.87
ACES	0.61	7.31	3.14	2.36	0.60	1.99	0.97	0.75
ADA	0.99	45.65	19.64	9.91	0.91	40.21	14.89	8.51
ADA17	0.86	27.68	10.80	6.58	0.84	19.39	9.63	6.07
ADRB1	0.93	15.44	9.16	8.83	0.58	8.94	3.24	2.75
ADRB2	0.89	16.52	8.66	8.19	0.60	2.61	1.47	1.00
AKT1	0.68	0.34	0.34	0.31	0.48	3.77	1.57	1.30
AKT2	0.84	4.32	4.97	5.05	0.68	12.10	4.97	3.17
AMPC	0.86	1.00	1.00	1.00	0.86	0.04	1.00	1.00
ALDR	0.73	23.98	6.67	4.03	0.70	30.92	8.69	4.97
ANDR	0.75	18.30	6.92	4.46	0.76	19.79	7.21	4.61
AOFB	0.52	3.28	1.64	1.81	0.54	0.82	1.97	1.40
BACE1	0.81	6.36	11.03	6.19	0.68	3.55	2.41	2.12
BRAF	0.75	35.59	7.78	4.87	0.73	32.55	8.30	4.48
CAH2	0.88	5.29	5.21	5.53	0.51	1.22	1.55	1.40
CASP3	0.72	3.02	5.93	4.37	0.60	9.55	3.22	2.46
CDK2	0.77	15.24	5.66	3.92	0.73	7.81	4.69	3.52
COMT	0.99	71.53	18.06	9.28	0.93	62.44	18.10	9.27
CP2C9	0.60	5.88	3.34	2.09	0.55	5.88	2.17	1.42
CP3A4	0.68	10.61	4.24	3.83	0.57	5.31	2.35	1.71
CSF1R	0.80	23.60	6.51	4.16	0.75	27.83	7.48	4.34
CXCR4	0.80	25.47	7.00	5.75	0.81	25.47	9.01	6.00

DEF	0.99	54.27	19.27	9.62	0.95	41.18	15.88	8.63
DHI1	0.68	9.71	4.55	3.73	0.65	13.04	4.91	3.09
DPP4	0.83	19.91	9.98	6.76	0.72	12.59	5.56	4.00
DRD3	0.75	0.83	1.79	2.00	0.55	1.04	0.92	0.81
DYR	0.88	62.51	14.72	7.62	0.81	35.60	10.13	5.93
EGFR	0.91	38.41	14.84	7.71	0.87	32.13	11.22	6.70
ESR1	0.90	24.09	8.36	6.29	0.85	44.25	12.49	6.95
ESR2	0.91	28.10	9.81	6.70	0.85	35.19	11.83	6.57
FA7	0.93	51.12	15.81	7.99	0.89	26.44	11.60	6.50
FA10	0.68	8.22	3.02	2.72	0.61	10.47	4.32	2.68
FABP4	0.91	13.02	11.12	7.03	0.85	34.74	8.56	6.61
FGFR1	0.51	1.70	1.03	1.23	0.49	0.00	0.73	0.79
FAK1	0.94	50.66	18.03	9.10	0.93	39.52	15.83	8.90
FKB1A	0.87	11.01	12.03	6.82	0.68	3.67	3.10	2.27
FNTA	0.86	49.52	12.87	6.86	0.77	18.76	7.13	4.63
FPPS	1.00	99.31	20.01	10.00	1.00	99.31	20.01	10.00
GCR	0.68	22.17	6.29	3.80	0.67	22.17	6.13	3.72
GLCM	0.78	30.38	13.01	6.67	0.76	26.59	9.67	5.93
GRIA2	0.72	36.28	9.24	4.94	0.71	26.73	8.36	4.75
GRIK1	0.78	24.14	7.75	4.86	0.68	5.03	3.97	2.48
HDAC2	0.88	10.29	6.93	6.76	0.53	9.21	2.27	1.62
HDAC8	0.90	60.91	16.14	8.24	0.83	44.94	12.84	6.89
HIVINT	0.82	17.09	11.83	8.21	0.60	1.01	0.60	0.60
HIVPR	0.81	8.96	4.93	4.59	0.82	12.88	7.17	5.11
HIVRT	0.59	14.56	4.44	2.69	0.58	15.16	4.20	2.55
HMDH	0.90	48.38	14.60	8.12	0.89	45.43	14.01	7.77
HS90A	0.67	47.07	9.56	4.78	0.65	17.22	7.97	4.33
HXK4	0.99	45.98	16.57	9.79	0.98	48.17	16.35	9.68
IGF1R	0.82	36.68	12.44	6.42	0.62	14.26	3.38	2.03
INHA	0.59	7.02	5.15	2.80	0.54	9.36	3.75	1.86
ITAL	0.60	24.13	6.68	3.34	0.49	16.09	4.21	2.10
JAK2	0.90	40.78	16.49	8.32	0.72	19.92	6.56	4.12
KIF11	0.80	10.48	8.18	5.57	0.79	39.29	10.97	6.27
KIT	0.73	2.43	1.45	1.87	0.53	3.04	1.57	1.15
KITH	0.90	41.48	14.83	7.39	0.87	45.08	12.71	6.86
KPCB	0.76	52.03	10.83	5.93	0.70	37.91	9.49	5.70
LCK	0.81	16.44	10.72	5.72	0.65	7.39	2.67	2.29
LKHA4	0.91	38.87	14.28	7.78	0.91	33.57	13.11	7.37
MAPK2	0.89	42.59	14.26	7.53	0.87	31.69	12.28	6.93
MCR	0.72	13.26	3.70	2.83	0.72	13.26	3.70	2.83
MET	0.91	46.20	14.00	8.56	0.85	50.45	13.16	6.93
MK01	0.92	43.68	15.47	7.73	0.83	34.69	8.62	4.69
MK10	0.87	11.59	10.40	7.99	0.64	3.86	2.31	1.73
MK14	0.71	6.41	6.41	3.95	0.64	11.43	3.98	2.68

MMP13	0.93	55.57	16.20	8.46	0.87	38.56	12.21	6.92
MP2K1	0.75	20.85	12.42	6.28	0.55	19.18	4.14	2.23
NOS1	0.70	11.08	11.02	5.70	0.57	2.01	2.81	2.40
NRAM	0.97	52.78	17.33	9.49	0.91	18.63	11.35	7.43
PA2GA	0.74	5.08	4.05	3.74	0.85	40.68	12.16	6.57
PARP1	0.90	46.14	14.18	8.17	0.77	17.55	6.89	4.45
PDE5A	0.69	12.82	7.74	4.12	0.63	22.38	4.98	2.99
PGH1	0.68	6.70	2.16	1.39	0.62	2.06	1.95	1.64
PGH2	0.83	49.46	11.32	6.07	0.82	36.81	10.44	5.70
PLK1	0.88	29.24	13.84	7.67	0.66	6.60	2.43	1.59
PNPH	0.93	53.48	14.78	7.77	0.93	53.48	14.78	7.77
PPARA	0.89	16.36	9.33	6.38	0.89	24.94	10.41	6.49
PPARD	0.89	8.75	8.08	6.42	0.80	10.84	5.17	4.21
PPARG	0.83	10.35	6.98	4.90	0.82	21.53	8.60	5.48
PRGR	0.75	34.57	7.99	4.34	0.68	4.45	1.84	1.95
PTN1	0.61	14.64	8.78	4.92	0.58	10.02	3.85	2.31
PUR2	0.99	34.80	18.81	10.00	0.99	2.05	19.21	9.80
PYGM	0.83	7.95	7.79	7.40	0.50	1.33	1.04	0.52
PYRD	0.89	51.23	14.63	7.85	0.89	51.23	14.63	7.85
RENI	0.64	9.66	4.24	2.89	0.62	9.66	4.43	2.60
ROCK1	0.78	9.08	6.82	5.51	0.56	2.02	2.21	2.00
RXRA	0.94	34.41	14.22	8.63	0.93	39.00	15.45	8.32
SAHH	1.00	54.86	20.09	10.02	1.00	1.61	20.09	10.02
SRC	0.73	25.77	8.71	4.70	0.61	8.59	2.79	1.79
TGFR1	0.95	48.88	17.45	8.95	0.79	28.58	8.42	4.59
THB	0.95	35.31	15.96	8.36	0.96	70.62	16.73	8.55
THRB	0.86	32.16	12.45	6.64	0.65	5.43	3.43	2.43
TRY1	0.94	43.58	15.86	8.51	0.77	10.73	5.75	4.17
TRYB1	0.84	16.39	11.51	6.02	0.67	8.88	2.57	2.37
TYSY	0.84	23.88	9.55	6.98	0.85	37.66	12.49	7.26
UROK	0.94	30.25	14.94	9.45	0.54	2.47	2.47	1.85
VGFR2	0.76	11.75	6.80	4.45	0.70	8.81	3.77	2.67
WEE1	0.99	59.60	18.10	9.33	0.99	0.99	17.90	9.52
XIAP	0.96	44.57	18.43	9.40	0.93	44.58	16.61	8.49

Table S2. SMILES for all specified substructures for all targeted libraries

Target Name	SMILES for specified substructures
AA2AR	n12c(ncn1)ncnc2N c12c([nH]cn1)ncnc2N
ABL1	n1ccnc1N n1cnccc1N
ACE	C(=O)(O)CNC=O
ACES	c12c(cccc1)nccc2N

ADA	c1[nH]cnc1
ADA17	c1cc(ccc1O)S(=O)(=O)N
ADRB1	C(N)C(O)C
ADRB2	C(N)C(O)C
AKT1	c12c([nH]cc1)ncnc2 c12c([nH]cc1)ncec2 [nH]1c2c(nc1)ncec2
AKT2	C1CNCCC1 c1cccn1
ALDR	CC(=O)O
AMPC	C(=O)(CCS=O)O C(=O)(C[N]C=O)O
ANDR	C12(C(CCC1)C1C(CC2)C2(C(CC1)CCCC2)C)C
AOFB	NCC=C
BACE1	C(CNC=O)O
BRAF	C1CC(CCN1)n1ncec1 c1cc(ccn1)c1n[nH]cc1 c1cc(ccn1)c1[nH]cnc1
CAH2	S(=O)(=O)N
CASP3	C(=O)(O)CCC=O
CDK2	c1ncnc1
COMT	c1(c(cccc1N(=O)O)O)O
CP2C9	c1cc(ccc1)c1cccc1 c1cc(ccc1)c1cnc1
CP3A4	c1ncsc1 c1[nH]cnc1
CSF1R	c1cccc1N
CXCR4	C1CCCCC1N
DEF	N(C=O)O
DHI1	C1CNCCC1 C1CNCCN1 c1cccn1
DPP4	C1CNCCN1C=O C1CCCCN1C=O C1CCN(C1)C=O
DRD3	C1CCNC1 C1CNCCN1 C1CCCCN1
DYR	c1cnc(nc1N)N
EGFR	c1cncnc1N
ESR1	c1cccc1O
ESR2	c1cccc1O
FA7	c1cccc1C(=N)N
FA10	S(=O)=O

FABP4	OC(=O)C
FAK1	n1ccenc1N
FGFR1	c1ccccc1 C1CNCCN1 C1CCCCN1
FKB1A	C(=O)(O)CN
FNTA	c1[nH]cen1
FPPS	P(=O)(CP(=O)(O)O)(O)O
GCR	P(=O)(CP(=O)(O)O)(O)O
GLCM	OCC(CO)CO OCC(O)CCO
GRIA2	C1C[N]C(C([N]1)O)O C1C[N]C(C([N]1)O)[N]
GRIK1	C(N)C(=O)O
HDAC2	C(=O)NO C(=O)NCCN
HDAC8	C(=O)NO
HIVINT	C(=O)CCO C(=O)CC=O
HIVPR	c1ccccc1S(=O)=O
HIVRT	c1(=O)cc[nH]c(=O)[nH]1
HMDH	C(=O)(O)CC(CCO)O
HS90A	c12c([nH]cn1)c(nen2)N
HXX4	c1enc(s1)NC=O
IGF1R	c12c(nc[nH]1)cccc2 c12c(cc[nH]1)encn2 c12n(enc1)ccnc2
INHA	c1ccccc1NC=O
ITAL	c1ccc(cc1Cl)Cl
JAK2	c1cncn1 c1ncccn1
KIF11	C1CCNC1 c1cc[nH]c1 N1=CCCN1
KIT	c1cccc(c1)NC=O
KITH	c1(=O)[nH]c(=O)cc[nH]1
KPCB	C1CC(=O)NC1=O
LCK	c1cnccc1N n1cnccc1N n1ccccc1N
LKHA4	c1ccccc1O
MAPK2	C1CCCN1=O
MCR	C12(C(CCC1)C1C(CC2)C2(C(CC1)CCCC2)C)C
MET	c1cnccc1

MK01	[nH]1c(ccc1)C(=O)N
MK10	c1ncccc1 n1ccnc1
MK14	c1[nH]cnc1
MMP13	ONC=O
MP2K1	NCCC#N
NOS1	NC=N
NRAM	O=C(O)C
PA2GA	C(=O)(O)C
PDE5A	[N]1C[N]CCC1N
PARP1	c1ccccc1C(=O)N
PGH1	c1ccccc1O
PGH2	c1ccccc1S(=O)=O
PLK1	c1c[nH]nc1 c1[nH]cnc1
PNPH	c12c(cc[nH]1)nc[nH]c2=O
PPARA	C(=O)(CC)O C(=O)(CO)O C(=O)(CN)O
PPARD	C(=O)(CC)O C(=O)(CO)O C(=O)(CN)O
PPARG	C(=O)(CC)O C(=O)(CO)O C(=O)(CN)O
PRGR	CC#N
PTN1	OCC(=O)O
PUR2	C(=O)(CNC=O)O
PYGM	C(CO)CCO [N](C=O)CC=O
PYRD	c1cnccc1C(=O)O
RENI	CNC(=O)C
ROCK1	c1cnccc1
RXRA	OC(=O)C
SAHH	c12c(ncn1)ncnc2[N]
SRC	c1cnenc1N n1ccnc1N
TGFR1	c1c(nccc1)C
THB	C(=O)(CN)O C(=O)(CO)O
THRB	C(=N)(C)N C(=N)(N)N
TRY1	C(=N)(C)N C(=N)(N)N

TRYB1	C1CCCCN1C=O
	C1CNCCN1C=O
TYSY	C(=O)(O)CNC=O
UROK	NC(=N)N
	CC(=N)N
VGFR2	c1enccc1
WEE1	C1CC(=O)NC1=O
XIAP	C(=O)(N)CNC=O
